ASBIRES 2026

Proceedings of the 18th Applied Science, Business & Industrial Research Symposium

19th November 2026

Faculty of Applied Sciences
Wayamba University of Sri Lanka
Kuliyapitiya

Faculty of Applied Sciences

Wayamba University of Sri Lanka

Proceedings of the 18th Symposium

On

Applied Science, Business & Industrial Research

19th November 2026

ASBIRES – 2026

Faculty of Applied Sciences

Wayamba University of Sri Lanka

Kuliyapitiya

© Faculty of Applied Sciences,
Wayamba University of Sri Lanka, 2026

Catalog in Publication Data approved by
National Library and Documentation Services Board

Symposium on applied science, business & industrial research (18th: 19th November 2026 Kuliyapitiya, Sri Lanka)

Proceedings. - Kuliyapitiya: Wayamba University of Sri Lanka, 2026 Vol. 18. - 306p; 36cm.

eISBN 978-624-6674-56-4
eISSN 2989-0683

i. 607 DDC 23 ii. Title

Applied science - Research
Industrial management - Research
Mathematical sciences - Research
Computer and information system - Research

ISBN 978-624-6674-56-4
ISSN 2989-0683

Typesetting, Designing and Printing
at Faculty of Applied Sciences,
Wayamba University of Sri Lanka, 2026

This event is proudly organized by the
Faculty of Applied Sciences, Wayamba University of Sri Lanka

EDITORIAL BOARD

Mrs. S.D. Dahanayaka

Editor in Chief

Department of Mathematical Sciences

Members

Prof. PMN Dharmawardane	*Department of Mathematical Sciences*
Mr. T Arudchelvam	*Department of Computing and Information Systems*
Dr. (Mrs). NADN Napagoda	*Department of Mathematical Sciences*
Dr. AD Dharmawansa	*Department of Industrial Management*
Mr. WCC Premarathna	*Department of Computing and Information Systems*
Mrs. WHDT Karunarathna	*Department of Mathematical Sciences*
Mrs. MPDS Lakshmi	*Department of Mathematical Sciences*
Ms. WMTH Wijesundara	*Department of Industrial Management*

Language Editing

Department of English Language Teaching

Members

Mr. EMHJ Edirisinghe
Ms. DHH Liyanage
Ms. GN Madumali
Ms. HMI Nawodya
Ms. HMHS Herath
Ms. LNR Pietersz
Ms. MDT Wicramasinghe
Ms. NMPB Nawarathna
Ms. SPC Marasinghe

Editorial Assistant

Mr. SPGE Illukkumbura	Ms. NWP Nawarathna
Ms. EMDJ Ekanayaka	Ms. PHAI Palansooriya
Mr. GMSD Kanchana	Ms. PHAJ Sewwandi
Mr. GN Dilshan	Ms. PPVA Dias
Ms. HSDM Higgoda	Ms. PRD Fernando
Ms. HWS Kaushalya	Mr. RAML Rajapaksha
Ms. KMCG Karunasekara	Ms. WMRT Wijekoon
Mr. KPRK Jayawardhana	Ms. WPAG Rajapaksha
Ms. NAVN Kanthi	Ms. WS Dilshani

PANEL OF REVIEWERS

Prof. SR Kodituwakku	University of Peradeniya
Prof. (Mrs.) DDM Jayasundara	University of Kelaniya
Prof. (Mrs.) I Hewapathirana	University of Kelaniya
Prof. (Mrs.) Lakshika S Nawarathna	University of Peradeniya
Prof. M Siyamalan	University of Jaffna
Prof. RMKT Rathanyaka	Sabaragamuwa University of Sri Lanka
Prof. WADK Jayendrika Wanigasekara	Wayamba University of Sri Lanka
Dr. AR Adikari	Southwestern College, USA
Dr. BMB Gayani Balasooriya	Ottawa University, USA
Dr. EA Chayani Dilrukshi	Nagaoka University of Technology, Japan
Dr. Himali Gammanpila	Eastern Kentucky University, USA
Dr. Indu Rasika Churchill	State University of New York at Oswego, USA
Dr. (Mrs.) Anuradha Hewaarachchi	University of Kelaniya
Dr. (Mrs). B Mayurathan	University of Jaffna
Dr. BMS Dillina Herath	ESOFT UNI
Dr. DD Menaka Ranasinghe	Open University of Sri Lanka
Dr. Dushyanthi Udeshika Vidanagama	General Sir John Kotelawala Defence University
Dr. (Mrs.) EJKP Nandani	University of Ruhuna
Dr. EYA Charles	University of Jaffna
Dr. (Mrs.) KD Prasangika	University of Ruhuna
Dr. R Nagulan	University of Vavuniya
Dr. S Kirushanth	University of Vavuniya
Dr. SPDSSK Karunarathna	University of Kelaniya
Dr. T Ketheesan	University of Jaffna
Dr. (Mrs.) DMKN Seneviratna	University of Ruhuna
Dr. Hakim A Usoof	University of Peradeniya
Dr. (Mrs.) Hemalika Abeysundara	University of Peradeniya
Dr. KDUD Fernando	Sabaragamuwa University of Sri Lanka
Dr. PGRS Ranasinghe	University of Peradeniya
Dr. RA Manjula Madhuwanthi	ESOFT Metro Campus
Dr. Sachith Abeysundara	University of Peradeniya
Dr. SK Nishadi Prasangini	Rajarata University of Sri Lanka
Dr. (Mrs.) Dinusha Priyadarshani	Open University of Sri Lanka
Dr. DL Chamara Pramod Liyanage	University of Sri Jayewardenepura
Dr. (Mrs.) DMPV Dissanayake	University of Kelaniya
Dr. (Mrs.)HL Jayetileke	University of Ruhuna
Dr. JABU Jayasinghe	University of Moratuwa
Dr. (Ms.) J Samantha Tharani	University of Jaffna
Dr. S Shriparen	University of Jaffna
Dr. SVA Almeida	University of Kelaniya
Dr. Tharaga Sharmilan	University of Kelaniya

Dr. T Kokul	University of Jaffna
Dr. WAC Weerakoon	University of Kelaniya
Dr. AJ Herath	Wayamba University of Sri Lanka
Dr. Emil Uduwalage	Wayamba University of Sri Lanka
Dr. US Liyanaarachchi	Wayamba University of Sri Lanka
Dr. (Mrs.) WDNSM Tennakoon	Wayamba University of Sri Lanka
Dr. LRA Gunasekara	Lincoln University, New Zealand
Dr. Sachinthani Eranga Alahakoon	Gampaha Wickramarachchi University of Indigenous Medicine
Dr. (Mrs.) GP Madhusanka	University of Sri Jayewardenepura
Mr. N Edwin Linosh	University of Vavuniya
Ms. R Yasotha	University of Vavuniya
Mr. SS Suthaharan	University of Vavuniya
Mr. S Suthakar	University of Jaffna
Mr. T Jeyamugan	University of Vavuniya
Mr. Wiraj Wickramaarachchi	Rajarata University of Sri Lanka
Ms. DMTD Dissanayake	Wayamba University of Sri Lanka
Mrs. PABH Amarathunga	Wayamba University of Sri Lanka
Mrs. WDMBK Dissanayake	Wayamba University of Sri Lanka
Ms. Rasini Nuwangi Wijeweera	University of Ruhuna

TABLE OF CONTENTS

Editorial Board…… *i*

Panel of Reviewers…….. *ii*

Table of Contents ……. *iv*

Computing and Information Systems

Cascade-Based Machine Learning Framework for Predicting Spillway Gate Operations at Randenigala Reservoir
Tilakaratne DCI, Wanninayake WMLN……... 1

Comparison of Machine Learning Models for Next-Hour Carbon Monoxide Prediction in Kandy, Sri Lanka
Marasinghe MMMS, Wanninayake WMLN…….. 6

Emotion-Adaptive Gameplay Using Multimodal Emotion Recognition
Gunasekara BSA , Aarudchelvam T.……... 11

MediTrack: A Context-Aware Mobile Health Application for Medication Support and Emergency Medical Information Access
Amarasinghe KKG, Perera KTDTP, Vidanagama VGTN…….. 16

SmartSpend: A Cross-Platform Mobile Application for Personal Finance Management
Liyanagamage HW, Fernando LRN, Vidanagama VGTN……... 21

Machine Learning for Personalized Food Allergen Risk Prediction
*Panditha*ge *SD, Gunasekara RPTH*……. 26

PearlWay: A Context-Aware Mobile Application for Beach Recommendations in Sri Lanka
Kulathunga MGAD, Daranagama DADA, Vidanagama VGTN…….. 31

Artificial Intelligence-Assisted Context-Aware Micro-Learning Platform for Personalized Educational Video Learning
Handapangoda HGRI, Mahagedara HBMHS, Vidanagama VGTN…… 36

Zonely: Harnessing Context-Aware Computing to Restore Digital Wellbeing and Focus
Peiris TKSA, Thennakoon PGDS, Vidanagama VGTN……... 42

Adoption and Perception of Digital Tools for Public Transport Improvement in Sri Lanka
Wickramanayake WANN, Premarathna WCC……. 49

Smart Water Leak Detection and Consumption Monitoring System: An Adaptive Statistical Abnormal-Flow Detection Approach
Chandira HPS, Pathirana OPTP, Vidanagama VGTN……... 54

An Explainable Machine Learning Framework for Lead Conversion Prediction and Sales Visit Route Planning in the Automotive Industry
Siriwardhana A, Premarathna WCC……... 61

A Calibrated Hurdle-Based Machine Learning Framework for Multi-Horizon District-Level Rainfall Forecasting
Wijerathna WSMRN, Gunasekara RPTH …… 67

Examscope: AI-based syllabus coverage analysis system using a hybrid large language model and machine learning classification
Prathapage LL, Jayakody JRKC…….. 72

Investigating the Suitability of Machine Learning Classification of Large Language Model -Generated Descriptions for Real Time Accident, Crime, and Disaster Reporting
Bandara DMDA, Jayakody JRKC…… 79

A Comparative Evaluation of Large Language Models for Mathematical Concept and Relation Extraction from Calculus Text
Athapaththu SSS, Liyanage DLCP, Karunaratna DD, Ratnayake JK……. 85

Image Processing Approach to Grading Pepper Quality in Sri Lanka
Rajapaksha RWWMHM, Munasinghe B…….. 90

Computational Empowerment in Sri Lankan Information & Communication Technology and Related Education
Basnayaka PBN, Munasinghe B……... 95

Machine Learning-Based Recommendations for First-Move Advantage in Chess Using Rating Difference, Time Control, and Opening Classification
Gamage KPLT, Munasinghe B …… 100

Comparison of Convolutional Neural Network Models for Image-Based Detection of Sewing Defects and Efficiency Improvement
Perera WASM, Munasinghe B …..……. 105

Gender-Based Differences in Information Security Awareness and Vulnerable Online Practices Among Art-based University Undergraduates: Evidence from a State University in Sri Lanka
Jayaweera LP……... 110

Industrial Management

Generative AI and the Evolving Career Transition of QA Engineers in the Sri Lankan IT Industry
Erandi RHN, Jayasundara DRT, Jayasinghe JABU…… 116

The Impact of Green Supply Chain Management Practices on Organizational Performance: Evidence from Apparel Exporters in Sri Lanka
Mathotage RR, Jothirathne PAAU……... 121

Are Supplier Performance and Transportation Efficiency Associated with Logistics Costs? An Empirical Analysis Based on Asian Countries
Hansana JAP, Pallegedara A……. . 127

Impact of Social Media Usage and Online Professional Presence on Employment Outcomes of Sri Lankan Management Graduates
Siriwardana DN, Prabodanie RAR…….. 132

A Technology Acceptance-Based Empirical Study of Factors Associated with Citizen Adoption of GovPay and the e-Revenue License System in Sri Lanka
Kaumadi WAI, Deegahawatura MMDR…….. 137

Association of Gender, Education Level, and Residence Type with Cybercrime Victimization Among Emerging Adult Internet Users in Sri Lanka
Rathnayaka RMKGSB, Wattegama EJ …… 142

The Moderating Role of The Employment Sector on The Relationship Between Culture of Support Policies and Work-Life Balance in The Western Province, Sri Lanka
Mathotage RR, Jothirathne PAAU……... 148

Online Hotel Review Sentiment and Booking Intention: A Moderated Study Among Sri Lankan Consumers
K.P.K. Prabhash, Habaragoda BS……... 153

The Impact of Lean Manufacturing Practices on Operational Performance: Evidence from the Sri Lankan Beverage Industry
Athapattu ARNC, Jothirathne PAAU……... 159

Association between the Marketing Mix and Sri Lankan Audience Engagement with Sinhala Cinema
Madhuranga WADS, Wattegama EJ…… 164

Determinants of Work-Life Balance: A Case Study of The Furniture Manufacturing Sector in Sri Lanka
Dasanayaka TPNN, Deegahawatura MMDR…….. 172

Impact of Social Influence on Investment Decisions:The Mediating Role of Financial Literacy Among Employees in Sri Lanka
Madhuwanthi MPDP, Karunarathne EACP…… 177

Effect of Sustainable Development Goals on Tourism Growth: Evidence from Asian Developed Economies
Jayathunga PDR, Deegahawatura MMDR……. 182

How Consumers Value Traditional Drawings in Sri Lanka: The Role of Art-Related Characteristics in Perceived Value
Ariyawansha GVADD, Karunarathne EACP…….. 188

Geographic Variation in Category-Level Spending Across Outlet Locations in Sri Lankan Grocery Retailing
Perera VPM, Prabodanie RAR…… 194

Crafting An Ethical Business Climate: An Entrepreneurial Value Perspective on Product Harm Crisis Management
Sabarjah S, Herath HMSP……... 199

Predictors of Generation Z's Purchase Intention for Green Packaged Products in Sri Lanka: The Mediating Role of Perceived Value
Rasangani NWGS, Wattegama EJ……... 204

Mathematics & Mathematical Modelling

Phase-Specific Modeling of Dengue Transmission Dynamics and Temporal Surges in the Gampaha Medical Officer of Health Area, Sri Lanka
Fernando MWJJ, Dharmawardane PMN…… 209

Developing an RSA-AES Hybrid Encryption Framework for Secure Student Results: A Case Study of the Faculty of Applied Sciences, Wayamba University of Sri Lanka
Shreen MT, Dahanayaka SD…… 215

Comparative Study of Analytical and Numerical Methods for Some Trigonometric Integrals
Tharushika AHP, De Silva TMM……. 226

Examination Scheduling Using Integer Linear Programming: A Case Study of the Faculty of Applied Sciences, Wayamba University of Sri Lanka
Hashini AI, Dahanayaka SD…… 226

Statistics

Box–Jenkins Modeling and Forecasting of Monthly Solar Radiation: A Case Study in Sri Lanka
Imalsha RD, Abeygunawardana RAB…….. 232

Modeling Hospital-Based Diabetes Mortality in Sri Lanka
Wijekoon WMRT, Dharmawardane PMN, Dissanayaka DMDK…… 237

Predicting Daily Coconut Retail Prices in Major Markets of Sri Lanka
Wanigasekara WADS, Appuhamy PADAN……... 243

A Hybrid Deep Learning Model for Forecasting of Stock Market Prices in Colombo Exchange
De Silva RSA, Lakshmi MPDS……. 248

An Empirical Investigation of the Short-Run and Long-Run Dynamics of Electricity Consumption and Industrial Output in Sri Lanka: A VECM Approach
Anjalee WAH, Aponsu GMLM……. 253

Battery Remaining Useful Life Estimation in Electric Vehicles Using Multiple Linear Regression and Random Forest Models
Palansooriya PHAI, Francisco GS…….. 259

A Comparative Study of Black-Scholes and Binomial Tree Model and Monte Carlo Simulation Method for Pricing European Call Options on Netflix Stock
Ekanayaka EMDJ, Karunananda GACM…… 265

Assessment of Factors Affecting Soil Organic Carbon Content in Sri Lankan Tea-Growing Soils
Wijethunga WMYN, Karunananda GACM…….. 270

A Comparative Study of Glycemic Control in Gestational and Pre-Gestational Diabetes in Batticaloa District – A Case Study
Weerasinghe TK, Panahatipola PMOP……... 275

Modelling and Forecasting Daily USD/LKR Exchange Rate Dynamics Using ARIMA and EGARCH Models
Nawarathna NWP, Aponsu GMLM ……. 280

Hybrid DNN–FNN Neural Networks Model for Potato Yield Prediction: A Case Study in India
Warnakula WPM, Napagoda NADN…… 286

Exchange Rate Movements and Tourism Demand in Sri Lanka: Evidence from a Vector Error Correction Model
Sewwandi JAB, Aponsu GMLM……... 292

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Cascade-Based Machine Learning Framework for Predicting Spillway Gate Operations at Randenigala Reservoir

Tilakaratne DCI[1], Wanninayake WMLN[2]

[1,2]*Department of Computing and Information Systems, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*

chathunit2000@gmail.com[1] , *lakminiw@wyb.ac.lk*[2]

ABSTRACT

Reservoir spillway operation is a part of flood management and dam safety in Sri Lanka, where decisions are usually made using reservoir observations, set rules and operator experience. This study suggests a cascade-aware machine learning framework to predict events at the Randenigala Reservoir by adding hydrological data from the upstream Victoria and Kotmale Reservoirs. Based on daily hydrological data in five years (2020-2024), a set of 28 travel-time-lagged inflow and rainfall features for both upstream reservoirs and local features for Randenigala Reservoir were engineered. A Random Forest classifier trained on the cascade feature set was compared to the same classifier trained only on local features in the order of train-test splits reflecting real-life deployment conditions. The experiment revealed that including signals from the upstream reservoirs improves precision of prediction from 0.60 to 0.92 and F1 score from 0.71 to 0.89 while maintaining recall at 0.86 through reducing the number of false positives. Feature importance and lead-time analyses further reveal that local water-balance variables remain the dominant predictors, with upstream signals functioning as a corroborating rather than independently decisive input. These findings of this study offer an evidence-based framework for integrating cascade hydrology into operational decision support, with direct relevance to flood management across interconnected reservoir systems.

KEYWORDS: Machine Learning, Random Forest, Reservoir Management, Spillway Prediction

1 INTRODUCTION

The reservoirs of Sri Lanka are vital for its irrigation, electricity generation, domestic water supply and flood mitigation operations. There exists a cascade of interconnected reservoirs in the Mahaweli Development Project where hydrological conditions of the downstream reservoirs depend on the water releases made by the upstream reservoirs. Randenigala is one such multi-purpose reservoir and is impacted hydrologically by other upstream reservoirs like Victoria and Kotmale. The upstream releases are observed at the Randenigala reservoir with some delay due to the travel-time effect.

The conventional operation of spillway operation relies on the rule curves, the current reservoir observations and the experience of the operators. While these approaches have been quite helpful for spillway operations, sudden changes in rainfall and inflows can make timely decision-making rather challenging. It is possible that improper or delayed decisions can be detrimental to the downstream flooding, extra water releases and hydropower storage management.

Machine learning techniques provide a possibility to construct nonlinear relations between hydrological variables and spillway events. However, the majority of reservoir prediction models

use only local reservoir variables, disregarding the hydrological connections among cascades. The purpose of this research is to construct the cascade-aware Random Forest (RF) framework for Randenigala using local variables and lagged rainfall and inflow variables from Victoria and Kotmale in order to see whether such an approach, improves spillway event prediction.

2 LITERATURE REVIEW

2.1 Reservoir Management and Gate Operation

The management of a reservoir requires flood mitigation, irrigation needs, energy generation, and storage, among others. Classical reservoir management methods rely on pre-set operating policies and either simulation or optimization models (Yeh, 1985; Labadie, 2004). Hydrological variability and non-linearity in rainfall- runoff relations have necessitated the application of data-driven modelling.

2.2 Hydrological Forecasting Techniques

The complexity of hydrological processes and the non-linear nature of rainfall-runoff relations have restricted the efficacy of traditional rule-based and statistical hydrological forecasting techniques, driving the development of data-driven methods that can account for the complexities in hydrology without much physical parameterization.

2.3 Machine Learning for Water Resources Management

Data-driven machine learning techniques have increasingly been used for hydrological forecasting applications. Neural network has been shown to be able to simulate rainfall-runoff relations (Hsu et al., 1995). Random Forest is a powerful non-linear modelling method that also ranks variable importance (Breiman, 2001). Long Short-Term Memory network is another example of a machine learning technique that can be used for sequential hydrological prediction due to its temporal dependency capability (Hochreiter & Schmidhuber, 1997).

2.4 Multi-Reservoir and Cascade-Based Modelling

Upstream reservoir conditions can influence downstream inflows and reservoir operations. Yi and Yi (2024) highlighted the relevance of using information from multiple reservoirs, while Li et al. (2024) investigated a hybrid approach to cascade reservoir inflow forecasting.

2.5 Upstream Travel-Time Lags and Early Warning

Hydrological effects from upstream reservoirs may take time to reach downstream locations. Therefore, incorporating lagged rainfall and inflow variables may help represent delayed hydrological relationships. Li and Jun (2024) reviewed machine learning methods for river flood routing, providing relevant context for understanding flood movement through river systems.

3 METHODOLOGY

The methodology involved a sequential pipeline (Figure 1) that included cascade data acquisition and merging, pre-processing, time-based feature engineering, exploratory analysis, model creation, and model comparison along with lead time analysis.

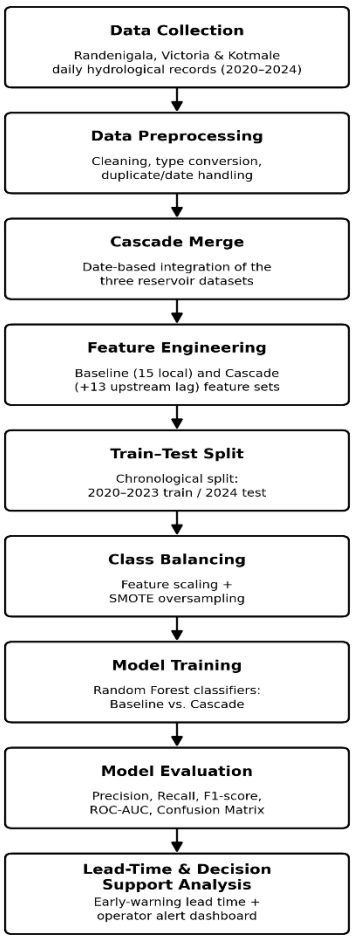


Figure 1: Methodology workflow adopted in the study.

The chronological train test split ensured that 2020-2023 data were used for training purposes while 2024 data were reserved for testing.

3.1 *Study Area and Data*

This research took into consideration the Kotmale-Victoria-Randenigala reach in the Mahaweli cascade. Water flows from Kotmale to Victoria and subsequently to Randenigala, where Rantambe is located below. The selection of the Randenigala reservoir as the target reservoir was due to the importance of the reservoir in hydropower generation, irrigation management, and flood control.

The daily data for hydrology of this research were collected from the Irrigation Department of Sri Lanka. The Randenigala dataset was collected from January 2020 to December 2024 and had 1,827 data points on water level, storage, inflow, rainfall, and releases variables. The Victoria and Kotmale datasets included upstream rainfall and inflow variables.

3.2 *Preprocessing and Feature Engineering*

Raw records were cleaned by standardizing variable names and numeric formats, removing duplicate dates, sorting chronologically, and aligning reservoir data by date. The target was defined as SPILLWAY_OPEN = 1 when daily spillway release was greater than zero, otherwise 0. Consecutive positive days were grouped as one event for lead-time analysis, while genuine zero rainfall and inflow values were retained.

Two feature sets were evaluated: 15 local Randenigala features and a cascade model with 13 additional upstream or combined features, totaling 28 predictors. These included local hydrological variables, lags, rolling means, water-level change, and calendar indicators. Victoria features used one- to three-day inflow lags, while Kotmale features used two- to four-day lags and rolling summaries. CASCADE_INFLOW_PRESSURE combined one-day-lagged Victoria inflow and two-day-lagged Kotmale inflow. To reduce same-day leakage, hydrological measurements and rolling summaries used information up to day $t-1$, while known calendar variables were retained.

3.3 *Model Development and Evaluation*

Random Forest was selected after preliminary comparison with Gradient Boosting. The baseline used 15 local features and the cascade model 28. Both used Random Forest Classifier

(n_estimators=200,class_weight='balanced', random_state=42), with other hyperparameters at library defaults. No systematic hyperparameter search was implemented; therefore, this configuration represents a fixed experimental setting rather than an optimized one.

Standard Scaler was fitted on training data only and applied to the holdout set. The Synthetic Minority Over-Sampling Technique (SMOTE) was then applied only to the scaled training data using random_state=42, with default sampling_strategy='auto' and k_neighbors=5. Resampling followed the chronological split, with no synthetic samples generated for the 2024 test set. The training set contained 1,441 observations (41 opening, 1,400 non-opening), becoming 1,400 per class after SMOTE, while the test set contained 366 observations (14 opening, 352 non-opening).

Performance was evaluated on the untouched 2024 test set using precision, recall, F1-score, the area under the receiver operating characteristic curve (ROC-AUC) from predicted probabilities, average precision (the area under the precision–recall curve, PR-AUC), and confusion-matrix counts. Default Random Forest predictions were used for threshold-dependent metrics, while lead-time analysis used a 0.4 probability threshold. Lead-time analysis was conducted to determine how early the model could provide an indication of an upcoming spillway-opening event. Consecutive spillway-opening days were grouped into individual events, and model predictions were examined within the seven days preceding each event. An alert was considered an advance warning only when the predicted probability exceeded the selected threshold of 0.4 before the event date. Chronological splitting prevented future-year training, and train-only SMOTE prevented synthetic test data.

4 RESULTS AND DISCUSSION

In the end the final dataset had 1,814 observations and 55 of those were spillway-opening days. This big gap between the numbers shows that I really need to use class-balancing techniques and look at more than accuracy to see if the model works. When I first compared Random Forest and Gradient Boosting, the Random Forest looked like a base classifier. Because of that, I used Random Forest for the controlled baseline-, versus-cascade comparison.

As shown in Table 1, the performance of the cascade model saw the improvement of its precision from 0.60 to 0.92 without altering its recall of 0.857. The value of F1-score rose from 0.706 to 0.889, while ROC-AUC score was raised from 0.988 to 0.996. The most significant operational improvement was the decrease in false positive predictions from eight to one. As both models predicted 12 of 14 spillway opening days for the 2024 test year, the inclusion of upstream variables in the prediction did not increase the number of correctly classified

Table 1: Baseline and cascade Random Forest performance.

Metric	Baseline RF	Cascade RF	Change
Precision	0.60	0.92	+0.32
Recall	0.857	0.857	0.00
F1-score	0.706	0.889	+0.183
ROC-AUC	0.988	0.996	+0.008

spillway opening events. It was used mainly for reducing the number of false alarm events. This means that the upstream cascade data worked as an additional signal that enabled the model to separate spillway-risk situations from similar local hydrological changes.

Based on the lead-time analysis, it is clear that the baseline and cascade models recorded identical lead times of 2.3 days for all the three spillway events of the year 2024. In this case, the cascade model did not add any extra day of early warning to the lead time. On the other hand, the cross-correlation analysis showed that there was an approximate peak lag of one day in the upstream and Randenigala inflow.

5 CONCLUSION

The paper designs and assesses the cascade-aware machine learning technique for the prediction of spillway-opening events in Randenigala Reservoir using local hydrological data along with the lagged rain and inflows from the upstream Victoria and Kotmale reservoirs.

In controlled experiments, the proposed cascade Random Forest improves precision from 0.60 to 0.92, F1-score from 0.706 to 0.889, and ROC-AUC from 0.988 to 0.996 while keeping the same recall of 0.857. Number of false positives is lowered from eight to one. Thus, it is possible to conclude that integration of the upstream cascade data into analysis may enhance the reliability of spillway risk classifying without lowering its sensitivity.

In summary, the paper presents a methodology for developing a cascade-aware reservoir decision-support system for Sri Lanka. Possible future studies may include longer historical datasets, sub-daily observations, input of the numerical weather prediction data, different sequence-based machine learning approaches, explainability methods, and uncertainty quantification. Extension of the approach to the whole Mahaweli cascade may help to understand if the basin-wide hydrological data can be used as predictive.

REFERENCES

- Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780
- Hsu, K., Gupta, H. V., & Sorooshian, S. (1995). Artificial neural network modeling of the rainfall-runoff process. *Water Resources Research*, *31*(10), 2517–2530.
- Labadie, J. W. (2004). Optimal operation of multireservoir systems: State-of-the-art review. Journal of Water Resources Planning and Management, 130(2), 93–111.
- Li, J., Dao, V., Hsu, K., Analui, B., Knofczynski, J. D., & Sorooshian, S. (2024). Improving cascade reservoir inflow forecasting and extracting insights by decomposing the physical process using a hybrid model. *Journal of Hydrology*, *630*, Article 130623.
- Li, L., & Jun, K. S. (2024). Review of machine learning methods for river flood routing. Water, 16(2), Article 364.
- Yeh, W. W.-G. (1985). Reservoir management and operations models: A state-of-the-art review. *Water Resources Research*, *21*(12), 1797–1818.
- Yi, S., & Yi, J. (2024). Reservoir-based flood forecasting and warning: Deep learning versus machine learning. Applied Water Science, 14, Article 237.

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Comparison of Machine Learning Models for Next-Hour Carbon Monoxide Prediction in Kandy, Sri Lanka

Marasinghe MMMS[1], Wanninayake WMLN[2]

[1,2]*Department of Computing and Information Systems, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*

malshamarasinghe44@gmail.com [1], *lakminiw@wyb.ac.lk* [2]

ABSTRACT

Air quality monitoring has become a critical concern in urban environments, where meteorological conditions and atmospheric pollutant concentrations contribute to short-term variations in air pollution. Carbon monoxide (CO) is an important pollutant with well-documented adverse effects on human health, making accurate short-term forecasting essential for effective air quality management. This study investigates the potential of machine learning for next-hour CO concentration prediction in Kandy, Sri Lanka, a city characterized by valley topography that can restrict pollutant dispersion. The study uses three months of hourly observations from January to March 2026 obtained from the Central Environmental Authority (CEA) of Sri Lanka. The analysis incorporates meteorological variables and ambient concentrations of CO, nitrogen dioxide (NO_2), nitrogen oxides (NOx), sulfur dioxide (SO_2), and ozone (O_3), with temporal dependencies represented using lag-based, rolling, and temporal features. The predictive performance of multiple regression-based and ensemble machine learning models was assessed using root mean squared error (RMSE), mean absolute error (MAE), and the coefficient of determination (R^2), with hyperparameter tuning performed using chronological validation. The results indicate that ensemble tree-based models generally outperformed linear and kernel-based models for short-term CO prediction. Among the evaluated approaches, the tuned CatBoost model achieved the strongest overall performance, with an R^2 of 0.7728, RMSE of 105.33 ppb, and MAE of 76.33 ppb on the unseen test set. These findings indicate the potential of machine learning, particularly ensemble tree-based methods, for next-hour CO prediction within the study setting in Kandy, Sri Lanka.

KEYWORDS: Air Quality, Carbon Monoxide, CatBoost, Machine Learning, Meteorological

1 INTRODUCTION

Kandy is a major city in Sri Lanka situated within a valley surrounded by mountain ranges, where restricted air circulation can contribute to pollutant accumulation. Carbon monoxide (CO), primarily associated with vehicular emissions and fuel combustion, is a harmful air pollutant that can affect human health by reducing the blood's oxygen-carrying capacity.

CO concentrations are influenced by meteorological and ambient air quality variables, making short-term prediction challenging. Although air pollution in Kandy has been studied, next-hour CO prediction has received limited attention. Machine learning provides a promising approach for capturing complex and nonlinear relationships among these variables.

This study aims to evaluate and compare multiple machine learning regression models for next-hour CO concentration prediction using three months of hourly meteorological and air quality observations collected by the Central Environmental Authority (CEA) in Kandy from January to March 2026. It further identifies the best-performing model and the most influential predictors for next-hour CO prediction.

The main contribution is a localized machine learning approach that integrates temporal, meteorological, and ambient air quality features for next-hour CO prediction in Kandy. The study provides a comparative evaluation of ten regression models and identifies the most influential predictors within the best-performing model.

2 LITERATURE REVIEW

2.1 Impact of Variations in Meteorological Parameters on Carbon Monoxide

Meteorological conditions influence CO concentrations through factors such as humidity, pressure, temperature, wind speed, and direction. These variables explained approximately 66.6% of CO variation in one study (Anjali, 2022), while temperature and wind conditions affect atmospheric stability, pollutant transport, and dilution (Masoudi & Gerami, 2017).

2.2 Influence of Air Quality Variables on Carbon Monoxide Prediction

Air quality variables such as nitrogen oxides (NOx), nitrogen dioxide (NO_2), and O_3 are important predictors of CO due to their links with traffic emissions and atmospheric processes (Wang, 2025). Nitric oxide (NO) was reported as the strongest pollutant influencing CO variation in a previous study (Abdul-Wahab et al., 2003).

2.3 Diurnal Variation of Carbon Monoxide Concentrations in Urban Areas

CO concentrations also exhibit temporal patterns in urban environments, with higher levels commonly observed during periods of increased traffic and weaker atmospheric mixing (Masoudi & Gerami, 2017).

2.4 Feature Engineering Techniques for Time-Series Carbon Monoxide Prediction

Feature engineering can help machine learning models capture these temporal dependencies. In particular, temporal features, cyclical encoding, lagged CO values, and rolling statistics have been shown to improve the prediction of hourly CO concentrations (Harun et al., 2025).

2.5 Machine Learning Approaches for Carbon Monoxide Prediction

Various machine learning approaches have been investigated for CO prediction, with performance differing across datasets and locations. Harun et al. (2025) reported XGBoost as the best-performing model among fourteen evaluated approaches using the University of California, Irvine (UCI) Air Quality dataset. In Malaysia, Almubaidin et al. (2024) found that a Matern 5/2 Gaussian Process Regression model achieved the best performance, while Latif et al. (2023) reported superior performance from an artificial neural network multilayer perceptron (ANN-MLP) model using air quality and meteorological data from Kuala Lumpur.

However, limited studies have explored next-hour CO prediction in Kandy, motivating the comparison of multiple machine learning models.

3 METHODOLOGY

3.1 Data Collection

Meteorological and air quality data were obtained from the CEA ambient air quality monitoring station in Kandy city (global positioning system (GPS) coordinates: 7.292651, 80.635649). Three

months of hourly observations, from January 01 to March 31, 2026, were used. The dataset included ambient temperature (AT, °C), relative humidity (RH, %), barometric pressure (BP, hPa), solar radiation (solar rad, W/m²), wind speed (WS, m/s), wind direction (WD, °), and gaseous pollutants CO, ozone (O_3), sulfur dioxide (**(**SO_2), NO_2 and NOx. All gaseous pollutant concentrations were measured in parts per billion (ppb).

3.2 Data Pre-processing

The data were ordered chronologically and checked for quality. Negative values in non-negative variables, including solar radiation and SO_2, were replaced with zero. The dataset was then divided chronologically into 80% training and 20% testing sets. Missing values were handled separately in the training and testing datasets using variable-specific imputation methods to prevent data leakage. Time-based interpolation was applied to pollutant variables, while linear interpolation was used for meteorological variables. Any remaining gaps after interpolation were filled using forward fill. Importantly, imputation was performed separately after the chronological split, so observations from the test period were not used to impute missing values in the training period. Outliers were retained, as extreme values may represent genuine pollution events relevant to CO prediction.

3.3 Feature Engineering

The training and testing datasets were temporarily combined in chronological order to create lagged and rolling features while preserving temporal continuity. These datasets were then separated before model training and evaluation. Lagged and rolling features were calculated using only observations available before the prediction time. CO concentrations at 1, 2, 3, and 6 hours prior and the 6-hour rolling standard deviation were used to capture short-term CO dynamics. The rolling statistic was calculated using preceding observations to avoid incorporating the current or future CO concentration. WD was transformed into sine and cosine components, while hour-of-day and day-of-week were cyclically encoded to represent temporal patterns. The target was defined as the next-hour CO concentration, CO(t+1).

3.4 Model Development and Evaluation

Ten regression models were evaluated: Linear Regression, Ridge, Lasso, Elastic Net, Kernel Ridge (degree-2 polynomial), Support Vector Regression (SVR), Random Forest, XGBoost, LightGBM, and CatBoost. Input features for the linear and kernel-based models were scaled using RobustScaler, where IQR is the interquartile range:

$$X_{\text{scaled}} = \frac{X - \text{Median}(X)}{\text{IQR}} \quad (1)$$

Model performance was evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and the coefficient of determination (R^2):

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|Y_i - \hat{Y}_i\right|, \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \hat{Y}_i\right)^2}, \quad (3)$$

$$R^2 = 1 - \frac{\text{SS}_{\text{res}}}{\text{SS}_{\text{tot}}}, \quad (4)$$

where N is the number of observations, Y_i is the actual value, $\hat{Y}_i$ is the predicted value, SS_{res} is the residual sum of squares, and SS_{tot} is the total sum of squares. The training set was further divided into 80% training and 20% validation subsets. CatBoost, LightGBM, XGBoost, and

Random Forest were tuned using grid search based on validation RMSE. The selected configurations were then used to retrain the models on the complete training set, and their performance was evaluated on the held-out test set.

4 RESULTS AND DISCUSSION

Table 1 summarizes the test-set performance of all ten models evaluated. Tree-based ensemble models generally outperformed the linear and kernel-based models in this dataset, with CatBoost achieving the best performance with an R^2 of 0.7719 and an RMSE of 105.55 ppb.

To ensure a fair comparison, CatBoost, LightGBM, XGBoost, and Random Forest were tuned using the same chronological validation set. As shown in Table 2, tuned CatBoost achieved the best performance (R^2 = 0.7728, RMSE = 105.33 ppb, MAE = 76.33 ppb), with

Table 1: Performance comparison of the evaluated regression models

Model	MAE (ppb)	RMSE (ppb)	R^2
CatBoost	77.608	105.548	0.7719
LightGBM	79.550	108.564	0.7586
XGBoost	79.327	109.865	0.7528
Random Forest	88.437	117.467	0.7174
Linear Regression	87.148	122.433	0.6930
ElasticNet	87.217	122.498	0.6927
Lasso (L1) Regression	87.246	122.534	0.6925
Ridge (L2) Regression	87.402	122.622	0.6921
Kernel Ridge (d2) Regression	102.853	141.678	0.5889
SVR	112.469	151.423	0.5304

Table 2: Performance comparison of the tuned regression models

Model	MAE (ppb)	RMSE (ppb)	R^2
CatBoost	76.330	105.330	0.7728
LightGBM	77.055	106.814	0.7663
XGBoost	78.826	107.380	0.7639
Random Forest	82.379	112.460	0.7410

RMSE reductions of 1.48, 2.05, and 7.13 ppb compared with LightGBM, XGBoost, and Random Forest, respectively.

Figure 1 shows that the CatBoost predictions generally followed the observed CO concentrations during the test period. However, some high-concentration peaks were underestimated, suggesting that short-lived extreme pollution events may be challenging to predict using the three-month dataset.

As shown in Figure 2, CO(t) was the most influential predictor, followed by NOx concentration and CO(t-1), while hour_sin and hour_cos also showed notable importance. This indicates that recent CO concentrations and temporal patterns contributed substantially to the CatBoost predictions for next-hour CO.

The better performance of tree-based ensembles may reflect their ability to capture nonlinear relationships and interactions among CO levels, co-pollutants, meteorological variables, and time of day. The underestimation of high

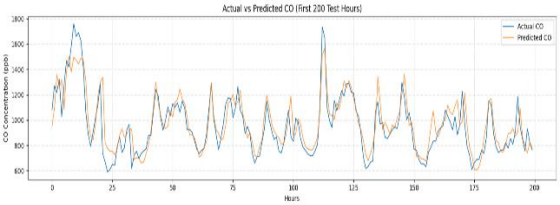


Figure 1: Actual vs Predicted CO Concentration

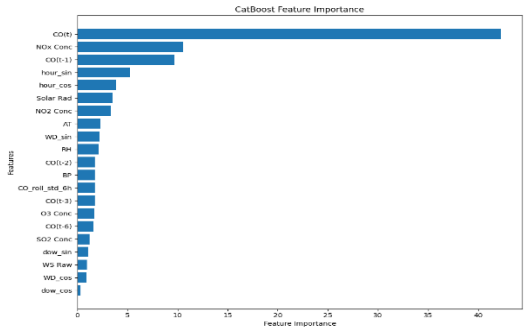


Figure 2: CatBoost Feature Importance

peaks may be related to their rarity in the three-month dataset and the difficulty of predicting extreme observations from limited training examples.

These findings are specific to a three-month dataset from a single monitoring station and may not generalize to other periods or locations. The absence of traffic and direct emission-related variables also limits the representation of factors influencing CO concentrations.

5 CONCLUSION

This study evaluated ten machine learning models for next-hour CO prediction in Kandy, Sri Lanka, using three months of hourly meteorological and air quality data from January to March 2026. Tree-based ensembles generally outperformed linear and kernel-based models, with tuned CatBoost achieving the best performance ($R^2 = 0.7728$, RMSE = 105.33 ppb, MAE = 76.33 ppb), although its advantage over LightGBM and XGBoost was small. These results indicate the potential of machine learning for short-term CO prediction.

The findings are limited by the short three-month monitoring period, the use of a single monitoring station, and the absence of traffic and other emission-related variables. Future studies should use longer datasets, include additional emission-related factors, and consider multiple monitoring locations to assess the robustness and generalizability of these results.

REFERENCES

- Abdul-Wahab, S. A., Al-Rubiei, R., & Al-Shamsi, A. (2003). A statistical model for predicting carbon monoxide levels. *International Journal of Environment and Pollution*, *19*(3), 209-224.
- Harun, H., Nsor, M., Akpan, E. E., Ogwo-Ude, U., Ogochukwu, G. N., George, B. W., & Tettey, J. N. A. (2025). Time-series forecasting of urban air pollutant levels using deep learning models. *PriMera Scientific Engineering*, *7*(4), 18–28.
- Masoudi, M., & Gerami, S. (2017). Status of CO as an air pollutant and its prediction, using meteorological parameters in Esfahan, Iran. *Pollution*, *3*(4), 527–537.
- Mohammad Abdullah Almubaidin, Nur, Sarmad Dashti Latif, Ali Najah Ahmed, Hayana Dullah, El-Shafie, A., & Sonne, C. (2024). MACHINE LEARNING PREDICTIONS FOR CARBON MONOXIDE LEVELS IN URBAN ENVIRONMENTS. *Results in Engineering*, *22*, 102114–102114.
- R. Anjali. (2022). Impact of Variations in Meteorological Parameters on Carbon Monoxide. *Research & Reviews: Journal of Physics*, *11*(1), 7–15.
- Sarmad Dashti Latif, Mustafa Almalayih, Ayman Yafouz, Ali Najah Ahmed, Nur'atiah Zaini, Irwan, D., Nouar AlDahoul, Sherif, M., & El-Shafie, A. (2023). Prediction of atmospheric carbon monoxide concentration utilizing different machine learning algorithms: A case study in Kuala Lumpur, Malaysia. *Environmental Technology and Innovation*, *32*, 103387–103387.
- Wang, Z. (2025). Research on Prediction of Air Quality CO Concentration Based on Python Machine Learning. *Advances in Internet of Things*, *15*(04), 87–95.

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research –2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Emotion-Adaptive Gameplay Using Multimodal Emotion Recognition

Gunasekara BSA[1], Arudchelvam T[2]

[1,2]*Department of Computing and Information Systems, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*

buddhimalsampath2001@gmail.com[1] , arul@wyb.ac.lk[2]

ABSTRACT

Traditional video games regulate difficulty through static curves or manual input, leaving the player's real-time emotional state unaddressed. Camera and microphone sensing offers a non-invasive alternative to physiological sensors for addressing this gap. This paper presents the design and technical evaluation of a prototype Two-Tier Emotion-Adaptive Gameplay system that integrates multimodal emotion recognition with an Unreal Engine 5 game. Facial expressions are classified by a Mini-Xception-based convolutional network, and vocal cues by a convolutional network operating on Mel-spectrograms. Their probability outputs are combined through confidence-weighted decision-level fusion and streamed to the game over Open Sound Control. A Micro-Adaptation layer adjusts entity speed and color, while a Macro-Adaptation layer evaluates a two-minute rolling emotional history to recommend calming or stimulating levels when negative states persist. On held-out validation data, the facial and vocal models, trained on public datasets, reached accuracies of 58.1% and 55.0% across six emotion classes. End-to-end latency was 1.57 s, and scripted tests confirmed that the Macro-Adaptation layer produced the expected recommendation for three emotion profiles. The effect of the adaptations on players was not evaluated. These results suggest that camera- and microphone-based emotion recognition can feasibly be integrated into a game engine, and they provide a baseline for a controlled player study.

KEYWORDS: Affective Computing, Dynamic Difficulty Adjustment, Multimodal Emotion Recognition, Unreal Engine.

1 INTRODUCTION

Player emotion is central to engagement, immersion, and well-being in interactive media, yet conventional games rely almost exclusively on explicit physical input to drive difficulty and pacing. Emotion recognition has been proposed to close this gap, but systems relying on a single modality - either facial expression or vocal prosody - are frequently unreliable under real gaming conditions: poor lighting and occlusion degrade Facial Emotion Recognition (FER), while game audio and background noise degrade Speech Emotion Recognition (SER) (Chen, Hirschberg, & Tsao, 2023).

To address these limitations, this paper presents the design and technical evaluation of a prototype Two-Tier Emotion-Adaptive Gameplay system that fuses facial and vocal emotion estimates using confidence-weighted decision-level fusion and couples the result to an Unreal Engine 5 (UE5) environment through an Open Sound Control (OSC) bridge. A Micro-Adaptation layer continuously adjusts gameplay variables, and a Macro-Adaptation layer evaluates a two-minute emotional history to recommend alternative levels. The objectives are to (i) implement a real-time multimodal recognition pipeline using a webcam and microphone, (ii) bridge the Python backend and UE5, (iii) implement and test the Two-Tier strategy, and (iv) evaluate

the recognition accuracy and end-to-end latency of the system.

2 LITERATURE REVIEW

FER using convolutional neural networks (CNNs) is accurate under controlled conditions but degrades with poor lighting, head-pose variation, and occlusion (Mollahosseini, Hasani, & Mahoor, 2019). Mini-Xception (Arriaga, Valdenegro-Toro, & Plöger, 2019), built on depthwise separable convolutions (Chollet, 2017), was designed for real-time emotion classification with few parameters. SER systems typically apply convolutional or recurrent networks to Mel-spectrograms and are similarly vulnerable to game audio and hardware noise (Chen et al., 2023).

Multimodal fusion is commonly grouped into early (feature-level), late (decision-level), and hybrid strategies (Lian et al., 2023). Feature-level fusion is computationally expensive and sensitive to temporal misalignment between audio and video, so decision-level fusion, which combines independent networks' outputs after inference, is generally preferred for real-time use. Weighting each modality by its confidence reduces the influence of an unreliable input.

Within games, Dynamic Difficulty Adjustment (DDA) has previously relied on invasive physiological sensors such as heart-rate variability, electroencephalography, or galvanic skin response (Oner & Kucukkulahli, 2025); while accurate, such sensors hinder mainstream adoption. Facial expression analysis has been used to adapt arcade-game difficulty during play (Blom et al., 2019), and recent work combined it with performance metrics to adjust difficulty and genre, noting multimodal integration as a next step (Saga & Sumi, 2026). Cameras and microphones have also been used to collect affective gameplay data (Bassano et al., 2019; Fanourakis & Chanel, 2025), and altering entity speed and ambient color can heighten immersion (Kutt, Ściga, & Nalepa, 2023). Integrating a multimodal deep-learning pipeline directly into a 3D game engine via OSC remains largely unaddressed, motivating this work.

3 METHODOLOGY

The system separates computation into a Python recognition backend and a UE5 frontend, connected asynchronously over OSC/UDP so that model inference does not block the game loop.

3.1 Visual modality

Webcam frames are converted to grayscale, and the face is located with OpenCV's Haar cascade detector, resized to 48×48 pixels, and classified into six emotions (angry, fear, happy, neutral, sad, surprise) by a Mini-Xception-based network of four residual modules of depth wise separable convolutions (16–128 filters), global average pooling, dropout (0.5), and a softmax output. It was trained on a FER-2013-derived face dataset (Goodfellow et al., 2015) from Kaggle, using 3,000 images per class (18,000 in total), an 80/20 train/validation split, flip, zoom, and rotation augmentation, and early stopping.

3.2 Acoustic modality

The player's voice is recorded in consecutive 2-second windows at 16 kHz, and silent windows are skipped. After trimming silence, a 128-band Mel-spectrogram is computed in decibels, scaled to 0–255, and resized to a 48×48 image, which the same Mini-Xception-

based architecture classifies into the same six classes. It was trained on the speech recordings of RAVDESS (Livingstone & Russo, 2018): 1,248 clips from 24 actors, with "calm" merged into "neutral", "disgust" excluded, and an 80/20 random split.

3.3 Confidence-weighted fusion

The two probability distributions are combined at the decision level, as in Equation (1),

$$P_{fused}(i) = W_v . P_v(i) + W_a . P_a(i) \quad (1)$$

where $P_v^{(i)}$ and $P_a^{(i)}$ are the visual and acoustic probabilities of class i, and the weights follow Equation (2),

$$W_v = C_v/(C_v + C_a) \quad (2)$$

$$W_a = 1 - W_v \quad (3)$$

where c_v and c_a are the highest-class probabilities (confidences) of the two models. If no face is detected or the player is silent, that modality's confidence is set to 0.05, so the other modality dominates. The class with the highest fused probability is taken as the current emotional state.

3.4 Communication bridge

After each processed webcam frame, the fused label is sent to UE5 over OSC/UDP, avoiding the handshaking overhead of HTTP or WebSocket.

3.5 Tier 1 – Micro-Adaptation

On each OSC update, positive or neutral states increase entity velocity to sustain challenge, while anger or fear reduce velocity and shift actor color toward calming tones.

3.6 Tier 2 – Macro-Adaptation

Each fused label is stored with its timestamp in a rolling 120-second history. Once at least 30 labels are available, the history is checked against fixed thresholds: anger above 50% recommends a calming "Zen Garden" level, and sadness plus neutrality above 60% recommends a high-arousal "Adrenaline" level. A non-intrusive prompt is shown rather than forcing a transition; the history is then cleared, and a 60-second cooldown follows.

4 RESULTS AND DISCUSSION

4.1 Classifier performance

Table 1 reports results on the 20% held-out subsets (3,600 face images; 249 audio clips). Both models are well above chance (16.7%) but only moderately accurate. The visual model missed most fear expressions (recall 0.29), and the acoustic model was weakest on sad and happy. Because these subsets also guided checkpoint selection, and the audio split was random across clips from the same actors, the figures are probably optimistic, particularly for unseen speakers. Fusion was not benchmarked against the single modalities, so its benefit remains to be established.

Table 1: Classifier performance on held-out validation data

Metric	Visual (n=3,600)	Acoustic (n=249)
Accuracy	0.581	0.550
Macro F1	0.574	0.537
Angry precision	0.49	0.71
Angry recall	0.62	0.41
F1: angry	0.55	0.52
F1: fear	0.36	0.66
F1: happy	0.76	0.41
F1: neutral	0.55	0.66
F1: sad	0.47	0.37
F1: surprise	0.75	0.61

4.2 System latency

End-to-end latency was measured from synchronized 30 fps video of the player's face and the monitor (33.33 ms per frame): 47 frames, or 1566.51 ms (1.57 s), elapsed between the onset of an emotional shift and the in-game adaptation. This delay combines camera capture, inference, network transmission, and the game's update cycle (spawn timer 2.0 s); a per-stage breakdown was not measured. It suits gradual changes to game variables, not frame-level reactions.

4.3 Macro-Adaptation test

The recommender was tested with three scripted 120-second emotion streams: a brief spike (16% angry, expected to be ignored), sustained frustration (54% angry, expected to trigger Zen Garden), and sustained boredom (62% sad/neutral, expected to trigger Adrenaline). All three produced the expected output, verifying the threshold logic but not the recognition accuracy.

4.4 Discussion and limitations

The pipeline runs end to end from webcam and microphone to in-game behaviour. The angry class drives the calming recommendation, yet about half of the visual model's "angry" predictions are wrong (precision 0.49), so a misclassified stretch could trigger a false recommendation; the 120-second window, non-forced prompt, and cooldown limit but do not remove this risk. Four limitations apply: the effect of the adaptations on players was not evaluated, so no claim is made that they regulate mood; the benefit of fusion over either modality alone was not benchmarked; the models were trained on web-sourced faces and acted studio speech, not gameplay data; and latency was not broken down by stage.

5 CONCLUSION

This paper presented the design and technical evaluation of a prototype Two-Tier Emotion-Adaptive Gameplay system that fuses facial and vocal emotion estimates through confidence-weighted decision-level fusion and couples the result to an Unreal Engine 5 environment via an OSC bridge. The pipeline ran from a webcam and microphone with an end-to-end latency of 1.57 s, the models reached moderate accuracies (58.1% and 55.0%), and the Macro-Adaptation layer behaved as specified in scripted tests, showing that multimodal emotion-adaptive gameplay is technically feasible. The effect of the adaptations on players and the benefit of fusion over single modalities remain open. Future work will benchmark the fusion against single modalities, evaluate the models on unseen speakers and in real gameplay, run a controlled study with a larger sample and self-reported measures, and extend the adaptations to dynamic lighting, spatial audio, and adaptive NPC behaviour.

REFERENCES

- Arriaga, O., Valdenegro-Toro, M., & Plöger, P. G. (2019). Real-time convolutional neural networks for emotion and gender classification. In *Proceedings of the 27th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2019)* (pp. 221–226). i6doc.
- Bassano, C., Ballestin, G., Ceccaldi, E., Larradet, F. I., Mancini, M., Volta, E., & Niewiadomski, R. (2019). A VR game-based system for multimodal emotion

data collection. In *Proceedings of MIG '19: Motion, Interaction and Games* (Article 38). Association for Computing Machinery.

- Blom, P. M., Methorst, S., Bakkes, S., & Spronck, P. (2019). Modeling and adjusting in-game difficulty based on facial expression analysis. *Entertainment Computing, 31,* Article 100307. https://doi.org/10.1016/j.entcom.2019.100307
- Chen, Y.-W., Hirschberg, J., & Tsao, Y. (2023). Noise robust speech emotion recognition with signal-to-noise ratio adapting speech enhancement. *arXiv*. https://arxiv.org/abs/2309.01164
- Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1251–1258). IEEE.
- Fanourakis, M., & Chanel, G. (2025). AMuCS: Affective multimodal Counter-Strike video game dataset. *Scientific Data, 12,* Article 1325.
- Goodfellow, I. J., Erhan, D., Carrier, P. L., Courville, A., Mirza, M., Hamner, B., Cukierski, W., Tang, Y., Thaler, D., Lee, D.-H., Zhou, Y., Ramaiah, C., Feng, F., Li, R., Wang, X., Athanasakis, D., Shawe-Taylor, J., Milakov, M., Park, J., … Bengio, Y. (2015). Challenges in representation learning: A report on three machine learning contests. *Neural Networks*, *64*, 59–63.
- Kutt, K., Ściga, Ł., & Nalepa, G. J. (2023). Emotion-based dynamic difficulty adjustment in video games. In *2023 IEEE 10th International Conference on Data Science and Advanced Analytics (DSAA)* (pp. 1–5). IEEE.
- Lian, H., Lu, C., Li, S., Zhao, Y., Tang, C., & Zong, Y. (2023). A survey of deep learning-based multimodal emotion recognition: Speech, text, and face. *Entropy*, *25*(10), Article 1440. https://doi.org/10.3390/e25101440
- Livingstone, S. R., & Russo, F. A. (2018). The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. *PLOS ONE, 13*(5), Article e0196391.
- Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2019). AffectNet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing, 10*(1), 18–31.
- Oner, A., & Kucukkulahli, E. (2025). Evaluating player stress and motivation through biofeedback-controlled dynamic difficulty adjustment. *Electronics, 14*(19), Article 3870.
- Saga, K., & Sumi, K. (2026). Emotion-driven adaptive game system: Design and evaluation. In *Proceedings of the Workshop on Computation: Theory and Practice (WCTP 2025)*. Atlantis Press. https://doi.org/10.2991/978-94-6239-638-8_14

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

MediTrack: A Context-Aware Mobile Health Application for Medication Support and Emergency Medical Information Access

Amarasinghe KKG[1], Perera KTDTP[2], Vidanagama VGTN[3]

[1,2,3] *Department of Computing and Information Systems, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*

kkgamarasinghe@gmail.com[1], tashmiprathibha@gmail.com[2] , tharinda@wyb.ac.lk[3]

ABSTRACT

Medication non-adherence, fragmented personal health records, and delayed access to essential information can complicate both routine self-care and emergency response. This paper presents MediTrack, a context-aware mobile health application designed to bring medication support, medical-record management, and emergency assistance into one patient-controlled platform. The system was implemented with Flutter and uses Firebase services for authentication and synchronized storage. Its principal functions include scheduled medication reminders, optical character recognition (OCR) for uploaded documents, multilingual interfaces in English, Sinhala, and Tamil, quick response (QR)-based access to selected emergency information, location-aware hospital check-in, and smartphone-sensor fall detection. Fall events are treated as suspected incidents rather than confirmed emergencies: accelerometer and gyroscope samples are evaluated as a temporal sequence comprising reduced acceleration, impact, and post-impact stillness, followed by a 20-second cancellation window before automatic emergency escalation. A preliminary evaluation comprised 48 usability task sessions with 8 participants and 32 controlled sensor trials on three Android devices. The prototype achieved an overall task completion rate of 91.7% (95% confidence interval (CI): [80.0%, 97.7%]), a median System Usability Scale (SUS) score of 82.5, and reminder delivery reliability of 97.9%. In controlled sensor trials, the fall detection algorithm achieved a sensitivity of 87.5% (95% CI: [61.7%, 98.4%]), a specificity of 93.8% (95% CI: [69.8%, 99.8%]). These findings show technical feasibility without wearable hardware, with some device-dependent variation. Before clinical deployment, the system requires validation with a representative participant sample, security review, and calibration across a wider range of devices and real-world conditions.

KEYWORDS: Digital Health, Emergency Support, Fall Detection, Medication Adherence, Mobile Application

1 INTRODUCTION

Long-term treatment often depends on actions performed by patients outside a clinical setting. Forgetting doses, keeping records in multiple locations, or being unable to communicate allergies and medication details during an emergency can reduce continuity of care. The World Health Organization identifies adherence as a multidimensional problem influenced by health-system, therapy, condition, socioeconomic, and patient-related factors (World Health Organization, 2003). Mobile interventions cannot remove all of these barriers, but they can place reminders, records, and communication tools on a device that is already carried by many users (Moon and Walsh, 2025).

Medication applications commonly focus on scheduling and notification. Existing research suggests that smartphone tools can support medication self-management, although application quality and functionality vary considerably (Lanke et al., 2025; Moon and Walsh, 2025). Emergency applications often treat fall detection,

location sharing, and medical-ID access separately. MediTrack explores whether these functions can be combined in a coherent, patient-controlled workflow.

The contribution of this work is threefold: (1) an integrated architecture for medication reminders, health records, and emergency information; (2) a context-aware workflow that connects fall sensing and hospital proximity with user confirmation; and (3) a reproducible evaluation protocol that distinguishes implemented functionality from outcomes that still require empirical measurement.

2 SYSTEM DESIGN AND IMPLEMENTATION

2.1 Architecture

MediTrack uses a layered client–service architecture implemented in Flutter. Application services coordinate authentication, notifications, optical character recognition (OCR), geolocation, quick response (QR) access, and records. Firebase Authentication handles user identity, Cloud Firestore stores structured data, and Firebase Storage holds files. Hospital discovery uses OpenStreetMap through the Overpass interface, with local caching for offline use.

Figure 1 summarizes the major data flows. Location monitoring, notifications, camera access, and record uploads are activated only through the relevant user workflows.

2.2 Medication and Record Support

Users can create active medication entries and schedule local notifications. For supported images, on-device OCR extracts Latin-script text that can be stored with the record to improve later retrieval. OCR output is treated as assistive text and must be verified against the original document.

2.3 Emergency and Context-Aware Support

The emergency profile contains information such as blood group, allergies, and emergency contacts. A QR workflow produces a link and fallback code for sharing permitted information. QR sharing can be disabled in the privacy settings. The hospital check-in module supports two modes: user-selected trusted hospitals and nearby mapped hospitals. An entry is recorded only when the previous position was outside the configured radius and the new one is inside, suppressing duplicate check-ins.

The fall detector samples acceleration and angular velocity at a target interval of 20 ms. For acceleration components $a_x(t)$, $a_y(t)$, and $a_z(t)$, the Signal Vector Magnitude (SVM) is evaluated as

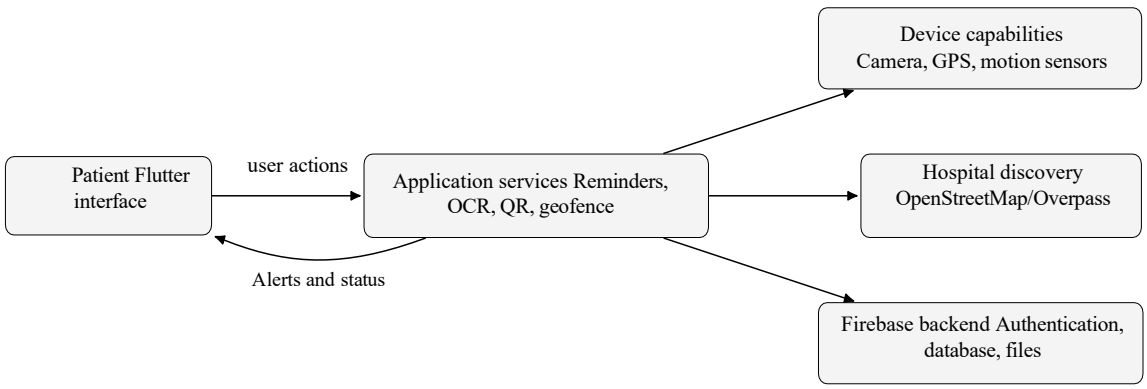


Figure 1: High-level architecture of the MediTrack prototype.

$$A(t)=\sqrt{a_x(t)^2+a_y(t)^2+a_z(t)^2} \quad (1)$$

Equation (1) is evaluated alongside gyroscope angular velocity $\omega(t)$through a multi-stage sequential state machine. The detector enters a fall-candidate state upon detecting either a free-fall drop ($A(t) < 5.5$ m/s^2) or a rapid rotational tumble ($\omega(t) \geq 2.2$ rad/s accompanied by an acceleration disturbance of $A(t) < 8.0$ m/s^{2}or $A(t) > 14.0$ m/s^2). An impact spike of at least $A(t) \geq 20.0$ m/s^{2}must occur within 1.2 seconds of the trigger. Following a 250 mspost-impact settling delay, the system evaluates stillness over a 1.2 second sliding window, requiring the acceleration-magnitude standard deviation to remain ≤ 1.0 m/s^{2}and mean angular velocity ≤ 1.0 rad/s, with overall confirmation concluding within 6 seconds of impact. Upon satisfying these configured criteria, a suspected fall is logged, initiating a 20 second cancellation window before automated emergency contact escalation. This threshold-based, sequential design is consistent with smartphone fall detectors reported in the literature (Stampfler et al., 2022).

2.4 Privacy and Safety Considerations

Electronic health data requires strict confidentiality and integrity controls (Tertulino et al., 2024). While MediTrack uses authenticated storage and configurable QR sharing, production deployment demands comprehensive security—including database rules, token revocation, audit logging, and encrypted transport. The prototype is a supportive tool and is not a substitute for professional medical diagnosis or emergency services.

3 METHODOLOGY

3.1 Evaluation Design

The implementation was evaluated through a mixed test programme containing automated functional tests, controlled sensor trials, location trials, and a task-based usability study. A total of 8 undergraduate students (N = 8; 4 male, 4 female, aged 21–25) from Wayamba University of Sri Lanka participated in the usability evaluation. All were active smartphone users, testing the app in Sinhala (n = 5), English (n = 2), and Tamil (n = 1). Informed consent was obtained prior to testing, and all data were anonymized.

Participants completed six core tasks: authentication, medication reminders, OCR prescription review, emergency profile setup, multilingual switching (English, Sinhala, Tamil), and hospital check-in, followed by the System Usability Scale (SUS) questionnaire (Brooke, 1996). Thirty-two sensor trials (N = 32) evaluated 16 simulated falls (forward, backward, lateral, syncope) and 16 daily activities (walking, stairs, abrupt sitting, table drops) across pocket, handheld, and bag placements on high-density safety mats. without exposing participants to actual falls. Geofence trials compared logged entry events with manually timestamped boundary crossings at two public hospital sites, testing multiple global positioning system (GPS) coordinates along the boundaries to verify entry detection.

3.2 Data Collection and Analysis

Table 1 defines the minimum measurements. For fall detection, sensitivity is TP/(TP + FN) and specificity is *TN/(TN+FP)*, where *TP, TN, FP* and *FN*

Table 1: Proposed evaluation measures

Component	Measures
Core tasks	Completion rate, time, errors
Reminders	Delivery rate, timing deviation
Fall sensing	Sensitivity, Specificity, false alarm rate (FAR)
Geofence	Entry detection, distance error, duplicates
Usability	SUS score and participant comments
Multilingual	Localization integrity

denote true positives, true negatives, false positives, and false negatives.

4 RESULTS AND DISCUSSION

Eight participants completed a total of 48 usability task sessions across three distinct device tiers: Google Pixel 7 (Android 14), Xiaomi Redmi Note 11 (Android 13 / MIUI 14), and Samsung Galaxy A07 (Android 15/ One UI 7). The overall task-completion rate was 91.7% (95% confidence interval (CI): [80.0%, 97.7%]), and the median SUS score was 82.5 (interquartile range [IQR]: 77.5–85.0). Minor OCR upload delays occurred mainly on the entry-level Samsung Galaxy A07, owing to limited processing memory.

Scheduled reminder evaluation yielded a 97.9% delivery reliability rate (47 out of 48 configured reminders triggered within ± 10 s). The single missed reminder occurred on the Xiaomi Redmi Note 11 due to aggressiveoriginal equipment manufacturer (OEM) background battery optimization (MIUI battery saver). In geofencing trials across two hospital sites (20 boundary crossings), entry detection was 100% with a mean boundary error of 11.4 ± 3.8 m. Redundant entries within 60 minutes were suppressed in 100% of trials.

To evaluate the fall detection algorithm independently of the usability tasks, 32 controlled fall and non-fall trials were conducted. Table 2 details the distribution of both the usability task sessions and the controlled sensor trials across the three testing devices.

Across the 32 controlled sensor trials, the fall detector achieved an overall sensitivity of 87.5% (14 out of 16 simulated falls detected; 95% CI: [61.7%, 98.4%]), a specificity of 93.8% (15 out of 16 daily activities correctly rejected; 95% CI: [69.8%, 99.8%]), and a FAR of 6.25% (1 out of 16). The two false negatives occurred during slow lateral sliding falls from a chair, where the impact acceleration peak failed to exceed the 20.0 m/s² threshold. The single false positive occurred on the Samsung Galaxy A07 when the phone was tossed forcefully onto a wooden desk. Sinhala, Tamil, and English labels also rendered cleanly on all tested devices.

The integrated design offers a practical advantage: medication information, uploaded records, and emergency profile data are managed through one interface. The two dashboard cards provide complementary workflows: Temporary Access limits the sharing window, while QR Access provides a reusable patient code and entry to the hospital QR scanner. Geofence records reflect location-triggered entries rather than completed QR check-in transactions. Results represent prototype feasibility rather than proof of clinical effectiveness, with generalizability bounded by sample size, scripted fall sequences, and the tested device set.

Table 2: Distribution of usability task sessions, sensor trials, and confusion matrix.

Device Model	Usability Task Sessions	Controlled Sensor Trials	TP	FN	TN	FP
Google Pixel 7	18	10	5	0	5	0
Xiaomi Redmi Note 11	18	11	5	1	5	0
Samsung Galaxy A07	12	11	4	1	5	1
Total	48	32	14	2	15	1

5 CONCLUSION

MediTrack consolidates medication reminders, health records, multilingual access, QR emergency profiles, hospital geofencing, and sensor-based fall detection into a unified mobile application without dedicated wearables. The multi-stage fall detection pipeline and 20-second cancellation window retain the patient in the decision loop. While initial feasibility was validated with 8 participants across 48 tasks and 32 sensor trials, future work requires longitudinal clinical trials with older populations, device calibration, and third-party security audits before operational deployment.

REFERENCES

- Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In Jordan, P.W., Thomas, B., McClelland, I.L., and Weerdmeester, B. (eds.), *Usability evaluation in industry* (pp. 189-194). Taylor and Francis
- Lanke, V., Trimm, K., Habib, B., and Tamblyn, R. (2025). Evaluating the effectiveness of mobile apps on medication adherence for chronic conditions: Systematic review and meta-analysis. *Journal of Medical Internet Research*, 27, e60822. https://doi.org/10.2196/60822
- Moon, Z., and Walsh, J. (2025). Digital interventions in medication adherence: A narrative review of current evidence and challenges. *Frontiers in Pharmacology*, 16, 1632474. https://doi.org/10.3389/fphar.2025.1632474
- Stampfler, T., Elgendi, M., Fletcher, R.R., and Menon, C. (2022). Fall detection using accelerometer-based smartphones: Where do we go from here? *Frontiers in Public Health*, 10, 996021. https://doi.org/10.3389/fpubh.2022.996021
- Tertulino, R., Antunes, N., and Morais, H. (2024). Privacy in electronic health records: A systematic mapping study. *Journal of Public Health*, 32(3), 435-454. https://doi.org/10.1007/s10389-022-01795-z
- World Health Organization. (2003). *Adherence to long-term therapies: Evidence for action*. World Health Organization.

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

SmartSpend: A Cross-Platform Mobile Application for Personal Finance Management

Fernando LRN[1], Liyanagamage HW[2], Vidanagama VGTN[3]
[1,2,3] *Department of Computing and Information Systems, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*
rebecanethmini6@gmail.com[1], hiruniwarsha01@gmail.com[2], tharinda@wyb.ac.lk[3]

ABSTRACT

In an increasingly fast-paced world, managing personal finances efficiently is a significant challenge for many individuals. Traditional methods of expense tracking often rely on manual data entry, which is tedious, time-consuming, and prone to human error. To address this, we propose SmartSpend, a cross-platform mobile application developed using the Flutter framework and powered by a Supabase backend. The application provides an intuitive platform for tracking expenses, managing savings goals, and visualizing financial data. A standout feature of SmartSpend is its automated PDF bill extraction capability. This Feature utilizes rule-based parsing algorithms to extract key information from digital receipts and invoices, such as merchant name, amount, and category. By automating data entry and providing real-time financial insights through interactive charts, SmartSpend empowers users to make informed financial decisions with minimal effort. This paper discusses the design, implementation, and evaluation of the SmartSpend application.

KEYWORDS: Automated Expense Tracking, Financial Data Visualization, Personal Finance Management, Rule-Based Parsing

1 INTRODUCTION

Personal finance management is a critical aspect of modern life, directly impacting an individual's financial stability and well-being. Despite its importance, many people struggle to maintain an accurate and up-to-date record of their daily expenses (Kim et al., 2022). The proliferation of digital transactions and e-receipts has created an opportunity to automate this process. However, many existing financial applications still require users to manually input transaction details or connect directly to bank accounts (Oney et al., 2019), which can raise privacy and security concerns.

SmartSpend was conceptualized to bridge the gap between manual tracking and intrusive bank integrations. The primary objective of this project is to develop a user-friendly, secure, and intelligent mobile application that simplifies expense tracking. By integrating document scanning and parsing features (such as PDF bill extraction), the app reduces the friction associated with data entry. Furthermore, the application aims to provide a holistic view of a user's financial health through interactive dashboards and goal-oriented savings trackers.

2 LITERATURE REVIEW

The domain of personal finance applications has seen significant growth over the past decade. Early applications primarily focused on manual ledger entries, acting as digital replacements for physical chequebook registers (Doe & Smith, 2023). Studies have shown that while these applications improve financial awareness, the high effort required for manual entry often leads to user fatigue and abandonment.

Recent advancements in Artificial Intelligence (AI) and Document Processing have paved the way for more automated solutions. Applications like Expensify and ReceiptLens have successfully demonstrated the utility in extracting data from images of physical receipts (Al-Quraishi & Elamvazuthi, 2020). However, these solutions are often tailored for enterprise users or require expensive subscriptions.

From a technical perspective, the rise of cross-platform frameworks like Flutter has revolutionized mobile app development, allowing for native performance on both iOS and Android from a single codebase (Wang & Chen, 2021). Coupled with Backend-as-a-Service (BaaS) platforms like Supabase, developers can now build scalable, real-time applications rapidly. SmartSpend builds upon these modern technologies to offer a robust, automated personal finance tool accessible to everyday users.

3 METHODOLOGY

The development of SmartSpend followed an Agile software development lifecycle, allowing for iterative improvements and rapid prototyping. The methodology consisted of several phases:

3.1 Requirement Analysis

Identified key user needs, including secure authentication, offline-capable data storage, automated receipt scanning, and clear visual analytics.

3.2 Technology Selection

- Frontend: Flutter (Dart) was chosen for its expressive UI capabilities, hot-reload feature, and cross-platform compatibility.
- Backend: Supabase was selected as the BaaS provider for its open-source nature, robust PostgreSQL database, and seamless authentication services.
- State Management: The Provider package was utilized for efficient and scalable state management within the Flutter app.
- PDF_Processing:The syncfusion_flutter_pdf package was chosen for reliable parsing and text extraction from PDF documents.

3.3 Prototyping & UI Design

Designed a modern, accessible interface adhering to Material Design guidelines, featuring a custom theme with dark mode support.

3.4 Implementation

Divided into sprints focusing on core modules (Auth, Dashboard, Scanner, Savings).

3.5 Testing & Deployment

Conducted unit and widget testing, followed by on-device testing across varying screen sizes.

4 SYSTEM DESIGN

The architecture of SmartSpend follows a classic client-server model, heavily utilizing cloud infrastructure for data persistence. The overall system architecture of SmartSpend, including its main layers and interactions, is illustrated in Fig 1.

4.1 Architecture Overview

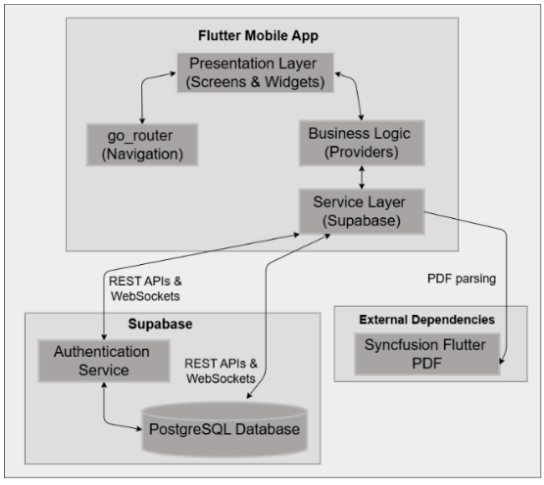


Figure 1: System Overview

- Presentation Layer: Built with Flutter widgets, responsible for rendering the UI and handling user interactions. Navigation is managed using the go_router package for a declarative routing approach.
- Business Logic Layer: Managed by the Provider architecture. Providers (e.g., ExpenseProvider) encapsulate the logic for fetching, adding, and updating data, decoupling the UI from direct backend calls.
- Service Layer: Contains dedicated services like SupabaseService for database interactions and Service for processing PDF files and extracting text.
- Data Access Layer / Backend: Hosted on Supabase, comprising a PostgreSQL database. Scalability is supported by PostgreSQL's robust architecture and Supabase's connection pooling. Data security and user privacy are strictly enforced via Row Level Security (RLS) policies.

4.2 Database Schema

The core database consists of interconnected tables:
Users: Stores user profiles and authentication tokens.
Expenses: Records individual transactions (ID, Amount, Merchant, Category, Date, User_ID).
Savings Goals: Tracks long-term financial targets (Target Amount, Current Saved, Deadline).

5 IMPLEMENTATION

The main user interfaces of SmartSpend, showcasing its key screens and functionalities, are presented in Fig 2. The system processes the file using `syncfusion_flutter_pdf` to extract raw text. Unlike machine learning approaches,



Figure 2: SmartSpend AI UI Collage

our system employs a deterministic, rule-based custom parsing algorithm. It utilizes Regular Expressions (Regex) and heuristic keyword matching (e.g., searching for terms like "Total:", "Amount Due:", "Date:") to identify and extract the Merchant Name, Total Amount, Date, and Expense Category.

5.1 Data Visualization

The dashboard utilizes the fl_chart library to render interactive pie charts and bar graphs. These visualizations aggregate the user's expense data by category and time period (e.g., weekly or monthly spend), providing immediate visual feedback on spending habits.

5.2 Backend Integration

SupabaseService initialize the connection to the Supabase project using a secure URL and Anon Key. All CRUD (Create, Read, Update, Delete) operations for expenses and savings goals are executed asynchronously via the Supabase Flutter client, ensuring data is instantly synchronized across devices.

6 RESULTS AND DISCUSSION

The implementation of SmartSpend successfully demonstrates a highly functional, automated personal finance tracker.

6.1 Extraction Performance and Accuracy

Table 1: Overall Extraction Accuracy

Metric	Correctly Identified	Accuracy
Total PDFs Processed	25	100%
Merchant Name	24	96%
Total Amount	25	100%
Date	24	96%
Expense Items / Category	22	88%

To evaluate the rule-based parsing algorithm, a dataset of 25 digitally generated PDF invoices was collected from various common merchants (e-commerce, utilities, subscriptions). Document selection criteria focused on standard, text-based PDFs to test the intended use case. The extraction results were manually annotated and verified by the authors. Table 1 presents the overall extraction accuracy. Processing time averaged under 1.5 seconds per document on a standard mid-range smartphone.

6.2 User Experience and Acceptance

User Acceptance Testing (UAT) was conducted with a group of 15 participants to assess the system's usability. Participants were tasked with onboarding, adding manual expenses, and using the PDF scanner. Feedback was generally positive regarding UI responsiveness and the convenience of the automated extraction. Table 2 summarizes key test scenarios.

Limitations: Currently, the extraction implementation relies on parsing digital PDFs. Processing physical receipts would require integrating an image-to-text machine learning model, which is a planned future enhancement.

Table 2: User Acceptance Testing Summary

Test ID	Module	Expected Outcome	Status
UT-01	Onboarding	Onboarding completed successfully	Pass
UT-02	Authentication	User redirected to dashboard after login	Pass
UT-03	Dashboard	Data displayed with a minor delay	Minor Issue
UT-04	Expense Management	Expense added and balance updated	Pass
UT-06	Admin Authentication	Admin redirected to dashboard	Pass

Additionally, the parsing algorithm's accuracy depends on the format of the invoice, and highly unstructured documents may require manual corrections by the user.

7 CONCLUSION

SmartSpend successfully addresses the complexities of personal finance management by offering a smart, automated, and visually intuitive mobile application. By leveraging modern development frameworks like Flutter and Supabase, the project provides a foundation for a scalable and performant solution. The integration of PDF document scanning for automated expense entry significantly lowers the barrier to maintaining accurate financial

records. Future iterations of the application aim to incorporate advanced Machine Learning models for predictive budgeting, anomaly detection in spending patterns, and broader support for image-based receipt scanning, further establishing SmartSpend as a comprehensive digital financial assistant.

REFERENCES

- Al-Quraishi, M. S., & Elamvazuthi, I. (2020). Optical character recognition (OCR) for invoice processing: A review. *IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, 120–125.
- Doe, J., & Smith, A. (2023). Evolution of digital ledgers and user fatigue in financial tracking. *International Journal of Financial Technology, 12*, 45–52.
- Kim, D., Lee, M., & Park, J. (2022). The effect of real-time expense tracking on consumer spending behavior. *Journal of Consumer Affairs, 56*(3), 1120–1138.
- Oney, S., Harrison, C., & Ouyang, J. (2019). Automating personal financial management through mobile applications. *International Journal of Human-Computer Studies, 124*, 56–68.
- Wang, Y., & Chen, H. (2021). Performance evaluation of cross-platform mobile app development frameworks: A case study of Flutter and React Native. *Journal of Systems and Software, 176*, 110931

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Machine Learning for Personalized Food Allergen Risk Prediction

Pandithage SD[1], Hansika Gunasekara[2]
[1,2]*Department of Computing and Information Systems, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*
senuripandithage31@gmail.com[1], hansikag@wyb.ac.lk[2]

ABSTRACT

Food allergy risk assessment remains challenging because responses to food ingredients differ between individuals, and relevant datasets are often highly imbalanced. In this study, a machine learning approach has been developed to estimate personalized food allergen risk using self-reported clinical, demographic, dietary, and family history information collected from 90 Sri Lankan participants. A positive target has represented a participant-reported allergic reaction, sensitivity, or perceived risk, rather than a clinically confirmed allergy diagnosis. Thirty-one candidate allergen targets have been identified, and 24 with at least 2 positive observations have been retained. Candidate features have been screened using Spearman correlation strictly within training folds to prevent test-set leakage. Logistic Regression, Random Forest, and XGBoost have been compared using 5-fold stratified cross-validation with imbalance-aware training. Accuracy, precision, recall, F1-score, and PR-AUC have been evaluated. On primary targets (>=5 cases), Random Forest achieved highest accuracy (80.65%), while Logistic Regression achieved 74.18% accuracy, 43.94% recall (77.35% with learned thresholds), and 0.418 PR-AUC. Logistic Regression has therefore been selected as the most suitable model for personalized risk prediction because positive-risk detection has been prioritized over accuracy alone. Apriori association rule mining has also been applied to identify strong co-occurrence patterns among self-reported allergen responses. The strongest observed rule, Mutton → Beef, has achieved 94.4% confidence and 3.15 lift. These associations have been treated as supplementary statistical evidence rather than clinically confirmed biological cross-reactivity. The findings have demonstrated the feasibility of personalized allergen-risk prediction and highlighted the need for larger, clinically validated datasets.

KEYWORDS: Association Rule Mining, Food Allergy, Logistic Regression, Machine Learning, Personalized Risk Prediction

1 INTRODUCTION

Food allergies continue to be a significant health issue since reactions to food components can vary from mild symptoms to life-threatening anaphylaxis (Sicherer & Sampson, 2009). Tailoring risk assessments has proven difficult due to the variability in individual reactions and the frequent imbalances in food allergy datasets.

Machine learning has increasingly been investigated for patient-level allergy-risk prediction. Ahmad *et al.* (2025) compared several classifiers for food-allergy detection and reported higher accuracy for tree-based models than Logistic Regression. However, accuracy can be misleading when positive cases are less frequent than negative cases. The present study has therefore focused on self-reported allergen-risk prediction using data from 90 Sri Lankan participants, with recall and F1-score considered alongside accuracy.

The study has compared Logistic Regression, Random Forest, and XGBoost and has selected the most suitable model for personalized risk prediction. Apriori association rule mining has also been used to identify

strong co-occurrence patterns among self-reported allergen responses.

2 LITERATURE REVIEW

Ahmad *et al.* (2025) compared Decision Tree, Random Forest, K-Nearest Neighbours, Support Vector Classifier, Naive Bayes, and Logistic Regression for food-allergy detection. Their results favoured tree-based classifiers in terms of accuracy, demonstrating that algorithm performance can vary across datasets. Landau *et al.* (2024) also demonstrated the use of patient-level electronic medical record data for food-allergy risk stratification and compared Logistic Regression with Random Forest. Their work supports the use of interpretable machine learning models for individualized risk assessment.

Association-rule mining provides a complementary approach to supervised classification. Apriori, introduced by Agrawal and Srikant (1994), identifies frequent item sets and association rules from transactional data. In the present study, it has been applied to allergen-response data to identify recurring co-occurrence patterns that can supplement the predictive model.

3 METHODOLOGY

3.1 Dataset and Target Definition

A structured, anonymized questionnaire has been used to collect demographic, dietary, personal-allergy, medical-condition, and family-history information from 90 Sri Lankan participants. The data represent self-reported experiences rather than clinically confirmed allergy diagnoses. Thirty-one candidate food or food-related targets have been identified. A value greater than zero has been encoded as 1 to represent a participant-reported allergic reaction, sensitivity, or perceived risk to that ingredient, while zero has been encoded as 0. Targets with fewer than two positive observations have been excluded, leaving 24 eligible targets for modelling.

3.2 Feature Preparation and Selection

Candidate predictors have included age, gender, and personal allergy history, frequency of outside food, lactose intolerance, medical conditions, family history, and family-reported allergy patterns. One-hot-encoded variables for province, blood type, dietary pattern, and work environment have also been included. Categorical variables have been converted using binary, ordinal, and one-hot encoding. Multi-response variables have been transformed into binary indicators, and continuous variables have been normalized using MinMaxScaler. The feature space contained 51 candidate predictors. Feature screening was performed per target using Spearman correlation (|rho| >= 0.15), retaining an average of 10.0 features per target without a universal pool. Feature selection and age normalization were conducted strictly inside each training fold, preventing test-set leakage.

3.3 Model Development and Evaluation

Stratified cross-validation was performed for each eligible target (5-fold CV for 17 primary targets with >=5 cases; adaptive folds for 7 exploratory targets with 2-4 cases). Logistic Regression, Random Forest, and XGBoost have been compared. Logistic Regression has been configured with balanced class weights, a maximum of 500 iterations, and random state 42.

Random Forest has been configured with 100 estimators, a maximum depth of 5, balanced class weights, and a random state of 42. XGBoost has been configured with 100 estimators, a maximum depth of 3, a learning rate of 0.05, and a random state of 42. SMOTE has been applied to XGBoost when at least four positive cases are available; otherwise, scale_pos_weight has been used. Accuracy, precision, recall, F1-score, PR-AUC, and Brier score were evaluated out-of-fold. Recall-oriented decision thresholds (F2 objective) were learned from out-of-fold validation predictions to minimize false negatives.

3.4 Association Rule Mining Using Apriori

Apriori association rule mining has been applied to the Boolean allergen-response matrix, following Agrawal and Srikant (1994). A minimum support of 0.10 has been used, requiring an itemset to occur in at least 10% of participants. Rule length has been restricted to two items, and rules with lift of at least 2.0 have been retained. Support, confidence, and lift have been used to interpret the rules. The resulting associations have been treated as statistical co-occurrence patterns within the study population. They have not been interpreted as proof of biological cross-reactivity or clinical allergy.

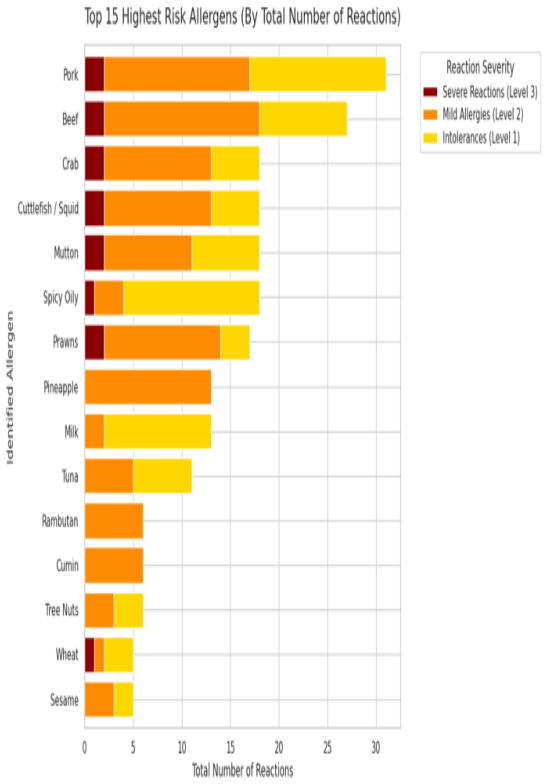


Figure 1: Most Frequently Reported Positive Food-Related Responses

4 RESULTS

4.1 Dataset Characteristics

90 Sri Lankan participants have been included in the dataset, and 21 have reported a pre-existing food allergy. The most frequently reported positive food-related responses have been pork, beef, crab, cuttlefish/squid, mutton, spicy oily foods, prawns, pineapple, milk, and tuna.

These values describe the study sample and have not been treated as population prevalence estimates.

Table 1: Mean Performance of the Evaluated Models

Model	Accuracy	Precision	Recall	F1-score	PR-AUC
Logistic Regression	74.18%	23.98%	43.94%	29.15%	0.418
Random Forest	80.65%	20.45%	22.54%	19.79%	0.398
XGBoost	72.75 %	23.65 %	46.75 %	28.94 %	0.385

Table 2: Strongest Pairwise Associations Identified Using Apriori

Antecedent	Consequent	Confidence	Lift	Support
Mutton	Beef	94.4%	3.15	18.9%
Mutton	Pork	94.4%	2.74	18.9%
Beef	Pork	85.2%	2.47	25.6%
Cuttlefish / Squid	Crab	83.3%	4.17	16.7%
Prawns	Cuttlefish / Squid	76.5%	3.82	14.4%

4.2 Model Performance

On primary targets, Random Forest achieved the highest accuracy (80.65%), while Logistic Regression achieved 74.18% accuracy, 43.94% recall, and highest PR-AUC (0.418). Applying out-of-fold learned F2 thresholds increased Logistic Regression recall to 77.35%. Across all 24 allergens, macro accuracy was 84.95% for Random Forest and 78.69% for Logistic Regression (33.21% default recall, 66.60% optimal recall). Logistic Regression was chosen because clinical sensitivity and PR-AUC were prioritized over raw accuracy.

4.3 Apriori Association Rules

Mutton → Beef and Mutton → Pork have achieved the highest confidence at 94.4%. Cuttlefish/Squid → Crab has produced the highest lift among the reported rules at 4.17. These results have indicated strong associations within the sample, but they have not established biological cross-reactivity.

5 DISCUSSION

The results have shown that the most accurate classifier is not necessarily the most suitable for personalized risk detection. Random Forest has achieved the highest mean accuracy, whereas Logistic Regression has achieved the highest mean recall and F1-score. This distinction has been important because the positive class has been comparatively rare. Accuracy alone could therefore give an overly favourable impression of performance.

The findings have differed from the accuracy-focused results reported by Ahmad *et al.* (2025), where tree-based classifiers achieved strong accuracy. The present findings have not demonstrated that tree-based methods are generally inferior. Instead, they have shown that model suitability has depended on the dataset, class distribution, and intended evaluation objective. In this dataset, class-weighted Logistic Regression provided the strongest precision-recall trade-off (PR-AUC 0.418, optimal recall 77.35%) and was therefore selected.

The result has also been consistent with the value of interpretable models identified by Landau *et al.* (2024). Logistic Regression has provided a comparatively simple modelling structure while achieving the strongest positive-case detection among the evaluated approaches.

Apriori has provided information that differs from supervised prediction. Logistic Regression has used participant characteristics to estimate allergen-specific risk, whereas Apriori has identified recurring relationships between self-reported allergen responses. The Mutton → Beef rule has shown a strong statistical relationship in the survey data, but it has not been sufficient to establish immunological cross-reactivity. The Apriori component has therefore been treated as supplementary evidence rather than a diagnostic mechanism.

6 LIMITATIONS AND FUTURE WORK

The study has several limitations. The sample size has been limited to 90 participants, while multiple features and allergen targets have been modelled. Several positive classes have also contained few observations, which has limited model stability. The target labels have been self-reported and do not represent clinically confirmed allergy diagnoses.

Although 5-fold stratified cross-validation, calibration, and PR curves were evaluated, external validation on an independent cohort is necessary. Longitudinal cohorts with clinical food challenges are also warranted. The Apriori rules have been derived from the same small survey dataset. Larger datasets and independent clinical or immunological evidence will be required before individual associations can be interpreted as biological cross-reactivity.

7 CONCLUSION

A personalized food allergen-risk prediction approach has been developed using self-reported data from 90 Sri Lankan participants. Comparing classifiers via stratified cross-validation, Logistic Regression achieved 43.94% recall (77.35% with learned thresholds) and 0.418 PR-AUC on primary targets, making it the most suitable model for personalized risk screening. Apriori has additionally identified strong co-occurrence patterns among self-reported allergen responses, providing supplementary evidence rather than clinically confirmed cross-reactivity. Larger datasets, clinically validated labels, and independent validation will be required before clinical use.

REFERENCES

- Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules in large databases. In *Proceedings of the 20th International Conference on Very Large Data Bases* (pp. 487–499). Morgan Kaufmann.
- Ahmad, Z., Farooq, A., Fuzail, M., Aziz, Y., & Aslam, N. (2025). Food allergy detection using a machine learning approach. *Kashf Journal of Multidisciplinary Research, 2*(4), 116–127.
- Landau, T., Gamrasni, K., Barlev, Y., Elizur, A., Benor, S., Mimouni, F., & Brandwein, M. (2024). A machine learning approach for stratifying risk for food allergies utilizing electronic medical record data. *Allergy*, *79*(2), 499–502.
- Sicherer, S. H., & Sampson, H. A. (2009). Food allergy: Recent advances in pathophysiology and treatment. *Annual Review of Medicine, 60*(1), 261–277.

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

PearlWay: A Context-Aware Mobile Application for Beach Recommendations in Sri Lanka

Daranagama DADA[1], Kulathunga MGAD[2], Vidanagama VGTN[3]
[1,2,3]*Department of Computing and Information Systems, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*
dasuniarchana02@gmail.com[1] , kulathungadileka@gmail.com[2] , tharinda@wyb.ac.lk[3]

ABSTRACT

Sri Lanka is a country which has more than 1340km of coastline and is visited annually by thousands of locals and international visitors. Tourists often face problems like difficulty accessing up-to-date safety data, lack of fully integrated tourism services and the inability to choose an appropriate beach location according to weather and individual preferences. The current tourism applications offer mostly static location information without taking into account contextual or personalized recommendations. In this research, context-aware mobile Geographic Information System (GIS) is proposed as a combination of Location-Based Services (LBS), real-time weather information, rule-based decision-making, and intelligent recommendation techniques to enrich the tourism experience in coastal areas. By taking user location, travel dates, weather conditions, monsoon seasons and preferences for activities into account, the system supports awareness of environmental conditions and provides personalized travel guidance from one mobile platform. Preliminary manual functional checks were conducted on five mid-range Android smartphones. Beach selection, nearby-beach display using the current location, and beach temperature display worked on these devices. These observations successfully demonstrate preliminary operation of selected functions. The recommendation engine ranks beaches according to users' selected activities, while a rule-based suitability score incorporates regional monsoon periods and forecast weather conditions. The application is intended to support context-informed beach selection rather than replace official coastal safety advice.

KEYWORDS: Coastal Safety, Context Awareness, Location-Based Services, Mobile GIS, Smart Tourism

1 INTRODUCTION

Tourism is one of the biggest sectors in the Sri Lankan economy and coastal tourism is one of the most appealing sectors in the country. Sri Lanka has numerous beaches which are good for surfing, snorkelling, diving, whale watching, family outings, and eco-tourism.

This research presents a context-aware coastal tourism platform, which is developed specifically for Sri Lanka, PearlWay.

2 LITERATURE REVIEW

Context-aware tourism applications rely on user information, including location, time, environment, and personal preferences, to offer adaptive and personalized services. This approach builds on established work on context-aware computing (Dey, 2001; Schilit et al., 1994). The proposed PearlWay system is informed by the following literature.

2.1 Location-Based Mobile Tourist Guide

Curran and Smith (2006) suggested a mobile tourist guide that would be based on location and use the Global Positioning System (GPS) technology to help tourists reach a new destination. A system that automatically learns where the user is and offers nearby points of interest as well as navigation instructions. The proposed Pearl

Way system uses a similar location-aware approach utilizing GPS to find nearby beaches, restaurants, hospitals, hotels and emergency services.

2.2 LOCUS: Personalized Tourism Recommendation System

AlSaeed (2023) presented a context-aware tourism recommender system, called LOCUS, which provides personalised recommendations by considering several context elements. In contrast to traditional tourism applications, which rely exclusively on user location, LOCUS incorporates three key context factors: user location, personal preferences, and past travel behavior. The system suggests attractions in accordance with these contextual factors and the optimal match with users' interests and travel patterns.

2.3 Mobile Tourism Application Typology

Gavalas et al. (2014) reviewed mobile recommender systems in tourism. Additionally, aspects such as usability, offline capabilities, ratings and user reviews play a significant role in user adoption and satisfaction.

Mateos and Bellogín (2025) reviewed 90 context-aware recommender studies, including datasets and evaluation. This motivates independent ranking evaluation for PearlWay; the effectiveness of location and weather integration requires further evaluation.

2.4 Research Gap

Previous research has proved the effectiveness of GPS based navigation and personalized recommendation systems, but there are still some limitations. The environment, including the weather, seasonal monsoon, high tides and coastal hazards, is important for beach tourism.

These are addressed by the proposed PearlWay platform, which links Location-Based Services (LBS), Mobile GIS, real-time weather information, rule-based coastal safety assessment, intelligent trip planning, emergency communication and planned offline mapping within a context-aware mobile application.

3 METHODOLOGY

The methodology contains six main modules, which include user context acquisition, activity recommendation, seasonal risk assessment, real-time weather integration, distance calculation, and emergency support.

3.1 User Context Acquisition

In order to acquire contextual information of the user and surrounding environment, the application collects in real-time user's GPS location, current system time, selected travel dates, preferred beach activities, weather forecasts, and beach information stored in the database.

3.2 Activity Recommendation Engine

The user's context information is gathered and the Activity Recommendation Engine uses this information to analyze the 47 beach records stored in the application catalogue and rank them based on the user's selected activities. This system is based on a rule-based scoring method, using explicit activity mappings and metadata-tag weights. For each selected activity a and beach b, the recommendation score is calculated as:

$$S(b) = \sum_{\{a \in A\}} \left[W_{\exp(a,b)} + W_{tag(a,b)} \right] \quad (1)$$

where the user has chosen the set of activities A. The explicit activity-mapping weight is given by:

$$W_{\exp(a,b)} = \{15 \; if \; b \in B_a; 0 \; otherwise\} \quad (2)$$

For instance, if the Surfing activity is chosen, beaches like Arugam Bay, Weligama and Hiriketiya get the explicit

activity boost. The weight (i.e. the metadata-tag weight) is defined as:

$$W_{\text{tag}}(a, b) = \{5 \text{ if a matched b; } 0 \text{ otherwise}\} \quad (3)$$

The resulting scores are used to rank the candidate beaches in descending order. The weights of 15 and 5 are manually chosen design parameters that require empirical optimization; their 3:1 ratio prioritizes explicit mappings. Independent labels and sensitivity analysis are needed to justify them.

3.3 Seasonal Risk Assessment

Sri Lanka experiences two main monsoon seasons which affect the sea state in various coastal areas. PearlWay checks if the dates the user has selected fall under a predetermined regional monsoon schedule. The rougher sea conditions in western and southern coastal areas during the Southwest Monsoon (May- September) place these areas under a potentially hazardous consideration, including Bentota, Hikkaduwa and Mirissa. The eastern and northern coastal regions (Trincomalee, Pasikudah, Nilaveli and Arugam Bay) are assessed for the possibility of higher risk during the application's October–January seasonal caution window.

The suitability score is initialized at 95 and bounded between 0 and 100 as follows:

$$Score_f = \max(0, \min(100,\ 95 - 25I_{monsoon} - 20I_{rain})) \quad (4)$$

where $I_{monsoon}$ is an indicator variable equal to 1 when the selected beach and travel month correspond to an affected monsoon region, and 0 otherwise. If the retrieved weather code falls between 61 and 99, a category covering rain, snow, showers, and thunderstorms (Open-Meteo, n.d.), I_{rain} is set to 1; otherwise, it is set to 0.

When a relevant monsoon condition is identified, 25 points are deducted from the baseline score of 95, giving a score of 70 before any weather penalty. The baseline and penalties are manually chosen and have not been calibrated. Seasonal windows and forecast codes cannot establish sea safety, tides, or rip-current risk; official marine advisories and local warnings must take precedence.

3.4 Real-Time Weather Information

In order to recommend travel plans that reflect actual conditions, PearlWay uses real-time weather data that is gathered through the Open-Meteo REST API. When it comes to the trips in the next seven days, current weather data, including temperature, rainfall, weather codes, and wind, is gathered for the application. Forecasts can change and may miss localized coastal conditions. The documented formula uses a weather-code penalty; calibrated wind or rainfall thresholds remain to be established. Trips outside the seven-day window and unavailable or stale data limit assessment (Open-Meteo, n.d.).

3.5 Distance Calculation

Distance between the user's current location and the beaches recommended by the system is calculated using geographical positioning system (GPS) coordinates. From the determined distance, an average driving speed and travel time for reaching the selected destination are estimated by the system. Coordinate-based distance represents geographical separation rather than road-route distance; assumed-speed travel time excludes traffic and road geometry. GPS errors and these approximations require validation.

3.6 Emergency Support

The application gives the user access to the necessary contacts during emergencies including emergency services, marine police, and nearest hospitals. Additionally,

the user can share his/her GPS coordinates during emergencies to help the emergency service personnel locate him/her as fast as possible. Android intents are used to initiate contact with the emergency contacts and location sharing during emergencies.

4 SYSTEM DESIGN

4.1 Presentation Layer

The Presentation Layer is built on Flutter and is responsible for providing the responsive and easy-to-use interface of the application.

4.2 Business Logic Layer

The Business Logic Layer is responsible for managing the key functionalities of the application including the recommendation engine, context-aware decision rules, weather analysis, and suitability score calculation.

4.3 Data Layer

The Data Layer uses Firebase Cloud Firestore database to store and manage application data including user profiles, trips history, favourites, and beach information. Although Firestore supports cached access (Firebase, n.d.), offline access remains an implementation limitation in this application. Offline records, map tiles and weather freshness therefore remain implementation limitations.

4.4 External Services Layer

The external services include the Open-Meteo Weather API for weather information, OpenStreetMap for mapping, the Geolocator package for GPS location, and Android native intents for functions such as making a call or sharing a help request.

Figure 1 shows the four-layer architecture. Figure 2 links user inputs and contextual data to scoring and trip planning. Its threshold of 70 is a design cutoff whose validity as a safety boundary requires experimental evaluation.

5 TESTING AND RESULTS

Preliminary manual functional checks were conducted on five phones: OPPO A5x and HONOR X6c (Android 16), and Samsung Galaxy A05, Samsung Galaxy A06 and Infinix HOT 60i (Android 15). Beach selection, nearby-beach display based on the current location, and beach temperature display worked on the tested phones.

5.1 Limitations

Offline access remains an implementation limitation. These were informal functional observations; structured quantitative evaluation and qualitative user studies remain future work. Supplied screenshots illustrate displayed information; its accuracy requires independent validation. One map screenshot shows API KEY REQUIRED, indicating incomplete map rendering in that session. Recommendation accuracy, coastal safety validity, scalability, response time

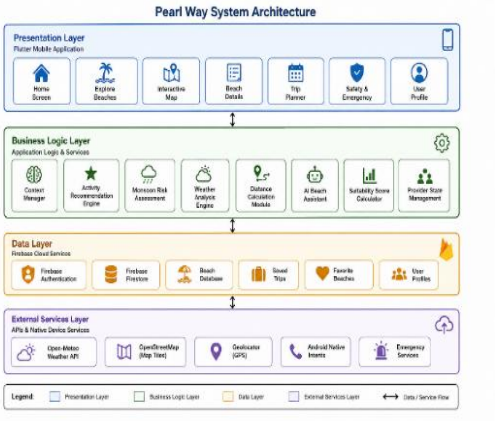


Figure 1: System Architecture

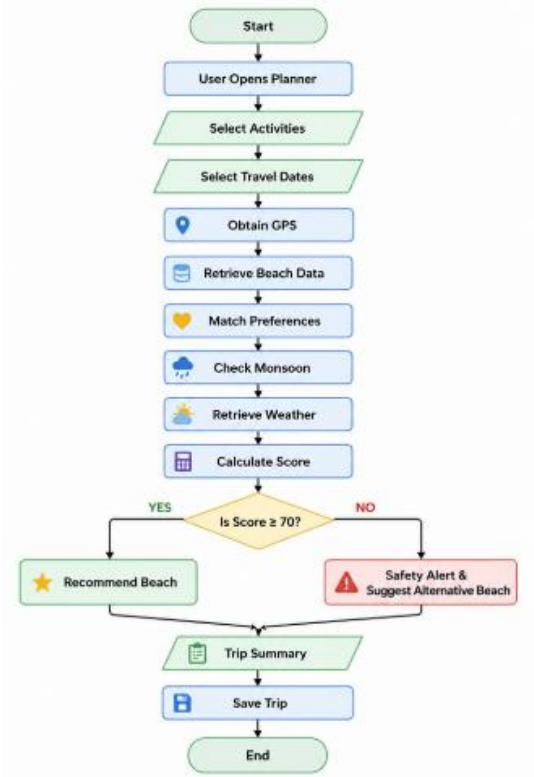


Figure 2: Flow Chart for Trip Planner

and user satisfaction remain unevaluated. Further validation is necessary before deployment. Further validation requires independently labelled activities, comparison with official environmental advisories, sensitivity analysis and user/system performance measurements.

6 CONCLUSION

PearlWay is a case study that highlights the integration of context-aware computing, LBS, and mobile GIS into one intelligent tourism platform.

Integration of Flutter, Firebase, OpenStreetMap, Open-Meteo API, and rule-based expert systems provides an integrated system that can be used for coastal tourism in Sri Lanka. Functional checks support selected functions on five Android devices; effectiveness, safety accuracy and scalability require further evaluation.

Future work will be done to integrate machine learning recommendation models, hotel reservations, crowd density estimation, multi-language support, and predictive coastal hazard analysis.

REFERENCES

- AlSaeed, D. (2023). LOCUS: A mobile tourism application and recommender system for personalized places. Informatica, 47(1).
- Curran, K., & Smith, K. (2006). A location-based mobile tourist guide. Tourism and Hospitality Research, 6(3), 200–209.
- Dey, A. K. (2001). Understanding and using context. Personal and Ubiquitous Computing, 5(1), 4–7.
- Firebase. (n.d.). Access data offline. https://firebase.google.com/docs/firestore/manage-data/enable-offline
- Gavalas, D., Konstantopoulos, C., Mastakas, K., & Pantziou, G. (2014). Mobile recommender systems in tourism. Journal of Network and Computer Applications, 39, 319–333.
- Mateos, P., & Bellogín, A. (2025). A systematic literature review of recent advances on context-aware recommender systems. Artificial Intelligence Review, 58, Article 20. https://doi.org/10.1007/s10462-024-10939-4
- Open-Meteo. (n.d.). Weather forecast API. https://open-meteo.com/en/docs
- Schilit, B. N., Adams, N., & Want, R. (1994). Context-aware computing applications. In Proceedings of the IEEE Workshop on Mobile Computing Systems and Applications (pp. 85–90).

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Artificial Intelligence-Assisted Context-Aware Micro-Learning Platform for Personalized Educational Video Learning

Handapangoda HGRI[1], Mahagedara HBMHS[2], Vidanagama VGTN[3]
[1,2,3] *Department of Computing & Information Systems, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*
ravinduinduwara13@gmail.com[1], hasindusahan613@gmail.com[2], tharinda@wyb.ac.lk[3]

ABSTRACT

The growth of digital learning platforms has improved access to educational content through mobile devices. However, learners still face challenges such as lengthy learning materials, difficulty finding relevant content, distractions from non-educational media, and limited personalization. This research proposes an Artificial Intelligence (AI)-assisted context-aware micro-learning platform that delivers educational content through a short-form video feed. The system uses learner profile information, including age and interests, together with watch history to recommend educational videos. It also integrates AI-generated quizzes based on watched educational content, an AI chatbot for learning assistance, and gamification features such as experience points, daily streaks, and a leaderboard. Additionally, the platform includes a video approval mechanism to support the controlled publication of educational content. The application was developed using Flutter and Dart, with Supabase as the backend. Functional testing was conducted to evaluate seven core components: personalized video recommendation, AI quiz generation, chatbot interaction, watch history, saved videos, gamification, and content approval. The observed results indicated that the tested components operated according to their intended functions. However, the evaluation was limited by AI Application Programming Interface (API) usage restrictions, limited Supabase storage capacity, and the availability of educational videos, preventing a broader user-based evaluation. Future work will focus on extending the system infrastructure and conducting user-based testing to investigate usability, recommendation relevance, user acceptance, and learning effectiveness.

KEYWORDS: Artificial Intelligence, Micro-Learning, Personalized Learning

1 INTRODUCTION

Mobile technology has changed how learners access educational resources, allowing them to learn through smartphones beyond traditional classrooms and schedules. However, the growing volume of online content creates challenges such as lengthy materials, difficult content discovery, distractions, and limited personalization. Many learning platforms rely on structured courses and long videos, which may not suit learners with limited free time who want to learn a specific concept within minutes. Large content collections also require users to search and compare resources, while general video platforms may contain unrelated content that causes distractions. Personalization is important because learners differ in age, interests, learning experience, and educational needs. Using learner profiles and previous interactions, such as watch history, can help recommend relevant educational videos. Context-aware computing further improves personalization by adapting system behavior to changing conditions. Therefore, this research presents an AI-powered context-aware micro-learning platform that combines personalized video recommendations, AI-based assessments, a chatbot and gamification.

The basic functions of the proposed system are:

- Recommend educational videos according to user interests, age, and learning history.
- Provide a short-form educational video feed for micro-learning.
- Generate quizzes using AI to evaluate learner knowledge.
- Provide AI-based learning assistance through a chatbot.
- Motivate learners through gamification features such as experience points (XP) for completing learning activities, daily streaks for maintaining regular learning activity, and a leaderboard for comparing progress with other learners.
- Allow users to save and revisit educational videos.
- Maintain an administrator approval process for uploaded educational content.

2 LITERATURE REVIEW

2.1 Mobile Microlearning

Mobile learning provides access to educational content beyond traditional classrooms. Mobile-based microlearning delivers content in small, manageable units, allowing learners to study during short periods. A systematic review found that it offers flexibility and accessibility across higher education and professional learning. However, microlearning must be carefully designed, as excessively small or disconnected units may lead to fragmented learning and fail to support meaningful learning objectives (Meng et al., 2016).

2.2 Short-Form Educational Video Learning

Short-form video platforms have significantly influenced how users consume digital information. The same interaction model can potentially be applied to educational content by presenting learning materials as concise videos that learners can consume sequentially. Recent research examining knowledge-based short-video platforms indicates that short videos can support microlearning in areas such as language learning, science communication, and vocational education (Sisi & Zainudin, 2025). However, conventional short-video platforms are primarily designed for broad content consumption rather than structured educational progression.

2.3 Personalized Learning and Recommendation Systems

Personalization is important in digital learning because learners differ in their interests, prior knowledge, learning progress, and preferred areas. Recent research shows increasing use of adaptive technologies and generative AI to support personalized learning. Educational recommender systems help filter learning resources and support personalization. However, systematic research has highlighted limitations in evaluating recommendation approaches and their educational benefits (Bin et al., 2024).

2.4 AI-Supported Assessment and Learning Support

Assessment is an important component of microlearning because watching educational content alone does not necessarily demonstrate whether the learner has understood the material. AI

can support formative assessment by generating questions from educational content and providing immediate feedback.

A systematic review of generative AI in educational assessment identified significant interest in using generative AI to support assessment practices while also highlighting the need for appropriate evaluation and responsible implementation. Recent research on AI-assisted personalized learning also reports positive effects across knowledge, competence, and emotional development, although the effectiveness depends on how AI is integrated into the learning process (Xia et al., 2024).

3 METHODOLOGY

This research follows a design and development methodology to create and evaluate an AI-powered context-aware micro-learning platform. The study began with a review of existing research on mobile micro-learning, educational recommendation systems, personalized learning, and AI-assisted assessment. The findings were used to identify current limitations and define the functional requirements of the proposed platform.

User-related data, including age, interests, and learning interactions, were selected as the primary inputs for personalization. Educational video metadata and user activity records were organized to support recommendation and learning analysis. A recommendation mechanism was then designed to suggest relevant educational videos, as in figure 1.

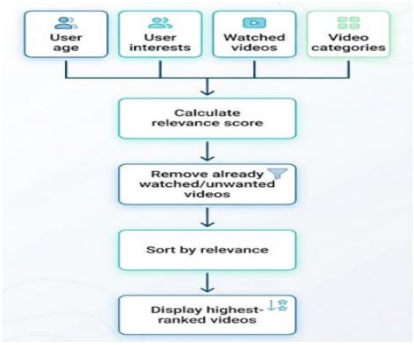


Figure 1: Recommendation Mechanism

To improve the learning experience under varying mobile conditions, a context-aware adaptation mechanism was developed to monitor network quality and adjust video resolution accordingly. In addition, generative AI was integrated to generate quizzes from learning topics and provide chatbot-based learning assistance.

The platform was implemented iteratively, with each component developed and tested individually before integration into the complete system. Functional testing was conducted to verify the operation of personalized recommendations, adaptive video delivery, quiz generation, chatbot interactions, and user engagement features.

4 SYSTEM DESIGN

4.1 Presentation Layer

The presentation layer is implemented using Flutter and provides the user interface of the mobile application. The main interfaces include the personalized video feed, user profile, saved videos, quiz interface, chatbot, leaderboard, and video upload and management screens.

The video feed uses a vertical scrolling interaction so that learners can move between short educational lessons with minimal navigation.

4.2 Application Logic Layer

The application logic layer manages the main learning functions of the system. It handles user interactions, recommendation requests, quiz workflows, XP calculation, streak management, saved videos, watched history, and content approval states.

The recommendation mechanism calculates a score using interest similarity, age suitability, and watch-history relevance:

$$R(u,v)=0.50S_I(u,v)+0.20S_A(u,v)+0.30S_H(u,v) \quad (1)$$

where:

- R(u,v) = final recommendation score for video vv for learner uu
- $S_I(u,v)$ = similarity between the learner's interests and the video's topic
- $S_A(u,v)$ = suitability of the video for the learner's age
- $S_H(u,v)$ = relevance based on the learner's previous watch history

Each component score ranges from 0 to 1. The weights represent heuristic design choices, not empirically optimized values. Videos are ranked by their scores and recommended accordingly. Recommendation quality requires separate evaluation based on relevance to user interests and viewing history.

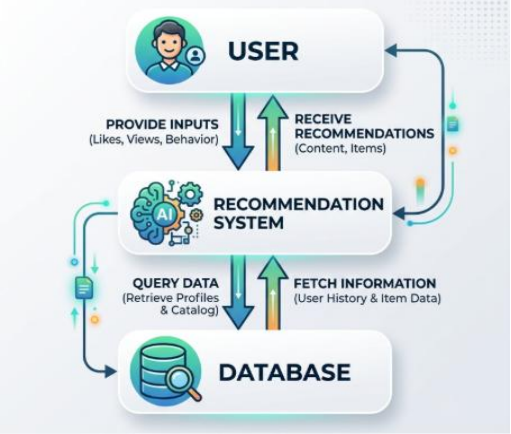


Figure 2: Relationship between User and the System

4.3 Data Layer

Supabase is used as the backend platform for storing and managing application data. User-related information, educational content, learning interactions, quiz-related data, saved videos, watched history, and other application data are managed through the backend. As illustrated in figure 2, the system maintains the relationship between users and their learning activities, enabling historical interactions to be used for personalization.

5 RESULTS AND DISCUSSION

5.1 Functional Testing

Functional testing was conducted to evaluate the main components of the proposed platform. The results are summarized in Table 1.

Table 1: Functional Testing Results

Test Function	Test scenario and observed result
Personalized recommendation	Technology-related interests and watch history resulted in recommendations for similar educational videos.
AI quiz generation	After watching an educational video, the user requested a quiz, and the system generated related questions.
AI chatbot	Learning-related questions were submitted to the chatbot. The system returned responses to the submitted questions.

Watch history	Educational videos were watched, and the corresponding videos appeared in the user's watch history.
Saved videos	An educational video was saved, and it remained accessible through the saved videos section.
XP and streak	Learning activities were completed, and the corresponding XP or streak information was updated as implemented.
Content approval	A video was uploaded and submitted for approval. Its availability changed according to the administrator's approval action.

5.2 Discussion

The functional testing results showed that the main components of the proposed platform operated according to their intended functions. The recommendation mechanism used user profile information and watch history to provide personalized educational videos, AI-based quiz generation and chatbot assistance also operated as intended. The testing confirms the functional feasibility of integrating personalized recommendation, AI-assisted learning, and gamification within a single mobile platform.

6 CONCLUSION

This research presented an AI-assisted context-aware micro-learning platform for personalized educational video learning. The system integrates personalized video recommendations, AI-generated quizzes, a chatbot, watch history, saved videos, gamification, and administrator-controlled content approval. It was developed using Flutter and Dart, with Supabase as the backend. Functional testing showed that the main features operated according to their intended functions. Recommendations reflected learner interests and watch history, quizzes were generated based on watched videos, and the chatbot responded to learning-related questions. Other tested features, including saved videos, XP, streaks, and content approval, also worked as expected. However, the evaluation did not measure recommendation accuracy, user acceptance, or improvements in learning outcomes. Testing was limited by AI API usage restrictions, Supabase storage capacity, and the availability of educational videos. Future work will focus on improving infrastructure, expanding educational content, and conducting user-based evaluations to assess usability, recommendation relevance, user satisfaction, quiz quality, and learning effectiveness. Further testing is also needed to evaluate chatbot responses and video delivery under different network conditions.

REFERENCES

- Bin, Q., M. F. Zuhairi, and J. Morcos, “A Comprehensive Study On Personalized Learning Recommendation In e-Learning System,” IEEE Access, pp. 1–1, Jan. 2024, doi: 10.1109/access.2024.3428419.
- Meng, J., Z. Wang, and Z. Li, “Application of Cognitive Load Theory in Mobile Micro-learning,” Proceedings of the 2016 International Conference on Management Science and Innovative Education, 2016, doi: 10.2991/msie-16.2016.110.
- Sisi, Z. and Z. Zainudin, “Systematic Review on Utilization of Douyin Knowledge-Based Short Video in

Education," Asian Pendidikan, vol. 5, no. 1, pp. 42–50, 2025, doi: 10.53797/aspen.v5i1.4.2025.

- Xia, Q., X. Weng, F. Ouyang, T. J. Lin, and T. K. F. Chiu, "A scoping review on how generative artificial intelligence transforms assessment in higher education," International Journal of Educational Technology in Higher Education, vol. 21, no. 1, May 2024, doi: 10.1186/s41239-024-00468-z.

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Zonely: Harnessing Context-Aware Computing to Restore Digital Wellbeing and Focus

Peiris TKSA[1], Thennakoon PGDS[2], Vidanagama VGTN[3]
[1,2,3]*Department of Computing and Information Systems, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*
shamilkapeiris@gmail.com[1], dilwansthennakoon@gmail.com[2], tharinda@wyb.ac.lk[3]

ABSTRACT

Smartphones support learning and productivity, yet context-blind interfaces and notifications contribute to habitual pickups and digital distraction among students. Rigid app limits create friction that users can override, and many existing automations change isolated settings rather than the working environment. This paper presents Zonely, a context-aware mobile application built with Flutter and Supabase that adapts the smartphone interface to the user's physical venue instead of restricting functionality. A custom geofencing engine based on the haversine formula runs entirely on the device and switches between four modes (Home, School, Library and Gym), each with a dedicated interface; on Android, entering academic venues also enables Do Not Disturb. Raw location coordinates are never uploaded to the cloud. Functional testing with simulated locations confirmed that all five tested features behaved as specified. A formative survey of ten university students found that study environments were commonly reported as distracting, that eight of nine respondents who had not used the app rated the concept of location-adaptive interfaces four or five out of five, and that the mean intention to switch from current screen-time methods was 4.10 out of 5. These findings indicate technical feasibility and perceived usefulness only. The study did not measure screen time, focus or wellbeing outcomes and did not include a field or pilot user study; these limitations and a planned evaluation are discussed.

KEYWORDS: Context-aware computing, Digital wellbeing, Geofencing, Mobile application, Smartphone distraction

1 INTRODUCTION

Mobile technology has become indispensable to modern life, serving as the primary medium for communication, education, and productivity. Learners and professionals rely on smartphones for academic resources, schedule management, and social connection. Yet these same devices have become major sources of distraction and cognitive overload.

Contemporary mobile platforms suffer from "context blindness" treating interfaces, notifications, and interactions uniformly regardless of whether the user is in a lecture, a library, a gym, or at home. This uniformity drives several measurable problems. Reflexive pickups habitual, unconscious unlocking during moments of boredom interrupt deep work and study loops. Notification fatigue, driven largely by social media and non-essential alerts, breaks concentration and is reported as an hourly distraction by many users. (American Brain Foundation, 2026) Additionally, roughly thirty-one percent of adults engage in doomscrolling, contributing to cognitive burnout and shortened attention spans. (Al-Mamun et al., 2025)

Existing screen-time interventions fall into three categories, each with distinct limitations. App time locks create psychological friction that users often override during low-willpower states. Feature phone downgrades limit personal efficiency and cut users off from essential productivity tools. Minimalist launchers often lack deep Operating System (OS)

integration, causing instability and friction in everyday workflows.Context-aware computing offers a way past these limitations by letting applications adapt behavior to the user's current environment. Rather than restricting device use, such a system can reshape the device's identity to match its surroundings.

This research therefore presents Zonely, a location-aware, context-driven mobile application that replaces rigid prohibition with empathetic adaptation.

2 LITERATURE REVIEW

2.1 Smartphone Addiction and Digital Wellbeing

The psychological dependency on smartphones has been extensively documented in contemporary literature. Nomophobia, defined as the discomfort, anxiousness, and worry caused by separation from a mobile device, represents a growing behavioral health concern. A systematic review and meta-analysis conducted across Indian populations found a pooled nomophobia prevalence of 77.6%, with the burden concentrated heavily among university students. (Pieh, C. et al., 2025) The same review noted that duration of mobile phone usage and smartphone ownership were significantly associated with higher nomophobia scores. Internationally, systematic reviews have reported global nomophobia prevalence ranging from 29.5% to 100% across different demographic groups. (Chang et al., 2023)

The relationship between excessive screen time and mental health has been established through randomized controlled trials. Research demonstrates that smartphone screen time reduction produces significant post-intervention effects on wellbeing, depressive symptoms, sleep quality, and stress, with small to medium effect sizes observed in healthy student populations. (Scribd. (n.d.). Haversine, 2026) However, screen time frequently returns to baseline levels rapidly after intervention cessation, suggesting that sustainable solutions must integrate seamlessly into daily routines rather than imposing temporary restrictions.

2.2 Context-Aware Mobile Computing

Context-aware computing refers to systems that can sense, interpret, and respond to environmental and user-related contextual information. In mobile environments, context awareness enables applications to adapt functionality based on location, time, activity, and surrounding conditions. The theoretical foundation for such systems rests on the premise that human attention requirements vary significantly across physical and social environments; the focus demands inside a campus library differ entirely from those at home or in a fitness center.

Research in location-based services has evolved considerably, with ongoing investigations into spatial marketing, behavioral research, and autonomous environment control. The integration of context awareness into personal wellbeing applications remains relatively underexplored, particularly regarding empathetic adaptation rather than prohibition-based intervention. (Laws, J., 2025)

2.3 Digital Wellbeing Interventions and Screen Time Management

Current digital wellbeing strategies include app-based monitoring, tech-free

zones, group detox programs, and built-in operating system screen time trackers. Evidence suggests that app-based interventions can effectively minimize screen time and promote mindful technology engagement. However, many existing tools function as external monitors or rigid gatekeepers rather than intelligent assistants that adapt to contextual needs.

The Pomodoro technique, a time-structured intervention involving focused work intervals followed by short breaks, has demonstrated consistent improvements in focus, reduced mental fatigue, and enhanced sustained task performance. (Priya, A. et al., 2026) Comparative studies indicate that systematic break structures outperform self-paced break-taking, making such techniques particularly suitable for integration into context-aware study environments. (Ogut, E., 2025)

2.4 Geofencing and Location-Based Services

Geofencing technology enables the creation of virtual boundaries around physical locations, triggering automated actions when a device enters or exits the defined perimeter. Applications of geofencing span marketing, logistics, behavioral research, and health intervention. In the domain of behavioral health, researchers have developed zero-permission geofencing applications to alleviate gambling disorders by triggering contextual interventions when users approach high-risk locations.

Privacy concerns represent a significant barrier to location-based service adoption. Users frequently express apprehension regarding continuous location tracking, data transmission, and server-side storage of movement patterns. Consequently, privacy-preserving architectures that perform location calculations strictly on-device while limiting cloud storage to anonymized or non-sensitive data are critical for user trust and regulatory compliance.

3 METHODOLOGY

This research follows a design and development methodology to create and evaluate a context-aware digital wellbeing platform. The study began with a comprehensive review of existing research on smartphone addiction, nomophobia, context-aware computing, digital wellbeing interventions, and geofencing technologies. The findings were used to identify current limitations in existing solutions and define the

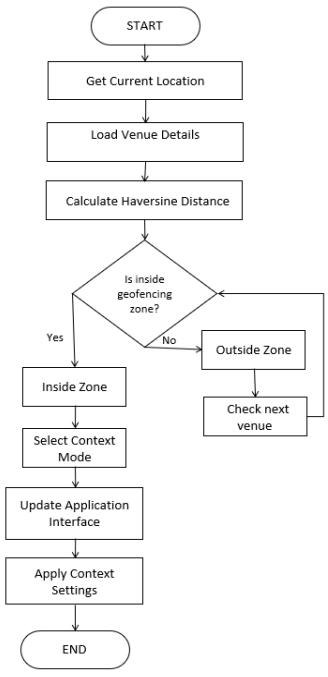


Figure 1: Design and Development Methodology

functional requirements of the proposed platform.

User-related contextual data, specifically physical venue definitions and associated behavioral modes, were selected as the primary inputs for system adaptation. Venue records containing GPS coordinates, contextual labels, and user-defined radii were organized to support geofence boundary detection and mode transition logic. A custom geofencing engine was then designed to evaluate device proximity against saved venues using haversine-based distance calculations performed entirely on-device. (Supabase. (n.d.)., 2026)

$$a = sin\text{^}2(\frac{\Delta\phi}{2}) + cos(\phi_1)cos(\phi_2)sin\text{^}2\left(\frac{\Delta\lambda}{2}\right) \quad (1)$$

To improve user acceptance and avoid the psychological friction associated with rigid app locks, an empathetic adaptation mechanism was developed. This mechanism transforms the application interface and system behavior based on detected context rather than disabling device functionality. In addition, native operating system integration was implemented to automate Do Not Disturb state changes during academic venue entry.

The platform was implemented iteratively, with each contextual mode and core system component developed and tested individually before integration into the complete system. Functional testing was conducted to verify geofence boundary detection accuracy, mode transition responsiveness, notification suppression, local data persistence, and cross-platform User Interface (UI) consistency. The functional testing results are presented and discussed in Section 5.

Table 1: Functional test results (simulated inputs)

Feature	How it was Tested	Pass/Fail
Geofencing & automatic mode switching	Simulated Global Positioning System (GPS) at Home, School etc. to verify automatic switching.	Pass
Native Do Not Disturb (DND) auto-toggle	Tested geofence entry/exit and verified DND activation and deactivation.	Pass
Pomodoro Timer	Tested timer start, pause, resume, and completion alerts.	Pass
Academic Timetable	Added classes and verified scheduled notifications.	Pass
Study Flashcards	Generated flashcards and verified display and randomization.	Pass

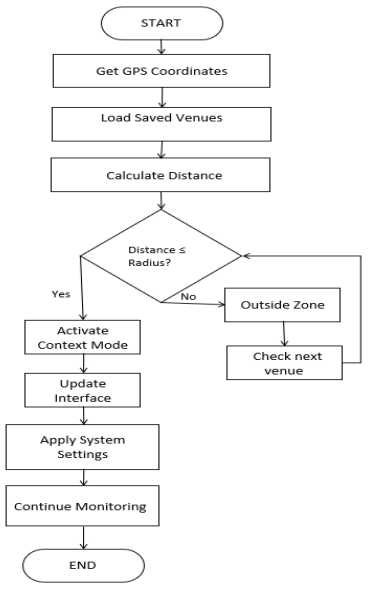


Figure 2: Empathetic Adaptation Mechanism

4 SYSTEM DESIGN

4.1 Presentation Layer

The presentation layer is implemented using Flutter and provides the user interface of the mobile application. The main interfaces include the dynamic context mode dashboard, focus analytics visualization, venue management screen, interactive timetable, Pomodoro clock, study flashcards, hydration tracker, and workout timer.

The application employs a responsive rendering architecture optimized for dynamic layouts across mobile devices. The interface automatically transitions between four dedicated contextual modes based on geofence triggers. In Home Mode, the interface serves as a personal command center displaying focus analytics, an Aura Assistant for mindfulness prompts, and a smart hydration tracker. In School Mode, the interface presents an interactive timetable with automated pre-class notifications and dynamic study flashcards. In Library Mode, a maximized Pomodoro clock and study planner dominate the interface, accompanied by immersive ambient focus audio controls. In Gym Mode, a high-contrast dark aesthetic presents a precision workout timer with lap recording and integrated media controls.

4.2 Application Logic Layer

The application logic layer manages the main contextual functions of the system. It handles user interactions, geofence polling, distance calculations, mode transition triggers, focus analytics aggregation, hydration logging, timetable events, and flashcard randomization.

The geofencing logic connects device location data with saved venue definitions through a central Venue Provider. The engine periodically polls device coordinates and evaluates haversine distance against each saved venue's radius. When distance falls within the defined threshold, the application transitions to the corresponding contextual mode and updates system settings accordingly; otherwise, it defaults to Home Mode. The context-aware adaptation component determines the appropriate interface configuration and, through Flutter Method Channel integration, triggers native Android Kotlin code to toggle system-level Do Not Disturb states when entering study zones.

4.3 Data Layer

Supabase is used as the backend platform for storing and managing application data. User account credentials, anonymized aggregate focus metrics, user profiles, and saved venue definitions are managed through the backend infrastructure, which maintains International Organization for Standardization (ISO) 27001 certification and Row Level Security policies. (Schmidtke, H. R., 2020)

user_id	name t...	category	latitude f.	longitude
2ace36...	Indoor Stadium	gym	7.4638712	80.0178999
2ace36...	College	university	NULL	NULL
866ec8...	Gym	gym	7.4642096	80.0177323
866ec8...	Wayamba	university	7.4647471	80.018737
2ace36...	Wayamba Univ	university	7.4649241	80.0186111

Figure 3: Database

Local persistence is achieved through Shared Preferences, caching operational state on-device including daily water levels, custom timetable events, and study flashcards to ensure offline availability. The privacy architecture strictly mandates that all geofence boundary checks,

distance evaluations, and location calculations take place on-device within the local Dart and Kotlin runtime. Raw GPS telemetry coordinates are never uploaded, transmitted, or logged to cloud servers, ensuring that continuous location tracking remains strictly local. Figure 3 shows that the User Identifiers (IDs) stay encrypted inside the supabase database and the table only saves latitudes and longitudes of saved places but not the current user location.

5 CONCLUSION

Zonely presents a context-aware smartphone prototype that adapts its interface according to the user's physical venue instead of relying on strict application blocking. Functional testing using simulated GPS locations demonstrated that the main features, including geofence-based mode switching and contextual interface adaptation, operate as intended. In addition, feedback from ten university students suggested that the concept is perceived as useful and relevant to common smartphone distraction problems. However, the current study does not demonstrate that Zonely reduces screen time, involuntary pickups, distraction, or improves focus and wellbeing. Important limitations include simulated geofence testing, a small and uniform sample, only one participant with prior app experience, limited personalisation, Android-specific DND functionality, and the absence of a sleep mode despite sleep being identified as a major distraction context. Future work should include real-world GPS testing, a larger and more diverse pilot study, behavioural measurements, validated wellbeing scales, improved personalisation, and additional contextual modes.

REFERENCES

- Al-Mamun, F., Mamun, M. A., Kaggwa, M. M., Mubarak, M., Hossain, M. S., ALmerab, M. M., Muhit, M., Gozal, D., Griffiths, M. D., & Sikder, M. T. (2025). The prevalence of nomophobia: A systematic review and meta-analysis. Psychiatry Research, 349, 116521. https://doi.org/10.1016/j.psychres.2025.116521
- American Brain Foundation. (2026, July 15). Doomscrolling and its effects on the brain. https://www.americanbrainfoundation.org/doomscrolling-and-its-effects-on-the-brain/
- Chang, X.-J., Hsu, F.-H., Liang, E.-C., Chiou, Z.-Y., Chuang, H.-H., Tseng, F.-C., Lin, Y.-H., & Chang, Y.-J. (2023). Not merely deemed as distraction: Investigating smartphone users' motivations for notification-interaction. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (pp. 1–17). Association for Computing Machinery. https://doi.org/10.1145/3544548.3581146
- Laws, J. (2025, May 13). Gen Z has a doomscrolling problem. Newsweek. https://www.newsweek.com/gen-z-doomscrolling-problem-2065999
- Ogut, E. (2025). Assessing the efficacy of the Pomodoro technique in enhancing anatomy lesson retention during study sessions: A scoping review. BMC Medical Education, 25(1), 1440. https://doi.org/10.1186/s12909-025-08001-0
- Pieh, C., Humer, E., Hoenigl, A., Schwab, J., Mayerhofer, D., Dale, R., & Haider, K. (2025). Smartphone screen time reduction improves mental health: A randomized controlled trial. BMC

Medicine, 23(1), 107. https://doi.org/10.1186/s12916-025-03944-z

- Priya, A., Sinha, R., Sethia, S., Rahman, M. H. U., & Melwani, V. (2026). Nomophobia in India: A systematic review and meta-analysis. Indian Journal of Medical Research, 163(5), 631–639. https://doi.org/10.25259/IJMR_2163_2025
- Schmidtke, H. R. (2020). Location-aware systems or location-based services: A survey with applications to CoViD-19 contact tracking. Journal of Reliable Intelligent Environments, 6(4), 191–214. https://doi.org/10.1007/s40860-020-00111-4
- Scribd. (n.d.). Haversine formula [PDF]. Retrieved September 5, 2026, from https://www.scribd.com/document/321848480/Haversine-formula-pdf
- Supabase. (n.d.). The Postgres development platform. Retrieved September 5, 2026, from https://supabase.com/

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Adoption and Perception of Digital Tools for Public Transport Improvement in Sri Lanka

Wickramanayake WANN[1], Premarathna WCC[2]
[1,2] *Department of Computing and Information Systems, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*
Nawoda.wickramanayake@gmail.com[1], chamidu@wyb.ac.lk[2]

ABSTRACT

The growing use of technologies has opened new ways to make public transportation better. Mobile applications, online ticketing, cashless payments, and real-time tracking are examples of tools that can enhance public transport services. Even though these digital tools are slowly starting to appear in Sri Lanka not many people are using them. This study looks at what makes people in Sri Lanka use transport tools. It focuses on things like knowing about the tools, how they see them, their education and where they live. The study uses two theories: one about how people accept technology and another about how people resist ideas. A cross-sectional quantitative survey was conducted using a structured questionnaire with five-point Likert-scale items. A total of 397 answers were looked at using a program called IBM SPSS Statistics. The data were analyzed using descriptive statistics, Pearson correlation, multiple regression, and one-way ANOVA. The results showed that people have an opinion about digital transport tools, but they don't know much about them, and they don't use them much. Knowing about the tools was the important factor for using them. Having an opinion also helped, but not as much. The model used explained half of why people use or don't use the tools. There were differences in how much people used the tools based on their education and where they lived. The study shows that making people know more helping them understand technology better making the tools easier to use and having them available in places is important, for getting more people to use digital transport tools in Sri Lanka.

KEYWORDS: Adoption, Awareness, Digital Transport Tools, Perception, Public Transport

1 INTRODUCTION

Digital technologies are changing how people use transportation. Things like apps, online ticketing, cashless payments and real-time tracking are becoming more common. These tools can make travelling easier, reduce confusion, and help people find transport information faster.Just because these digital services are available doesn't mean people will use them. In Sri Lanka digital transport options are slowly being introduced,. A lot of people still use old ways. Factors like how much people know about these tools what they think about them their education and where they live can affect whether they choose to use them or not.

Even though some studies have looked at transport use in other countries there is not much research about how it works in Sri Lanka. This study fills that gap by looking at how Sri Lankan people use transport tools. It looks at how awareness and opinions influence people's choices and whether adoption differs according to education level and residence area. The goal of the study is to collect data that can help create better and fairer digital transport plans, in Sri Lanka.

2 LITERATURE REVIEW

Digital transformation in transportation consists of mobile applications, electronic ticketing, cashless payments, and real-time passenger information systems.

Previous research indicates that users adopt these technologies when they are aware of them perceive them as useful and when they can access services. Reliable and useful digital information can make travel more convenient. It Can reduce uncertainty (Toth et al., 2022).

This study relies mainly on the Technology Acceptance Model (TAM) and the Innovation Resistance Theory (IRT). TAM states that users adopt technology readily when they believe the technology is useful and easy to use. (Sekiou & Dahmani, 2024) found that usefulness and ease of use also positively impact transformation adoption. IRT explains why users resist technologies; (Labiba et al., 2025) identified information quality and usage barriers as factors that affect public transport app adoption.

Demographic factors also influence adoption. (Durand et al., 2022) Demonstrated that differences in education and in residence area relate to inequality and to access to transport technologies. Yet little research has examined awareness, perception, education level and residence area, within the Sri Lankan public transport context. This study fills that gap.

3 METHODOLOGY

This study used a deductive quantitative approach to examine the relationships between awareness, perception, education level, residence area, and the adoption of digital transport tools among public transport users in Sri Lanka.A cross-sectional, descriptive and explanatory research design was used. Data were collected from public transport users using an online questionnaire sent through Google Forms. The study used convenience sampling and 397 valid responses were included in the analysis. The questionnaire consisted of questions and five-point Likert-scale items measuring awareness, perception and adoption of digital transport tools. Education level and residence area were also considered as variables. The collected data were analyzed using IBM SPSS Statistics. Descriptive statistics were used to identify the levels of awareness, perception and adoption of transport tools. Cronbach's Alpha was used to assess reliability while skewness and kurtosis were used to examine normality. Pearson correlation and multiple regression analysis were used to examine the relationships between awareness, perception, and the adoption of digital transport tools while one-way ANOVA was used to examine differences, in adoption of digital transport tools according to education level and residence area.

4 RESULTS AND DISCUSSION

397 valid responses were included in the analysis. The respondents came from a variety of residence areas. Had different levels of education. The respondents from residence areas made up the largest group, accounting for 43.6 percent of the total (n = 173). The respondents from residence areas represented 28.7 percent of the sample (n = 114).

Table 01: Distribution of Respondents by Residence Area

Residence Area	Frequency	Percentage (%)
Rural	110	27.7
Semi-Urban	173	43.6
Urban	114	28.7
Total	**397**	**100.0**

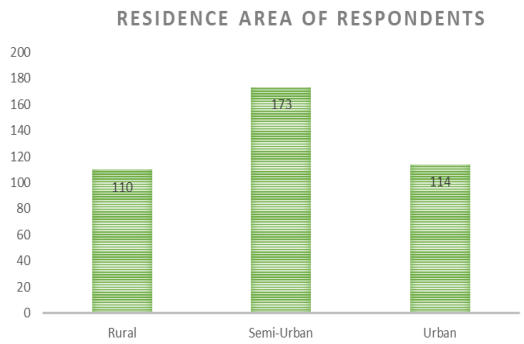


Figure 1: Residence Area of Respondents

The respondents, from rural residence areas accounted for 27.7 percent of the sample (n = 110).

In terms of education most people had finished G.C.E. Advanced Level with 230 people saying that. Next were people with a Bachelor's Degree, which had 54 people. Then came G.C.E. Ordinary Level with 46 people. After that was Diploma, with 40 people. Lastly, the Postgraduate Degree had 27 people. The numbers show that the group had people from different levels of education.

4.1 Reliability and Descriptive Results

The measurement instrument showed internal consistency with an overall Cronbach's Alpha of 0.834. Each construct also showed reliability with Alpha values of 0.812 for awareness 0.845 for perception and 0.801 for adoption.

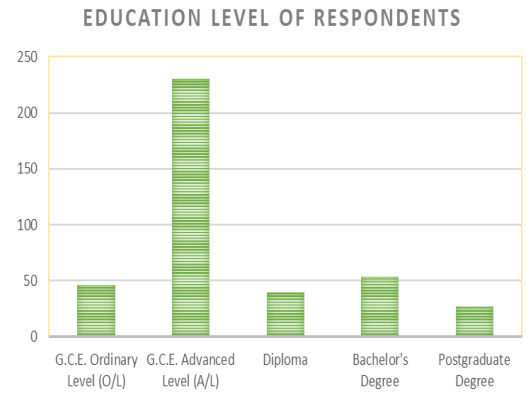


Figure 2: Education Level of Respondents

Descriptive analysis showed a gap between respondents attitudes and their real use of digital transport tools.Perception had an average score of 4.09 showing that respondents see digital transport technologies very positively.Awareness was at a level, with an average of 3.11.Adoption had an average score of 2.30.This suggests that although respondents generally view transport tools positively the positive view of digital transport tools has not yet turned into a similar level of real adoption.

4.2 Hypothesis Testing

The multiple regression model was statistically significant. Accounted for 41.1 percent of the variation in adoption ($R^2 = 0.411$). Awareness had a positive and statistically significant impact on adoption ($\beta = 0.623$, $p < 0.001$) making it the most powerful predictor in the model. Perception also had a significant positive influence on adoption ($\beta = 0.079$, $p = 0.044$) but its effect was much smaller compared to awareness.

Table 02: Distribution of Respondents by Education Level

Education Level	Frequency	Percentage (%)
G.C.E. Ordinary Level (O/L)	46	11.6
G.C.E. Advanced Level (A/L)	230	57.9
Diploma	40	10.1
Bachelor's Degree	54	13.6
Postgraduate Degree	27	6.8
Total	**397**	**100.0**

The one-way ANOVA results revealed a statistically significant difference in adoption across education levels ($F = 12.961$, $p < 0.001$) and across residence areas ($F = 11.446$ $p < 0.001$). Therefore, all four hypotheses proposed in the study were supported.

4.3 Discussion

The findings show a gap between how much respondents think digital transport tools are good and how much they actually use them. Even though users usually see the benefits of digital transport tools they still use them only a little.

Awareness was the factor that predicted whether people will use digital transport tools. This means that just having an opinion about digital transport tools is not enough. People also need to know enough and to see the transport services that are available.The strong effect of perception matches the Technology Acceptance Model, which says that people judge technology by how useful it's how easy it is to use.

The large differences that were found between people with education levels and those living in different areas show that not all public transport users adopt digital transport tools in the same way. Variations in literacy in how people can access technology in the infrastructure and in how often people see digital transport services may explain these differences. These findings show that we need to raise awareness more and make digital transport tools more accessible to people, in all kinds of demographic and geographical groups.

Table 03: Hypothesis Testing Results

Hypothesis	Statistical Result	Decision
H1:Awarenes→Adoption	β=0.623, p<0.001	Supported
H2: Perception → Adoption	β=0.079, p=0.044	Supported
H3: Education Level → Adoption	F=12.961,p< 0.001	Supported
H4: Residence Area → Adoption	F=11.446, p<0.001	Supported

5 CONCLUTION

This study examined the relationships between awareness, perception, education level, residence area, and the adoption of digital transport tools among public transport users in Sri Lanka.The results showed that respondents had a positive view of digital transport tools. Awareness was moderate and actual use was quite low. Awareness turned out to be the factor that predicts whether people will use these tools. Perception also had a positive effect. The study found that people with different education levels and those living in different areas had different rates of adoption. These findings point out that we need to raise awareness improve digital skills make tools easier to use and provide more digital transport services in all parts of the country. Transport authorities and service providers should therefore aim to increase user awareness and make digital services simple, easy to reach and inclusive.

This study has some limits. Using convenience sampling and collecting data online may mean the results do not fully represent all Sri Lankan public transport users. The cross-sectional design captured answers at one moment, in time. The study focused only on awareness, perception, education level, and residence area and did not examine other

factors that may be associated with the adoption of digital transport tools. Future research could use representative sampling across a wider range of areas and include additional technological and behavioral factors to provide a more comprehensive understanding of digital transport adoption in Sri Lanka.

REFERENCES

- A. W. Prananta, W. E. Kuswandoro, M. Afifuddin, P. D. Rahma and H. Mulyaningsih, "Digital Transformation in Industrial Technology and Its Social Impact on Online Public Transportation," Join: Journal of Social Science, *vol. 1, no. 3, pp. 291-312, 2024*Figures and Tables
- S. Toth, B. Bittner, T. Z. Kovacs and A. S. Nagy, "Digital transformation possibilities in public transportation in Debrecen," *Issues in Information Systems,* vol. 22, no. 3, pp. 305-319, 2022.
- J. Weber, "Evidence Triangulation for Public Transport Innovation: A Mixed-Methods Study on Real-Time Information Systems," *International Journal of Computational and Experimental Science and Engineering (IJCESEN),* vol. 11, no. 3, pp. 4580-4586, 2025.
- A. Wilson, S. Rodger, S. Bowen and M. Tewdwr-Jones, "Public engagement, digital technology and transport: engaging through open, early and experience-centred perspectives at scale," *Contemporary Social Science,* vol. 20, no. 1, p. 7–26, 2025.
- M. N. Labiba, D. R. Mutiara, R. Shadrina, P. W. Handayani and N. C. Harahap, "Exploring the Barriers to Public Transport App Adoption Using Innovation Resistance Theory," *Journal of Information Systems Engineering and Business Intelligence,* vol. 11, no. 2, p. 293–310, 2025.
- D. Marikyan and S. Papagiannidis, "Technology Acceptance Model: A Review," in *TheoryHub Book*, S. Papagiannidis, Ed., Newcastle upon Tyne, UK, TheoryHub, 2025.
- A. Sekiou and R. Dahmani, "Digitalization of the Algerian Transportation Sector," *Uluslararası Yönetim Akademisi Dergisi (International Management Academy Journal),* vol. 7, no. 3, p. 764–773, 2024.
- A. Durand, T. Zijlstra, N. van Oort, S. Hoogendoorn-Lanser and S. Hoogendoorn, "Access denied? Digital inequality in transport services," *Transport Reviews,* vol. 42, no. 1, p. 32–57, 2022.
- J. Pallant, SPSS Survival Manual: A Step by Step Guide to Data Analysis Using IBM SPSS Statistics, 7 ed., Abingdon, Oxon: Routledge, 2020.
- G. Hatem, J. Zeidan, M. Goossens and C. Moreira, "Normality Testing Methods and the Importance of Skewness and Kurtosis in Statistical Analysis," *BAU Journal - Science and Technology,* vol. 3, no. 2, 2022.
- C. Jacob, *Statistical Power Analysis for the Behavioral Sciences*, Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1998.
- J. F. Hair Jr, W. C. Black, B. J. Babin and R. E. Anderson, *Multivariate Data Analysis*, Andover, Hampshire, United Kingdom (EMEA edition): Cengage Learning, 2019.
- A. Field, Discovering Statistics Using IBM SPSS Statistics, London: SAGE Publications Ltd, 2018.

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Smart Water Leak Detection and Consumption Monitoring System: An Adaptive Statistical Abnormal-Flow Detection Approach

Chandira HPS[1], Pathirana OPTP[2], Vidanagama VGTN[3]
[1,2,3]*Department of Computing and Information Systems, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*
tharinduprasad824@gmail.com[1] , senathchandira6932@gmail.com[2], tharinda@wyb.ac.lk[3]

ABSTRACT

Water leakage is an important source of residential water wastage and may remain unnoticed for long periods. This research presents an Internet of Things (IoT)-based water consumption monitoring and abnormal-flow detection system integrating an ESP32 microcontroller, a water flow sensor, Firebase Realtime Database, a Python-based Context Engine, and a Flutter mobile application. The Context Engine analyses valid historical water-flow records by calculating the mean and standard deviation to generate an adaptive threshold using T = Mean + (2 × Standard Deviation). The evaluation dataset contained 1,000 historical water-flow records, and 10 statistical abnormal-flow alert events were recorded. For an example baseline, a mean flow rate of 7.31 L/min and a standard deviation of 4.52 L/min produced a threshold of 16.35 L/min. When the current flow exceeded the learned threshold, the system created an alert in Firebase and delivered a Firebase Cloud Messaging (FCM) notification to the mobile application. However, independently labelled ground-truth data for normal usage and actual leaks were unavailable; therefore, quantitative detection accuracy could not be reliably calculated. The results demonstrate the implementation of adaptive threshold-based abnormal-flow detection and real-time alert delivery, rather than validated physical leak detection.

KEYWORDS: Adaptive Threshold, ESP32, Firebase, Internet of Things (IoT), Statistical Anomaly Detection

1 INTRODUCTION

Freshwater is one of the most precious natural resources, but much of it is wasted every year due to undetected water leaks. Household leaks can remain unnoticed for long periods, causing high water bills, property damage, and water wastage. Conventional water monitoring techniques mainly use manual checking or fixed thresholds to detect abnormal water usage. However, these methods do not adapt to different household consumption patterns and can produce incorrect results. With the rapid development of Internet of Things (IoT) technology, household utilities can now be monitored using sensors, wireless communication, and cloud computing techniques at relatively low costs. For example, smart water monitors can give information regarding water consumption and alert the users in case of any abnormalities. Yet most of the IoT solutions are based on some predetermined thresholds which do not take into account the behavior of users.

To overcome these limitations, this research proposes a Smart Water Leak Detection and Consumption Monitoring System integrating IoT hardware, Firebase, Flutter, and a Python-based Context Engine for adaptive statistical abnormal-flow detection. The system uses an ESP32 connected to a water flow sensor to collect real-time consumption data and send it to Firebase Real time Database. The Python Context Engine analyses the data and triggers alerts when irregularities are detected. Firebase Cloud Messaging (FCM) delivers notifications through the Flutter mobile application, which provides real-time monitoring, historical data, reports, and leak alerts.

The main aims are real-time water monitoring, adaptive statistical abnormal-flow detection, real-time notifications, and adaptive threshold generation based on historical household water usage.

2 LITERATURE REVIEW

The rapid advancement of Internet of Things (IoT) technologies has significantly improved the development of intelligent water monitoring systems. Modern IoT-based solutions utilise sensors, wireless communication, and cloud computing to continuously monitor water consumption and provide users with real-time information. These technologies help reduce water wastage by enabling early leak detection and efficient water resource management.

Several researchers have proposed smart water monitoring systems using microcontrollers such as Arduino and ESP32 integrated with flow sensors. These systems transmit sensor readings to cloud platforms where users can monitor water consumption through web or mobile applications. Although these systems provide real-time monitoring, many rely on predefined threshold values to identify abnormal water flow. Since household water consumption varies depending on the number of occupants and daily activities, a threshold selected for one household may not represent the normal usage pattern of another household.

Recent research has investigated machine-learning and statistical techniques for detecting anomalies in water-consumption data. Machine-learning approaches can provide flexible anomaly detection, but they generally require suitable training data and additional computational resources compared with simpler statistical methods. For a small-scale household IoT system, a statistical approach can therefore provide a simpler method for generating an adaptive threshold from available historical usage data.

Statistical anomaly detection offers a simpler alternative for analysing historical water usage data. It can calculate the mean and standard deviation to establish a normal consumption baseline. New sensor readings can then be compared with an adaptive threshold derived from these values, allowing the threshold to be generated from the household's historical usage pattern without requiring complex machine-learning model training.

Previous water-leak detection studies have also investigated approaches based on minimum night flow and continuous water-consumption patterns. Minimum night flow methods analyse water consumption during periods of low household activity to identify unexpected flow that may indicate leakage. Other smart-meter approaches consider continuous water flow over a period of time as an indication of possible leakage. These approaches show the importance of considering household water-consumption patterns and temporal behaviour when detecting abnormal water usage.

Firebase is widely used in IoT applications because of its real-time database, authentication services, and cloud messaging capabilities. Combined with Flutter, it supports cross-platform mobile applications for live sensor data, historical records, and instant notifications.

Although many systems provide real-time monitoring and alert generation, few integrate adaptive statistical analysis, mobile applications, and cloud-based IoT architectures. Many also lack user-specific settings, context-aware notifications, and historical data analysis within a single

platform. Therefore, this research proposes an IoT-based smart water monitoring system that combines ESP32 hardware, Firebase cloud services, Flutter mobile technology, and a Python-based Context Engine implementing statistical anomaly detection. The proposed system generates adaptive thresholds using historical water-consumption data to identify abnormal flow while providing real-time monitoring, cloud synchronisation, and instant mobile notifications.

3 METHODOLOGY

The proposed system integrates IoT hardware, cloud computing, statistical data analysis, and mobile technology to provide intelligent household water monitoring. Figure 1 illustrates the overall architecture of the proposed system.

3.1 Data Collection

A water flow sensor is installed on the household water supply line to measure the real-time water flow rate. The ESP32 microcontroller continuously reads sensor pulses, converts them into flow rate and water usage values, and transmits the data to Firebase Realtime Database through a Wi-Fi connection. The transmitted data include Flow rate (L/min), Daily water usage, Total water usage, abnormal-flow status, and Timestamp.

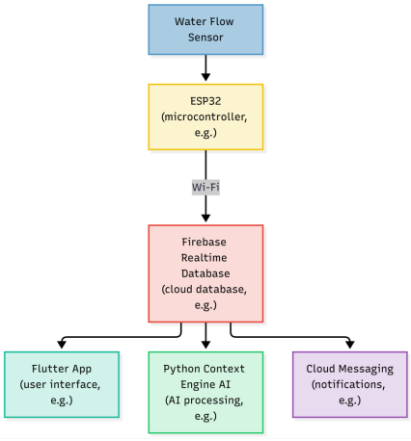


Figure 1: Overall architecture of the proposed system.

3.2 Historical Data Processing

The Python Context Engine retrieves historical water consumption records from Firebase. Before performing statistical analysis, the system filters the collected data by removing Invalid sensor readings, missing or corrupted values, flow values below the minimum valid sensor-flow limit, and previous abnormal-flow alert records. Only valid normal water consumption records are used to build the *statistical baseline.*

3.3 Statistical Anomaly Detection

The filtered historical data are analysed to calculate the statistical baseline of normal household water consumption. The system computes the following statistical parameters Mean (μ), Median, Standard Deviation (σ). The adaptive abnormal-flow threshold is calculated using

$$T = \mu + 2\sigma \quad \textbf{(1)}$$

Whenever the current water flow exceeds this adaptive threshold, the system identifies the reading as an abnormal event and classifies it as an abnormal-flow event that may indicate a potential water leak.

3.4 Alert Generation

When an abnormal-flow event is detected, the Context Engine creates an alert record in Firebase Realtime Database containing the current flow rate, calculated threshold, event time, and severity level.

The Notification Service sends a Firebase Cloud Messaging (FCM) push notification to all registered mobile devices,

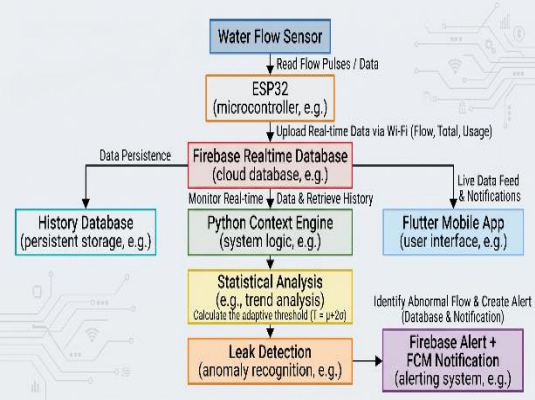


Figure 2: Overall system workflow

enabling immediate abnormal-flow alerts even when the application runs in the background.

3.5 Mobile Application

The Flutter mobile application provides an interactive interface for monitoring household water consumption, supporting Secure user authentication, Real-time dashboard, Water usage history, Statistical reports, Alert management, User settings, and Push notifications. Multiple users can create individual accounts using Firebase Authentication while accessing shared water monitoring data from the single installed IoT device.

4 SYSTEM DESIGN AND IMPLEMENTATION

The proposed Smart Water Monitoring System consists of four major components: the IoT sensing layer, cloud platform, Python Context Engine, and Flutter mobile application. Figure 2 illustrates the overall system workflow

4.1 IoT Data Collection

The proposed system employs an ESP32 microcontroller integrated with a YF-S201 water flow sensor to continuously monitor household water consumption. The sensor measures the flow rate by generating pulses proportional to the volume of water passing through the pipeline. The ESP32 converts these pulses into water flow measurements and uploads the collected information to the Firebase Realtime Database through a Wi-Fi connection. The uploaded data include the current flow rate (L/min), daily water usage (L), total water consumption (L), abnormal-flow status generated by the Context Engine, and the timestamp of the latest reading. The ESP32 continuously updates Firebase, allowing the system to monitor water consumption in real time.

4.2 Firebase Cloud Platform

Firebase serves as the central cloud platform for storing and synchronizing data. Firebase Realtime Database stores live sensor data, historical water usage, alert records, user settings, and device notification tokens. Firebase Authentication provides secure user access, while Firebase Cloud Messaging (FCM) delivers real-time push notifications for abnormal water flow events.

4.3 Python Context Engine

The Python Context Engine acts as the intelligent processing unit, continuously monitoring water flow data in Firebase and performing adaptive statistical analysis to identify abnormal water consumption.

It retrieves historical water consumption records, removes invalid sensor readings and previous abnormal-flow alert records, learns normal household water usage patterns, calculates an adaptive statistical threshold, compares current flow rates, creates Firebase alert records, and sends FCM push notifications. The Context Engine periodically refreshes its statistical baseline using newly collected historical data so that the detection threshold can reflect changes in household water-usage patterns.

4.4 Flutter Mobile Application

The Flutter mobile application provides an interface for monitoring household water consumption through Dashboard, History, Reports, Alerts, Profile and Settings modules. It listens to Firebase Realtime Database and automatically updates with new sensor readings and alert records. Users receive real-time push notifications through Firebase Cloud Messaging when abnormal water consumption is detected.

5 STATISTICAL ABNORMA FLOW DETECTION PROCESS

The proposed detection mechanism uses statistical anomaly detection rather than a manually fixed threshold. After valid historical samples are obtained, the system calculates the Mean (μ) and Standard Deviation (σ) and generates an adaptive threshold using Equation (1). The current water flow is then compared with this threshold. If the current flow exceeds the threshold, the system classifies the event as a potential water leak. A Firebase alert record is created, and Firebase Cloud Messaging immediately sends a push notification to the user's mobile device. The Context Engine continues monitoring new sensor readings while preventing duplicate notifications for the same abnormal-flow event using hysteresis and state tracking.

6 EXPERIMENTAL RESULTS AND DISCUSSION

The proposed system was tested using an ESP32 microcontroller connected to a water flow sensor under various household water usage conditions. The Flutter mobile application successfully displayed real-time water consumption, historical records, statistical reports, and alert notifications through Firebase Real time Database.

During normal household water usage, the Python Context Engine learned a statistical baseline from valid historical flow-rate samples, allowing the adaptive threshold to represent the household's normal usage pattern. When water flow exceeded the calculated adaptive threshold, the Context Engine classified the event as an abnormal-flow event that may indicate a potential water leak. A new alert record was created in Firebase, and Firebase Cloud Messaging successfully delivered a push notification to the registered mobile device. The implemented hysteresis mechanism prevented repeated notifications for the same abnormal-flow event until the water flow returned below the recovery threshold.

The implemented statistical method generated the abnormal-flow detection threshold from valid historical flow-rate samples by calculating the mean (μ) and standard deviation (σ), as defined in Equation (1). For example, a mean flow rate of 7.31 L/min and a standard deviation of 4.52 L/min produced a threshold of 16.35 L/min. Based on the exported Firebase data, the evaluation dataset contained 1,000 historical water-flow records, from which 10 statistical abnormal-flow alert events were recorded. The recorded alert events corresponded to flow readings classified by the Context Engine as abnormal relative to the learned thresholds. The learned thresholds varied according to the historical flow-rate data rather than using a single manually assigned value. These results demonstrate that the system can adapt its detection threshold according to historical household water consumption rather than relying on a single manually

fixed value. However, true-positive, false-positive, and false-negative rates could not be independently calculated from the exported dataset because it does not contain manually labelled ground-truth leak and normal-usage test results. Therefore, further controlled normal-usage and simulated-leak tests are required to calculate the detection rate and other classification performance measures.

7 CONCLUSION

This research presents a Smart Water Leak Detection and Consumption Monitoring System integrating IoT, cloud computing, mobile technology, and statistical anomaly detection. The implemented system provides real-time household water monitoring and adaptive abnormal-flow detection using historical consumption patterns. The Flutter mobile application provides monitoring, records, reports, settings, and Firebase Cloud Messaging (FCM) notifications. Experimental observations demonstrated abnormal-flow alert generation, adaptive threshold calculation, and real-time notification delivery.

8 LIMITATIONS

The proposed system has several limitations that should be considered when evaluating its performance. The current implementation uses a single water flow sensor; therefore, it monitors the total household water supply but cannot identify the exact pipe, outlet, or appliance where a leak occurs. The system also depends on a stable Internet and Wi-Fi connection for communication between the ESP32, Firebase, Python Context Engine, and Flutter mobile application.

Statistical leak detection requires sufficient valid historical water usage samples to establish a reliable baseline. Therefore, a newly installed system may require an initial learning period before the adaptive threshold becomes available. The method defined in Equation (1) may also classify unusually high but legitimate water usage as abnormal, while small or slow leaks may remain below the threshold.

Sensor accuracy, calibration, and installation may affect flow-measurement reliability. The evaluation was limited by the experimental dataset and lack of labelled ground-truth data; therefore, detection accuracy, precision, recall, and error rates could not be reliably calculated. The results should be interpreted as abnormal-flow detection rather than validated physical leak detection.

REFERENCES

- Atzori, L., Iera, A., & Morabito, G. (2010). The internet of things: A survey. Computer networks, 54(15), 2787-2805.
- Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. Future generation computer systems, 29(7), 1645-1660.
- Pevnev, V., Zemlianko, H., & Tsuranov, M. (2026). Smart water metering and leak detection systems. In Encyclopedia of Smart and Green Cities (pp. 1–9). Singapore: Springer Nature Singapore.
- Sarker, I. H. (2022). AI-based modeling: techniques, applications and research issues towards automation, intelligent and smart systems. SN computer science, 3(2), 158.
- Tiwari, S., Botre, B. A., Arundhathi, B., Sridhar, S., & da Silva Santos, C. M. Smart Water Metering and Leak Detection. In Smart Technologies for Water Sustainability and Policy (pp. 78-96). CRC Press.

- Zanella, A., Bui, N., Castellani, A., Vangelista, L., & Zorzi, M. (2014). Internet of things for smart cities. IEEE Internet of Things Journal, 1(1), 22–32.

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research - 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

An Explainable Machine Learning Framework for Lead Conversion Prediction and Sales Visit Route Planning in the Automotive Industry

Siriwardhana HKMRA[1], Premarathna WCC[2]
[1,2] *Department of Computing and Information Systems, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*
amasiriwardhana210@gmail.com[1] , chamidu@wyb.ac.lk[2]

ABSTRACT

Automotive sales teams must identify promising enquiries and plan field visits within a fixed working day, yet conventional lead labels and manually prepared routes treat these decisions separately. This study presents an explainable decision-support framework that combines leakage-controlled lead-conversion prediction, route-grouped travel-time prediction, conversion/travel-aware scheduling, and action-oriented explanations. The lead dataset contains 10,000 enquiries, including 3,047 conversions, and the route dataset contains 15,089 visit records from 2,460 officer-day routes. Six classifiers were evaluated using a stratified 80/20 partition; XGBoost achieved the highest holdout ROC-AUC of 0.8193, while Logistic Regression achieved the highest F1-score of 0.6198. Three travel-time regressors were evaluated with routes kept intact across partitions; XGBoost achieved a mean absolute error of 4.355 minutes and an R-squared value of 0.8888. The deployed prototype ranks reachable leads by conversion potential per predicted travel-time unit, generates road-network routes, and supplies feature-attribution and counterfactual explanations. The framework is intended to support, rather than automate, final sales decisions. The contribution is the operational integration of established methods rather than a new algorithm. Current evidence demonstrates technical feasibility only. Repeated cross-validation with uncertainty estimates, quantitative scheduler baselines, ablation tests, subgroup fairness analysis, explanation usability assessment, and a controlled field study are required before claiming statistical superiority or improved sales outcomes.

KEYWORDS: Explainable Artificial Intelligence, Lead Scoring, Route Planning, Travel-Time Prediction, XGBoost

1 INTRODUCTION

Automotive dealerships often receive more enquiries than field officers can visit. Manual Hot/Warm/Cold labels vary by officer, and manually planned routes may minimize neither travel time nor expected sales value. Predictive lead scoring can rank prospects consistently, but a score alone is operationally weak when a customer cannot be reached within the available day. Machine-learning studies show the value of lead prioritization while observing that scoring is rarely connected to downstream sales action (González-Flores et al., 2025; Wu et al., 2023).

This study addresses that gap through one workflow that links conversion probability, predicted travel time, constrained scheduling, and explanation. Its contribution is framework-level and application-level integration: leakage-controlled lead modelling, route-grouped travel evaluation, joint conversion/travel-aware scheduling, action-oriented explanations, and deployment through web and mobile interfaces. The individual algorithms are established; novelty lies in their operational composition for automotive field sales.

2 LITERATURE REVIEW

Tree ensembles remain strong on heterogeneous tabular data because they represent nonlinear relationships and tolerate irrelevant features (Breiman, 2001; Chen & Guestrin, 2016; Grinsztajn et al., 2022). Logistic Regression remains a useful transparent baseline, and probability calibration matters when predicted probabilities determine resource allocation (Niculescu-Mizil & Caruana, 2005).

Vehicle routing, by contrast, is a constrained optimization problem in which route quality depends on travel estimates, working-time limits, and the value assigned to candidate visits (Toth & Vigo, 2014).

SHAP provides global and local feature attribution (Lundberg & Lee, 2017), whereas counterfactual methods identify feasible input changes that would alter a model output (Mothilal et al., 2020). Explanations can improve traceability but do not establish causality, fairness, or appropriate use. Human review remains necessary when financial or personal attributes influence access to sales attention (National Institute of Standards and Technology (NIST), 2023).

3 METHODOLOGY

3.1 Data and leakage control

Masked operational data were provided by a capital leasing and finance company following a research data request. The lead dataset contains 10,000 automotive enquiries, including 3,047 conversions. Predictors include demographic, financial, product, and engagement variables. Identifiers, stored conversion probability, conversion date, and manual Hot/Warm/Cold labels were excluded.

Missing numerical and categorical values were median and mode-imputed, categorical variables one-hot encoded, numerical variables were standardized, and affordability and engagement features were engineered, producing 72 transformed features. A stratified 80/20 split created 8,000 training and 2,000 test records.

The route dataset contains 15,089 visits from 2,460 officer-day routes and 42 transformed features. Explicit post-visit and post-route outcome variables were excluded from travel-time modelling. Entire routes were assigned to train or test sets, yielding 12,065 training and 3,024 test records, with no route overlap. Retained route-level and environmental descriptors were used as supplied. Their historical pre-departure availability could not be independently verified.

3.2 Models, scheduling, and explanation

Lead conversion was modelled with Logistic Regression, Random Forest, XGBoost, shallow and deep multilayer perceptron, and a tabular residual network. Model selection was performed within the training partition; classical models used cross-validation and neural models used an internal validation split. Binary classification metrics used a fixed 0.50 threshold. Travel-time was modelled with XGBoost, a multilayer perceptron, and a tabular residual network and assessed using mean absolute error (MAE), root mean squared error (RMSE), and R^2.

Starting from the district starting location and remaining workday budget, the scheduler scores each unvisited lead as predicted conversion probability divided by predicted travel time.

At each step, candidates whose predicted travel time plus the configured meeting duration exceed the remaining budget are treated as infeasible for that step; the highest-scoring feasible lead is appended, and the calculation repeats.

Meeting duration is user-configurable from 15 to 90 minutes (40 minutes by default). Travel times are point estimates; prediction intervals or uncertainty buffers are not used. Lead predictions are accompanied by feature attribution and scenario-based counterfactual suggestions, which describe model behavior rather than causal recommendations. Fig. 1 illustrates the integrated framework for lead-conversion prediction, travel-time estimation, scheduling, and explanation.

4 RESULTS AND DISCUSSION

4.1 Lead-conversion performance

Table 1 reports performance on the 2,000-record stratified holdout. XGBoost ranked first by ROC-AUC (0.8193), but Logistic Regression followed closely (0.8158) and achieved the highest F1-score (0.6198). All classification metrics used a fixed 0.50 threshold. On the same holdout, Brier scores were 0.1739 for XGBoost and 0.1783 for Logistic Regression; 10-bin reliability analysis gave ECE values of 0.1370 and 0.1472, respectively, and indicated systematic probability overprediction. Thus, XGBoost had modestly lower probability error, but neither model is established as adequately calibrated. The ROC-AUC difference of 0.0035 is too small to establish superiority without repeated evaluation.

Fig. 2 compares the holdout ROC-AUC and F1-scores across the evaluated lead-conversion models.

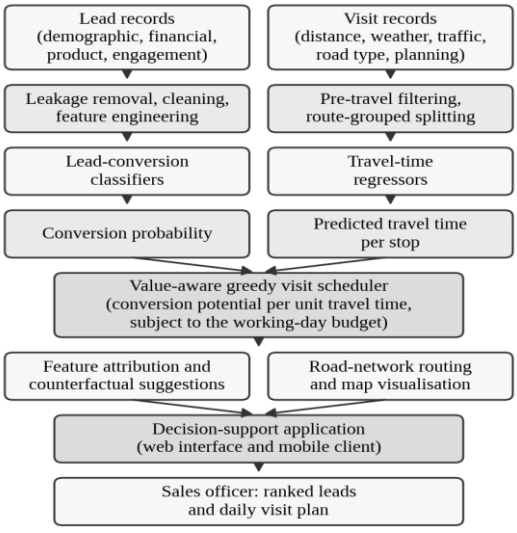


Figure 1: Integrated lead-conversion, travel-time, scheduling, explanation, and deployment framework.

Table 1: Lead-conversion holdout results

Model	ROC-AUC	F1	Recall
XGBoost	0.8193	0.6169	0.6995
Logistic Reg.	0.8158	0.6198	0.7455
Random Forest	0.8106	0.6107	0.6749
MLP-Shallow	0.8076	0.6148	0.7209
MLP-Deep	0.8065	0.6191	0.7274
TabResNet	0.7994	0.5901	0.8013

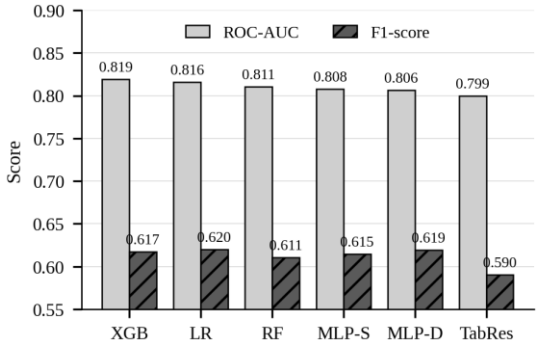


Figure 2: Comparison of holdout ROC-AUC and F1-score across lead-conversion models.

4.2 Travel-time and framework performance

Under the route-grouped split, XGBoost achieved MAE = 4.355 minutes, RMSE = 6.842 minutes, and R^2 = 0.8888. The multilayer perceptron was close (MAE = 4.479; RMSE = 6.849; R^2 = 0.8886), while TabResNet was weaker (MAE = 4.946; RMSE = 7.282; R^2 = 0.8740).

XGBoost was selected because it had the lowest observed error and supported exact tree-based attribution. Table 2 summarizes the route-grouped travel-time prediction results for the evaluated regressors.

The prototype generated feasible daily sequences, feature explanations, road-network routes, and mobile-accessible plans. In a retrospective replay of all 127 eligible district-follow-up-date instances in the saved database, all 502 candidates entering scheduling were placed within the predicted 540-min. workday budget using the default 40-min. meeting duration; no candidates remained budget infeasible at termination. Candidate scheduling coverage, the all-candidates scheduled rate, and predicted time-budget feasibility were therefore 100% under these defaults.

Fig. 3 compares the travel-time prediction errors of the three evaluated

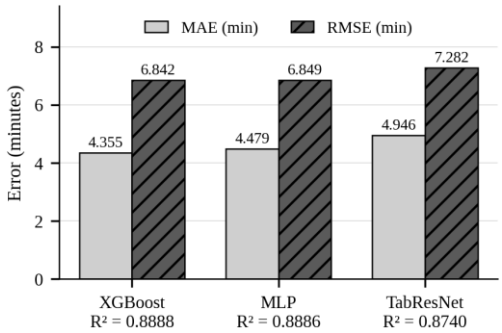


Figure 3: Route-grouped travel-time error for the evaluated regressors.

Table 2: Route-grouped travel-time results

Model	MAE	RMSE	R^2
XGBoost	4.355	6.842	0.8888
MLP	4.479	6.849	0.8886
TabResNet	4.946	7.282	0.8740

regression models. These are model-based planning metrics, not observed visit-completion rates; the evaluation used point travel-time predictions and did not include return-to-base travel, breaks, cancellations, or unforeseen delays.

5 CONFIRMATORY EVALUATION

5.1 Repeated validation and statistical comparison

A confirmatory experiment should repeat the complete preprocessing, feature engineering and hyperparameter-tuning pipeline with ten repetitions of five-fold stratified cross-validation, with all preprocessing and model selection performed within the appropriate training folds to prevent data leakage. It should report the mean, standard deviation, and 95% confidence interval for ROC-AUC, F1, precision, and recall. XGBoost and Logistic Regression should then be compared using a corrected paired test across matched folds. This design would determine whether the small observed ROC-AUC difference is consistent across data partitions rather than a consequence of one split.

5.2 Scheduler baselines and business outcomes

Matched officer-day candidate sets should be scheduled using nearest-neighbor, probability-only, distance-only, the proposed probability/travel-time ratio, and an exact method on small instances.

All scheduling methods should use consistent starting locations, working time budget, candidate sets, and meeting duration assumptions to ensure fair comparison. The current scheduler uses configurable meeting durations and point predictions of travel time without uncertainty intervals or additional time buffers. Future experiments should therefore evaluate sensitivity to meeting duration, travel-time prediction errors, and unexpected delays. Outcomes should include expected conversions per day, completed visits, total travel time, mean travel time per visit, expected conversion value, and percentage of working time spent travelling. Reporting the trade-off frontier is preferable to declaring one policy universally best.

5.3 Ablation, fairness, and user evaluation

Ablation should assess engineered features, engagement variables, and scheduler components, including route-feature availability. Subgroup ROC-AUC, recall, and error rates should be evaluated to identify potential fairness concerns. Sales-officer studies should assess explanation usefulness and inappropriate reliance. Sensitive-data handling should follow Sri Lanka's Personal Data Protection Act (Personal Data Protection Act, 2022).

5.4 Field-evaluation and responsible operation

A controlled field study should compare the framework with existing dealership practices, measuring verified conversions, completed visits, travel time, and schedule adherence. Unreachable leads, cancellations, and delays should be recorded to distinguish predicted feasibility from actual outcomes. Officers should retain control over recommendations. Deployment monitoring should assess calibration, travel-time errors, fairness, and data drift, with safeguards for sensitive data and model review (NIST, 2023).

6 LIMITATIONS

Lead results are single-holdout estimates, and the calibration analysis is descriptive rather than evidence of calibration under temporal or geographic shift. The retrospective scheduler replays measures predicted time budget feasibility, not actual visit completion or business benefit. Travel time scheduling uses point prediction without uncertainty intervals, and the historical pre-departure availability of some retained route-level and environmental descriptors could not be independently verified. Explanation usefulness has not been tested with sales officers. The current evidence therefore supports technical feasibility only; it does not support claims of statistical superiority, improved sales, causal recommendations, or equivalent performance across demographic and socioeconomic groups.

7 CONCLUSION

The study integrates leakage-controlled lead scoring, route-grouped travel-time prediction, conversion/travel-aware scheduling, and action-oriented explanation in one deployable framework. XGBoost obtained a holdout ROC-AUC of 0.8193 for conversion and a route-grouped MAE of 4.355 minutes for travel time. The main contribution is the operational composition of established methods. Repeated validation,

comparative scheduling experiments, ablation, fairness analysis, explanation usability assessment, and a controlled field study are the required next steps.

REFERENCES

- Breiman, L. (2001). Random forests. *Machine Learning, 45(1),* 5-32.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* 785-794.
- González-Flores, L., Rubiano-Moreno, J., & Sosa-Gómez, G. (2025). The relevance of lead prioritization: A B2B lead scoring model based on machine learning. *Frontiers in Artificial Intelligence, 8,* Article 1554325.
- Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on typical tabular data? *Advances in Neural Information Processing Systems, 35.*
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30,* 4765-4774.
- Mothilal, R. K., Sharma, A., & Tan, C. (2020). Explaining machine learning classifiers through diverse counterfactual explanations. *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency,* 607-617.
- National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework (AI RMF 1.0).* NIST AI 100-1.
- Niculescu-Mizil, A., & Caruana, R. (2005). Predicting good probabilities with supervised learning. *Proceedings of the 22nd International Conference on Machine Learning,* 625-632.
- Personal Data Protection Act, No. 9 of 2022, Parliament of the Democratic Socialist Republic of Sri Lanka. (2022).
- Toth, P., & Vigo, D. (2014). *Vehicle routing: Problems, methods, and applications* (2nd ed.). Society for Industrial and Applied Mathematics.
- Wu, M., Andreev, P., & Benyoucef, M. (2023). The state of lead scoring models and their impact on sales performance. *Information Technology and Management.* https://doi.org/10.1007/s10799-023-00388-w

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

A Calibrated Hurdle-Based Machine Learning Framework for Multi-Horizon District-Level Rainfall Forecasting

Wijerathna WSMRN[1], Hansika Gunasekara[2]
[1,2] *Department of Computing and Information Systems, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*
ruvininwijerathna@gmail.com[1] , hansikag@wyb.ac.lk[2]

ABSTRACT

Rainfall forecasting is challenging because rainfall is zero-inflated, right-skewed, and governed by nonlinear temporal dependencies. This study proposes a calibrated hurdle-based CatBoost framework for district-level rainfall forecasting, decomposing prediction into a binary occurrence classifier and a conditional amount regressor for wet-day observations (rainfall >1mm, 58.3% of the 127,825-record dataset). The framework uses fourteen years (2010–2023) of daily climate data from all 25 Sri Lankan districts, compiled from the Department of Census and Statistics and NASA POWER, a satellite-derived reanalysis product. Nineteen forecast horizons (1–130 days) were modelled with isotonic regression and Platt scaling applied by horizon, and CatBoost was compared against Random Forest and XGBoost. On the held-out 2023 test set, the calibrated framework achieved a rainfall-occurrence area under the receiver operating characteristic curve AUC of 0.855 and F1-score of 0.846 at one day, remaining at or above 0.78 across all horizons; the amount regressor achieved a more modest R^2 of 0.17–0.33, reflecting the inherent difficulty of predicting exact quantities from a zero-inflated distribution. A separate, uncalibrated walk-forward validation across five splits produced a more conservative one-day AUC of 0.847 ± 0.021, stabilising at ~0.75 between 14 and 90 days. Decomposing zero-inflated rainfall prediction and integrating horizon-aware calibration provide a robust approach to multi-horizon, district-level rainfall forecasting.

KEYWORDS: CatBoost, Hurdle model, Multi-horizon forecasting, Probability calibration Rainfall forecasting

1 INTRODUCTION

Rainfall is among the most important variables for agricultural planning, water resource management, and disaster preparedness in Sri Lanka, and among the most difficult to forecast. Daily rainfall is zero-inflated and right-skewed, and wet-day amounts depend on strongly nonlinear atmospheric interactions. Conventional agricultural calendars rely on fixed seasonal averages, which smooth over the week-to-week variability that determines whether a field receives adequate water at the right time. District-level, horizon-specific probabilities are therefore needed instead of a single historical norm. Gradient boosted decision trees suit tabular climate data, but most existing models target a single horizon, use a single weather station, or treat rainfall amount as an ordinary continuous target, ignoring its zero-inflated structure. This study addresses these gaps with a calibrated, hurdle-based CatBoost framework that (i) separates rainfall occurrence from rainfall amount, (ii) is trained and tested across nineteen horizons from 1 to 130 days, (iii) calibrates probabilities and selects decision thresholds per horizon, and (iv) is validated across all 25 districts using fourteen years (2010–2023) of daily data from the Department of Census and Statistics and NASA POWER.

2 LITERATURE REVIEW

Machine learning has previously been applied to climate forecasting in Sri Lanka, but only with a narrow scope. Hennayake et al. (2022) used long short-term memory networks to forecast temperature and precipitation for a single Colombo station, a useful proof of concept that does not establish performance across districts or longer horizons. Crop-simulation studies of sowing-date timing, such as Amarasingha et al. (2014), use process-based models at specific research sites; this study instead compiles district-level daily climate records, including paddy statistics from the Department of Census and Statistics Sri Lanka (2024), across all 25 districts.

Gradient boosting has repeatedly proven effective for rainfall prediction. Abdullah and Said (2025) found that CatBoost outperformed Random Forest and Support Vector Regression for daily and monthly rainfall in India, while Hsu et al. (2024), in a single-station comparison in Taiwan, reported rain-occurrence AUC of 0.77 for Random Forest versus 0.73 for CatBoost. CatBoost's ordered boosting and categorical-feature handling (Prokhorenkova et al., 2018) suit the mixed features used here. These studies support CatBoost as a strong rainfall classifier, but none addresses multiple horizons, calibrates probabilities, or tests generalisation across a wide spatial area.

Daily rainfall is zero-inflated and therefore unsuitable for modelling as an ordinary continuous variable. A two-part "hurdle" model (Mullahy, 1986) handles this: one model predicts whether it rains, and a second, applied only to wet days, predicts how much. Raw classifier scores are generally poorly calibrated, so this study applies Platt scaling (Platt, 1999) and isotonic regression (Zadrozny & Elkan, 2002), the two most common methods. Few operational forecasting studies in South Asia combine a hurdle structure with horizon-specific calibration extending into the sub-seasonal range (Vitart et al., 2017); this study, tested across all 25 districts and nineteen horizons, addresses that gap.

3 METHODOLOGY

3.1 Data and Feature Engineering

Daily climate data for all 25 districts (2010–2023) were compiled from the Department of Census and Statistics and NASA POWER, yielding 127,825 district-day records across 16 variables. A wet day was defined as rainfall exceeding 1 mm; across the full cleaned dataset, 58.3% of district-days were wet and 41.7% dry. Leakage-safe feature engineering ensured that no feature used information unavailable at the forecast issue date, producing lag features, rolling-window statistics, wet-spell and seasonal indicators, and district-level climatological normal. The categorical rain-intensity feature was derived from the previous day's rainfall, so it carries no same-day information. The final dataset comprised 817 input features across 118,700rows, partitioned chronologically into training (2011–2021, eleven years), validation (2022) and held-out test (2023) sets; the fourteen-year span (2010–2023) therefore refers to the full dataset and eleven years to the training partition.

3.2 Hurdle-Based Forecasting Framework

Random Forest, XGBoost and CatBoost were first compared on rain occurrence classification at four

representative horizons; CatBoost achieved the highest validation AUC and was adopted for the full framework. At each of the nineteen horizons (1–130 days), rainfall prediction was structured as a two-part hurdle model (*Mullahy, 1986*): a binary classifier estimated the probability of rainfall occurrence, and a separate conditional regressor, trained only on wet-day observations, predicted the corresponding amount. Because heavier rainfall events are considerably less frequent than light ones, sample weighting (1×, 3× and 8×) prevented the model from underrepresenting high-magnitude cases during training. Raw classifier probabilities tend to be poorly calibrated, particularly at longer horizons, so calibration was horizon-dependent: isotonic regression *(Zadrozny & Elkan, 2002*) up to 30 days and Platt scaling (Platt, 1999) beyond. The decision threshold was selected separately for each horizon by maximising the validation F1-score instead of using the conventional 0.5 cut-off.

3.3 Validation Strategy

Model performance was first assessed on a chronologically held-out 2023 test set (9,125 district-day rows). To confirm that the reported performance was not an artefact of a single favourable train-test split, an independent walk-forward validation was subsequently performed across five chronological splits, evaluating the rain-occurrence classifier at a fixed, uncalibrated threshold of 0.50 on each split.

4 RESULTS AND DISCUSSION

4.1 Model Selection and Forecasting Performance

CatBoost was selected over Random Forest and XGBoost on rain-occurrence AUC at four representative horizons on the 2022 validation set (Table 1), achieving the highest AUC at all four. On the held-out 2023 test set (9,125 district-day rows), the calibrated, horizon-optimised framework achieved a rain-occurrence AUC of 0.855 and an F1-score of 0.846 at one day, remaining at or above approximately 0.78 across all nineteen horizons (Table 2). The rainfall-amount regressor achieved a more modest R^2 of 0.17–0.33, generally declining with horizon, consistent with the inherent difficulty of predicting exact daily amounts from a zero-inflated distribution at longer lead times; correlation and mean absolute error are more informative than R^2 here. The continuous amount output used in downstream ranking should therefore be read as directionally informative rather than precise. This is a genuine limitation, but it reflects a well-documented property of zero-inflated rainfall data rather than a deficiency specific to this model.

Table 1: Model Selection Validation-Set AUC by Algorithm and Horizon

Horizon (days)	Random Forest	XGBoost	CatBoost
1	0.7898	0.8103	0.8188
7	0.7134	0.7265	0.7390
30	0.7286	0.7250	0.7341
85	0.7136	0.7117	0.7264

Table 2: Model Performance Results for Rainfall

Horizon (days)	Rain AUC	Rain F1	Rain RMSE	Rain R^2
1	0.855	0.846	8.113	0.328
7	0.792	0.815	8.917	0.193
30	0.792	0.831	9.205	0.171
85	0.798	0.843	9.428	0.178
130	0.834	0.853	9.66	0.207

4.2 Walk-Forward Validation

Independent walk-forward validation across five chronological splits, using a fixed, uncalibrated decision threshold of 0.50 at each horizon, confirmed that one-day rain-occurrence AUC was reasonably stable across different training periods (0.847 ± 0.021), stabilising around 0.75 between the 14- and 90-day horizons (Table 3). This more conservative, uncalibrated figure should not be read as conflicting with the calibrated, threshold-optimised test-set AUC of 0.855 reported in Section 4.1: the two use deliberately different evaluation conditions, and the walk-forward figure is the more conservative estimate, included specifically to avoid overstating performance from a single favourable train-test split.

4.3 Benchmark Comparison

In contextual comparison against two published studies, the proposed framework achieved lower error than Hennayake et al. (2022) on Colombo district (rainfall RMSE 8.59 mm versus 16.91 mm) and a higher rain-occurrence AUC than Hsu et al. (2024) across all 25 districts (0.855 versus 0.77 and 0.73 on their single-station dataset). These comparisons are contextual rather than like-for-like: the compared studies differ from this one in study period, spatial coverage, forecast horizon, and evaluation protocol, and the figures should be read as indicating that the proposed framework is competitive with, rather than strictly superior to, prior work.

4.4 Feature Importance

For the one-day-ahead rain-occurrence classifier, the rain-intensity feature is the most important predictor (22.5% of total importance), more than three times the next-ranked feature, the training-period climatological rainfall mean (6.2%). Because this feature is built from the previous day's rainfall rather than same-day values (Section 3.1), its importance does not reflect target leakage. Cyclical day-of-year encoding, wind direction and season together contribute about 9%, and the remainder is spread thinly across the lag, rolling and interaction features.

5 CONCLUSION

This study addressed the challenge of rainfall prediction in Sri Lanka's operational climate forecasting, which is difficult because of zero-inflated data, skewed rainfall magnitudes and complex atmospheric dynamics. Whereas most

Table 3: Walk-Forward Validation Rain Occurrence AUC

Horizon (days)	Split 1	Split 2	Split 3	Split 4	Split 5	Mean ± SD
1	0.867	0.869	0.838	0.839	0.819	0.847 ± 0.021
3	0.786	0.825	0.792	0.770	0.759	0.786 ± 0.025
7	0.752	0.806	0.776	0.755	0.726	0.763 ± 0.030
14	0.754	0.797	0.763	0.757	0.726	0.760 ± 0.025
30	0.737	0.787	0.748	0.752	0.733	0.751 ± 0.021
60	0.725	0.797	0.718	0.763	0.730	0.747 ± 0.033
90	0.753	0.798	0.704	0.761	0.729	0.749 ± 0.035

existing work focuses on single locations or a single horizon, this study evaluated a calibrated, hurdle-based CatBoost framework across all 25 districts, using fourteen years (2010–2023) of daily climate data and nineteen horizons from 1 to 130 days. Decomposing prediction into a binary occurrence classifier and a conditional amount regressor handled the zero-inflated data effectively, although the amount regressor's modest R^2 (0.17–0.33) remains a genuine limitation. With horizon-aware calibration, the model achieved a one-day rain-occurrence AUC of 0.855 on the test set and stayed above approximately 0.78 up to 130 days. A stricter walk-forward validation (uncalibrated, fixed threshold) gave a more conservative but stable one-day AUC of 0.847 ± 0.021, settling around 0.75 between 14 and 90 days. These results are comparable to two published single-station benchmarks. Future work should add climatology and persistence baselines with Brier-score verification, live climate feeds, irrigation covariates, and agronomic decision rules to help farmers in rainfed districts time sowing.

REFERENCES

- Abdullah, M., and Said, S. (2025). Performance evaluation of machine learning regression models for rainfall prediction. Iranian Journal of Science and Technology, Transactions of Civil Engineering, 49, 4231–4250.
- Amarasingha, R. P. R. K., Galagedara, L. W., Marambe, B., Silva, G. L. L. P., Punyawardena, R., Nidumolu, U., Howden, M., and Suriyagoda, L. D. B. (2014). Aligning sowing dates with the onset of rains to improve rice yields and water productivity. Tropical Agricultural Research, 25(3), 277–286.
- Department of Census and Statistics Sri Lanka. (2024). Paddy statistics: 2023/2024 Maha season. Department of Census and Statistics.
- Hennayake, K. M. S. A., Dinalankara, R., and Mudunkotuwa, D. Y. (2022). Machine learning-based weather prediction model for short term weather prediction in Sri Lanka. International Journal of Multidisciplinary Studies, 9(1), 65–87.
- Hsu, S., Sharma, A. K., Tanone, R., and Ye, Y. (2024). Predicting rainfall using Random Forest and CatBoost models. Proceedings of the 9th World Congress on Civil, Structural, and Environmental Engineering, Paper ICGRE 146.
- Mullahy, J. (1986). Specification and testing of some modified count data models. Journal of Econometrics, 33(3), 341–365.
- Platt, J. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In A. Smola, P. Bartlett, B. Schölkopf, & D. Schuurmans (Eds.), Advances in large margin classifiers (pp. 61–74). MIT Press.
- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. Advances in Neural Information Processing Systems, 31, 6639–6649.
- Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., Déqué, M., Ferranti, L., Fucile, E., Fuentes, M., et al. (2017). The Subseasonal to Seasonal (S2S) Prediction Project database. Bulletin of the American Meteorological Society, 98(1), 163–173.
- Zadrozny, B., & Elkan, C. (2002). Transforming classifier scores into accurate multiclass probability estimates. Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 694–699

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

ExamScope: AI-Based Syllabus Coverage Analysis System Using a Hybrid Large Language Model and Machine Learning Classification

Prathapage LL[1], Jayakodi JRKC[2]
[1,2] *Department of Computing and Information Systems, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*
lahirulakshithaprathapage@gmail.com[1] , kithsirij@wyb.ac.lk[2]

ABSTRACT

Checking whether examination papers properly reflect the course syllabus has long been a tedious, manual effort, one that remains inherently subjective and hard to maintain across multiple academic departments. To solve this, we introduce ExamScope, a hybrid AI system built to streamline and automate syllabus coverage analysis. Our approach pairs the flexible document-parsing strengths of a Large Language Model (LLM) with the deterministic consistency of a supervised, pairwise Multi-Layer Perceptron (MLP) classifier. Using one-shot prompting, the LLM converts unstructured syllabus text into a clean, machine-readable topic hierarchy while segmenting past exam papers into discrete questions. Candidate question–topic pairs are transformed into dense vector representations via a pre-trained sentence-transformer and scored by the trained pairwise MLP matcher. By decoupling document layout parsing from the final classification step, ExamScope produces auditable, reproducible results that can be independently evaluated. When tested on a ground-truth dataset of 398 questions across 9 academic subjects, the pairwise matcher reached an 87.2% pair-classification accuracy, a ROC-AUC of 0.918, and a top-1 topic ranking accuracy of 65.0%. In end-to-end evaluations, the system's calculated coverage percentages aligned perfectly with manual benchmark assessments. Processing takes roughly 10 seconds for native digital PDFs, extending to 40 seconds for scanned documents requiring optical character recognition. Instead of serving as an unyielding automated judge, ExamScope acts as a human in the loop decision support tool, surfacing transparent, confidence scored analytics that help educators refine their assessments. This work delivers a fully functional software system, a pairwise matching framework, and a reusable hybrid architecture for educational text analysis.

KEYWORDS: Large Language Models, Multi-Layer Perceptron, Natural Language Processing, Sentence Embeddings, Syllabus Coverage Analysis

1 INTRODUCTION

Examination papers remain one of the most fundamental tools for measuring how well students have grasped a specific academic curriculum. While a syllabus clearly outlines the expected learning outcomes and the intended distribution of topics, ensuring that actual exams faithfully reflect that blueprint is still surprisingly difficult. Traditionally, verifying this alignment is a purely manual task and a painfully slow one. It requires curriculum reviewers to painstakingly check past exam questions against syllabus entries by hand. This subjective process frequently leads to inconsistent evaluations, with results varying depending on who is doing the review or which academic year is being analyzed. This issue becomes particularly acute during accreditation reviews, when institutions face immense pressure to produce rigorous, reproducible evidence that their assessments actually match their stated curricula.

On the surface, recent leaps forward in Natural Language Processing (NLP) and Large Language Models (LLMs) seem like an obvious solution, given their

impressive ability to parse unstructured text. However, relying purely on generative LLMs for final decision-making is problematic. These models are inherently stochastic, meaning they can provide different answers on different runs, which makes their outputs hard to audit and impossible to "retrain" on specialized institutional data. On the other hand, older techniques like keyword matching or basic vector-space models simply aren't sophisticated enough. They consistently struggle to handle the semantic nuances of academic questioning, often failing to distinguish between conceptually similar subtopics. To bridge this gap, this paper introduces ExamScope, a hybrid AI system built specifically to automate syllabus coverage analysis. ExamScope strategically merges these two technologies: it uses an LLM (guided by one-shot prompting) to handle the messy work of parsing documents and segmenting questions, and then hands off to a supervised Multi-Layer Perceptron (MLP) classifier, which operates on dense sentence embeddings to make reliable, deterministic topic assignments.

2 LITERATURE REVIEW

Most research in automated examination processing has centered on generating questions automatically or grading student answers (Bhadouria et al., 2024; Ayaan & Ng, 2025). By contrast, checking syllabus coverage remains a surprisingly neglected application of natural language processing, traditionally relying on crude keyword matching or basic vector-space similarity (Lukwaro et al., 2024). While large-scale curriculum mapping projects have proven that pedagogical content can be successfully extracted at scale, they consistently point to a major bottleneck: the difficulty of assembling gold-standard datasets for precise semantic alignment (Javadian Sabet et al., 2024).

A practical coverage system must handle both native digital PDFs and scanned, image-based documents. Modern document processing increasingly pairs Optical Character Recognition (OCR) with LLM-based interpretation to parse intricate layouts (Paudel et al., 2024). Furthermore, studies in prompt engineering demonstrate that supplying even a single example one-shot prompting dramatically cuts down structural extraction errors compared to zero-shot instructions when converting unstructured prose into clean JSON formats.

For classifying short text segments, pre-trained models like all-MiniLM-L6-v2 generate dense 384-dimensional sentence embeddings that far surpass old bag-of-words approaches at capturing underlying context (Reimers & Gurevych, 2019). Attaching a shallow neural network, such as a Multi-Layer Perceptron (MLP), to frozen sentence embeddings yields a system that is fast, fully auditable, and deterministic at inference time, crucial qualities for academic assessment (Piao, 2021). Finally, research into human factors underlines that educational AI tools must remain transparent and reviewable if they are to gain genuine trust among educators (Al-Abdullatif, 2024; Ching & Jamaludin, 2025).

3 METHODOLOGY

The research follows a design and development methodology to build the ExamScope artefact. The process is divided into seven major phases, moving from raw document data to the final evaluation. This section explains the detailed procedures for text extraction, the hybrid strategy for document structuring and segmentation, and the design and training of the core Multi-Layer Perceptron (MLP) classification pipeline.

3.1 System Architecture Overview

Before diving into the specific phases, Figure 1 provides a block diagram illustrating the integration and data flow across the proposed ExamScope system. This architecture is designed for security and scalability, separating native digital processing (client-side) from resource-intensive Optical Character Recognition (OCR) fallback operations (server-side).The overall data flow proceeds through these core stages: users upload documents, the extraction layer converts files to plain text, the structuring and segmentation layer creates machine-readable units (topics and questions), the MLP topic matcher performs the final classification, and the analysis layer computes and presents the coverage metrics.

3.2 Document Ingestion and Extraction

The primary inputs required are a syllabus document and one or more past examination papers for a given course, provided in PDF format. The system supports two distinct extraction paths based on the document's nature.

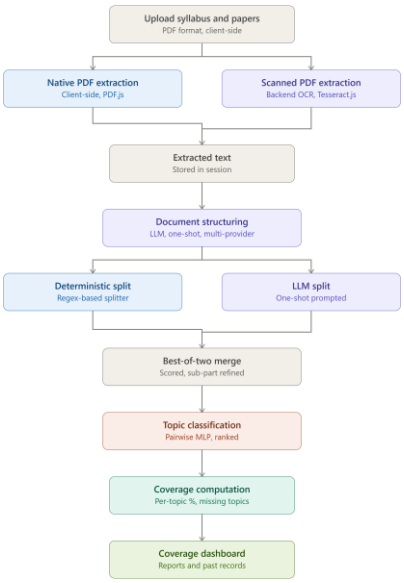


Figure 1: Proposed System Architecture Integration Block Diagram

Native PDF Path (Client-Side): For native, text-based PDFs, text extraction occurs entirely in the user's browser using the PDF.js library. The system reads the PDF as an array buffer and iterates through each page, concatenating the text layer and normalizing whitespace. This approach ensures raw document files for native PDFs never leave the user's browser, supporting the system's security model.

Scanned Document Fallback (Server-Side): When a PDF page does not contain an extractable text layer, it is base64-encoded and transmitted to the backend. A server-side Python rendering script converts the page to an image, and Tesseract.js (the Node.js port of the Tesseract OCR engine) is invoked to perform Optical Character Recognition. The recovered text is returned to the client and persisted only transiently on the server to prevent data leakage. Only the extracted plain text is stored on the server as an Extraction Session, not the original PDF files.

3.3 Syllabus and Question Structuring (LLM with One-Shot Prompting)

To convert unstructured PDF text into machine-readable JSON schemas, ExamScope utilizes an LLM (Google Gemini by default, with Groq and DeepSeek fallbacks). Rather than zero-shot prompting, which often yields unpredictable keys or extra conversational prose we employ one-shot prompting by providing an explicit system rule along with a single input-output example.

Syllabus Parsing Prompt & Example:

- System Directive: "Extract syllabus units, topic names, and subtopics into JSON. Output strictly valid JSON with no markdown wrapping or preamble."
- Demonstration Input: "Unit 2: Data Structures. Cover Arrays, Linked Lists, and Binary Search Trees."
- Demonstration Output: {"topics": [{"name": "Data Structures", "subtopics": ["Arrays", "Linked Lists", "Binary Search Trees"]}]} Question Segmentation Prompt & Example:
- System Directive: "Segment past paper text into distinct questions and sub-questions. Preserve question numbers and mark allocations."
- Demonstration Input: "1. (a) Define a binary search tree. [5 Marks] (b) Differentiate between array and linked list implementation. [5 Marks]"
- *Demonstration Output:* [{"id": "1a", "text": "Define a binary search tree.", "marks": 5}, {"id": "1b", "text": "Differentiate between array and linked list implementation.", "marks": 5}]

To maximize segmentation reliability, this LLM output is run in parallel with a deterministic regex splitter. Both outputs are scored based on question count and sub-part separation, selecting the optimal result via a best-of-two merge strategy.

3.4 ML Pipeline Design - Pairwise MLP Topic Matcher

The core of the methodology is the supervised question-to-topic classification pipeline, trained on the specific subjects (e.g., Business and Industry Law) the system will analyze. This is implemented as a binary pairwise matcher rather than a traditional multi-class softmax classifier. This architectural choice is made for flexibility: it simplifies the task to a match/non-match probability per candidate topic, smoothing the class imbalance found in syllabus datasets.

Feature Extraction (Sentence Embeddings): Before classification, each question and each candidate topic/subtopic text are converted into a fixed-length numerical vector using the pre-trained Sentence-Transformer model (sentence-transformers/all-MiniLM-L6-v2). This generates normalized 384-dimensional dense semantic embeddings, capturing meaning beyond exact keyword overlap.

Pairwise MLP Architecture: For a given question u and a candidate topic v, a single 1,536-dimensional pairwise feature vector is constructed by concatenating four components: the question embedding, the topic embedding, their element-wise absolute difference, and their element-wise product.

$$[\, u \mid v \mid |u-v| \mid u \odot v \,]$$

This combined vector is fed into a shallow Multi-Layer Perceptron (MLP Classifier) trained using the scikit-learn library.

Offline Training and Dataset Construction: A ground-truth training dataset was manually labelled from real academic past papers and syllabus topics. Positive pairwise training examples consist of (Question, True Topic) pairs. For each positive pair, multiple negative training examples are generated by pairing the same question against other topics within the same subject. Training is conducted with a GroupShuffleSplit (80/20 train/test ratio) grouped by the original question text to ensure no data leakage between splits.

Live System Inference and Topic Selection: During analysis, every segmented question is matched against every possible syllabus topic/subtopic candidate for the subject. Each candidate pair is scored by the trained MLP. To improve resilience, this match probability is blended with raw cosine similarity between the embeddings. The candidate with the highest blended score is selected. If the selected candidate is a theory/subtopic, it is mapped back to its parent syllabus topic for final covered/missing status reporting.

4 RESULTS AND DISCUSSION

The pairwise MLP classifier was evaluated on a labelled dataset of 398 questions across 9 academic subjects (split into 2,242 held-out test pairs and 80 top-1 ranking questions). To enhance system generalizability:

Current Scope: High accuracy is demonstrated within technical CS/IT course modules.

Dataset Expansion: Ongoing work focuses on increasing labelled question-topic pairs across non-STEM disciplines to improve cross-field generalizability and reduce dataset scarcity.

4.1 Classification Performance

The pair-level binary classification achieved an overall accuracy of 87.2% and a ROC-AUC score of 0.918 across 2,242 held-out test pairs. Detailed performance metrics are summarized in Table 1.

Table 1. Pairwise classification metrics (held-out test set)

Metric	Non-Match	Match
Precision	0.870	0.878
Recall	0.963	0.650
F1-Score	0.914	0.747
Support	1,590	652

Top-1 topic ranking accuracy reached 65.0% on a held-out evaluation sample of 80 questions. While pairwise scoring discriminates strongly between match and non-match pairs (ROC-AUC = 0.918), top-1 ranking represents a significantly harder task: correctly ranking the true topic above all subject candidates. This 65.0% accuracy indicates that roughly one in three questions is classified under a different topic than a human labeler would assign, reinforcing that the system should not be treated as a fully autonomous grader. Furthermore, because top-1 evaluation was restricted to 80 held-out questions due to dataset availability, these figures serve as an initial benchmark rather than a statistical guarantee across all academic domains.

4.2 Pipeline Execution, Case Analysis, and Latency

Latency benchmarks showed that native digital PDFs process end-to-end in approximately 9.58 seconds, whereas scanned papers requiring backend OCR take roughly 40.45 seconds due to image rendering overhead.

In initial end-to-end case trials on representative papers (e.g., Machine Learning and Information Security modules), automated topic coverage percentages matched human ground truth with a topic-set Jaccard similarity of 1.00. This indicates that aggregate coverage computation is somewhat resilient to minor individual question misclassifications, provided at least one question per topic is correctly identified. However, these preliminary case trials represent ideal, high-coverage conditions; performance on papers with heavily skewed topic distributions or ambiguous multi-part questions requires broader validation.

4.3 Limitations and Generalizability

Several methodological and practical limitations must be noted:

Training Data Scarcity and Domain Bounds: The classifier was trained on 398 questions spanning 9 technical subjects. Performance is heavily dependent on topic sample density. For subjects outside the training distribution or fields with highly overlapping subtopic terminology (e.g., adjacent IT modules), classification confidence decreases.

Evaluation Sample Size: Top-1 ranking performance was evaluated on 80 held-out questions. While pair-level evaluation was conducted on 2,242 pairs, broader cross-institutional validation with larger evaluation sets is necessary to confirm domain-independent generalizability.

Topical vs. Pedagogical Alignment: ExamScope measures topical coverage presence (whether a topic is tested) but does not currently evaluate cognitive depth (e.g., Bloom's Taxonomy levels) or answer quality

5 CONCLUSION

ExamScope demonstrates that combining an LLM for flexible document parsing with a supervised pairwise MLP classifier provides an evaluable, efficient framework for automated syllabus coverage analysis. Achieving an 87.2% pair classification accuracy and a 65.0% top-1 ranking accuracy on held-out test data, the system successfully reduces manual cross-referencing workload while maintaining transparency. Given the sample constraints of 80 held-out evaluation questions and limited course trials, ExamScope is positioned as a human-in-the-loop decision-support tool rather than an autonomous judge. Future work will focus on expanding multi-institutional dataset sizes, fine-tuning domain-adapted sentence embeddings, implementing independent subtopic metrics, and incorporating cognitive depth analysis.

REFERENCES

- Al-Abdullatif, A. M. (2024). Modeling teachers' acceptance of generative artificial intelligence use in higher education: The role of AI literacy, intelligent TPACK, and perceived trust. *Education Sciences*, 14(11), Article 1209. https://doi.org/10.3390/educsci14111209
- Ayaan, A., & Ng, K.-W. (2025). Automated grading using natural language processing and semantic

analysis. *MethodsX*, 12, Article 103395. https://doi.org/10.1016/j.mex.2025.103395

- Bhadouria, A., Gupta, P., Bindal, P., Madan, K., & Sonal, S. (2024). Automated examination system using machine learning and natural language processing. In *Proceedings of the 2024 Sixteenth International Conference on Contemporary Computing (IC3-2024)* (pp. 1–10). Association for Computing Machinery. https://doi.org/10.1145/3675888.3676144
- Ching, K. W., & Jamaludin, K. A. (2025). Understanding school teachers' acceptance of AI in education: Insights from the Technology Acceptance Model (TAM). *International Journal of Academic Research in Progressive Education and Development*, 14(3), 564–579. https://doi.org/10.6007/IJARPED/v14-i3/22415
- Javadian Sabet, A., Bana, S. H., Yu, R., & Frank, M. R. (2024). Course-Skill Atlas: A national longitudinal dataset of skills taught in U.S. higher education curricula. *Scientific Data*, 11, Article 1086. https://doi.org/10.1038/s41597-024-03931-8
- Lukwaro, E. A. E., Kalegele, K., & Nyambo, D. G. (2024). A review of NLP techniques and associated challenges in extracting features from education data. *International Journal of Computing and Digital Systems*, 16(1), 961–979. https://doi.org/10.12785/ijcds/160170
- Paudel, P., Khadka, S., G.C., R., Shah, R., & Joshi, B. (2024). *Optimizing Nepali PDF extraction: A comparative study of parser and OCR technologies* (arXiv Preprint No. arXiv:2407.04577). arXiv. https://doi.org/10.48550/arXiv.2407.04577

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research –2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Investigating the Suitability of Machine Learning Classification of Large Language Model -Generated Descriptions for Real-Time Accident, Crime, and Disaster Reporting

Bandara DMDA[1], Jayakody JRKC[2]
[1,2]*Department of Computing and Information Systems, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*
ashandissanayaka2000@gmail.com[1], kithsirij@wyb.ac.lk[2]

ABSTRACT

Emergency response in Sri Lanka, as in many developing countries, still depends on a witness telephoning an authority to report an accident, a crime, or a disaster. Under stress, callers find it difficult to describe the situation and its location accurately, which delays the response of the police, hospitals, and city authorities. This study investigates whether machine learning classifiers are suitable for categorizing natural-language incident descriptions generated by a vision-language large language model (LLM) from a single photograph. In the proposed pipeline, a mobile client sends the photograph to the Google Gemini model through a secure API call and receives a text description; a back-end service then classifies the description as Accident, Crime, or Disaster. Five classifiers (K-Nearest Neighbours, Naïve Bayes, Support Vector Machine, Decision Tree, and Random Forest) were trained on TF-IDF features extracted from a labelled dataset of 1,500 incident descriptions (500 per class) and evaluated on a stratified 20% hold-out set. Random Forest performed best, with an accuracy of 0.87 and an F1-score of 0.88, and the complete pipeline from image capture to nearest-responder lookup averaged 1.84 seconds. These results indicate that pairing a vision-language model with a lightweight text classifier is technically feasible for automated incident reporting. They were obtained on a curated dataset and without end-user testing, so real-world effectiveness remains to be established.

KEYWORDS: Classification, Gemini Vision-Language, Large Language Model, Random Forest, Term Frequency–Inverse Document Frequency

1 INTRODUCTION

In Sri Lanka, as in many developing countries, emergency reporting still relies on a member of the public describing an incident and its location to an authority by telephone. High stress impairs eyewitness recall (Deffenbacher et al., 2004), which is precisely the condition of a person reporting a crash, a crime, or a collapsing hillside. Location is a second difficulty: many rural roads and settlements, including those in landslide- and flood-prone areas repeatedly affected by severe floods (Post-Disaster Needs Assessment Secretariat, 2025), have no formal street names. Smartphones with cameras, the global positioning system (GPS), and internet access are now widespread, and cloud-hosted multimodal models such as Gemini (Gemini Team, 2023) can turn a photograph into a natural-language description within seconds. A phone can therefore supply both the location and a description of the scene.

Existing automated approaches address only parts of this problem. Camera-based detectors identify a single event type, such as weapons (Pasi et al., 2025) or road collisions (Tamagusko et al., 2022); sensor-based systems need dedicated vehicle hardware (Pathik et al., 2022); and crisis-informatics research analyses social-media text and images after an event (Alam et al., 2018; Imran et al., 2015). Few studies bring large language

model (LLM) generated scene descriptions and incident classification together. This study aims at filling this gap by analyzing how machine learning algorithms can classify incident descriptions provided by a vision-language model from a single image. The contributions are a five-step reporting pipeline and a controlled comparison of five classifiers on a common feature space, addressing two research questions: (RQ1) which of five widely used classifiers most accurately assigns a Gemini-generated description to Accident, Crime, or Disaster, and (RQ2) whether the complete pipeline runs fast enough for real-time use.

2 LITERATURE REVIEW

2.1 Conventional and infrastructure-based detection

Only for fixed roadside cameras. Pathik, Traditional reporting needs no extra infrastructure but depends on the caller's ability to say what occurred and where. Closed-circuit television (CCTV) with object detection removes part of this burden but works only where cameras are installed. Pasi et al. (2025) used the You Only Look Once (YOLO) framework (Redmon et al., 2016) to detect weapons and alert law enforcement, and Tamagusko et al. (2022) applied transfer learning to traffic-camera images to detect
Collisions; each address one event type. Pathik et al. (2022) combined vehicle Internet of Things (IoT) sensors with deep learning but depend on equipped vehicles, and Nakayenga et al. (2024) observed that most artificial intelligence (AI) public-safety systems address a single hazard.

2.2 Citizen-generated and multimodal crisis data

Imran et al. (2015) surveyed the processing of social-media messages in mass emergencies, and Alam et al. (2018) released CrisisMMD, a multimodal Twitter dataset from natural disasters. This work processes content already posted online and does not support direct reporting to a responder. Multimodal models learn joint image–text representations, as in Contrastive Language–Image Pre-training (CLIP) (Radford et al., 2021), and models such as Gemini can generate free-text image descriptions, avoiding a dedicated image classifier for each incident type.

2.3 Text classification methods

Automated text categorization is a mature field (Sebastiani, 2002). Text is commonly converted to term frequency–inverse document frequency (TF-IDF) weights (Salton & Buckley, 1988). Naïve Bayes assumes word independence (McCallum & Nigam, 1998) and trains quickly; K-Nearest Neighbours (KNN) (Cover & Hart, 1967) labels an example by its nearest neighbours but scales poorly at prediction time; Support Vector Machines (SVM) (Cortes & Vapnik, 1995) maximize the margin between classes and suit high-dimensional text (Joachims, 1998); a Decision Tree (Quinlan, 1986) is easy to interpret but prone to overfitting, which Random Forest (Breiman, 2001) addresses by combining many trees.

3 METHODOLOGY

3.1 System Architecture

The pipeline has five sequential steps, shown in Fig. 1. In Step 1



Fig. 1: Five-step system architecture and data flow.

(Capture), a Flutter mobile client takes a photograph and, with the user's permission, reads the device's GPS coordinates. In Step 2 (Describe), the photograph is sent through a Python (Flask) back end, over a secure application programming interface (API) call, to the Gemini vision model, which returns a natural-language description. In Step 3 (Classify), the description is converted to a TF-IDF vector (5,000 features, English stop words removed) and classified as Accident, Crime, or Disaster by the Random Forest model, serialized with Joblib. In Step 4 (Locate responder), the coordinates are sent to the Google Maps Places API to find the nearest police station, hospital, and city council office; all three are returned regardless of the predicted category, so the user can still choose the right responder if the classification is wrong. In Step 5 (Review and dispatch), a dashboard shows the description, category, and suggested responder; the user may edit them and then confirms, and the payload is sent by an API call where the responder's infrastructure allows it, or otherwise as a short message service (SMS) message through the Twilio messaging API. Human review is deliberate: every AI output is treated as a draft.

3.2 Dataset description

The models were trained and tested on a labelled dataset of 1,500 incident records, each with a title, a free-text description, and a label (Accident, Crime, or Disaster). The classes are balanced at 500 records each (Table 1).

3.3 Classification models and training

Five classifiers were trained: KNN, Naïve Bayes, SVM, Decision Tree, and Random Forest. All were implemented with scikit-learn (Pedregosa et al., 2011) on the same TF-IDF feature space with identical vectorization parameters, so performance differences can be attributed to the classifier alone. Accuracy, precision, recall, and F1-score were used, because accuracy alone can hide uneven performance across classes.

3.4 sssssSelection of the vision-language model

Gemini was compared with DeepSeek as the description source. Both received the same inputs, and the outputs were scored by the Flesch Reading Ease

Table 1: Dataset summary

Property	Value
Total records	1,500
Classes	Accident, Crime, Disaster
Records per class	500 (balanced)
Input text	Title + description, merged
Training set (80%)	1,200 (400 per class)
Test set (20%)	300 (100 per class)
Split method	Stratified sampling

Table 2: Gemini versus DeepSeek descriptions

Model	Mean Flesch	Mean perplexity
Gemini	47.6	199.8
DeepSeek	27.7	114.1

score (Flesch, 1948; higher is easier to read) and by perplexity under a pre-trained Generative Pre-trained Transformer 2 (GPT-2) model (Radford et al., 2019; lower is more predictable). Gemini scored higher on readability (47.6 vs 27.7) and also on perplexity (199.8 vs 114.1), as Table 2 shows. Because a human checks the output in Step 5, readability was weighted more heavily and Gemini was selected.

4 RESULTS AND DISCUSSION

Table 3 and Figs. 2 and 3 present the results on the held-out test set. Random Forest and SVM achieved the same accuracy (0.87), but Random Forest had higher precision (0.89 vs 0.87) and F1-score (0.88 vs 0.87), while SVM had slightly higher recall (0.88 vs 0.87). The difference is small on a 300-record test set and should not be over-interpreted

Table 3: Comparative performance of classification models

Model	Accuracy	Precision	Recall	F1-score
K-Nearest Neighbours	0.81	0.82	0.81	0.81
Naïve Bayes	0.80	0.82	0.78	0.79
Support Vector Machine	0.87	0.87	0.88	0.87
Decision Tree	0.66	0.65	0.65	0.65
Random Forest	0.87	0.89	0.87	0.88

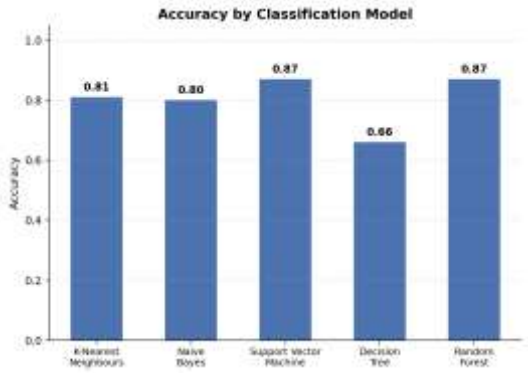


Fig. 2: Accuracy by Classification Model.

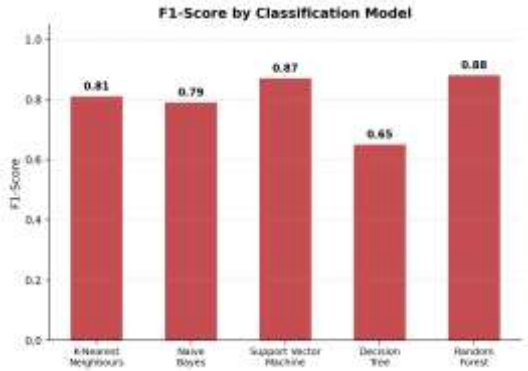


Fig. 3: F1-Score by Classification Model.

. A plausible reason is that Gemini phrases the same event in varied ways, and Random Forest's many trees on random feature subsets are comparatively robust to such variation.

Naïve Bayes and KNN performed worse, probably because Naïve Bayes assumes word independence while Accident and Disaster descriptions share vocabulary such as "damage", "injured", and "wreckage", and because short, machine-generated descriptions from different categories lie close together in TF-IDF space, which hampers KNN. The Decision Tree's low accuracy (0.66) is consistent with its tendency to overfit. These explanations are untested interpretations.

The full pipeline averaged 1.84 seconds per report (near-instantaneous for a user). Two conclusions follow. First, an ensemble on TF-IDF features is a practical way to classify vision-language output, so a custom image classifier may be unnecessary. Second, a sub-two-second pipeline suggests that a citizen-operated tool of this kind could shorten the delay between an incident and an alert reaching the correct authority.

5 CONCLUSION

This study examined how suitable machine learning classifiers are for categorizing LLM-generated incident descriptions within a five-step citizen reporting pipeline. Random Forest performed best among the five classifiers (F1-score 0.88, accuracy 0.87), and the complete pipeline averaged 1.84 seconds, so the approach is technically feasible, with scene interpretation delegated to Gemini and categorization to a lightweight classifier. Future work should (i) evaluate the classifier on Gemini-generated descriptions of real, anonymized photographs, (ii) conduct a user study with a standard instrument such as the System Usability Scale (Brooke, 1996), (iii) develop an offline fallback, since the pipeline depends on an active internet connection, (iv) extend capture to short video, and (v) integrate directly with responder systems through APIs.

REFERENCES

- Alam, F., Ofli, F., & Imran, M. (2018). CrisisMMD: Multimodal Twitter datasets from natural disasters. Proceedings of the 12th International AAAI Conference on Web and Social Media (ICWSM), 465–473.
- Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324
- Brooke, J. (1996). SUS: A quick and dirty usability scale. In P. W. Jordan et al. (Eds.), Usability evaluation in industry (pp. 189–194). Taylor & Francis.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297. https://doi.org/10.1007/BF00994018
- Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13(1), 21–27.
- Deffenbacher, K. A., Bornstein, B. H., Penrod, S. D., & McGorty, E. K. (2004). A meta-analytic review of the effects of high stress on eyewitness memory. Law and Human Behavior, 28(6), 687–706.
- Flesch, R. (1948). A new readability yardstick. Journal of Applied Psychology, 32(3), 221–233.
- Gemini Team, Google. (2023). Gemini: A family of highly capable multimodal models. arXiv. https://arxiv.org/abs/2312.11805
- Imran, M., Castillo, C., Diaz, F., & Vieweg, S. (2015). Processing social media messages in mass emergency: A survey. ACM Computing Surveys, 47(4), Article 67.
- Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. Proceedings of ECML-98, Lecture Notes in Computer Science, 1398, 137–142.
- McCallum, A., & Nigam, K. (1998). A comparison of event models for Naive Bayes text classification. AAAI-98 Workshop on Learning for Text Categorization, 41–48.
- Nakayenga, R., Ojo, M. O., & Kasule, K. (2024). Artificial intelligence in public safety and emergency management: A

systematic review of threat detection and hazard prediction. Journal of Emergency Management and Technology, 18(2), 112–129.

- Pasi, S., Kumar, R., & Sharma, A. (2025). Real-time crime and weapon detection using the YOLO framework for automated law enforcement alert systems. IEEE Transactions on Human-Machine Systems, 55(1), 45–56.
- Pathik, B. B., Rahman, M. M., & Hossain, M. S. (2022). Smart city emergency response: Integrating IoT vehicle sensors and visual deep learning for rapid accident detection. IEEE Internet of Things Journal, 9(14), 12040–12052.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.
- Post-Disaster Needs Assessment Secretariat. (2025). Sri Lanka 2025 floods: Estimated population exposure and rapid damage report. Ministry of Disaster Management, Sri Lanka.
- Quinlan, J. R. (1986). Induction of decision trees. Machine Learning, 1(1), 81–106.
- Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. Proceedings of the 38th International Conference on Machine Learning, 139, 8748–8763.
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI Technical Report.
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 779–788.
- Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. Information Processing & Management, 24(5), 513–523.
- Sebastiani, F. (2002). Machine learning in automated text categorization. ACM Computing Surveys, 34(1), 1–47.
- Tamagusko, T., Correia, M. G., Huynh, M. A., & Ferreira, A. (2022). Deep learning applied to road accident detection with transfer learning and synthetic images. Transportation Research Procedia, 64, 90–97.

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

A Comparative Evaluation of Large Language Models for Mathematical Concept and Relation Extraction from Calculus Text

Athapaththu SSS [1], *Liyanage DLCP* [2], *Karunaratna DD* [3], *Ratnayake JK* [4]
[1,4] *Department of Mathematics, Faculty of Applied Sciences, University of Sri Jayewardenepura*
[2] *Department of ICT, Faculty of Technology, University of Sri Jayewardenepura*
[3] *Department of Communication and Media Technologies, University of Colombo School of Computing, University of Colombo,*
sanjulaathapaththu@sjp.ac.lk [1], *dlchamara@sjp.ac.lk* [2], *ddk@ucsc.cmb.ac.lk* [3], *jratnayake@sjp.ac.lk* [4]

ABSTRACT

Automatically extracting concepts and relations from mathematical text can support knowledge-graph construction, semantic retrieval, and educational applications, but complex notation and implicit dependencies make this task challenging. This study comparatively evaluates ChatGPT, DeepSeek, and Qwen for concept and relation extraction from calculus text. The corpus comprised 20 short text units selected from the OpenStax textbook Calculus Volume 1 and converted into LaTeX to preserve mathematical notation. Each model processed the units using the same structured zero-shot prompt, while a subject expert independently annotated the concepts and relations. Concept extraction was evaluated using precision, recall, F1-score, and hallucination rate; relations were assessed using a graded correspondence scheme. ChatGPT achieved the highest recall (79.70%) but the lowest precision (55.08%) and highest hallucination rate (44.91%). Qwen obtained the highest precision (68.89%) and lowest hallucination rate (31.11%), while DeepSeek showed intermediate precision and recall. Despite these differences, all three models produced similar F1-scores, ranging from 65.15% to 65.78%. Relation extraction showed substantially weaker agreement with the human annotation, particularly in relation type, direction, granularity, and the representation of complex meanings. These findings suggest that LLMs can provide useful preliminary mathematical knowledge extraction, but reliable relation extraction requires clearer relation definitions, improved prompting, and expert validation.

KEYWORDS: Knowledge-graph, Large Language Models, Mathematical Concept Extraction, Relation Extraction

1 INTRODUCTION

This study compares three large language models (LLMs) - ChatGPT, DeepSeek, and Qwen - in extracting mathematical concepts and identifying relations among them from a calculus textbook. Their outputs are evaluated against human annotations using precision, recall, F1-score, and hallucination rate for concept extraction, together with an evaluation of the identified relations. In a similar study (de Paiva et al., 2023) evaluated ChatGPT alone for extracting mathematical terms from category-theory abstracts. We aim to broaden this work by extending extraction from isolated terms to relations and by comparing multiple LLMs. The study aims to determine the relative capabilities and limitations of the selected LLMs and assess their usefulness in developing mathematical knowledge resources.

In mathematics, definitions introduce concepts; theorems express implications and dependencies; and examples, explanations, and procedures demonstrate how concepts are applied. Thus, mathematical text can be structured into concepts and relations that support knowledge representation, evaluation, and semantic reasoning. Such representations can ground LLMs in a

specific context and help validate outputs and mitigate hallucinations (Amayuelas et al., 2025; Dehal et al., 2025). They also have pedagogical applications: for example, (Liyanaarachchi et al., 2021) analyzed students' concept structure and learning using such a representation.

Difficulty of extracting mathematical concepts and relations can be addressed by the semantic reasoning capabilities of LLMs; however, their outputs may be incomplete, inconsistent, and sensitive to prompt design. Evaluation against expert annotations is therefore essential.

2 RELATED WORKS

MathKG (Wang, 2022) used BERT-based models for entity and relation extraction, while MathGloss (Horowitz & Paiva, 2023) employed spaCy to extract mathematical concepts. More recent systems, such as AutoMathKG (Bian et al., 2025), have incorporated LLMs, including Llama-2-7B, for relation identification. Also, (de Paiva et al., 2023) further demonstrated that an LLM can provide useful preliminary concept extraction, although it cannot replace human expertise. However, comparative evidence across LLMs remains limited, particularly for the joint extraction of concepts and relations. The present evaluation addresses this gap and examines the extent to which current LLMs can provide reliable inputs for downstream tasks.

3 METHODOLOGY

This study evaluates the capabilities of three LLMs to extract mathematical concepts and relations from a calculus textbook (Herman & Strang, 2016), compared to those of a human annotator. The methodology consisted of four main stages: corpus selection and segmentation, LLM prompting and sconcepts and relations extraction, human annotation and comparison, and evaluation of extracted concepts and relationships.

3.1 Corpus selection and segmentation

The study used selected material from the OpenStax textbook Calculus Volume 1 (Herman & Strang, 2016). The corpus was drawn from Sections 2.1, "A Preview of Calculus," and 2.2, "The Limit of a Function," in Chapter 2. The source text is in LaTeX to preserve its mathematical notation and expressions. Twenty short text units (S01–S20), comprising concepts, definitions, theorems, explanations, and examples, were then selected for analysis.

3.2 LLM Prompting and, Concepts and Relations Extraction

Three Large Language Models were selected for the comparative evaluation: ChatGPT, DeepSeek and Qwen (Table 1). A structured zero-shot prompt was given to extract mathematical concepts and identify relations in each segment (Table 2). Two tables (A and B) were constructed with columns as instructed in the prompt.

3.3 Human Annotation

Human annotation was performed by a subject expert, who independently reviewed each of the 20 segments and performed the same task.

A master concept list was constructed by grouping all the concepts according to their semantic similarity, by assigning closely related mathematical meanings to a common group. The resulting semantic groups were converted into a binary representation, with rows corresponding to semantic concepts and The extracted

Table 1: LLMs used

LLM	Version	Platform	Is Paid
ChatGPT	ChatGPT-5.6-Luna	Web Interface	Paid
DeepSeek	DeepSeek-V3	Web Interface	Free
Qwen	Qwen-3.7-Plus	Web Interface	Free

3.4 Preparing Master Concept List

columns representing the human annotator and LLMs. A value of 1 indicated that a concept was identified, and a value of 0 indicated that it was not.

3.5 Evaluation Metrics

concepts of each LLM were compared with the human annotated results. The binary representation was used to determine the number of true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN).

Table 2: Zero-shot LLM prompts

Instructions:
1. Extract only concepts that are explicitly present or clearly implied by the passage. 2. Do not add outside concepts that are not needed to understand the passage. 3. Classify each concept as one of: object, process, property, relation, theorem/rule, representation, example. 4. State whether the concept is explicitly defined in the passage. 5. Identify relations only from this list: is-a, defined-using, prerequisite-of, used-for, property-of, represented-by. 6. Mark each relation as explicit or inferred. Return the answer in two tables: A. Concept table: Concept | Type | Explicitly defined? | Importance (Central,Supporting or Peripheral) B. Relation table: Source concept | Relation | Target concept | Explicit/Inferred Passage: (S01 - S20)

Table 3: Scoring scheme for relations

Score	Meaning
1.0	All correct
0.5	Correct source and target concept, imprecise relation/ Correct relation type, partially correct source and target
0.375	Correct source concept, correct relation type, unsupported target concept
0.25	Partially correct source and target concept, imprecise relation type
0.125	Reversed source and target concepts
0.0	Unsupported relation

Hallucinated concepts (HC) are the number of relations with zero score. These values were used to calculate:$precision = P = TP / (TP + FP)$, $recall = R = TP / (TP + FN)$, $F1 = 2PR / (precision + recall)$, $Hallucination\ Rate = HC / Total$.

3.6 Evaluation of Identified Relations

The identified relations were evaluated by manually comparing the relations generated by the three LLMs with the corresponding relations identified by the human annotator. Each relation was represented as a triplet, consisting of a source concept, relation

type, and target concept. Some LLMs included derived relations that the human-annotator excluded due to their elementary nature. These relations were excluded from the scoring. A graded scoring scheme was applied to quantify the degree of correspondence (Table 3).

4 RESULTS AND DISCUSSION

An example of concepts extracted from selected segments by the human annotator, ChatGPT, DeepSeek, and Qwen are presented in Table 4.

Based on the values for TP, FP, FN and TN, precision, recall, F1-score, and hallucination rate were calculated. The values obtained are reported in Table 5.

ChatGPT achieved the highest recall, as it identified the largest proportion of the concepts annotated by the expert. However, its precision was comparatively lower, and it had the highest hallucination rate. In contrast, Qwen achieved the highest precision and the lowest hallucination rate, indicating that its extracted concepts were more closely aligned with the human annotator. DeepSeek performance is average compared to ChatGPT and Qwen. Noticeably, their overall F1-score performance was similar, although the models differed considerably in their precision and recall. Table 6 summarizes the results of relation evaluation. The evaluation also revealed several challenges. ChatGPT identified the relation Tangent line → property-of → Function f(x), whereas the human annotator identified Tangent line → is-a → Linear function. Therefore, the intended usage of each relation type needs to be clearly specified. The results indicate a noticeable deviation (score close to zero) from the human annotator. The study also revealed differences in the level of detail captured by the LLMs and the human annotator. This suggests that the LLMs may extract more fine-grained relationships than those considered relevant by the human annotator for a particular segment.

Table 5: Precision (P), Recall (R), F1-Score and Hallucination Rates (HR).

LLM	P (%)	R (%)	F1 (%)	HR (%)
ChatGPT	55.08	79.7	65.15	44.91
DeepSeek	64.22	66.5	65.34	35.78
Qwen	68.89	62.94	65.78	31.11

Noticeably, complex concept descriptions were not always extracted as a single coherent relation by the LLMs. For example, for the concept of the *non-existence of a limit*, an LLM generated several separate relations. In contrast, the human annotator represented the idea as a single relation. This shows that LLMs may decompose complex mathematical definitions into multiple shorter relations, which may not preserve the complete semantic relationship expressed in the passage.

Table 6: Scoring results for relations extracted by the three LLMs

	ChatGPT	DeepSeek	Qwen
1.0	9	6	5
0.5	30	27	31
0.375	16	9	1
0.25	24	28	10
0.125	29	22	11
0.0	52	48	34
Excluded	134(46 %)	79(36 %)	55(37%)
Total	294	219	147

5 CONCLUSION

LLM outputs are comparable; however, ChatGPT achieved the highest recall, while Qwen achieved the highest precision and the lowest hallucination rate, indicating differences in the extraction behavior of the models.

The LLM-generated relations showed deviations from the human-annotated results in relation type, level of detail, and representation. Some relations generated by the LLMs were more fine-grained than those identified by the human annotator. These findings indicate that relation extraction requires more careful prompting, clearly defined relation types, and human validation. LLMs can be considered useful supporting tools for mathematical knowledge extraction, but they cannot yet fully replace human expertise.

6 ACKNOWLEDGEMENT

This research was funded by University of Sri Jayewardenepura research grant RC/URG/SCI/2026/27.

REFERENCES

- Amayuelas, A., Sain, J., Kaur, S., & Smiley, C. (2025). Grounding LLM Reasoning with Knowledge Graphs (arXiv:2502.13247).
- Bian, R., Geng, Y., Yang, Z., & Cheng, B. (2025). AutoMathKG: The automated mathematical knowledge graph based on LLM and vector database. *Computational Intelligence* (41).
- de Paiva, V., Gao, Q., Kovalev, P., & Moss, L. S. (2023). *Extracting Mathematical Concepts with Large Language Models.* Conference on Intelligent Computer Mathematics.
- Dehal, R. S., Sharma, M., & Rajabi, E. (2025). Knowledge Graphs and Their Reciprocal Relationship with Large Language Models. *Machine Learning and Knowledge Extraction*, *7*(2), 38.
- Herman, E., & Strang, G. (2016). *Calculus: Volume 1*. OpenStax, Rice University.
- Horowitz, L., & Paiva, V. de. (2023). MathGloss: Building mathematical glossaries from text. Conference on Intelligent Computer Mathematics.
- Liyanaarachchi, B. T., Gunawardene, S., Ratnayake, J. K., & Jayewardene, R. (2021). Representing the Formal Structure of Mathematical Concepts Incorporating Learning Taxonomies. ICMAS (2021) 112.
- Wang, J. (2022). *Math-KG: Construction and Applications of Mathematical Knowledge Graph* (arXiv:2205.03772). arXiv.

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research –2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Image Processing Approach to Grading Pepper Quality in Sri Lanka

Rajapaksha RWWMHM[1], Munasinghe B[2]
[1,2]*Department of Computing & Information Systems, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*
hansikamadushani688@gmail.com[1], bhagya@wyb.ac.lk[2]

ABSTRACT

Pepper quality grading is an important process for determining the quality and market value of pepper products. However, conventional grading practices mainly rely on visual inspection and human judgement, which may lead to inconsistent results and require considerable time. This study presents an image-based deep learning system for automated pepper quality grading and price estimation. A dataset of 3,038 labelled images representing five black and white pepper quality grades was prepared and divided into training, validation, and testing sets using a 70:20:10 ratio. Images were resized to 224 × 224 pixels and augmented to improve data diversity. EfficientNetB0, ResNet50, and Visual Geometry Group 16 (VGG16) were trained using transfer learning and comparatively evaluated. The VGG16 model was subsequently fine-tuned to improve its classification performance, achieving a test accuracy of 92.81%. The trained model was integrated into a Streamlit based web application that allows users to upload pepper images and obtain predicted grades, confidence scores, and a grade-based price reference. A preliminary user evaluation with seven respondents was conducted to assess the developed application, and the findings indicated positive user feedback. Due to the limited number of respondents, the user evaluation results are considered preliminary. The developed system demonstrates the potential of deep learning to provide a faster and more consistent approach to pepper quality assessment.

KEYWORDS: **Automated Pepper Grading, Deep Learning, Image Classification, Pepper Quality Classification, Transfer Learning**

1 INTRODUCTION

Pepper is an economically important agricultural crop, and the quality of pepper products directly influences their market value and acceptance. Quality grading is therefore an important activity for farmers, buyers, and other stakeholders involved in the pepper supply chain. However, conventional pepper grading commonly depends on visual inspection and human judgement, which can be time-consuming and may produce inconsistent results. Previous studies have demonstrated the potential of image processing and deep learning for automated agricultural product quality assessment, including pepper quality classification using transfer learning and convolutional neural networks.

Deep learning-based image classification can learn relevant visual features directly from images and reduce the need for manually designed features. Transfer learning is particularly useful when the available dataset is limited, as pre-trained convolutional neural networks can be adapted to a specific classification task. Previous research has also shown that fine-tuning pre-trained models such as VGG16 and ResNet can improve performance in agricultural image classification tasks.

This study focuses on the development of an automated image-based system for classifying black and white pepper according to predefined quality grades. A dataset containing 3,038 labelled pepper images was used to train and compare three transfer learning

models: EfficientNetB0, ResNet50, and VGG16. The best-performing model was subsequently fine-tuned and integrated into a web-based application for practical use. The developed application provides predicted pepper grades, confidence scores, from uploaded images, together with a grade-based price reference to provide indicative price information. A preliminary user evaluation was also conducted to assess the usability and practical value of the developed system.

2 LITERATURE REVIEW

Deep learning has been widely used for agricultural image classification and quality assessment because it can automatically learn visual features from images. Transfer learning allows pre-trained deep learning models to be adapted to specific agricultural classification tasks and is useful when task-specific datasets are limited (Pan & Yang, 2010). Data augmentation can further improve the diversity of training images and help reduce overfitting (Shorten & Khoshgoftaar, 2019).

Previous studies have demonstrated the use of deep learning and transfer learning for pepper quality assessment. Ren et al. (2021) investigated external quality detection of pepper using deep learning and transfer learning, showing the potential of pre-trained models for automated pepper quality analysis. Other agricultural image classification studies have also reported successful use of architectures such as VGG16, ResNet, and EfficientNet.

However, limited attention has been given to an integrated system that performs multi-grade black and white pepper classification together with estimated price information through a web-based application. Therefore, this study compares EfficientNetB0, ResNet50, and VGG16, fine-tunes the selected model, and integrates it into a web application for automated pepper grading and price estimation.

3 METHODOLOGY

The proposed system was developed through a sequence of dataset preparation, image preprocessing, data augmentation, deep learning model development, model evaluation, and web application development stages. A dataset of 3,038 labelled pepper images representing five quality grades—Black Pepper Grade I, II and III, and White Pepper Grade I and II—was used. The dataset was divided into training, validation, and testing subsets using a 70:20:10 ratio, resulting in 2,125 training images, 607 validation images, and 306 testing images.

Before model training, all images were resized to 224 × 224 pixels and prepared for model input. Data augmentation techniques, including rotation, flipping, zooming, and brightness adjustment, were applied to increase image diversity and improve model generalization.

Three pre-trained convolutional neural network architectures, EfficientNetB0, ResNet50, and VGG16, were developed using transfer learning and evaluated using accuracy, precision, recall, F1-score, and confusion matrices. Based on the comparative results, VGG16 was selected for further fine-tuning. The first 15 layers were frozen while higher layers were retrained using batch normalization, dropout, stronger augmentation, and early stopping.

The final fine-tuned VGG16 model was integrated into a web-based

application. The application accepts an uploaded pepper image, preprocesses the image, predicts its quality grade, and provides a confidence score. In addition, a grade-based price reference was incorporated into the application to provide indicative price information according to the predicted grade. This price reference is presented as supporting information and was not evaluated as an independent price-prediction model. The application was developed using Python, TensorFlow/Keras, Streamlit, OpenCV and web technologies including HTML, CSS, and JavaScript.

Finally, a user evaluation survey with seven respondents was conducted after demonstrating the developed application to assess its usability, usefulness, and practical acceptance.

4 RESULTS AND DISCUSSION

4.1 Dataset and Experimental Setup

The developed system was trained using a dataset containing 3,038 labelled pepper images representing five quality classes: Black Pepper Grade I, Black Pepper Grade II, Black Pepper Grade III, White Pepper Grade I, and White Pepper Grade II. The dataset was divided into training, validation, and testing sets using a 70:20:10 ratio, resulting in 2,125 training images, 607 validation images, and 306 testing images. The images were resized to 224 × 224 pixels and augmented during training to improve the diversity of the training data.

4.2 Comparative Model Performance

Three transfer learning models, EfficientNetB0, ResNet50, and VGG16, were trained and evaluated using the same pepper classification task. EfficientNetB0 achieved a test accuracy of 69.93%, while ResNet50 achieved 81.05%. The initial VGG16 model achieved the highest accuracy among the three models, with a test accuracy of 81.37%.

Table 1: Comparison of Test Accuracy of the Deep Learning Models

Model	Test Accuracy
EfficientNetB0	69.93%
ResNet50	81.05%
VGG16	81.37%

The results indicate that VGG16 provided the best initial performance and was therefore selected for further fine-tuning. The relatively lower performance of EfficientNetB0 compared with the other models suggests that model architecture and learned feature representations can influence classification performance for the given pepper image dataset.

4.2 Fine-Tuned VGG16 Performance

The selected VGG16 model was further fine-tuned by freezing the first 15 layers and retraining the higher layers. Batch normalization, dropout, stronger augmentation, and early stopping were incorporated to improve generalization. Following fine-tuning, the model achieved a test accuracy of 92.81%, with precision of 0.93, recall of 0.92, and F1-score of 0.92.

The improvement from 81.37% to 92.81% indicates that fine-tuning was effective in adapting the pre-trained VGG16 model to the characteristics of the pepper dataset. The results demonstrate

Table 2: Performance of the Fine-Tuned VGG16 Model

Metric	Score
Accuracy	**92.81%**
Precision	0.93
Recall	0.92
F1-score	0.92

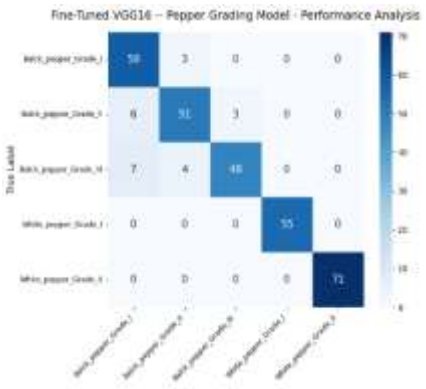


Figure 1: Confusion Matrix of the Fine-Tuned VGG16 Model

that selective retraining of higher-level features can improve the model's ability to distinguish between the different pepper quality grades.

4.3 Developed Web Application

The fine-tuned VGG16 model was integrated into a web-based pepper grading application. The application allows users to upload a pepper image and receive the predicted quality grade together with a confidence score. A grade-based price reference was also incorporated to provide indicative price information alongside the classification result. The system follows the sequence of image upload → preprocessing → VGG16 prediction → grade and price output.

The integration of the trained model into a web interface makes the classification system more accessible to non-technical users. In addition to automated grading, the inclusion of an estimated price provides information that can support practical quality and trading decisions.

4.5 User Evaluation of the Developed System

After development of the web application, a user evaluation was conducted with seven respondents to assess the practical usefulness and usability of the system. The evaluation showed positive feedback regarding the ability of the application to reduce difficulties associated with manual pepper grading. All respondents either agreed or strongly agreed that the web application could reduce difficulties in manual grading, with 85.71% agreeing and 14.29% strongly agreeing.

These findings provide preliminary evidence of positive user feedback towards the developed application. The application provides automated grade prediction and confidence scores through a single interface, while a grade-based price reference is provided as supporting information. However, the evaluation involved only seven respondents; therefore, the findings are considered preliminary and cannot be generalised to the wider population of pepper farmers, sellers, or other stakeholders. In addition, the grade-based price reference was not independently evaluated as a price-prediction model. Larger-scale user evaluations and independent validation of the price component would be required in future work to establish stronger evidence of the system's practical effectiveness.

5 CONCLUSION

This study developed an image-based deep learning system for automated grading of black and white pepper and estimated price prediction. A dataset of 3,038 labelled pepper images representing five quality grades was used to train and evaluate three transfer learning models: EfficientNetB0, ResNet50, and VGG16. Among the initial models, VGG16 achieved the highest test accuracy of 81.37%. Further fine-tuning improved the VGG16 performance to 92.81% test accuracy, with a precision of 0.93, recall of 0.92, and F1-score of 0.92. The final model was integrated into a web-based application that provides automated grade prediction, confidence scores, and a grade-based price reference as supporting information.

A preliminary user evaluation involving seven respondents indicated positive feedback towards the developed application, suggesting its potential usefulness for supporting pepper grading activities. However, the small number of respondents limits the generalisability of the user evaluation findings, and the grade-based price reference was not independently evaluated as a price-prediction model. Overall, the findings demonstrate that the proposed system can provide a faster and more consistent approach to pepper quality assessment while supporting users with additional price-related information. Further evaluation with a larger and more diverse user group and dataset would be beneficial to improve the reliability and practical applicability of the system.

REFERENCES

- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778. https://doi.org/10.1109/CVPR.2016.90
- Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191
- Ren, Y., Zhang, Y., Sun, H., & Gao, X. (2021). External quality detection of pepper based on deep learning and transfer learning. Computers and Electronics in Agriculture, 189, 106412.
- Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6, 60. https://doi.org/10.1186/s40537-019-0197-0
- Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. International Conference on Learning Representations (ICLR).
- Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. Proceedings of the 36th International Conference on Machine Learning, 97, 6105–6114.

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research - 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Computational Empowerment in Sri Lankan Information & Communication Technology and Related Education

Basnayaka PBN[1], Munasinghe B[2]
[1,2]*Department of Computing and Information Systems, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*
nimansabasnayaka0220@gmail.com[1] , bhagya@wyb.ac.lk[2]

ABSTRACT

Sri Lanka's national digital transformation agenda targets a USD 15 billion digital economy by 2030, built on platform-first Digital Public Infrastructure such as digital identity, payment, data-exchange, network and cloud platforms. These platforms rely on a computationally capable workforce, yet little empirical work has examined whether Sri Lankan Information and Communication Technology (ICT) and related education actually build Computational Empowerment (CE) — the confidence and competence to create technology rather than merely consume it. This cross-sectional study measured self-reported CE among four purposively sampled groups connected to the ICT education-to-digital-economy pipeline: ICT students at Ordinary and Advanced Level (O/L and A/L), undergraduates and fresh graduates, ICT teachers and lecturers, and government ICT employees (N = 161). A structured, conditionally branched five-point Likert questionnaire was used. CE showed a moderately positive grand mean of 3.50 (standard deviation, SD = 0.66), with good internal consistency in three of the four respondent groups (α = .83 to .86), though the Teacher/Lecturer CE battery showed poor reliability (α = .18). CE did not differ significantly across the four groups (p = .090), but among 148 respondents with complete data it was strongly correlated with Transformation Agency (r = .746, p < .001) and was its strongest regression predictor. These findings indicate association, not causation. Overall, the results suggest a meaningful but incomplete level of perceived computational capability among respondents, and that hands-on skill-building in coding, cloud computing and artificial intelligence literacy merits priority alongside general awareness messaging, subject to the study's limitations.

KEYWORDS: Computational Empowerment, Computational Thinking, Digital Skills, Information & Communication Technology Education, Sri Lanka

1 INTRODUCTION

The world is undergoing rapid digital transformation, and international organizations such as the Organization for Economic Co-operation and Development (OECD) increasingly emphasize that this transformation depends not only on technological infrastructure but also on citizens who possess the knowledge, skills and confidence to actively participate in digital environments (OECD, 2019). One construct that captures this participatory capability is Computational Empowerment (CE): equipping individuals with computational thinking, digital problem-solving ability, programming knowledge and the confidence to create technology rather than merely consume it (Smith et al., 2020).

Sri Lanka has embarked on a comprehensive national digital transformation journey led by the Ministry of Digital Economy (MODE) and the Information and Communication Technology (ICT) Agency (ICTA), aiming to grow the digital economy to approximately USD 15 billion by 2030 (MODE, 2024) through a platform-first approach comprising Digital Public Infrastructure (DPI) components such as the Sri Lanka Unique Digital Identity (SL-

UDI), the National Data Exchange (NDX) and GovPay. There is limited empirical evidence on whether Sri Lankan ICT and related education actually develop CE. This paper addresses that gap by assessing the level of CE reported by learners across the education-to-workforce pipeline, and by examining how CE is associated with learners' broader sense of digital agency.

2 LITERATURE REVIEW

Computational Empowerment extends beyond computer literacy and computational thinking by emphasizing the meaningful use of computational knowledge to create technological solutions and address real-world challenges.

The literature identifies several recurring dimensions of CE: computational thinking, programming skills, data literacy, systems thinking, artificial intelligence (AI) readiness and computational problem-solving. Smith et al. (2023) proposed a three-tier research agenda linking participatory practices, educational principles and transparent ("glass-boxed") digital tools. Dindler et al. (2020) distinguished between "coding" — the manual design and building of digital artifacts — and "decoding" — the critical analysis of how developers embed values into systems. Tissenbaum et al. (2019) showed that resolving genuine community problems with code produces an identity shift toward viewing oneself as an active technical creator rather than a passive consumer. Van Mechelen et al. (2022), reviewing 109 peer-reviewed studies over a decade, found that emerging-technology instruction is often dominated by short-term tool use with little long-term formative assessment of whether computational capability actually develops.

Within Sri Lanka, previous ICT-education research has concentrated on literacy, technology acceptance, e-learning and infrastructure (Karunathilake & Vidanagama, 2021; Gamage, 2020), with little attention paid to whether education develops Computational Empowerment specifically.

3 METHODOLOGY

A quantitative, cross-sectional survey design was adopted. A structured questionnaire using conditional branching in Google Forms was distributed to four purposively sampled participant groups directly connected to the ICT education-to-digital-economy pipeline: ICT students at Ordinary Level (O/L) and Advanced Level (A/L), undergraduates and fresh graduates, ICT teachers and lecturers, and government ICT employees. Data collection closed once 161 valid responses had been received (22 students, 100 undergraduates/fresh graduates, 22 teachers/lecturers and 17 government ICT employees). All 161 respondents answered every CE item, so CE descriptive, reliability and group analyses use N = 161. Correlation and regression used the 148 respondents with complete scores on all five constructs; the 13 excluded students completed the Sinhala-language form, whose Sinhala-labeled responses were not scored in the thesis analysis and are re-scored in a sensitivity analysis (Section 4). This paper reports the Computational Empowerment findings from a broader dataset analyzed in full in the first author's final-year thesis (Basnayaka & Munasinghe, 2026), which also examined Transformation Agency (TA).

Computational Empowerment was operationalized as one of five composite constructs derived from the questionnaire's

Likert-item batteries (the others being Awareness, Access, Transformation Agency and Barriers). CE items, tailored to each respondent group through conditional branching, captured perceived computational skill and exposure to coding, application programming interfaces (APIs), cloud computing and AI tools, each rated on a five-point scale (1 = Strongly Disagree to 5 = Strongly Agree). The CE composite was the item mean (three items for students and teachers/lecturers, six for undergraduates/fresh graduates, four for government employees). Validity rested on content validity (items reviewed by the supervisor); no factor analysis was performed. Internal consistency of the CE composite was assessed with Cronbach's Alpha ($\alpha \geq .70$ considered acceptable). Descriptive statistics, a one-way analysis of variance (ANOVA) across the four groups, Pearson correlation with the other constructs, and a multiple regression with CE as a predictor of Transformation Agency were computed in IBM SPSS Statistics.

4 RESULTS AND DISCUSSION

Across the full sample (N = 161), Computational Empowerment had a grand mean of 3.50 (standard deviation [SD] = 0.66) on the 1–5 scale, indicating that respondents lean toward agreement that their ICT education is building computational skills, with room for improvement. Table 1 presents CE by respondent group.

poor for the Teacher/Lecturer group (α = .18, based on only three items and n = 22); this result is treated as a measurement limitation rather than evidence that teachers lack computational empowerment, and is flagged for future instrument refinement. subgroups (n = 17–22) limit the power to detect differences. Reliability was good for three of the four groups (α = .83–.86) but

Pearson correlation analysis (N = 148) showed that CE was significantly and positively correlated with every other construct in the study: Awareness (r = .428), Access (r = .517), Transformation Agency (r = .746) and Barriers (r = .497), all $p < .001$. A multiple regression with

Welch's ANOVA (p = .082) and the Kruskal–Wallis test (p = .075) agreed. This suggests that computational empowerment is reported comparably across the school-to-workforce pipeline, although the small across the four respondent groups, F (3, 157) = 2.20, p = .090, η^2 = .04. Levene's test indicated unequal variances (p = .003), but

Table 1: Computational Empowerment by Respondent Group

Group	N	CE Mean	SD	Cronbach's α
ICT Student (O/L or A/L)	22	3.27	0.88	0.825
Undergraduate / Fresh Graduate	100	3.60	0.63	0.851
Teacher / Lecturer	22	3.32	0.45	0.182
Government ICT Employee	17	3.49	0.67	0.856
Overall	161	3.50	0.66	—

Note. CE = Computational Empowerment; SD = standard deviation. Item sets differ by group; the student row includes 13 Sinhala-language respondents.

A one-way ANOVA found no statistically significant difference in CE Transformation Agency as the dependent variable and CE, Awareness, Access and Barriers as predictors showed that CE was the strongest statistical predictor ($B = 0.562$, $p < .001$), with the overall model explaining 64.2% of the variance in Transformation Agency ($R^2 = .642$, $F(4,143) = 64.07$, $p < .001$), while Awareness and Access were not significant once CE was accounted for; Barriers was also positively associated with TA ($B = 0.293$, $p < .001$). Diagnostics supported the model (variance inflation factors 1.42–1.64, Durbin–Watson = 2.09, Breusch–Pagan $p = .511$, maximum Cook's D = 0.11); residuals departed mildly from normality (Shapiro–Wilk $p = .009$), but heteroscedasticity-consistent (HC3) standard errors gave the same significance pattern. Re-scoring the 13 Sinhala-language students ($N = 161$) gave $r = .646$ and $R^2 = .577$, with CE still significant ($B = 0.449$, $p < .001$).

These findings are consistent with the theoretical argument advanced by Iivari et al. (2024) and Smith et al. (2023) that computational empowerment underpins a learner's broader sense of agency to create and lead digital change; however, this cross-sectional design shows association only, so neither its direction nor the role of a shared third factor can be established. The findings also suggest that Sri Lanka's digital transformation communication may have comparatively limited association with learners' sense of digital agency unless it is paired with hands-on computational skill-building of the kind described by Tissenbaum et al. (2019) and Dindler et al. (2020). *Limitations.* The cross-sectional design supports association, not causation. The purposive sample was unevenly distributed across groups (22, 100, 22 and 17) and relied on self-reported perceptions, so results should not be generalized to all Sri Lankan learners and educators. Validation relied on content validity and Cronbach's alpha only, and the Teacher/Lecturer mean is uncertain ($\alpha = .18$). The 148-case regression excluded 13 students, and its coefficients shift when they are included.

5 CONCLUSIONS

Respondents across the school-to-workforce pipeline reported a moderate, fairly consistent level of Computational Empowerment, and CE showed the strongest association with Transformation Agency among the constructs studied. Given the cross-sectional, self-reported and non-probability nature of the data, these findings are preliminary and do not show that CE causes agency. Curricula at every level may therefore benefit from prioritizing project-based, hands-on computational skill-building — coding, APIs, cloud platforms, open data and AI tools — explicitly connected to Sri Lanka's national digital platforms. Future instruments for teachers' computational empowerment should be revised and piloted with a larger, validated item set.

REFERENCES

- Basnayaka, P. B. N., & Munasinghe, B. (2026). Transformation Agency & Computational Empowerment in Sri Lankan ICT & Related Education - Final Year Project Theses, Department of Computing and Information Systems, Wayamba University of Sri Lanka.
- Dindler, C., Smith, R. C., & Iversen, O. S. (2020). Computational empowerment: Participatory design in education. CoDesign, 16(1), 66–80.

https://doi.org/10.1080/15710882.2020.1722173

- Gamage, S. N. (2020). ICT in education in Belt and Road Initiative countries: Sri Lanka chapter. In ICT in Education in Belt and Road Initiative Countries (pp. 1–12). LIRNEasia / UNESCO.
- Iivari, N., Iversen, O. S., Smith, R. C., Schaper, M.-M., Ventä-Olkkonen, L., Hartikainen, H., Sharma, S., Kinnula, M., Lehto, E., Holappa, J., & Molin-Juustila, T. (2024). Transformative agency – The next step towards children's computational empowerment. In Proceedings of the 2024 ACM Interaction Design and Children Conference (IDC '24), 322–337. ACM. https://doi.org/10.1145/3628516.3655806
- Karunathilake, T. S. L., & Vidanagama, V. G. T. N. (2021). Identification of a framework for implementing ICT in Sri Lankan secondary education system. Journal of Education and Learning (EduLearn), 15(2), 242–250. https://doi.org/10.11591/edulearn.v15i2.19601
- Ministry of Digital Economy (MODE). (2024). National Digital Strategy 2030. Government of Sri Lanka.
- OECD. (2019). Going Digital: Shaping Policies, Improving Lives. OECD Publishing. https://doi.org/10.1787/9789264312012-en
- Smith, R. C., Bossen, C., Dindler, C., & Iversen, O. S. (2020). When participatory design becomes policy: Technology comprehension in Danish education. In Proceedings of the 16th Participatory Design Conference 2020, Vol. 1, 148–158. ACM. https://doi.org/10.1145/3385010.3385011
- Smith, R. C., Schaper, M.-M., Tamashiro, M. A., Van Mechelen, M., Petersen, M. G., & Iversen, O. S. (2023). A research agenda for computational empowerment for emerging technology education. International Journal of Child-Computer Interaction, 38, Article 100616. https://doi.org/10.1016/j.ijcci.2023.100616
- Tissenbaum, M., Sheldon, J., & Abelson, H. (2019). From computational thinking to computational action. Communications of the ACM, 62(3), 34–36. https://doi.org/10.1145/3265747
- Van Mechelen, M., Smith, R. C., Schaper, M.-M., Tamashiro, M., Bilstrup, K.-E. K., Lunding, M., Petersen, M. G., & Iversen, O. S. (2022). Emerging technologies in K–12 education: A future HCI research agenda. ACM Transactions on Computer-Human Interaction, 30(3), Article 47. https://doi.org/10.1145/3569897

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Machine Learning-Based Recommendations for First-Move Advantage in Chess Using Rating Difference, Time Control, and Opening Classification

Gamage KPLT[1], Munasinghe B[2]
[1,2] *Department of Computing and Information Systems, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*
kushankapabodi@gmail.com[1], bhagya@wyb.ac.lk[2]

ABSTRACT

White always moves first in chess, but players cannot choose that colour assignment because it is determined by tournament pairing or the online platform. This study therefore does not recommend who should move first; it develops context-aware opening recommendations after a player's colour is known. Using the Chess Digits 200K dataset of rated Lichess games from May 2019, 199,995 cleaned games were analyzed across Bullet, Blitz, Rapid, and Classical time controls. The main features were rating difference, time-control category, and Encyclopedia of Chess Openings (ECO) code. Logistic Regression was evaluated with three binary targets, while Light Gradient Boosting Machine (LightGBM) and Categorical Boosting (CatBoost) were used for three-class outcome predictions. The White Non-Loss Logistic Regression model, where a White win or draw is favourable and a White loss is unfavourable, achieved 57.85% accuracy and a macro F1-score of 0.58, compared with an approximately 52.75% majority-class baseline. LightGBM achieved 52.15% accuracy on the three-class task, whereas balanced CatBoost achieved 38.09% and over-predicted draws. The model probability was converted into an opening recommendation score by combining 70% historical White score and 30% predicted White non-loss probability. The proposed framework is intended for players, coaches, analysts, and training systems to compare openings for an assigned colour, rating context, and time control. The recommendations are decision-support rankings rather than guarantees of game outcomes.

KEYWORDS: Chess Analytics, First-Move Advantage, Machine Learning, Opening Recommendation, Rating Difference

1 INTRODUCTION

Chess is a two-player strategy game in which White moves first and Black responds. The colour is not normally chosen by the player; it is assigned by tournament pairing rules or by an online platform. Therefore, the present study does not recommend whether a player should move first. Instead, it studies the practical first-move advantage and recommends opening choices after the assigned colour is known.

Player ratings represent relative playing strength, time control determines how much thinking time is available, and the Encyclopaedia of Chess Openings (ECO) groups opening families. These variables are available before or at the opening stage and can be used to build an interpretable decision-support model. Ribeiro et al. (2013) showed that White's advantage is measurable but limited, while Glickman (2025) showed that order effects should be interpreted together with player strength and ties.

The aim of this research is to develop and evaluate a context-aware opening recommendation framework that uses pre-game metadata to rank opening choices conditional on assigned colour, rating context, and time control. The intended beneficiaries are chess players preparing an opening repertoire, coaches supporting player preparation, analysts studying opening performance, and training sys-

tems that need data-driven opening suggestions. The value of the recommendation is therefore in opening preparation after colour assignment, not in deciding who moves first.

2 LITERATURE REVIEW

First-move advantage depends on context rather than colour alone. Time pressure affects strategic risk-taking in professional chess (Carow & Witzig, 2025), supporting the use of time-control category. Player strength is also important because a White win may reflect a rating advantage rather than the first move. Omori and Tadepalli (2024) showed that rating can be estimated from move and clock information, demonstrating the close relationship between strength and game behaviour.

Opening choice affects the type of position that develops. De Marzo and Servedio (2023) quantified similarity and complexity in online chess openings, supporting ECO as a compact opening representation. For modelling, Light Gradient Boosting Machine (LightGBM) provides efficient gradient-boosted trees for structured data (Ke et al., 2017), while Categorical Boosting (CatBoost) is designed to handle categorical features effectively (Prokhorenkova et al., 2018). Samara et al. (2026) compared machine-learning approaches for chess outcome classification and highlighted the usefulness of rating and game-dynamic information as well as the difficulty of separating outcome classes.

The Chess Digits page is the direct data repository for the 200K file used in this study (Chess Digits, n.d.), rather than a prior recommendation study. Within the literature reviewed for this manuscript, no study was identified that combines these elements to convert a White non-loss probability into an opening ranking conditioned on assigned colour, rating group, time control, and ECO using this dataset. This combination of interpretable prediction and conditional opening ranking is the specific research gap and contribution addressed here.

3 METHODOLOGY

3.1 Data Source and Analytical Sample

The direct source was the Chess Digits 200K dataset, containing 200,000 rated Lichess games from May 2019 across Bullet, Blitz, Rapid, and Classical time controls. The downloadable data also include moves, clock information, and Stockfish evaluations (Chess Digits, n.d.). Only metadata required for this study were retained. After cleaning, 199,995 valid games remained. The key modelling variables were WhiteElo, BlackElo, ECO, opening name, result, time control, and category.

White won 100,351 games (50.18%), Black won 94,499 (47.25%), and 5,145 games (2.57%) were draws. The very small Draw class created class imbalance, which was considered during model evaluation.

Outcome distribution: White win - 100,351 games (50.18%); Draw - 5,145 games (2.57%); Black win - 94,499 games (47.25%).

3.2 Features, Targets, and Model Evaluation

Rating difference was calculated as White player's Elo rating (WhiteElo) minus Black player's Elo rating (BlackElo); a positive value means White was higher rated. Category represents Bullet, Blitz, Rapid, or Classical play, and ECO represents the opening family produced by both

players. Three Logistic Regression targets were evaluated: White Non-Loss (White win or draw versus White loss), White Win (win versus draw/loss), and Draw-versus-Loss among non-winning White games. LightGBM and CatBoost were additionally used for Black Win, Draw, and White Win classification.

The Logistic Regression workflow used a stratified train-test split, categorical encoding, and a fixed random state for reproducibility. Models were evaluated using accuracy, precision, recall, macro F1-score, and confusion matrices. The majority-class accuracy for White Non-Loss was approximately 52.75%, providing a minimum benchmark for the main model.

3.3 From Prediction to Recommendation

Prediction and recommendation are separate stages. The machine-learning model first estimates the probability that White avoids defeat for a given rating difference, time-control category, and ECO code. That probability is then combined with the historical average White score for the corresponding context. The resulting value is a ranking score used to compare candidate openings after a player's colour is already known.

Let R_W denote the White Recommendation Score, S_W the average historical White Score, P_NL the model's White Non-Loss probability, and R_B the Black Recommendation Score. Groups with fewer than 30 games were excluded from leading recommendations to reduce instability from very small samples.

$$R_W = 0.70\, S_W + 0.30\, P_NL \quad (1)$$

$$R_B = 1 - R_W. \quad (2)$$

The 70/30 weighting gives greater emphasis to observed historical performance while using the model probability as supporting evidence. It is a researcher-defined design choice, not an optimized constant, and is identified as a limitation for future validation.

4 RESULTS AND DISCUSSION

4.1 Model Results

White's average score was approximately 0.5146 when a win was counted as 1, a draw as 0.5, and a loss as 0. This is about 1.46 percentage points above an even score of 0.5 and is interpreted only as a descriptive association. The White Non-Loss Logistic Regression model achieved 57.85% accuracy and a macro F1-score of 0.58, approximately 5.10 percentage points above the majority-class baseline.

The White Win model achieved 57.38% accuracy. The Draw-versus-Loss model achieved 54.50% accuracy with a macro F1-score of 0.41. Among the three-class models, LightGBM achieved 52.15% accuracy and a macro F1-score of 0.3866, while balanced CatBoost achieved 38.09% accuracy and a macro F1-score of 0.3296. Balanced CatBoost increased attention to draws but produced many false Draw predictions. Binary and three-class accuracies are not directly comparable because they answer different prediction questions.

Model summary: White Non-Loss Logistic Regression - 57.85% accuracy, macro F1 = 0.58; White Win Logistic Regression - 57.38%, macro F1 approximately 0.57; Draw-versus-Loss Logistic Regression - 54.50%, macro F1 = 0.41; LightGBM three-class - 52.15%, macro

F1 = 0.3866; balanced CatBoost three-class - 38.09%, macro F1 = 0.3296.

4.2 Worked Recommendation Example

The following illustrative example shows how a prediction becomes a recommendation. Consider three candidate ECO openings for the same Rapid rating context after the player has already been assigned White. The values below are calculation examples used to demonstrate the ranking process; they are not additional empirical results reported from the dataset.

Illustrative candidate values: A (ECO C41): S_W = 0.72, P_NL = 0.74; B (ECO C40): S_W = 0.66, P_NL = 0.69; C (ECO B00): S_W = 0.60, P_NL = 0.62.

For Candidate A, (1) gives R_W = 0.70(0.72) + 0.30(0.74) = 0.726. The same calculation gives 0.669 for Candidate B and 0.606 for Candidate C, so Candidate A ranks highest for White in this illustrative context. Using (2), the complementary Black scores are 0.274, 0.331, and 0.394 respectively. Thus, the system does not choose the player's colour or first move; it ranks opening conditions conditional on the colour already assigned.

4.3 Practical Value, Beneficiaries, and Limitations

The practical beneficiaries are players, coaches, analysts, and training platforms. A player can use the ranking while preparing an opening repertoire for White or Black after pairing information is known. A coach can use it to identify opening families that performed favourably for players of a similar rating and time control. Analysts and training systems can use the same framework to present context-specific opening comparisons rather than a universal 'best opening'.

The moderate accuracy is expected because the models use only rating difference, broad time-control category, and opening code. They do not observe tactical mistakes, board positions, exact clock usage, or engine evaluations. The data also represent online chess and may not fully generalize to professional over-the-board play. The reported experiments used a single train-test split; future work should add repeated cross-validation, player- or time-based validation, probability calibration, external datasets, and systematic testing of the 70/30 weighting.

5 CONCLUSION

This research developed an interpretable framework for opening recommendations conditional on assigned player colour. It does not recommend who should make the first move, because White and Black are determined by tournament pairing or the playing platform. Instead, once colour is known, the framework combines historical performance with a modelled White Non-Loss probability to rank opening contexts by rating group, time control, and ECO.

The White Non-Loss Logistic Regression model was selected because its 57.85% accuracy and 0.58 macro F1-score aligned best with the main objective and exceeded the approximately 52.75% majority baseline. LightGBM was the stronger three-class model, while balanced CatBoost was less reliable because of draw over-prediction. The recommendations are most useful to players and coaches as preparation support and should be interpreted as context-specific rankings rather than guaranteed outcomes.

6 DATA AVAILABILITY

The direct dataset is the Chess Digits 200K dataset, publicly available at https://web.chessdigits.com/data. It contains 200,000 rated Lichess games from May 2019. The cleaned analytical dataset used in this study contains 199,995 games. Lichess is the original playing platform, while Chess Digits is the direct data source.

REFERENCES

- Carow, J., & Witzig, N. M. (2025). Time pressure and strategic risk-taking in professional chess. Journal of Economic Behavior & Organization, 238, 107218. https://doi.org/10.1016/j.jebo.2025.107218
- Chess Digits. (n.d.). Data: 200k Lichess games from May 2019. Retrieved August 6, 2026, from https://web.chessdigits.com/data
- De Marzo, G., & Servedio, V. D. P. (2023). Quantifying the complexity and similarity of chess openings using online chess community data. Scientific Reports, 13, 5327. https://doi.org/10.1038/s41598-023-31658-w
- Glickman, M. E. (2025). Paired comparison models with strength-dependent ties and order effects. Statistical Modelling. Advance online publication. https://doi.org/10.1177/1471082X251400474
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 30, 3146-3154. https://proceedings.neurips.cc/paper/6907-lightgbm-a-highly-efficient-gradient-boosting-decision-tree
- Omori, M., & Tadepalli, P. (2024). Chess rating estimation from moves and clock times using a CNN-LSTM. arXiv. https://doi.org/10.48550/arXiv.2409.11506
- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. Advances in Neural Information Processing Systems, 31, 6638-6648. https://proceedings.neurips.cc/paper/2018/hash/14491b756b3a51daac41c24863285549-Abstract.html
- Ribeiro, H. V., Mendes, R. S., Lenzi, E. K., del Castillo-Mussot, M., & Amaral, L. A. N. (2013). Move-by-move dynamics of the advantage in chess matches reveals population-level learning of the game. PLOS ONE, 8(1), e54165. https://doi.org/10.1371/journal.pone.0054165
- Samara, K., Antreassian, A., Klug, M., & Hasan, M. S. (2026). Machine learning approaches for classifying chess game outcomes: A comparative analysis of player ratings and game dynamics. Electronics, 15(1), Article 1. https://doi.org/10.3390/electronics15010001

Proceedings of the18th Symposium on Applied Science, Business & Industrial Research –2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Comparison of Convolutional Neural Network Models for Image-Based Detection of Sewing Defects and Efficiency Improvement

Perera WASM[1], Munasinghe B[2]
[1,2]*Department of Computing and Information Systems, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*
malshashalani95@gmail.com[1], bhagya@wyb.ac.lk[2]

ABSTRACT

Manual inspection of garments for sewing defects is time-consuming, inconsistent across inspectors and shifts, and difficult to scale in high-volume production lines, directly limiting production efficiency. This study presents an automated image classification system that detects and diagnoses sewing defects originating from needle and bobbin mechanism failures, a defect category that has received limited attention relative to general fabric-surface defect detection. A dataset of 5,455 images across six classes (bobbin thread pulling up, broken stitch, crooked seam, needle mark, skipped stitch, and normal) was constructed from the StitchingNet repository and expanded through offline augmentation. Three pre-trained convolutional neural network architectures, Visual Geometry Group 16 (VGG16), EfficientNetB3, and MobileNetV2, were fine-tuned under identical two-phase transfer learning conditions. VGG16 achieved the highest test accuracy of 89.94% and a macro-averaged F1-score of 0.9004, outperforming EfficientNetB3 (85.21%) and MobileNetV2 (79.52%). Across all three models, needle mark and normal classes were most frequently confused, indicating a visual ambiguity rather than a model-specific weakness. The best-performing model, VGG16, was deployed in a Streamlit web application backed by a Supabase cloud database, supporting shift-based batch inspection, defect-cause diagnosis, and trend reporting for factory quality management. Validation on independently captured phone images yielded 91.67% accuracy, closely matching the benchmark test accuracy and confirming that the model generalizes beyond its curated training source. This study also identifies a limitation in the evaluated deployment point and proposes extending this work toward real-time defect detection at the point of stitching, catching defects as they occur.

KEYWORDS: **Convolutional Neural Network, Deep Learning, Sewing Defect Detection, Transfer Learning**

1 INTRODUCTION

Garment manufacturing is a labour-intensive process in which product quality directly determines customer satisfaction and manufacturer competitiveness. Sewing defects are among the most common sources of quality loss, with broken and skipped stitches alone responsible for a substantial share of reported sewing problems in industrial settings (T. Islam, 2017). Traditional quality control relies on manual visual inspection, which is slow, inconsistent across inspectors and shifts, and difficult to scale as production volumes grow.

Deep learning-based image classification offers a route to automating this inspection task. Convolutional Neural Networks (CNNs) have demonstrated strong performance in general fabric and surface defect detection, but most existing work treats sewing defects as a single generic category rather than tracing them to their mechanical origin (Choudhary et al., 2018). Defects such as skipped stitch, needle mark, and broken stitch arise specifically from needle and bobbin mechanism failures, and diagnosing them at this level of granularity allows corrective action to be targeted at the

actual faulty component rather than treated as a generic quality failure.

This paper addresses this gap by comparatively evaluating three transfer learning architectures, VGG16, EfficientNetB3, and MobileNetV2, for classifying six defect categories tied to needle and bobbin mechanisms. The best-performing model is deployed in a cloud-connected web application that supports shift-based batch inspection and trend reporting, and its generalization is validated on independently captured images beyond the training data source.
Because a labelled dataset was available only at the quality-checking point rather than at the sewing machine itself, the system evaluated here classifies static images from that stage; beyond these results, this paper also examines this placement as a practical limitation and proposes a concrete direction for moving detection to the sewing operator's workstation, where defects could be identified in real time as they occur during stitching.

2 LITERATURE REVIEW

Sewing defects arise mainly from the interaction between the needle, the bobbin, machine settings, and fabric properties (Choudhary et al., 2018). Needle-related failures such as skipped stitches, needle marks, and crooked seams typically result from incorrect needle selection or misalignment, while bobbin-related failures such as thread pulling up and broken stitches commonly result from improper thread tension (M. Islam 2021). CNNs have become the standard approach for automated visual inspection, extracting hierarchical features directly from images without manual feature engineering. Fabric surface defects have likewise been linked to apparel manufacturing efficiency (Kalkancı, 2020). Kim et al. (2022) used CNN feature-map analysis to detect broken stitches, showing that convolutional layers effectively capture the spatial irregularities associated with sewing flaws. Jung et al. (2025) introduced the StitchingNet dataset, spanning ten sewing-related defect categories. This study applies deep transfer learning to a subset of these categories, focused specifically on needle- and bobbin-related defects.Despite this progress, existing literature treats needle and bobbin-caused defects as part of a broader, undifferentiated taxonomy alongside mechanically unrelated flaws such as staining or fabric pinching, rather than isolating them for dedicated, mechanically diagnostic classification (Kim et al. (2022) ; Jung et al., 2025). Controlled comparisons of multiple transfer learning architectures under identical conditions on such a dataset remain rare, and most studies report accuracy only on held-out data from the same curated source, without validating performance on independently captured, real-world images. Most also stop at model evaluation without producing a deployable, production-facing tool for factory quality teams.

3 METHODOLOGY

3.1 Dataset and Preprocessing

The dataset was derived from StitchingNet (Jung et al., 2025), which contains 14,565 images. Only the five defect classes caused mainly by needle and bobbin failures (bobbin thread pulling up, broken stitch, crooked seam, needle mark, and skipped stitch) and the normal class were selected, as other defects

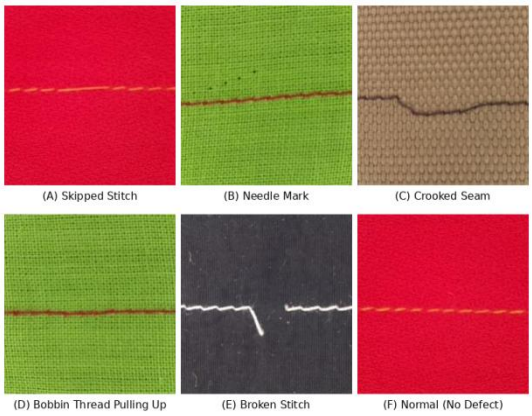


Figure 1: Visual characteristics of the six sewing defect categories used for classification

cannot be traced to these components. These 2,728 images were expanded to 5,455 through offline augmentation (brightness/contrast, rotation, flipping, cropping) using Albumentations, then split 70:15:15 into training, validation, and test sets with a fixed seed, using the same split across all three architectures.

3.2 Transfer Learning Architectures

VGG16, EfficientNetB3, and MobileNetV2 were adapted with a shared classifier head (Global Average Pooling, Batch Normalization, two dense layers with dropout, softmax output) and trained in two phases: the pre-trained base was frozen while only the classifier head was trained, then the deepest layers were unfrozen and fine-tuned at a reduced learning rate. All models used the Adam optimizer, categorical cross-entropy loss, and identical early stopping and checkpointing.

3.3 Deployment and Evaluation

The best-performing model was deployed in a Streamlit web application with a Supabase backend, supporting shift-based batch upload, defect diagnosis with severity and probable cause, and daily, weekly, and monthly trend

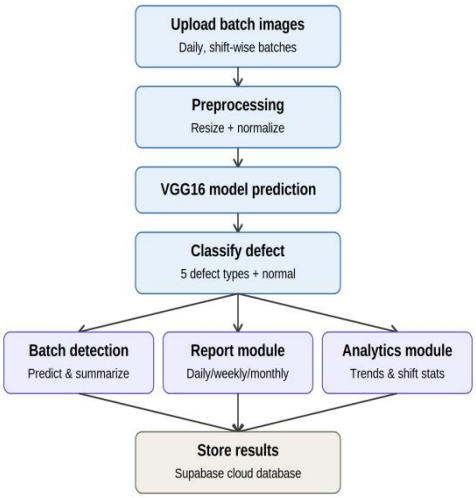


Figure 2: Workflow of the deployed sewing defect classification system

reporting. Performance was evaluated using accuracy, precision, recall, F1-score, and confusion matrices on the test set, and generalization was validated on 24 independently captured phone images against a matched Kaggle subset.

4 RESULTS AND DISCUSSION

VGG16 achieved the highest test accuracy of 89.94% and macro-averaged F1-score of 0.9004 on the 825-image test set, outperforming EfficientNetB3 (85.21%, F1 = 0.8521) and MobileNetV2 (79.52%, F1 = 0.7980), as summarized in Table 1. This ordering is consistent with VGG16's deep, uniform stack of 3×3 convolutional layers, which appears better suited to the fine-grained texture discrimination that sewing-defect classification requires, even though its effective deployed parameter count (≈15.1M, after removing the ImageNet classification head) is close to that of EfficientNetB3 (≈12M) and substantially larger than MobileNetV2 (≈3.05M).

Across all three architectures, bobbin thread pulling up, broken stitch, and crooked seam were classified with F1-

Table 1: Comparative performance of the three transfer learning models on the test set

Model	Test Acc.(%)	Macro F1	Params
VGG16 (selected)	89.94	0.9004	15.1M
EfficientNet B3	85.21	0.8521	12.0M
Mobile NetV2	79.52	0.7980	3.05M

scores above 0.90, while needle mark and normal were consistently the weakest classes, with VGG16's confusion matrix showing 34 of 134 needle mark images misclassified as normal. Because this pattern recurred independently across three architectures with different inductive biases, it is attributable to genuine visual ambiguity between faint needle marks and defect-free fabric rather than to any single model's limitations.

Real-world validation of the deployed VGG16 model (Table 2) showed that accuracy on independently captured phone images (91.67%) closely tracked the benchmark test accuracy, with only a 4.16 percentage-point gap relative to the matched Kaggle subset. Average confidence was consistently lower for incorrect predictions (43.5%) than for correct ones (71.5%) on the phone images, indicating that the model's confidence score remains a meaningful reliability signal even under domain shift from curated to real-world imaging conditions. These results support the deployment of VGG16 in a shift-based inspection tool: the model not only achieves the highest accuracy among the three architectures tested, but also maintains reliable, well-calibrated performance when applied to images collected under conditions that differ from its training source, an important requirement for practical factory-floor use.

Table 2: Accuracy comparison between the curated test set and independently captured images

Dataset	Images	Acc. (%)	Avg. Conf. (Correct)	Avg. Conf. (Incorrect)
Full Kaggle test set	825	89.94	--	--
Kaggle test subset (matched)	24	95.83	94.0	65.0
Self-collected phone images	24	91.67	71.5	43.5

5 CONCLUSION

This study developed and compared three transfer learning architectures for classifying sewing defects originating from needle and bobbin mechanism failures, addressing a gap between generic fabric-defect research and the specific diagnostic needs of garment quality control. VGG16 was identified as the most effective architecture, achieving 89.94% test accuracy and a macro F1-score of 0.9004, and was deployed in a cloud-connected web application supporting shift-based batch inspection and trend reporting. Independent validation on real-world images confirmed that the model generalizes beyond its curated training source, supporting its practical use as a production-facing quality-control tool. By replacing slow, manual, error-prone inspection with automated classification and diagnosis, the system evaluated in this

study contributes directly toward the production efficiency improvement motivating this research. A recurring limitation, the confusion between needle mark and normal classes across all three architectures, and a deployment-placement limitation identified through practical consideration of the system's use in a factory setting, together motivate the future work discussed below.

6 FUTURE WORK

The system evaluated here is designed around the quality-checking point, where inspectors upload shift-wise batches of images after already identifying defects manually, an added step that becomes impractical when a shift can involve hundreds or over a thousand garments. A more effective placement is at the sewing operator's workstation: a camera mounted close to the needle could observe the stitching process directly and flag a defect such as a skipped stitch or needle mark as it begins to form, allowing the operator to correct it immediately rather than accumulating defective units for later inspection. Realising this would require a labelled video dataset captured at the point of sewing.

Consistent confusion between needle marks and normal classes across all three models highlights the need for more data covering faint or partially visible needle marks under varied fabric and lighting conditions.

REFERENCES

- Choudhary, A. K., Sikka, M. P., & Bansal, P. (2018). The study of sewing damage and defects in garments. Research Journal of Textile and Apparel, 22(2), 109–120. https://doi.org/10.1108/RJTA-08-2017-0041
- Islam, M. (2021). Identification of different types of sewing defect and causes with controlling methods. ResearchGate.https://www.researchgate.net/publication/351838737
- Islam, T. (2017). Analysis of major defects, position, and percentage in sewing lines of a garments factory with the help of Pareto chart, cause-effect diagram and sigma level. International Journal of Scientific and Engineering Research, 8(6). https://www.researchgate.net/publication/321085010
- Jung, W.-K., Kang, J., Kwon, W., & Kim, H. (2025). StitchingNet and deep transfer learning method for sewing stitch defect detection. Journal of Computational Design and Engineering, 12(4), 140–154. https://doi.org/10.1093/jcde/qwaf037
- Kalkancı, M. (2020). Investigation of surface defects and apparel manufacturing efficiency of fabrics woven from recycled cotton and blends. Industria Textila, 71(3), 266–274.
- Kim, H., Jung, W.-K., Park, Y.-C., Lee, J.-W., & Ahn, S.-H. (2022). Broken stitch detection method for sewing operation using CNN feature map and image-processing techniques. Expert Systems with Applications, 188, Article 116014. https://doi.org/10.1016/j.eswa.2021.116014

Proceedings of the18th Symposium on Applied Science, Business & Industrial Research –2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Gender-Based Differences in Information Security Awareness and Vulnerable Online Practices Among Art-based University Undergraduates: Evidence from a State University in Sri Lanka

Jayaweera LP[1]
[1]*Department of Information technology, Faculty of Dance and Drama, University of the Visual and Performing arts, Colombo, Sri Lanka*
lahiruprabashan@gmai.com[1], lahiru.pj@vpa.ac.lk[1]

ABSTRACT

Information security has become an essential skill for university students in the present situation. Universities itself being directed to protect the online resources belongs to university by providing various information security training to staff and students. When designing an information security program for university students who are from none-IT background, it's still doubtful that demographic factors like gender are affecting information security awareness and, facing common vulnerabilities. Prior studies keep unresolved question whether male or female university students show the strong cyber security awareness. This paper investigates gender-based differences in three information security awareness dimensions such as computer efficacy, internet literacy, and network knowledge, along with a composite measure of common cybersecurity vulnerabilities. The study focused on second-year students enrolled in an art-based state university in Sri Lanka. Second-year students were selected because they are expected to have substantial engagement with the internet and digital platforms compared with students in other academic years. The Morgan table was used to select the sample size. Stratified random sampling technique was used for select the sample. Data were collected among 251 undergraduates (144 female, 107 male) at a university using a structured, five-point Likert-scale online questionnaire and analyzed using independent-samples t-tests to compare male and female students on each measure. The outcome of this research indicates that even male students show more awareness than female students, but it is not significant for all three dimensions and security vulnerabilities. Therefore, universities should be design cybersecurity training, regardless of gender and address all students equally.

KEYWORDS: Cybersecurity vulnerability, Gender differences, Information security awareness, University undergraduates

1 INTRODUCTION

In contemporay university education, students are highly dependent on the internet and digital platforms. Students use digital platforms for academic, social, administrative, financial etc. purposes in different directions. After the COVID pandemic, online education has become the most effective option for the teaching and learning process. With the present situation, university students have become one of the most vulnerable groups to cyber threats (Arpaci et al., 2025).Therefore, Information security awareness (ISA) becomes a necessary skill for students to identify and prevent cyber threats while browsing and using digital platforms. The role of demographic factors such as gender regarding cyber threat awareness and facing vulnerabilities is subject to debate. The study about "Gender-based pilot analysis of cybersecurity awareness in a digitizing Indian society" by Karthikeyan, males have slightly stronger awareness of basic cybersecurity concepts and security indicators, while females appear to be more cautious when engaging in privacy-related online activities. These findings highlight the importance of developing cybersecurity training and awareness

programs that take into account the different needs, behaviours, and levels of awareness among diverse demographic groups. (Karthikeyan et al., 2026). If a gender gap exists, universities need to focus on gender targeted awareness programs in order to secure online resources related to the university. This study is carried out to examine whether male and female university undergraduates differ in core information security awareness and vulnerability to common cybersecurity threats. The study was carried out to assess whether there are significant differences between male and female students in computer efficacy, internet literacy, and networking knowledge and to compare male and female undergraduate students' exposure to common cyber security vulnerabilities.

2 LITERATURE REVIEW

The existing studies about information security awareness and vulnerabilities are questionable to provide direct answer whether gender factor is affecting to identify and mitigating the cyber-attacks. Majority of existing research focuses on few variables to identify the awareness level of the sample. Nagahawatta (2020), for focuses at Sri Lankan university students and found that male students came out ahead, showing stronger overall cybersecurity awareness than their female peers, even though most students in that study were at least broadly familiar with basic tools like antivirus software, which seems to support the common assumption that male students are somehow "more tech-savvy" by default. The study conducted by Du and Chintakovid (2023), based on the population of Yunnan University of Finance and Economics in China, and result was completely opposite of Nagahawatta (2020). In their sample, female students actually scored higher across several areas, including general awareness, technical knowledge, privacy habits, and password management. Another study conducted by Alqahtani (2022) where sample contains majority of female students' shows comparatively weaker cybersecurity awareness than male students, particularly around browser security habits and how carefully they managed their passwords. McCormac et al. (2017) also examined individual differences in information security awareness and considered gender, personality and risk-taking propensity as possible factors related to information security awareness. Similarly, a study on gender differences in information security and privacy reported that gender differences can vary across particular security and privacy behaviors rather than appearing equally in every behavior (McGill & Thompson, 2021). Recent research on university students has also developed measures covering personal cybersecurity awareness and secure online behaviors, showing that gender differences can occur in some specific behavior areas while not necessarily appearing across all dimensions (Bognar & Bottyan, 2024). Based on above studies could not be identified exactly whether the gender factor is important or not for focusing on cyber security awareness. As mentioned earlier, the existing studies focus on limited variables to measure the cyber security awareness level, such as browser security, using anti-virus software, etc. The present research focuses on 3 variables, which are computer efficacy, internet literacy and networking knowledge to measure cyber security

awareness and 10 common vulnerability experiences faced by the population.

3 METHODOLOGY

The second-year students of a state university were selected. The selected students were from an art-based university context. They were selected, because they actively engaging with different digital platforms compared with other years. There were 708 students in second year according to sample frame; student registration list was available in student welfare division. The Morgan table was used to select the sample size as its established method to get the sample size for a small or a medium-size population. According to the Morgan table, sample size is 251 and stratified random sampling techniques were used to select the sample in-order to reduce the bias. An online questionnaire was distributed using various channel and 251 records use to analyze the data. The questionnaire was consisted three main sections for gathering general information, information security awareness level and common vulnerability practices. Section 1. Gathering general information, including Gender, section 2. Measures awareness levels across specific domains using a 5-point Likert scale ranging from 'Poor' (1) to 'Excellent' (5), and section 3. Measures the frequency or agreement with unsafe online behaviors using a 5-point Likert scale ranging from 'Strongly Disagree' (1) to 'Strongly Agree' (5). The questionnaire was developed specifically for this study based on the research objectives and reviewed to ensure that the questions were understandable and relevant before it was distributed to the participants. No separate pilot survey was conducted prior to the main data collection.

4 DATA ANALYSIS

The analysis was performed using Python with advanced statistical libraries. A significance level of 0.05 was used for the statistical tests. Welch's Independent-Samples t-test was performed due to unequal group size which does not assume equal variances; 144 females and 107 male. Holm-Bonferroni Correction: Applied to adjust p-values for multiple comparisons across multiple outcomes such as Overall Awareness, Computer Efficacy, Internet Literacy and Networking Knowledge and the ten vulnerability items. As the sample size is moderate to quantify the magnitude of differences between genders Hedges' g: was used.

5 RESULT AND DISCUSSION

The first step is to check the reliability of the data set for further processing. The Cronbach's alpha was 0.957, indicating very high internal consistency across all dimensions measured. The following table represents the statistical summary of the above test.

The table below shows the results of Hedges' g which represent the magnitude of differences between genders.

Table 1 shows the Welch t-test value, raw p value which generate from Welch-T test after multiple comparison correction. According to above two tables Overall awareness, slightly higher in male students (M = 2.98, SD = 0.71) compared to female students (M = 2.84, SD = 0.66). The Welch's t-test yielded t = -1.63, with a raw p-value of .105. After applying the Holm-Bonferroni correction , the adjusted

Table 1. Statistical Summary

Outcome	Female M (SD)	Male M (SD)	Welch T	Raw p	Holm p
Overall awareness	2.84 (0.66)	2.98 (0.71)	-1.63	.105	.526
Computer efficacy	2.80 (0.72)	2.92 (0.76)	-1.23	.219	.565
Internet literacy	2.89 (0.71)	3.02 (0.73)	-1.45	.147	.565
Networking knowledge	2.78 (0.93)	2.87 (0.92)	-0.77	.445	.565

Table 2. Hedges' g effect sizes

Outcome	Female M(SD)	Male M(SD)	Hedges' g
Overall awareness	2.84 (0.66)	2.98 (0.71)	-0.21
Computer efficacy	2.80 (0.72)	2.92 (0.76)	-0.16
Internet literacy	2.89 (0.71)	3.02 (0.73)	-0.19
Networking knowledge	2.78 (0.93)	2.87 (0.92)	-0.10

p-value was .526. The effect size suggests a small magnitude difference that does not reach statistical significance.

5.1 Computer efficacy

This section of the questionnaire was designed to assess the students' ability to manage security tasks on their computers. Mean scores were 2.80 (SD = 0.72) for females and 2.92 (SD = 0.76) for males. The difference was not significant

5.2 Internet literacy

Internet literacy section determines the students' ability to recognize secure websites and manage social media privacy and browse internet resources securely. Females scored an average of 2.89 (SD = 0.71), while males scored 3.02 (SD = 0.73).

5.3 Networking knowledge

This section is for assessment of Wi-Fi, secure internet connectivity, etc., and yielded the lowest differences among domains. Mean scores were 2.78 (SD = 0.93) for females and 2.87 (SD = 0.92) for males

5.4 Vulnerable online practices

10 risky behaviors were used to assess engagement in vulnerable practices in the sample. The overall mean for female students was 2.83 (SD = 0.74) and for male students was 2.97 (SD = 0.74). The statistical analysis (Welch t = -1.48, p_raw = .141, p_holm = .565) confirmed that gender does not significantly impact the total vulnerability score (g = -0.19 The following table represents each item used to assess the common vulnerabilities. According to below table 3 male students are comparatively high use of simple guessable password, malware incidents and using advance security features. Other vulnerable practices are almost similar for both parties.

The results of this study indicate there is no high degree of difference between female and male students regarding their information-security awareness and common vulnerable practices. The overall awareness level is characterized as "average" for both groups. Although male students reported slightly higher mean awareness, they also reported slightly higher involvement in vulnerable practices (M = 2.97 vs. 2.83), however, this was not statistically significant.

6 CONCLUSION

The study determines that gender is not a significant factor in determining information-security awareness or vulnerable practices among art-based

Table 3. Mean values of each vulnerability for gender

Vulnerable practices	Female	Male
Password hygiene	2.82	3.15
Malware incidents	2.82	3.11
Security app management difficulties	2.69	2.80
Phishing recognized	3.17	3.14
Software download securely	2.56	2.79
Use public Wi-Fi	2.82	2.85
Public device logout	2.71	2.78
Advanced security features of mobile/laptop	2.95	3.13
Account synchronizing	3.14	2.91
Delaying software update	2.78	2.90

university students in this specific Sri Lankan context. Therefore, university should provide education for information security regardless of gender based and must focusing on special area where both parties are equally lack of knowledge such as secure software download, secure Wi-Fi usage, use of advance security features etc. However, there are a number of limitations in this study such as the selected sample being second-year students of an art-based university in Sri Lanka. The study did not focus on institutional culture. This study captures a snapshot on time, which does not account for change of knowledge and behavior of sample. As the study was conducted among selected art-based undergraduates in one state university, the findings may not be generalized to students from other academic disciplines or universities. Future studies can consider students from different academic disciplines and universities to provide a broader understanding of gender differences in information security awareness and vulnerable online practices.

REFERENCE

- Alqahtani, M. A. (2022). Factors affecting cybersecurity awareness among university students. Applied Sciences, 12(5), 2589. https://doi.org/10.3390/app12052589
- Arpaci, I., Aslan, O., & Oner, I. E. (2025). Cybersecurity awareness scale (CSAS) for social media users: Development, validity and reliability study. Information Development. https://doi.org/10.1177/02666669251336562
- Bognar, L., & Bottyan, L. (2024). Evaluating online security behavior: Development and validation of a personal cybersecurity awareness scale for university students. Education Sciences, 14(6), 588. https://doi.org/10.3390/educsci14060588
- Du, X., & Chintakovid, T. (2023). A survey of cybersecurity awareness among undergraduate students at Yunnan University of Finance and Economics in China. In Proceedings of the 2023 4th International Conference on Education, Knowledge and Information Management (ICEKIM 2023) (pp. 740–753). Atlantis Press. https://doi.org/10.2991/978-94-6463-172-2_78
- Karthikeyan, S., Umadevi, K., & Wan, M. (2026). Gender-based pilot analysis of

cybersecurity awareness in a digitizing Indian society. Proceedings of the Indian National Science Academy. Advance online publication. https://doi.org/10.1007/s43538-025-00678-7

- McCormac, A., Zwaans, T., Parsons, K., Calic, D., Butavicius, M., & Pattinson, M. (2017). Individual differences and information security awareness. Computers in Human Behavior, 69, 151–156. https://doi.org/10.1016/j.chb.2016.11.065
- McGill, T., & Thompson, N. (2021). Exploring potential gender differences in information security and privacy. Information & Computer Security, 29(5), 850–865. https://doi.org/10.1108/ICS-07-2020-0125
- Nagahawatta, R., Warren, M., & Yeoh, W. (2020). A study of cybersecurity issues in Sri Lanka. International Journal of Cyber Warfare and Terrorism, 10(3), 59–72. https://doi.org/10.4018/IJCWT.2020070105

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Generative AI and the Evolving Career Transition of QA Engineers in the Sri Lankan IT Industry

Erandi RHN[1], Jayasundara DRT[2], Jayasinghe JABU[3]
[1,2,3]*Department of Mathematics, Faculty of Engineering, University of Moratuwa*
nipuniranepura@hotmail.com [1], ravindij@uom.lk[2], bimalij@uom.lk[3]

ABSTRACT

Career transition ambition among Quality Assurance (QA) engineers in the Sri Lankan IT industry can depend on the use of Generative AI in QA engineering roles, task engagement, and reskilling behaviour. To study this, a sample of 104 respondents was used to collect data using a structured online questionnaire. Descriptive statistical techniques summarize demographic characteristics, while odds ratios were used to evaluate the association between emerging QA roles and willingness to reskill. In this study group 80% of respondents fall within the 25–35 age group. Since younger professionals are often more open to adopting new tools and learning emerging technologies such as Generative AI, this would have contributed to the relatively high level of AI usage observed in this study. The sampled QA engineers further show their highest engagement in core QA tasks, including bug detection and reporting, test case generation and test reporting, and documentation, while comparatively low engagement is observed in advanced technical tasks, such as automation script creation, performance test analysis, and CI/CD pipeline integration assistance. The findings reveal that hybrid QA roles dominate the workforce, with AI tools widely used in test case generation, bug detection, and documentation. Through the odds ratios, it is revealed that QA professionals preparing to transition into advanced roles such as AI/ML Quality Specialist, Test Data Engineer, and Quality Architect are more likely to engage in reskilling activities. Thus, reskilling should be more focused on advanced engineering capabilities, AI, and analytical competencies when focusing on career transitioning.

KEYWORDS: Career Development, Generative AI, IT Quality Assurance, Reskilling Willingness

1 INTRODUCTION

In the context of software development, the integration of Generative AI is reshaping the Software Development Life Cycle (SDLC), particularly in the field of Quality Assurance (QA) (Pressman & Maxim, 2020; Sommerville, 2016; Durrani el al., 2024). QA plays a critical role in ensuring software reliability, functionality, and performance. Traditionally, QA engineers have been responsible for manual testing, test case design, defect identification, and documentation. However, with the increasing adoption of AI-powered tools, many of these tasks are becoming automated or augmented. Tools powered by Generative AI can now generate test cases, predict potential defects, analyze test coverage, and even produce detailed reports, thereby significantly improving efficiency and reducing manual effort (Almalki & Javaid, 2022; Zhang et al., 2022; Amalfitano et al., 2023). Rather than focusing solely on repetitive manual tasks, QA professionals are increasingly required to engage in more analytical, strategic, and technical activities (Baker et al., 2023; Khan & Khan, 2022). The emergence of hybrid roles, which combine manual testing with automation and AI-assisted processes, reflects this

transition. As a result, QA engineers must adapt to new technologies, acquire advanced technical skills, and continuously update their knowledge to remain relevant in an evolving industry (Garousi et al., 2024).

In the Sri Lankan IT industry, which is experiencing rapid growth, the integration of AI technologies is becoming increasingly prominent. Organizations are investing in automation and AI-driven solutions to remain competitive in the global market. However, there is limited empirical research that specifically examines how these changes affect QA engineers in this regional context. Understanding these dynamics is essential for organizations, policymakers, and educational institutions to develop strategies that support workforce adaptation and sustainable growth (Garousi et al., 2024; Dishan et al., 2024).

Furthermore, the concept of reskilling has become a critical factor in the AI-driven workplace. As job roles evolve, professionals must acquire new competencies to meet changing demands. The willingness of QA engineers to reskill is influenced by various factors, including career aspirations, exposure to AI technologies, and perceived benefits of adopting new roles. In this context, analyzing the relationship between emerging QA roles and reskilling behavior provides valuable insights into workforce readiness and adaptability.

This study aims to identify the tendency to use Generative AI in QA tasks in the IT industry, to assess the trends of transitioning into new QA roles and reskilling willingness, and to recognize the trending professional skills among QA roles which will be highly beneficial to the growing workforce focusing on career transitioning towards career development.

2 METHODOLOGY

The structured questionnaire included a range of close-ended questions covering demographic characteristics, professional roles, level of engagement in QA-related tasks, and the extent of AI usage in daily work activities. In addition, the questionnaire assessed respondents' preparation for transitioning to emerging QA roles and their willingness to acquire new skills to adapt to AI-driven changes in the industry. The survey was administered through an online platform, allowing respondents to participate conveniently and anonymously. The target population consisted of QA engineers with varying levels of experience and involvement in different testing roles, ensuring a diverse representation of perspectives related to AI adoption.

A total of 104 valid responses were obtained and used for the analysis. The study adopted convenience and snowball sampling techniques, and the questionnaire was shared through professional networks and internal communication channels within the organizations. The questionnaire was divided into five sections covering demographic and professional characteristics, exposure to Generative AI (GenAI), career and skills transition, the perceived impact of GenAI on QA skill development, and open-ended feedback. The key variables and their corresponding measurement scales are summarized in Table 1.

Table 1: Variables and Measurement Scales

Variable	Measurement Scale
Demographic	Categorical
QA involvement	0%–100%
GenAI usage	Yes / No
GenAI usage frequency	Ordinal
GenAI in QA tasks	0%–100%
Perceived AI effectiveness	5-point Likert
Reskilling need	Yes / No / Maybe
Potential QA role transition	Selected / Not selected
Essential future QA skills	Selected / Not selected
Future QA role perceptions	5-point Likert
AI impact on QA skills	−100% to +100%

Two main analytical methods were employed in this study: descriptive statistical analysis and Odds Ratio analysis**.** Odds Ratio (OR) analysis was applied to examine the association between QA roles that professionals are preparing to transition into and their willingness to reskill.

$$OR = \frac{ad}{bc} \quad (1)$$

a = number preparing to transition into a role and willing to reskill, b = number preparing to transition but not willing to reskill, c = number not preparing to transition into the role but willing to reskill, d = number not preparing to transition into the role and not willing to reskill

3 RESULTS

3.1 Descriptive Analysis

Table 1 presents the results obtained from the data analysis, focusing on descriptive statistics and Odds Ratio analysis. The findings provide insights into the demographic characteristics of the respondents, their professional roles, patterns of AI usage, engagement in QA tasks, and the relationship between emerging QA roles and reskilling willingness.

3.2 Job role distribution

The analysis of job roles indicates that hybrid QA roles, which combine both manual and automation testing, represent the largest proportion of respondents (approximately 59%). This highlights a significant shift from traditional role specialization toward more integrated and flexible roles. The dominance of hybrid roles reflects an ongoing transition in the QA industry. Organizations appear to be adopting a balanced approach by combining manual testing expertise with automation and AI-assisted techniques. This approach allows teams to maintain testing quality while gradually integrating advanced technologies. It also indicates that QA professionals are expected to possess a broader skill set, including both traditional testing knowledge and modern technical capabilities.

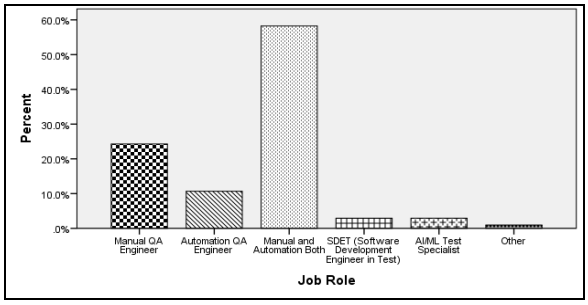


Figure 1: Job Role of QA Engineers

Table 2: QA Workforce Characteristics

Variable	Key Findings
Age	~80% in 25–35 age group
Gender	63% Female, 37% Male
Work Experience	2–5 years (45%) dominant
Designation	>60% QA Engineers / Senior QA Engineers
Job Roles	59% Hybrid (Manual + Automation) <5% (SDET, AI/ML Testers

3.3 Task Engagement Levels

The disparity in task engagement levels indicates that while AI is helping automate routine tasks, there is still a need for QA professionals to develop advanced technical skills.

3.4 AI Usage by QA Job Role and Age

Manual QA Engineers represent about 25%, suggesting that AI is being actively adopted even in traditional manual testing tasks such as test case generation, defect analysis, and documentation. Automation QA Engineers account for around 10% of the sample and show moderate usage, typically on a weekly or occasional basis, implying that AI plays more of a supporting role alongside established automation frameworks.

Specialized roles like SDET and AI/ML Test Specialist contribute less than 5% each and show limited variation in usage frequency, due to small samples.

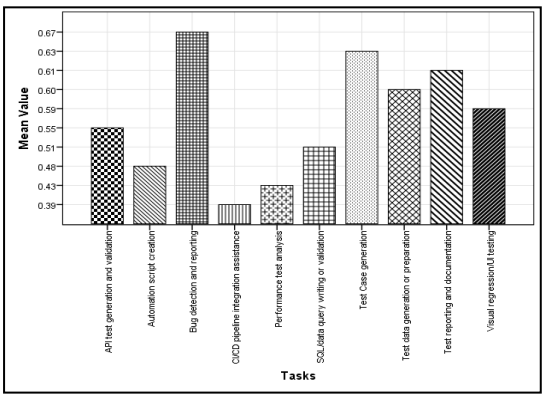


Figure 2: Average proportional engagement in day-to-day tasks

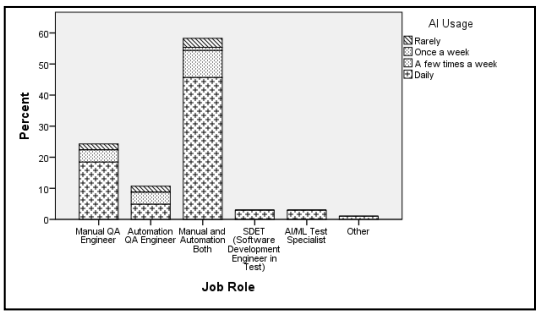


Figure 3: AI Usage Frequency Across Software Testing Job Roles

The Other category is minimal. In terms of age distribution, 89 of the 104 respondents (85.6%) were below 35 years of age. Among these respondents, 76% reported using AI tools on a daily basis, indicating a high level of routine AI adoption among the younger segment of the QA workforce.

3.5 Odds Ratio Analysis

The Odds Ratio (OR=2.08) for respondents preparing to transition into AI/ML Quality Specialist roles highlights that individuals planning to move into AI-focused roles are approximately twice as likely to engage in reskilling compared to those who are not preparing for such roles. The OR (2.38) for respondents preparing to transition into the Test Data Engineer role indicates a positive association with willingness to reskill. This suggests that individuals preparing for data-related roles are more than twice as likely to engage in reskilling activities. However, the 95% confidence interval (0.88, 6.45) includes the value 1, indicating that the statistical significance of this result can be changed. The OR (4.13) for respondents preparing to transition into the Quality Architect role indicates a strong positive association with reskilling willingness. This suggests that these individuals are more than four times as likely to engage in reskilling compared to those who are not preparing for this role.

Table 3: Odds Ratio Analysis of Future Transitioning QA Roles and Reskilling Willingness

Future QA Role	OR	90% CI
AI/ML Quality Specialist	2.08	1.02 – 4.25
Test Data Engineer	2.38	0.88 – 6.45
Quality Architect	4.13	1.52 – 11.21

Key: OR > 1 indicates a positive association

Importantly, the 90% confidence intervals, which do not include the value 1, indicate that the results are statistically significant. These findings emphasize that professionals aiming for strategic and leadership-oriented roles are highly motivated to develop advanced skills, including expertise in automation frameworks, quality strategy, and AI integration.

4 CONCLUSION

QA professionals preparing to transition into advanced roles such as AI/ML Quality Specialist, Test Data Engineer, and Quality Architect are more willing to engage in reskilling activities. Among these roles, the Quality Architect and AI/ML Quality Specialist positions showed the strongest and most statistically significant association with reskilling willingness. Thus, it can be concluded that when QA engineers have career transition intentions, it is crucial to be motivated in developing advanced engineering competencies and AI and analytical competencies. Furthermore, suggest that the high level of AI usage among younger QA professionals highlights the need to strengthen AI related competencies to support future career transitions.

5 ACKNOWLEDGEMENT

The authors gratefully acknowledge those involved in data collection and all participants for their valuable contribution to this study.

REFERENCES

- Almalki, M., & Javaid, N. (2022). Machine learning techniques for software testing: A systematic literature review. IEEE Access, 10, 71834–71861.
- Amalfitano, D., Faralli, S., Hauck, J. C. R., Matalonga, S., & Distante, D. (2023). Artificial intelligence applied to software testing: A tertiary study. *ACM Computing Surveys*, *56*(3), 1–38. doi.org
- Baker, A., Habib, M., & Khan, S. (2023). The impact of artificial intelligence on software quality assurance practices. International Journal of Software Engineering and Applications, 1–14.
- Dilshan, A., Wijayanayake, J., Asanka, D., & Karunarathna, C. (2024). Adoption success of using generative AI apps for the e-commerce platforms in Sri Lanka. In *Proceedings of the 3rd International Conference on Sustainable & Digital Business (ICSDB 2024)*. doi.org
- Durrani, U. K., Akpinar, M., Adak, M. F., Kabakus, A. T., Öztürk, M. M., & Saleh, M. (2024). A decade of progress: A systematic literature review on the integration of AI in software engineering phases and activities (2013–2023). *IEEE Access*, *12*, 171185–171204. doi.org
- Garousi, V., Joy, N., & Keleş, A. B. (2024). AI-powered test automation tools: A systematic review and empirical evaluation. *arXiv*. doi.org
- Khan, A., & Khan, R. (2022). Software testing evolution: From manual testing to AI-based testing. *International Journal of Computer Applications*, 7–12.
- Pressman, R. S., & Maxim, B. R. (2020). *Software engineering: A practitioner's approach* (9th ed.). McGraw-Hill Education.
- Sommerville, I. (2016). *Software engineering* (10th ed.). Pearson Education.
- Zhang, Y., Harman, M., & Jia, Y. (2022). Machine learning driven test automation: A survey. *IEEE Transactions on Software Engineering*, *48*(9), 3512–3530.

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

The Impact of Green Supply Chain Management Practices on Organizational Performance: Evidence from Apparel Exporters in Sri Lanka

Maleesha AS[1], Jothirathne PAAU[2]

[1,2]*Department of Industrial Management, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*

sahanmaleesha4@gmail.com[1], anoja@wyb.ac.lk[2]

ABSTRACT

The apparel industry faces increasing pressure to improve environmental sustainability while maintaining operational and financial performance. This study examines the effects of Internal Environmental Management, Green Manufacturing, Green Purchasing and Cooperation with Customers on Environmental, Operational and Perceived Financial Performance among apparel exporters in Sri Lanka. A quantitative cross-sectional survey was conducted using a structured questionnaire. After screening, 98 responses representing 98 different firms were analyzed using descriptive statistics, correlation and multiple regression. GSCM adoption was moderate. Internal Environmental Management significantly predicted Environmental Performance ($\beta = 0.489$, $p < 0.001$; $R^2 = 0.615$). Internal Environmental Management, Green Manufacturing and Green Purchasing significantly predicted Operational Performance ($R^2 = 0.746$), while Green Purchasing, Green Manufacturing and Internal Environmental Management significantly predicted Perceived Financial Performance ($R^2 = 0.764$). Cooperation with Customers showed a significant effect only on Environmental Performance. The findings indicate that GSCM practices contribute differently across performance dimensions and should therefore be prioritized according to organizational objectives.

KEYWORDS: Environmental performance, Green purchasing, Green supply chain management, Operational performance, Organizational performance

1 INTRODUCTION

The apparel industry faces significant environmental challenges from resource consumption, waste and pollution, and Green Supply Chain Management (GSCM) has emerged as an approach for integrating environmental considerations into supply chain activities (Green et al., 2012; Fang & Zhang, 2018). Sri Lanka's export-oriented apparel industry operates in global supply chains with rising sustainability expectations. Previous studies have examined GSCM and sustainability performance in Sri Lanka, but little evidence shows how individual GSCM practices affect different dimensions of organizational performance among apparel exporters (Herath & Gunawardana, 2024; Susitha & Nanayakkara, 2023). This study therefore examines the effects of Internal Environmental Management (IEM), Green Manufacturing (GM), Green Purchasing (GP) and Cooperation with Customers (CC) on three dimensions of organizational performance: Environmental (EP), Operational (OP) and Perceived Financial Performance (FP), and identifies the level of GSCM adoption. The findings aim to help managers and policymakers prioritize green investments under cost and buyer-compliance pressures.

2 LITERATURE REVIEW

GSCM integrates environmental considerations into core supply chain activities with the aim of reducing waste, emissions and resource consumption while supporting competitive and

operational goals (Feng et al., 2024; Al-Sheyadi et al., 2019). Research consistently shows that GSCM improves environmental performance and can generate operational and financial benefits though the strength of these effects varies across performance dimensions and industry contexts (Green et al., 2012; Fang & Zhang, 2018). This is especially relevant in apparel and textile supply chains, where export-oriented production faces high environmental burdens, strong buyer scrutiny and implementation barriers related to cost, technology, and supplier capability (Susitha & Nanayakkara, 2023).

Literature distinguishes internal GSCM practices, including environmental management systems and top management commitment from external practices that extend environmental requirements to suppliers and customers through purchasing criteria and collaboration (Fang & Zhang, 2018; Al-Sheyadi et al., 2019). These practices are related but not interchangeable and their performance effects differ by outcome and context. A key theme in recent research is that GSCM should not be treated as a single undifferentiated construct, as individual dimensions contribute differently depending on whether the outcome is environmental, operational or financial (Çankaya & Sezen, 2019; Sahoo & Vijayvargy, 2021).

This study focuses on four practices. IEM often functions as an enabling capability by supporting the execution of other green supply chain practices (Khan & Yu, 2020; Sahoo & Vijayvargy, 2020). GM involves cleaner production and waste minimization, consistently linked to improved efficiency and resource (Kalyar et al., 2020). GP extends environmental criteria to supplier selection, though its effects are less consistent, depending on supplier capability and implementation depth (Çankaya & Sezen, 2019). CC involves buyer collaboration on sustainability requirements while it supports market access, its direct performance effects remain mixed (Laari et al., 2016; Ahmad et al., 2022).

3 METHODOLOGY

3.1 Research Design

Internal Environmental Management (IEM) represents the organizational commitment, formal environmental systems, top management support and cross-functional coordination required to implement environmental initiatives.

As an enabling capability, IEM strengthens environmental control (H1a) and supports the implementation of other GSCM practices, thereby improving operational (H2a) and perceived financial (H3a) performance (Sahoo & Vijayvargy, 2021; Al-Sheyadi et al., 2019).

Green Manufacturing (GM) involves cleaner production, waste minimization, resource efficiency and process improvement. These practices can reduce

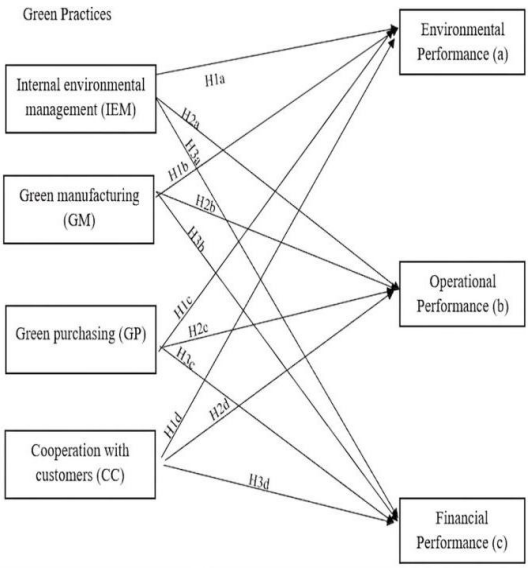


Figure 1: Conceptual framework

material and energy consumption (H1b), improve process efficiency and operational outcomes (H2b), and generate financial benefits through cost reduction and productivity gains (H3b) (Green et al., 2012; Kalyar et al., 2020).

Green Purchasing (GP) builds environmental criteria into supplier selection, evaluation and purchasing decisions. By extending requirements upstream, it can improve supplier-related environmental performance (H1c), material quality and resource efficiency (H2c) and related costs (H3c), although its effects depend on supplier capability and implementation depth (Çankaya & Sezen, 2019; Sahoo & Vijayvargy, 2021).

Cooperation with Customers (CC) involves collaboration with buyers on environmental requirements, information sharing, product specifications and sustainability objectives. It can support environmental improvement (H1d), operational alignment with customer expectations (H2d) and market access (H3d), although its direct effects on performance vary across contexts (Laari et al., 2016; Ahmad et al., 2022).

Each hypothesis states a significant positive impact (H1 = EP, H2 = OP, H3 = FP; a to d = IEM, GM, GP, CC).

A quantitative, cross-sectional survey design was adopted. Figure 1 presents the conceptual model and the twelve hypotheses. For each performance dimension, the following regression model was tested. Performance = $\beta_0 + \beta_1(IEM) + \beta_2(GM) + \beta_3(GP) + \beta_4(CC) + \varepsilon$

3.2 Population and Sample

The target population comprised 155 registered apparel exporting firms identified through the Sri Lanka Export Development Board (EDB) exporter information. Purposive sampling was employed because respondents needed sufficient knowledge of GSCM practices and organizational performance. Questionnaires were distributed online and printed formats. Of 109 responses received, 11 were excluded during screening, leaving 98 valid responses. Each came from a different firm, so the sample comprises 98 firms with one respondent per firm.

3.3 Data Collection and Measurement

Each GSCM practice (IEM, GM, GP and CC) was measured with four items and each performance dimension (EP, OP and FP) with five items, giving 31 items rated on a five-point Likert scale (1 = strongly disagree, 5 = strongly agree). Items were adapted from validated GSCM and performance scales, mainly Sahoo and Vijayvargy (2021), Kalyar et al. (2020) and Çankaya and Sezen (2019), and contextualized to the Sri Lankan apparel export setting. Participation was voluntary and confidential.

3.4 Data Analysis

Data was analyzed using SPSS. Cronbach's alpha assessed internal consistency; skewness and kurtosis assessed approximate normality and Principal Component Analysis assessed construct validity. Pearson correlation examined bivariate relationships and multiple regression assessed unique predictor effects. Multicollinearity was assessed using tolerance and VIF, while residual independence was examined using Durbin-Watson statistics.

4 RESULTS AND DISCUSSION

4.1 Reliability Analysis

Cronbach's alpha values exceed the commonly accepted threshold of 0.70, indicating satisfaction with excellent internal consistency for these constructs.

The reported KMO value was 0.868, while Bartlett's test was statistically significant (χ^2 = 629.994, df = 21, p < 0.001). These results indicate that the analyzed correlation matrix was suitable for factor analysis.

4.2 Descriptive Statistics

Descriptive results indicated a moderate overall level of GSCM adoption (mean = 3.43 - 3.60 across the four practices), with Internal Environmental Management (IEM) the most established practice. Among performance outcomes, Environmental Performance recorded the highest mean (3.75), while Operational (mean=3.39) and Perceived Financial Performance (mean=3.27) were moderate.

4.3 Correlation Analysis

All correlations were positive and significant at p < .001. Associations between GSCM practices and performance ranged from r = 0.601 to 0.799. The strongest relationships were GP–FP (r = 0.799), GM–OP (r = 0.773) and IEM–EP (r = 0.721).

Table 1: Reliability analysis

Construct	Items	Cronbach's α
IEM	4	0.886
GM	4	0.901
GP	4	0.891
CC	4	0.896
EP	5	0.893
OP	5	0.888
FP	5	0.925

Table 2: Descriptive Statistics of Study

Variable	Mean	SD	Interpretation
IEM	3.60	1.11	Moderate
GM	3.45	1.11	Moderate
GP	3.45	1.15	Moderate
CC	3.43	1.19	Moderate
EP	3.75	1.00	High
OP	3.39	1.04	Moderate
FP	3.27	1.23	Moderate

Table 3: Pearson correlation matrix

	IEM	GM	GP	CC	EP	OP	FP
IEM	1.00	0.62	0.57	0.53	0.72	0.71	0.69
GM	0.62	1.00	0.62	0.79	0.64	0.77	0.73
GP	0.57	0.62	1.00	0.75	0.60	0.70	0.79
CC	0.53	0.79	0.75	1.00	0.64	0.74	0.72
EP	0.72	0.64	0.60	0.64	1.00	0.75	0.67
OP	0.71	0.77	0.70	0.74	0.75	1.00	0.87
FP	0.69	0.73	0.79	0.72	0.67	0.87	1.00

Table 4: Summary of Regression Results (Standardized β Coefficients)

Predictor	EP	OP	FP
IEM	0.489***	0.312***	0.240 *
GM	0.090 ns	0.311***	0.293 **
GP	0.071 ns	0.219 *	0.492***
CC	0.255 *	0.166 ns	-0.006 ns
R^2	0.615	0.746	0.764

*Note: * p < .05, ** p < .01, *** p < .001, ns = not significant.*

4.4 Regression Results

The EP model was significant, F(4,93) = 36.845, p < .001, explaining 61.5% of variance. IEM was significant (β = .489, p < .001) and CC was marginally significant (β = .255, p = .049). The OP model explained 74.6% of variance; IEM, GM and GP were significant, whereas CC was not. The FP model explained 76.4% of variance, with GP the strongest predictor, followed by GM and IEM; CC was not significant. These results support H1a, H1d, H2a–H2c and H3a–H3c, while H1b, H1c, H2d and H3d were not supported.

4.5 Discussion

The differential effects support treating GSCM as a set of distinct practices. IEM was the strongest predictor of EP, consistent with evidence that management commitment and internal environmental systems provide the foundation for environmental improvement (Sahoo & Vijayvargy,

2020; Al-Sheyadi et al., 2019). Its significant effects on OP and FP also suggest that internal environmental capabilities can support process and resource management.

GM significantly predicted OP and FP but not EP. This pattern is consistent with research linking cleaner production and resource efficiency to operational and financial benefits (Green et al., 2012; Kalyar et al., 2020). GP showed the strongest financial effect and a significant operational effect, differing from studies reporting weaker direct effects (Çankaya & Sezen, 2019).

In the buyer-driven apparel context, environmental supplier requirements may also support material quality, compliance and supply reliability.

CC was significant only for EP and non-significant for OP and FP. This is consistent with evidence that customer-driven GSCM can support environmental outcomes while direct operational and financial effects vary (Laari et al., 2016; Ahmad et al., 2022). Customer cooperation may therefore be more closely linked to buyer sustainability expectations in this context; this is a possible explanation rather than a directly measured reason

Table 5: Summary of Regression Results

Hypothesis	Relationship	Decision
H1a	IEM → EP	S
H1b	GM → EP	NS
H1c	GP → EP	NS
H1d	CC → EP	S
H2a	IEM → OP	S
H2b	GM → OP	S
H2c	GP → OP	S
H2d	CC → OP	NS
H3a	IEM → FP	S
H3b	GM → FP	S
H3c	GP → FP	S
H3d	CC → FP	NS

Note: S = Supported; NS = Not Supported.

5 CONCLUSION

This study examined how four GSCM practices relate to organizational performance among 98 apparel exporting firms in Sri Lanka. The findings show that GSCM is not a one-size-fits-all strategy. Internal Environmental Management was the strongest predictor of Environmental Performance, while Operational Performance was predicted by Internal Environmental Management, Green Manufacturing and Green Purchasing. Green Purchasing was the strongest predictor of Perceived Financial Performance ($\beta = 0.492$, $p < 0.001$), which contrasts with studies reporting weaker links between purchasing and performance (Çankaya & Sezen, 2019). Because the study is cross-sectional and financial performance is perceptual, this indicates a strong positive association, not that green purchasing causes financial gains. Cooperation with Customers was only weakly associated with performance, consistent with evidence that customer-driven sustainability is often compliance-oriented (Fang & Zhang, 2018).

Theoretically, the results support treating GSCM as a set of distinct practices with different performance associations rather than a single capability. Managerially, they support a targeted, practice-specific approach in this export-driven setting (Susitha & Nanayakkara, 2023): firms seeking better perceived financial outcomes may consider prioritizing green purchasing and internal environmental management over customer cooperation. Limitations include the cross-sectional design, purposive sampling, a single self-reported

respondent per firm (with possible common-method bias), perceptual financial measures, a modest sample of 98 firms, and a focus on Sri Lankan apparel exporters, which limit causal inference and generalizability. Future research could use probability-based sampling, multiple respondents per firm, longitudinal designs, objective performance data and cross-industry samples.

REFERENCES

- Ahmad, A., Ikram, A., Rehan, M. F., & Ahmad, A. (2022). Going green: Impact of green supply chain management practices on sustainability performance. Frontiers in Psychology, 13, 973676.
- Al-Sheyadi, A., Muyldermans, L., & Kauppi, K. (2019). The complementarity of green supply chain management practices and the impact on environmental performance. Journal of Environmental Management, 242, 186–198.
- Çankaya, S., & Sezen, B. (2019). Effects of green supply chain management practices on sustainability performance. Journal of Manufacturing Technology Management, 30(1), 98–121.
- Fang, C., & Zhang, J. (2018). Performance of green supply chain management: A systematic review and meta-analysis. Journal of Cleaner Production, 183, 1064–1081.
- Feng, T., Qamruzzaman, M., Sharmin, S. S., & Karim, S. (2024). Bridging environmental sustainability and organizational performance: The role of green supply chain management in the manufacturing industry. Sustainability, 16(14), 5918.
- Green, K. W., Zelbst, P. J., Meacham, J., & Bhadauria, V. S. (2012). Green supply chain management practices: Impact on performance. Supply Chain Management: An International Journal, 17(3), 290–305.
- Herath, H. M. R. P., & Gunawardana, K. (2024). Impact of green supply chain management practices on environmental performance of the apparel industry in Sri Lanka. CINEC Academic Journal, 6(2), 121.
- Kalyar, M. N., Shoukat, A., & Shafique, I. (2020). Enhancing firms' environmental performance and financial performance through green supply chain management practices and institutional pressures. Sustainability Accounting, Management and Policy Journal, 11(2), 451–476.
- Laari, S., Töyli, J., Solakivi, T., & Ojala, L. (2016). Firm performance and customer-driven green supply chain management. Journal of Cleaner Production, 112(3), 1960–1970.
- Sahoo, S., & Vijayvargy, L. (2021). Green supply chain management practices and its impact on organizational performance: Evidence from Indian manufacturers. Journal of Manufacturing Technology Management, 32(4), 862–886.
- Susitha, E., & Nanayakkara, M. (2023). Impact of green supply chain management practices on the triple bottom line: A study on apparel manufacturers of Sri Lanka. Journal of Asia Business Studies, 17(6), 1228–1249.

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Are Supplier Performance and Transportation Efficiency Associated with Logistics Costs? An Empirical Analysis Based on Asian Countries

Hansana JAP[1] , Pallegedara A[2]
[1,2] *Department of Industrial Management, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*
pawanhansana13@gmail.com [1], asankap@wyb.ac.lk[2]

ABSTRACT

This study examines whether supplier performance and transportation efficiency indicators are associated with logistics costs across 25 Asian economies. Four World Bank Logistics Performance Index (LPI) waves (2007, 2010, 2012 and 2014) were matched with World Bank Doing Business cost and time data, the period with a single comparable cost methodology (DB06–15). Eight hypotheses on average logistics cost per container (US$) were examined with panel regression, composite indices, a Random Forest and K-Means clustering. The hypothesized signs were supported in the preferred individual specifications, although significance varied across estimators. Time to export (BCT) and time to import (DCT) carried most of the Random Forest's predictive importance (86.7%), which is not an effect size, and cost and time come from the same Doing Business indicators, so part of their association may be mechanical. When the composite indices were entered together, the supplier performance index lost significance, consistent with shared variance. Clustering produced three groups, including a high-cost landlocked group whose profile may reflect geography and income as well as export and import procedures. The panel is short and unbalanced (N = 94–97), the design is observational and the Random Forest was validated on one small test set, so the findings are associations and the policy implications are tentative.

KEYWORDS: Logistics Cost, Logistics Performance Index, Panel Regression, Supplier Performance, Transportation Efficiency

1 INTRODUCTION

Supply chain efficiency is important for trade competitiveness, while high logistics costs remain a challenge for many developing Asian economies (Arvis et al., 2018). The World Bank's Logistics Performance Index (LPI) rates six dimensions of logistics capability, but most studies use LPI scores or trade outcomes rather than direct cost measures, leaving a measurement gap (perception scores rather than cost), a methodological gap (panel estimation rarely combined with machine learning) and a regional gap (limited Asian evidence).

This study uses container cost as the outcome and applies panel regression, composite indices, Random Forest and K-Means to 25 Asian economies to examine how supplier performance and transportation efficiency are associated with logistics costs and to identify economies with similar profiles. The design is observational, so results are associations, not causal effects.

2 LITERATURE REVIEW

Logistics performance is the efficiency of a country's logistics system, measured by the perception-based LPI (Arvis et al., 2018); logistics cost covers moving, storing and managing goods and is hard to compare across countries (Rantasila & Ojala, 2012). Here, supplier performance and transportation efficiency are the authors' operational constructs, built from national LPI and Doing Business indicators (Section 3); supplier

performance is a national, perception-based proxy for supplier service quality, not a firm-level measure. Koh et al. (2018) relate corruption to LPI in 26 Asian countries over 2007–2014, without measuring cost.

Evidence on cost is indirect: Arvis et al. (2018) link better logistics performance to lower trade-related costs; Banomyong et al. (2023) report a moderate negative relationship between logistics cost-to-GDP and LPI dimensions in ASEAN; Sénquiz-Díaz (2021) and Wang et al. (2024) link road and port quality and logistics competence to exports and trade volume. None examines container cost, so they motivate rather than test the hypotheses, which rest on transaction-cost reasoning: friction that adds delay, uncertainty or handling raises the cost of moving a container. Higher LPI scores should lower cost (H1–H6: efficient clearance cuts fees and waiting; competent providers and visibility reduce handling losses; reliable delivery reduces buffer stocks; better infrastructure and shipment arrangement lower unit cost), and longer export and import times (BCT, H7; DCT, H8) should raise it through storage and handling charges.

3 METHODOLOGY

3.1 Variables and Model Specification

The panel covers 25 Asian economies chosen for consistent LPI and Doing Business data availability (Japan and South Korea excluded, Singapore kept as a benchmark; listed in Section 4.3) over four LPI waves (2007, 2010, 2012, 2014), merged by country and same-year observation. LPI scores come from the World Bank LPI database (World Bank, 2026) and cost and time from the Doing Business historical data (World Bank, 2020). Of 100 observations, 97 have cost data (Myanmar lacks three) and 94 also have complete LPI scores (Bangladesh 2012, Iran 2014 and Maldives 2007 missing). Observations were deleted listwise: Models 1–6 and the Random Forest use N = 94, and the BCT and DCT models N = 97.

The dependent variable is log(AVG_COST), where AVG_COST is the mean of the Doing Business export and import cost of a standard container (US$). The log reflects right-skewed cost, and percentage changes are $(e^{\beta} - 1) \times 100$ per one-point LPI rise or per additional day of BCT (time to export) or DCT (time to import). Supplier performance is measured by customs efficiency (CUS), logistics competence (LCS), tracking and tracing (TTS) and timeliness (TIM); transportation efficiency by infrastructure (INF), international shipments (ISS), BCT and DCT.

SPI = (zCUS + zLCS + zTTS + zTIM)/4 and TEFI = (zINF + zISS − zBCT − zDCT)/4, with z a z-score and equal weights; BCT and DCT enter with reversed sign, so higher TEFI means lower friction. Z-scores were computed in Stata over each variable's available observations; the composite-index Random Forest standardised the 94 complete observations.

3.2 Sample, Data and Analytical Approach

Pooled OLS, fixed-effects (FE) and random-effects (RE) models were estimated in Stata, with the Hausman (1978) test guiding the choice, as is standard in LPI panel research (Jayathilaka et al., 2022) (classical p-

values 0.002–0.252; RE is preferred except in Models 3, 4 and 8, and Models 1, 6 and 7 are borderline at 0.080–0.089). Standard errors are clustered by economy (Cameron & Miller, 2015; 25 clusters, so inference is approximate); no year effects were included. As LPI items are highly correlated, each enters a separate model controlling for BCT and DCT, which are also estimated alone.

The Random Forest (Breiman, 2001) used the eight indicators and, separately, SPI and TEFI, with one random split of the 94 observations (75 training, 19 test), 500 trees, maximum depth 5, minimum leaf size 3, seed 42, and no cross-validation or tuning. Importance is impurity-based (refit on all 94 observations), which is biased when predictors are correlated (Strobl et al., 2007). K-Means used the eight indicators and log(AVG_COST), standardised, as country means over available waves; k = 3 was kept for interpretability although the silhouette (Rousseeuw, 1987) is higher for k = 2 (0.464) than k = 3 (0.389), and cost is an input, so the "high-cost" label is partly by construction.

4 RESULTS AND DISCUSSION

4.1 Descriptive Statistics and Correlation

Average logistics cost (N = 97) ranges from USD 401 to USD 15,433 per container (mean USD 2,547); BCT and DCT range from 6–89 and 4–117 days and are strongly correlated (r = 0.93). LPI items correlate negatively with cost (r = −0.66 to −0.72, p < 0.001), BCT and DCT positively (r = 0.89, 0.88), and LPI items are highly intercorrelated (r = 0.79–0.95; mean VIF = 11.2, up to 15.7). Cost, BCT and DCT describe the same procedures, so part of their association is mechanical.

4.2 Panel Regression Results

Every hypothesized sign was obtained under each preferred estimator (Table 1), so H1–H8 are supported in these individual specifications, but support is weaker across estimators: in fixed-effects Models 1–6, DCT is never significant (p = 0.56–0.78) and is negatively signed, and BCT is significant at 5% only in Model 4. Most explained variation lies between economies, so these are cross-sectional associations, not within-economy effects. Percentage changes come from separate models and estimators with omitted correlated items, so they are not ranked, and a one-point LPI change is not comparable with a per-day effect.

4.3 Composite Index and Machine Learning Evidence

SPI and TEFI each predict cost alone, although SPI is significant only at 10% under fixed effects. Together, SPI loses significance and changes sign (RE +0.107, FE +0.004), while TEFI remains

Table 1: Hypotheses and Preferred Estimation Results

Hyp.	Est.	β	SE	Effect
H1 CUS	RE	−0.240***	(0.072)	−21.3%
H2 LCS	RE	−0.295***	(0.074)	−25.5%
H3 TTS	FE	−0.205***	(0.072)	−18.6%
H4 TIM	FE	−0.143**	(0.062)	−13.3%
H5 INF	RE	−0.271***	(0.082)	−23.8%
H6 ISS	RE	−0.167***	(0.064)	−15.4%
H7 BCT	RE	0.032***	(0.005)	+3.3%
H8 DCT	FE	0.016***	(0.005)	+1.6%

*Note: RE/FE = preferred estimator; clustered SEs; N = 94 (H1–H6), 97 (H7–H8); R^2 0.79–0.85 (RE overall), 0.20–0.31 (FE within); ***p < 0.01, **p < 0.05; effect = (e^β − 1) × 100 per LPI point (BCT, DCT: per day).*

significant (RE −0.800, $p < 0.01$; FE −0.396), and the Hausman test favours FE for this model; the indices are strongly correlated ($r = 0.88$), so this reflects shared variance and possible statistical suppression, SPI's positive sign should not be read as a cost-raising effect, and TEFI's prominence is partly mechanical because it contains BCT and DCT.

The eight-indicator Random Forest reached a test R^2 of 0.846 (training 0.942), but on one split with 19 test observations, no cross-validation, and the same economy in both sets, which inflates R^2; it shows predictive fit only. BCT (43.4%) and DCT (43.3%) hold 86.7% of impurity-based importance, followed by ISS (3.4%), INF (3.0%) and the four supplier items (1.2–2.3% each); with the composites, TEFI holds 92.2% and SPI 7.8%. Importance is not an effect size or causal evidence.

K-Means produced three clusters (Fig. 1): 12 medium-performance economies (Bangladesh, Cambodia, India, Indonesia, Iran, Jordan, Maldives, Myanmar, Pakistan, Philippines, Sri Lanka, Vietnam), four high-performance economies (China, Malaysia, Singapore, Thailand) and nine low-performance, high-cost landlocked economies (Afghanistan, Bhutan, Kazakhstan, Kyrgyz Republic, Lao PDR, Mongolia, Nepal, Tajikistan, Uzbekistan), whose BCT averages 58.1 days against 14.0 in the high-performance cluster. A silhouette of 0.389 shows moderate, not clear, separation.

4.4 Discussion

Because the low-performance cluster coincides exactly with landlocked status, export and import procedures cannot be separated from geography, income, trade structure or dependence on neighbours' transit infrastructure; this is consistent with United Nations and World Bank (2022) evidence on landlocked economies but does not show that export and import procedures cause the cost differences.

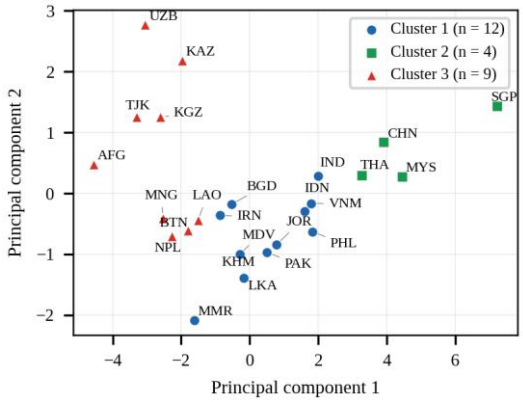


Figure 1: K-Means country clusters (first two principal components; ISO country codes). Source: Author's analysis.

5 CONCLUSION AND RECOMMENDATIONS

In the preferred specifications for 2007–2014, supplier performance indicators are associated with lower logistics cost and export and import times with higher cost, although not uniformly across estimators. Export and import times carry most of the Random Forest's predictive weight, while supplier performance loses significance when modelled jointly with transportation efficiency. Consistent with, but not demonstrated by, these associations, measures that shorten export and import times, such as customs modernisation, electronic documentation and coordinated border management, merit consideration, particularly in landlocked economies.

Limitations: a short, unbalanced panel (N = 94–97); omitted variables and no year effects; city-level proxies for five economies; non-annual LPI waves

matched to annual Doing Business data; overlapping cost and time measures; correlated predictors; and a small machine-learning sample without cross-validation. Future work needs longer panels, more controls, validated machine learning and mediation analysis.

REFERENCES

- Arvis, J.-F., Ojala, L., Wiederer, C., Shepherd, B., Raj, A., Dairabayeva, K., and Kiiski, T. (2018). *Connecting to compete 2018: Trade logistics in the global economy*. World Bank.
- Banomyong, R., Varadejsatitwong, P., and Julagasigorn, P. (2023). Exploring the relationship between national logistics cost and the logistics performance index: An ASEAN perspective.
- Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.
- Cameron, A. C., and Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources, 50*(2), 317–372.
- Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica, 46*(6), 1251–1271.
- Jayathilaka, R., Jayawardhana, C., Embogama, N., Jayasooriya, S., Karunarathna, N., Gamage, T., and Kuruppu, N. (2022). Gross domestic product and logistics performance index drive the world trade: A study based on all continents. *PLOS ONE, 17*(3), e0264474.
- Koh, B. H., Wong, W. P., Tang, C. F., and Lim, M. K. (2018). The double-edge resource-based view of logistics performance and governance in Asian countries. *Asia Pacific Journal of Marketing and Logistics, 30*(3).
- Rantasila, K., and Ojala, L. (2012). *Measurement of national-level logistics costs and performance* (ITF Discussion Paper No. 2012/04). OECD Publishing.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53–65.
- Sénquiz-Díaz, C. (2021). Transport infrastructure quality and logistics performance in exports. *Economics, 9*(1), 107–124.
- Strobl, C., Boulesteix, A.-L., Zeileis, A., and Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics, 8*, Article 25.
- Wang, W., Wu, Q., Su, J., and Li, B. (2024). The impact of international logistics performance on import and export trade: An empirical case of the "Belt and Road" initiative countries. *Humanities and Social Sciences Communications, 11*, Article 1028.
- World Bank. (2020). *Doing Business historical data* [Data set]. World Bank.
- World Bank. (2026). *World Development Indicators: Logistics performance index* [Data set]. World Bank DataBank.

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Impact of Social Media Usage and Online Professional Presence on Employment Outcomes of Sri Lankan Management Graduates

Siriwardana DN[1], Prabodanie RAR[2]
[1,2] *Department of Industrial Management, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*
dasuninaveesha@gmail.com[1], ranga@wyb.ac.lk[2]

ABSTRACT

The increasing use of social media in graduate recruitment has changed how graduates search for employment and present themselves to employers. This study examined the impact of Social Media Usage (SMU) and Online Professional Presence (OPP) on Employment Status and Self-Perceived Employability (SPE) among management graduates in Sri Lanka. A quantitative, deductive, cross-sectional design was adopted, with data collected from 520 respondents using a structured questionnaire and convenience sampling. SMU and OPP were independent variables, while Employment Status (objective) and SPE (subjective) were dependent variables. Data were analysed using descriptive statistics, correlation analysis, binary logistic regression, and multiple linear regression in Stata. SMU and OPP were positively and significantly associated with both outcomes, even after controlling for age, gender, degree, specialization, work experience, hometown, and university. The logistic model was significant (McFadden $R^2 = 0.368$), and the linear model explained 61.7% of the variance in SPE, with OPP ($B = 0.844$, $p < .001$) showing a stronger association than SMU ($B = 0.547$, $p < .001$). Graduates who use social media strategically for professional networking and maintain stronger online professional profiles were more likely to be employed and more confident about their employability. The study contributes empirical evidence from the Sri Lankan management-graduate context, where research on digital employability remains limited.

KEYWORDS: Career Readiness, Digital Employability, Graduate Recruitment, Online Professional Presence, Social Media Usage, Sri Lanka

1 INTRODUCTION

Social networking platforms have become important to graduate job searches and career management, giving graduates access to professional contacts, employer visibility, and channels to present their qualifications beyond a conventional curriculum vitae. Nikolaou (2014) identified social networking websites as relevant recruitment channels, Utz (2016) and Davis et al. (2020) highlighted career benefits of LinkedIn networking, and Laukkarinen (2023) linked strategic social media behaviour to job attainment. Social Capital and Weak-Tie Theories explain how networks provide employment information unavailable through formal channels (Granovetter, 1973; Lin, 2001), while Signalling and Impression Management Theories explain how online profiles are deliberately constructed to convey competence and professionalism (Spence, 1973; Goffman, 1959).

This study therefore examines whether purposeful professional use of social media and a strong online professional presence are associated with employment outcomes among management graduates in Sri Lanka. Objective Employment Status (employed or unemployed) is distinguished from subjective Self-Perceived Employability (SPE), and "employment outcomes"

refers to both. The primary objective is to examine the relationship of Social Media Usage (SMU) and Online Professional Presence (OPP) with both outcomes; a secondary objective is to determine whether these relationships remain significant after controlling for demographic and background characteristics.

2 LITERATURE REVIEW

Social Capital Theory holds that social networks provide access to job information and referrals (Lin, 2001), while Weak-Tie Theory argues that weak ties give broader access to novel opportunities (Granovetter, 1973). Following Nikolaou (2014), SMU is treated as a multidimensional construct capturing engagement with social networking websites, job-search usage, time spent, and frequency of visits (Davis et al., 2020; Utz, 2016).

Online Professional Presence (OPP) reflects the quality and professionalism of a graduate's digital profile, explained by Signalling Theory (Spence, 1973) and Impression Management Theory (Goffman, 1959), and is conceptualised through profile completeness, network size, and active presence (Davis et al., 2020; Utz & Breuer, 2019; Laukkarinen, 2023).

Evidence remains limited for Sri Lankan management graduates, and objective employment status and subjective employability are rarely examined together. Using the above theories as the base framework, this study tests two hypotheses. H1: SMU is positively associated with Employment Status and SPE among management graduates. H2: OPP is positively associated with Employment Status and SPE among management graduates.

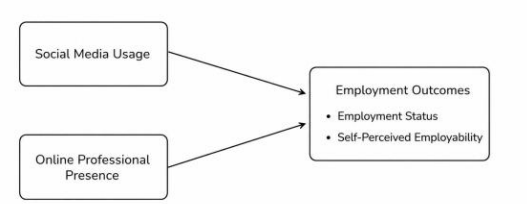


Figure 1: Conceptual framework

Figure 1 describes the conceptual framework used for the analysis.

3 METHODOLOGY

A quantitative, deductive, cross-sectional design was adopted. The target population was recent management graduates of Sri Lankan universities (hereafter, management graduates). Data were collected through a structured questionnaire from 520 respondents using convenience sampling, distributed through university networks and alumni associations; the sample is therefore not statistically representative of the target population. It exceeds the minimum of 384 respondents suggested by Krejcie and Morgan (1970) at a 95% confidence level, used only as an adequacy benchmark and not for random-sampling inference.

Social Media Usage (SMU) was measured across eight items (Q1–Q8) capturing engagement, job-search use, time spent, and frequency of visits on LinkedIn and Facebook, averaged into a composite score. The platforms were combined because SMU is conceptualised as overall purposeful networking engagement rather than platform-specific behaviour: Nikolaou (2014) examined both within one job-search framework, and Social Capital and Weak-Tie Theories treat both as channels to job information. As the platforms serve different career functions (Utz, 2016), the composite does not separate platform-specific effects. Online Professional Presence (OPP) was measured across 16 items (Q9–Q24) capturing profile

completeness, network size, and active presence, averaged into a composite score. All SMU, OPP, and SPE items used five-point Likert-type scales. Employment Status (Q25) was coded 1 = employed and 0 = unemployed for a binary logistic regression. Self-Perceived Employability (SPE) was formed from four items (Q29–Q32) and used as the outcome in a multiple linear regression. Age, gender, degree, specialization, work experience, hometown, and university were control variables in both models, and internal consistency was assessed using Cronbach's alpha.

4 DATA COLLECTION AND ANALYSIS

The dataset contained 520 valid observations with no missing values or duplicate respondent identifiers. Employment status was almost evenly split: 273 respondents (52.5%) were employed and 247 (47.5%) unemployed. Mean age was 25.03 years (SD = 2.44) and mean work experience was 2.08 years (SD = 1.86). The sample comprised 276 male (53.1%) and 244 female (46.9%) respondents, mainly from BBA, MBA and MSc Management programmes.

4.1 Descriptive Statistics and Reliability

Table 2 summarizes the descriptive statistics and reliability of the principal constructs.

The principal constructs showed moderate levels. Reliability was acceptable to good for all three composite scales (α = 0.711 for SMU, 0.778 for OPP, 0.806 for SPE), supporting the use of all items without data-driven item removal (Taber, 2018; Hair et al., 2019).

Table 2: Descriptive statistics and reliability (Source: Author compiled)

Variable	Mean	SD	α
SMU	2.997	0.625	0.711
OPP	2.980	0.536	0.778
SPE	3.026	0.892	0.806
Job Offers	1.410	1.159	—
Interviews	2.208	1.502	—

4.2 Correlation Analysis

Table 3 presents the correlation matrix. SMU and OPP were positively correlated with both Employment Status and SPE, and moderately with each other (r = 0.330). These bivariate relationships provide initial support for H1 and H2, which the regression models below test with controls.

5 RESULT AND DISCUSSION

5.1 Regression Results

The binary logistic regression model predicting Employment Status was statistically significant, LR $\chi^2(32)$ = 264.82, $p < .001$, McFadden pseudo-R^2 = 0.368, with both SMU ($p < .001$) and OPP ($p < .001$) as significant positive predictors. Work experience (p = 0.194) and age (p = 0.295) were not significant, nor was any individual category of gender, degree, specialization, hometown, or university.

Table 3: Correlation matrix of principal variables (Source: Author compiled)

	SMU	OPP	SPE	Emp.
SMU	1.00	0.330	0.570	0.500
OPP		1.00	0.631	0.496
SPE			1.00	0.485
Emp.				1.00

Table 4: Comparison of the Employment Status and SPE models (Source: Author compilation)

Feature	Logistic (Emp.)	Linear (SPE)
SMU	+, $p < .001$	$B = 0.547$, $p < .001$
OPP	+, $p < .001$	$B = 0.844$, $p < .001$
Fit	McFadden $R^2 = 0.368$	$R^2 = 0.617$

The multiple linear regression predicting SPE was also significant, $F(32, 487) = 24.57$, $p < .001$, $R^2 = 0.617$, adjusted $R^2 = 0.592$, with positive and significant (unstandardised) coefficients for SMU ($B = 0.547$, $p < .001$) and OPP ($B = 0.844$, $p < .001$). Work experience was also a significant positive predictor ($B = 0.083$, $p < .001$), while age ($p = 0.276$) and the remaining controls were not. The Breusch–Pagan test showed no heteroskedasticity ($p = 0.805$), supporting the OLS specification. Table 4 summarises the two models.

Both models converge: SMU and OPP are positively and significantly related to employment outcomes, supporting H1 and H2. The two models' R^2 values are not directly comparable, but the convergence across an objective and a subjective outcome strengthens the findings.

OPP showed a stronger association than SMU with both outcomes, particularly SPE, suggesting that the quality and professionalism of a graduate's online representation may matter more than frequency of use alone.

This is consistent with Signalling Theory (Spence, 1973), whereby employers read profile completeness and networking activity as signals of competence, and with Social Capital Theory (Lin, 2001) and Weak-Tie Theory (Granovetter, 1973), whereby connections provide valuable resources and novel employment information. The simultaneous significance of SMU and OPP suggests that networking behaviour and professional self-presentation (Goffman, 1959) are complementary rather than redundant.

Work experience being positively associated with SPE is consistent with Rothwell and Arnold's (2007) view that self-perceived employability reflects confidence in competing in the labour market. Demographic and background controls were generally not significant once SMU and OPP were included, suggesting that digital professional behaviour may matter more than demographic background within this sample.

6 CONCLUSION

This study examined whether Social Media Usage and Online Professional Presence are associated with employment outcomes among 520 management graduates in Sri Lanka. Both were positively and significantly associated with Employment Status and Self-Perceived Employability, even after controlling for demographic characteristics, consistent with the Social Capital, Weak-Tie, Signalling, and Impression Management theories. Limitations include the cross-sectional design and convenience sampling, which limit causal inference and generalisability; possible reverse causation; self-reported measures open to social desirability bias; the focus on LinkedIn and Facebook only; and construct validity assessed only through Cronbach's alpha. Future research should employ probability sampling, longitudinal

designs, confirmatory factor analysis, separate analysis of LinkedIn and Facebook, and moderators such as digital literacy and career self-efficacy. Universities, career development centres, employers, and management graduates should place greater emphasis on digital career readiness, professional networking, and online personal branding.

REFERENCES

- Davis, J., Wolff, H.-G., Forret, M. L., & Sullivan, S. E. (2020). Networking via LinkedIn: An examination of usage and career benefits. *Journal of Vocational Behavior, 118*, Article 103396. https://doi.org/10.1016/j.jvb.2020.103396
- Goffman, E. (1959). *The presentation of self in everyday life*. Doubleday.
- Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology, 78*(6), 1360–1380. https://doi.org/10.1086/225469
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.
- Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. *Educational and Psychological Measurement, 30*(3), 607–610. https://doi.org/10.1177/001316447003000308
- Laukkarinen, M. (2023). Social media as a place to see and be seen: Exploring factors affecting job attainment via social media. *The Information Society, 39*(4), 199–212. https://doi.org/10.1080/01972243.2023.2199418
- Lin, N. (2001). *Social capital: A theory of social structure and action*. Cambridge University Press.
- Nikolaou, I. (2014). Social networking web sites in job search and employee recruitment. *International Journal of Selection and Assessment, 22*(2), 179–189. https://doi.org/10.1111/ijsa.12067
- Rothwell, A., & Arnold, J. (2007). Self-perceived employability: Development and validation of a scale. *Personnel Review, 36*(1), 23–41. https://doi.org/10.1108/00483480710716704
- Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics, 87*(3), 355–374. https://doi.org/10.2307/1882010
- Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2
- Utz, S. (2016). Is LinkedIn making you more successful? The informational benefits derived from public social media. *New Media & Society, 18*(11), 2685–2702. https://doi.org/10.1177/1461444815604143
- Utz, S., & Breuer, J. (2019). The relationship between networking, LinkedIn use, and retrieving informational benefits. *Cyberpsychology, Behavior, and Social Networking, 22*(3), 180–185. https://doi.org/10.1089/cyber.2018.0294

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

A Technology Acceptance-Based Empirical Study of Factors Associated with Citizen Adoption of GovPay and the e-Revenue License System in Sri Lanka

Kaumadi WAI[1], Deegahawatura MMDR[2]
[1,2] *Department of Industrial Management, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*
kaumadiedirisinghe@gmail.com[1], dharsana@wyb.ac.lk[2]

ABSTRACT

Digital government platforms succeed only when citizens actually adopt them, yet the factors driving citizen adoption of Sri Lanka's newly introduced digital payment platforms remain under-researched. Reported usage among those surveyed was comparatively low (mean = 2.42 of 5; Behavioral Intention = 4.05). This study examined the factors influencing citizen adoption of GovPay and the e-Revenue License System in Sri Lanka, applying the Technology Acceptance Model (TAM) extended with trust. The objective was to examine how Perceived Ease of Use, Perceived Usefulness, and Trust influence citizens' Attitude toward these platforms, how Attitude influences Behavioral Intention, and how Behavioral Intention influences Actual Usage. Five hypotheses were formulated and tested. Primary data were collected through a self-administered bilingual (Sinhala and English) online questionnaire distributed via Google Forms, yielding 121 valid responses from Sri Lankan citizens aged eighteen and above. The two platforms were analysed jointly as one context. Data were analysed using Partial Least Squares Structural Equation Modelling. All five hypotheses were supported. Perceived Usefulness was the strongest predictor of Attitude, followed by Perceived Ease of Use, while Trust exerted a significant but comparatively small effect. Attitude strongly predicted Behavioral Intention, and Behavioral Intention significantly predicted Actual Usage, though it explained only a small proportion of its variance ($R^2 = 0.075$). This suggests a gap between stated intentions and reported usage in this sample. Unmeasured factors, such as facilitating conditions and digital literacy, may help explain actual adoption and warrant testing. The study recommends that policymakers emphasise the practical benefits of these platforms, that developers attend to usability and reliability, and that future research incorporate additional variables and longitudinal designs.

KEYWORDS: Behavioral Intention, E-Government Adoption, Perceived Usefulness, Technology Acceptance Model, Trust

1 INTRODUCTION

The rapid advancement of Information and Communication Technology (ICT) has transformed how governments deliver public services, with e-government initiatives increasingly used to improve efficiency, accessibility, and transparency. Sri Lanka has embraced this transformation through platforms such as GovPay, a centralised digital government payment system, and the e-Revenue License System, which enables citizens to renew vehicle revenue licenses online.

However, technology availability alone does not guarantee adoption. The Technology Acceptance Model (TAM) developed by Davis (1989) explains adoption through Perceived Ease of Use (PEOU) and Perceived Usefulness (PU), while Gefen et al. (2003) and Carter and Bélanger (2005) extended TAM with Trust to better explain adoption of online systems involving financial transactions and government institutions.

Although e-government adoption has been widely studied internationally, there is limited empirical evidence on citizens'

adoption of Sri Lanka's newly introduced digital payment platforms, and few studies have simultaneously examined the full chain from Perceived Ease of Use, Perceived Usefulness, and Trust through Attitude and Behavioral Intention to Actual Usage. This study addresses this gap by applying an extended TAM to examine how these factors jointly shape citizens' Attitude, Behavioral Intention, and Actual Usage of GovPay and the e-Revenue License System.

2 LITERATURE REVIEW

E-government refers to the use of digital technologies by government institutions to deliver information, services, and communication facilities to citizens (Carter & Bélanger, 2005).

The Technology Acceptance Model (Davis, 1989) proposes that Perceived Ease of Use and Perceived Usefulness jointly determine users' Attitude toward a technology, which in turn shapes Behavioral Intention and ultimately Actual Usage.

Because digital government services often require citizens to exchange personal and financial information, later research incorporated trust into the TAM framework. Gefen et al. (2003) demonstrated that users are more willing to adopt online systems when they believe the system is secure and reliable, while Carter and Bélanger (2005) found that trust in both the Internet and government institutions significantly shapes citizens' acceptance of e-government services. AbdulKareem and Oladimeji (2024) similarly found that trust and digital literacy significantly contribute to citizens' willingness to adopt digital government services, and Wu et al. (2026), in a systematic review of e-government acceptance studies, identified TAM as one of the most widely applied frameworks, with trust consistently among the most significant predictors. Lamidi et al. (2023) similarly support TAM in Nigerian e-filing.

Despite this body of work, most existing studies were conducted outside Sri Lanka, and few have examined digital payment platforms specifically rather than general e-government information portals. In Sri Lanka, Rathnayake et al. (2025) found that Perceived Usefulness predicted Attitude but Trust did not, in a government AI chatbot study. Moreover, most prior research concludes its analysis at the level of Behavioral Intention rather than testing whether stated intentions translate into Actual Usage (Wu et al., 2026), leaving the intention-behaviour relationship in e-government adoption under-examined. Based on this gap, the present study proposes the following hypotheses:

H1: Perceived Ease of Use has a significant positive effect on citizens' Attitude toward adopting the platforms.

H2: Perceived Usefulness has a significant positive effect on citizens' Attitude toward adopting the platforms.

H3: Trust has a significant positive effect on citizens' Attitude toward adopting the platforms.

H4: Attitude has a significant positive effect on citizens' Behavioral Intention to use the platforms.

H5: Behavioral Intention has a significant positive effect on citizens' Actual Usage of the platforms.

3 METHODOLOGY

This study adopted a quantitative, deductive, cross-sectional research design. Primary data were collected

through a self-administered questionnaire distributed bilingually in Sinhala and English via Google Forms between June and July 2026. The target population comprised Sri Lankan citizens aged 18 years and above who were aware of, or had prior experience using, GovPay and/or the e-Revenue License System. Because items and the experience question referred to both platforms jointly, they were analysed as one pooled context, without platform-level comparison. Given the absence of a complete sampling frame, non-probability convenience and snowball sampling techniques were employed to recruit eligible respondents through online distribution channels. The questionnaire was shared with accessible respondents, and the recruitment network was extended through further sharing among eligible participants. Of the 130 responses collected, nine were excluded (five reported no awareness of either platform and four gave invariant responses), resulting in 121 valid responses for analysis.

The questionnaire measured six constructs on a five-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree): Perceived Ease of Use (5 items), Perceived Usefulness (5 items), Trust (6 items, covering both technology and government dimensions), Attitude (4 items), Behavioral Intention (7 items), and Actual Usage (7 items), adapted from Davis (1989), Gefen et al. (2003), and Carter and Bélanger (2005).

Data were analysed using Partial Least Squares Structural Equation Modelling (PLS-SEM). The measurement model was assessed for reliability, convergent and discriminant validity, and common method bias (Harman's test and full collinearity VIF). The structural model was then estimated using bootstrapping (5,000 subsamples) to test the five hypothesised relationships (with R^2, f^2 and 95% confidence intervals), and bootstrapped indirect effects were used to examine the sequential mediating roles of Attitude and Behavioral Intention.

4 RESULTS AND DISCUSSION

The final sample (n = 121) was broadly balanced by gender (52.1% male, 47.1% female, 0.8% prefer not to say) and concentrated in the 26-35 age group (43.0%). Most resided in the Western Province (53.7%); 85.2% used the internet daily and 99.2% had prior online payment experience. Regarding the platforms, 37.2% were aware but had never used them, 28.1% had used them at least once, 24.8% used them regularly, and 8.3% had tried but faced difficulties (61.2% had used or tried them; percentages are of all 121 respondents, 119 of whom answered this item). Mean scores were high for Attitude (4.26) and Behavioral Intention (4.05), moderate for Trust (3.01), and lowest for Actual Usage (2.42, SD = 1.27).

All constructs showed strong reliability and convergent validity (Table 1). Loadings exceeded 0.70 (Hair et al., 2019) except BI2 (0.618). All Heterotrait-Monotrait ratios were below 0.90 (Henseler et al., 2015). Common method bias appeared unlikely to be severe (Harman's first factor = 41.19%; full collinearity VIFs below 3.3, Kock, 2015, except Perceived Usefulness = 3.802, which warrants caution).

As shown in Table 2, all five hypotheses were supported. Perceived Usefulness emerged as the strongest predictor of Attitude ($\beta = 0.452$, $p < .001$),

Table 1: Measurement Model Reliability and Validity

Construct	α	CR	AVE
PEOU	0.925	0.944	0.769
PU	0.913	0.935	0.742
Trust	0.935	0.949	0.756
Attitude	0.920	0.943	0.807
Behavioral Intention	0.927	0.943	0.705
Actual Usage	0.950	0.959	0.770

Note. CR computed from outer loadings; PEOU = Perceived Ease of Use; PU = Perceived Usefulness.

Table 2: Hypothesis Testing Results (All Supported)

Path	β	SE	t	p	f²
H1 PEOU → ATT	0.301	0.117	2.566	.016	0.075
H2 PU → ATT	0.452	0.107	4.210	<.001	0.170
H3 Trust → ATT	0.121	0.042	2.850	.004	0.031
H4 ATT → BI	0.726	0.080	9.085	<.001	1.117
H5 BI → AU	0.274	0.066	4.135	.001	0.081

Note. ATT = Attitude; BI = Behavioral Intention; AU = Actual Usage. R^2 (adjusted): ATT = 0.584 (0.573); BI = 0.528 (0.524); AU = 0.075 (0.067).

followed by Perceived Ease of Use (β = 0.301, p = .016) and Trust (β = 0.121, p = .004). These predictors explained 58.4% of the variance in Attitude (adjusted R^2 = 0.573). This ordering is consistent with Davis (1989), who found perceived usefulness more strongly linked to technology use than ease of use. The modest effect of Trust diverges from AbdulKareem and Oladimeji (2024), who found trust strongly influenced perceived usefulness and ease of use in Nigerian e-filing.

Attitude strongly predicted Behavioral Intention (β = 0.726, p < .001), explaining 52.8% of its variance and representing the strongest relationship in the model, consistent with Davis (1989) and Wu et al. (2026). However, Behavioral Intention explained only 7.5% of the variance in Actual Usage (β = 0.274, p = .001; adjusted R^2 = 0.067). This weak intention-usage link is a notable finding of this study. This corresponds to the intention-behaviour gap noted by Wu et al. (2026). It appears descriptively (4.05 versus 2.42) but was not tested directly; facilitating conditions or first-use friction are possible but unmeasured explanations. Bootstrapped indirect effects (Table 3) were significant, with all 95% confidence intervals excluding zero, but the sequential effects on Actual Usage were small (0.024 to 0.090). Given the cross-sectional design, they indicate associations, not causal mechanisms.

5 CONCLUSION

This study examined the factors influencing citizen adoption of GovPay and the e-Revenue License System in Sri Lanka using an extended Technology Acceptance Model that incorporates Trust. All five hypotheses were supported: Perceived Usefulness, Ease of Use, and Trust each significantly predicted Attitude, which strongly predicted Behavioral Intention, which only weakly predicted Actual Usage. Attitudes and intentions were highly favourable while reported usage was low, so low usage is unlikely to reflect negative attitudes alone.

For government policymakers, communication strategies could emphasise the practical benefits of

Table 3: Bootstrapped Indirect Effects

Path	Effect	p	95% CI
PEOU → ATT → BI	0.218	.016	[0.035, 0.396]
PU → ATT → BI	0.328	<.001	[0.155, 0.494]
Trust → ATT → BI	0.088	.004	[0.028, 0.154]
ATT → BI → AU	0.199	.001	[0.101, 0.295]
PEOU → … → AU	0.060	.017	[0.008, 0.128]
PU → … → AU	0.090	.001	[0.034, 0.157]
Trust → … → AU	0.024	.004	[0.006, 0.048]

Note. = ATT → BI; 5,000 bootstrap subsamples.

GovPay and the e-Revenue License System, since Perceived Usefulness had the strongest association with Attitude.

This study is limited by its non-probability sampling, modest sample size (n = 121), cross-sectional design, and self-reported usage data. The sample was digitally active, so results cannot be generalised or used to estimate national usage, and the platforms were analysed jointly. Future research should assess each platform separately, incorporate additional variables such as facilitating conditions, digital literacy, and perceived risk, drawing on the Unified Theory of Acceptance and Use of Technology (Venkatesh et al., 2003), and should employ longitudinal designs to examine whether the intention-usage gap identified here narrows over time.

REFERENCES

- AbdulKareem, A. K., & Oladimeji, K. A. (2024). Cultivating the digital citizen: Trust, digital literacy and e-government adoption. *Transforming Government: People, Process and Policy, 18*(2), 270–286. https://doi.org/10.1108/TG-11-2023-0196
- Carter, L., & Bélanger, F. (2005). The utilization of e-government services: Citizen trust, innovation and acceptance factors. *Information Systems Journal, 15*(1), 5–25. https://doi.org/10.1111/j.1365-2575.2005.00183.x
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008
- Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping: An integrated model. *MIS Quarterly, 27*(1), 51–90. https://doi.org/10.2307/30036519
- Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review, 31*(1), 2–24. https://doi.org/10.1108/EBR-11-2018-0203
- Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8
- Kock, N. (2015). Common method bias in PLS-SEM: A full collinearity assessment approach. *International Journal of e-Collaboration, 11*(4), 1–10. https://doi.org/10.4018/ijec.2015100101
- Lamidi, W. A., Olowookere, J. K., Saad, R. A. J., & Ahmi, A. (2023). An assessment of the factors influencing adoption of e-filing system in Nigeria. *International Journal of Business Information Systems, 42*(1), 1–22. https://doi.org/10.1504/IJBIS.2023.128299
- Rathnayake, A. S., Nguyen, T. D. H. N., & Ahn, Y. (2025). Factors influencing AI chatbot adoption in government administration: A case study of Sri Lanka's digital government. *Administrative Sciences, 15*(5), 157. https://doi.org/10.3390/admsci15050157
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425–478. https://doi.org/10.2307/30036540
- Wu, X., Chong, A. Y. L., Peng, Y., & Bao, H. (2026). Predicting the acceptance of e-government: A systematic review. *Internet Research, 36*(1), 100–131. https://doi.org/10.1108/INTR-12-2022-0970

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Association of Gender, Education Level, and Residence Type with Cybercrime Victimization Among Emerging Adult Internet Users in Sri Lanka

Rathnayaka RMKGSB[1], Wattegama EJ[2]
[1,2] *Department of Industrial Management, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*
gayankasun779@gmail.com[1], erandiw@wyb.ac.lk[2]

ABSTRACT

Sri Lanka's rapid expansion of the Internet access has made emerging adults (aged 18 - 29) which is one of the most digitally active and most exposed segments of the population to become as victims for crimes that happening in the cyberspace. The role of basic personal background characteristics in cybercrime victimization within this group remains underexplored locally. This study reports findings from a cross-sectional survey of 311 emerging adult Internet users recruited through convenience sampling across multiple districts in Sri Lanka, examining whether gender, education level, and residence type are associated with cybercrime victimization and it has measured as a binary outcome (victimized versus not victimized). Data were analyzed using Chi-square tests of independence in IBM SPSS Statistics, with Cramér's V as the effect size. The results show a statistically significant association between gender and cybercrime victimization ($\chi^2 = 15.718$, $p < .001$, $V = 0.22$), with 64.2% of male respondents classified as victimized compared to 41.8% of female respondents. Education level and residence type show no statistically significant association with victimization. Within this sample, gender was the only personal background factor associated with higher observed levels of victimization. These exploratory findings should be read as associations rather than causal or population-level effects, and they suggest that awareness and prevention programs in the Sri Lankan context may consider a gender-responsive approach rather than relying solely on education level or residence type.

KEYWORDS: Cybercrime Victimization, Emerging Adults, Internet Users

1 INTRODUCTION

Sri Lanka's rapid expansion of the Internet and smartphone access has made emerging adults, defined as those aged 18 – 29 years, one of the most digitally active group (Arnett, Žukauskienė, & Sugimura, 2014). Digital adoption among this age group has outpaced the country's protective infrastructure and the findings of a research on Sri Lankan undergraduates shows that young Internet users are already exposed to cyber-related harm (for example, Nalaka & Diunugala, 2020). While behavioral correlates of victimization, such as, online engagement, safety behavior, and digital awareness, have received growing research attention, comparatively little is known about how basic personal background characteristics are associated with victimization within the Sri Lankan emerging-adult population specifically.

Although the demographic factors, such as, gender, education level, and residence type have frequently examined by the foreign cybercrime victimization research context, their relevance in the local context remains underexplored, particularly outside university-specific or social-media-specific samples (Nalaka & Diunugala, 2020). Understanding whether, and how strongly, these background characteristics are associated

with victimization has direct implications for how prevention and awareness initiatives are targeted, since interventions that overlook demographic differences may fail to reach the groups most exposed to harm.

The main objective of this research is to systematically examine the association between three personal background factors, namely gender, education level, and residence type, and the cybercrime victimization among emerging adult Internet users in Sri Lanka. Drawing on data from a cross-sectional survey of 311 emerging adult internet users, this research conducts separate Chi-square tests of independence for each of the three factors, reports effect sizes, and further discusses their implications for cybersecurity awareness and prevention practice in Sri Lanka.

2 LITERATURE REVIEW

Demographic characteristics are consistently examined as contextual correlates of cybercrime victimization, even in studies where behavioral or cognitive variables are the primary focus (Outlaw et al., 2002). Gender is among the most frequently reported correlates, though findings vary by the type of cybercrime considered. Sri Lankan evidence on cyberbullying among university students also points to gender differences in reported experiences (Inthusha & Kajananthan, 2023). This suggests that gender differences in victimization may reflect differences in online behavior and platform use rather than a uniform vulnerability of one gender over the other.

Age within the emerging-adulthood range has also been linked to higher victimization: a large Finnish population survey found that people aged 15 – 24 were more likely than older groups to be cybercrime victims, even after adjusting for other factors (Oksanen & Keipi, 2013). Education level presents a more ambiguous picture: although higher educational attainment is often assumed to build digital literacy, some studies find that any protective effect is offset by the greater online exposure and engagement that typically accompanies higher education (van Wilsem, 2013). Within the Sri Lankan setting, education has been examined mainly as a correlate of criminal involvement rather than victimization (Prabhashini, 2025).

Residence type, for example, living independently, in shared accommodation, in a hostel, or with family, is theorized to affect the level of informal supervision and guardianship a young adult experiences, which may in turn influence exposure to risky online situations. However, empirical evidence on residence type as a correlate of cybercrime victimization remains limited in the Sri Lankan context, and existing studies have tended to focus narrowly on university-based, social-media-related victimization rather than cybercrime victimization more broadly (Nalaka & Diunugala, 2020). This gap motivates the present study's examination of gender, education level, and residence type as personal background correlates of cybercrime victimization among a broader, multi-district sample of Sri Lankan emerging adults.

3 METHODOLOGY

In order to achieve the objective specifically, the study has tested the following hypotheses;

H1: There is a significant association between gender and cybercrime victimization.
H2: There is a significant association between education level and cybercrime victimization.
H3: There is a significant association between residence type and cybercrime victimization.

Accordingly, a quantitative, cross-sectional research design was adopted and the data were collected through a structured, self-administered online questionnaire which was distributed to emerging adult Internet users aged 18 – 29 residing across multiple districts of Sri Lanka, using the convenience sampling approach. A total of 318 responses were initially collected; following data screening, seven were excluded due to illogical reason and accordingly 311 usable responses were used for the analysis.

The questionnaire has captured three personal background factors, such as, gender (male / female), education level (school level, diploma, undergraduate, graduate, postgraduate), and residence type (living with family, in a hostel, shared accommodation or living alone), alongside a binary measure of cybercrime victimization (0 = not victimized, 1 = victimized) covering six binary (Yes/No) forms of cybercrime victimization, namely online fraud, phishing, unauthorized access, hacking, cyberbullying, and privacy violations past 12 months]. Respondents reporting two or more "Yes" responses were classified as victimized (1), and the rest as not victimized (0), consistent with the multiple-victimization criterion (Outlaw et al., 2002); the results therefore relate to multiple victimization.

Descriptive statistics were used to profile the sample and Chi-square tests of independence were conducted in IBM SPSS Statistics to examine whether each personal background factor was significantly associated with cybercrime victimization. Statistical significance was evaluated at the $p < .05$ level. Cramér's V was reported as the effect size, expected cell counts were checked, and findings are interpreted as associations only.

4 RESULTS AND DISCUSSION

Out of the 311 respondents, a slight majority were male (53.1%) compared to female (46.9%). The sample was heavily concentrated among undergraduates (76.5%), followed by graduates (13.5%), diploma holders (5.5%), school-level respondents (3.5%), and postgraduates (1.0%). Regarding living arrangements, 45.0% resided in a hostel, 38.6% lived with family, 10.6% lived in shared accommodation, and 5.8% lived alone. Overall, 167 respondents (53.7%) were classified as victimized, while 144 (46.3%) were classified as not victimized. Table 1 summarizes the demographic profile of the sample.

4.1 Gender and Cybercrime Victimization

Table 2 presents the crosstabulation of gender and cybercrime victimization. The Chi-square test revealed a statistically significant association between gender and cybercrime victimization ($\chi^2 = 15.718$, df =1, $p < 0.001$). Among male respondents (n = 165), 64.2% were classified as victimized, compared to 41.8% of female respondents(n = 146).

Table 1: Demographic Profile of Respondents

Variable	Category	Percentage (%)
Gender	Male	53.1
	Female	46.9
Education Level	Undergraduate	76.5
	Graduate	13.5
	Diploma	5.5
	School Level	3.5
	Postgraduate	1.0
Residence Type	Hostel	45.0
	With Family	38.6
	Shared Accommodation	10.6
	Alone	5.8
Cybercrime Victimization	Victims	53.7
	Non-victims	46.3

The effect size was small to moderate (Cramér's V = 0.22; odds ratio = 2.50, 95% CI [1.58, 3.96]), and all expected cell counts exceeded 5. This pattern is partly in line with prior Sri Lankan evidence of gender differences in cyber-related victimization (Inthusha & Kajananthan, 2023), and may reflect differences in the platforms used, risk tolerance, or online behavior between genders, although these mechanisms were not measured here.

4.2 Education Level and Cybercrime Victimization

The Chi-square test for education level was not statistically significant (χ^2 = 0.938, df = 4, p = 0.919, Cramér's V = 0.05), indicating no meaningful association between educational attainment and cybercrime victimization in this sample. Because the postgraduate and school-level groups were very small, two cells (20%) had expected counts below 5, so this result should be interpreted cautiously. This finding aligns with the view that formal education does not automatically confer protective digital literacy (van Wilsem, 2013) and is broadly consistent with prior Sri Lankan research on education and criminal involvement (Prabhashini, 2025).

Table 2: Gender × Cybercrime Victimization Crosstabulation

Gender	Not Victimized (0)	Victimized (1)	Total
Male	59 (35.8%)	106 (64.2%)	165
Female	85 (58.2%)	61 (41.8%)	146
Total	144 (46.3%)	167 (53.7%)	311

Table 3: Summary of Chi-Square Test Results for Personal Background Factors (at 5% Significance Level)

Personal Background Factor	Pearson Chi-Square (df)	p-value	Cramér's V	Result
Gender (H1)	15.718 (1)	<.001	0.22	Significant
Education Level (H2)	0.938 (4)	0.919	0.05	Not Significant
Residence Type (H3)	4.826 (3)	0.185	0.12	Not Significant

4.3 Residence Type and Cybercrime Victimization

The Chi-square test for residence type was also not statistically significant (χ^2 = 4.826, df = 3, p = 0.185, Cramér's V = 0.12), suggesting that living arrangement, whether with family, in a hostel, in shared accommodation, or alone, is not significantly associated with cybercrime victimization in this sample.

Table 3 summarizes the results of three Chi-square tests performed, with effect sizes (Cramér's V).

Taken together, gender was the only personal background factor significantly associated with cybercrime victimization in this sample, while education level and residence type showed no significant association. These results indicate that victimization in this sample is not evenly distributed across genders, and that male

emerging adults in particular may warrant closer attention in prevention efforts, although the cross-sectional, convenience-sample design means these are associations that should not be generalized or read as causal. At the same time, the lack of association for education level and residence type suggests that these two factors are unlikely to be useful criteria for targeting awareness or prevention programs on their own, and that behavioral factors, which were not measured here, may play a larger role than demographic markers such as academic attainment or living arrangement.

5 CONCLUSION

This study has examined the association between three personal background factors, gender, education level, and residence type, and cybercrime victimization among 311 emerging adult Internet users in Sri Lanka. Chi-square tests of independence showed that gender was significantly associated with victimization, with male respondents classified as victimized more often (64.2%) than female respondents (41.8%); the effect size was small to moderate (V = 0.22). Education level and residence type were not significantly associated with victimization.

These findings suggest that cybersecurity awareness and prevention initiatives, particularly those delivered through university orientation programs and counseling services, may consider a gender-responsive approach that acknowledges differing patterns of online exposure between male and female users, rather than assuming a uniform victim profile. Because education level and residence type were not significantly associated with victimization, broad demographic targeting based on academic attainment or living arrangement may be of limited value; given the convenience sample, these implications are tentative.

This study is limited by its cross-sectional design, convenience sampling, small subgroups, bivariate analysis, and self-reported data, which may affect causal inference, generalizability, and accuracy. Future research should use larger probability-based samples, examine specific cybercrime types, apply multivariate analyses (e.g., logistic regression), and explore gender differences in victimization.

REFERENCES

- Arnett, J.J., Žukauskienė, R., & Sugimura, K. (2014). The new life stage of emerging adulthood at ages 18–29 years: Implications for mental health. *The Lancet Psychiatry*, *1*(7), 569 - 576. https://doi.org/10.1016/S2215-0366(14)00080-7
- Inthusha, K., & Kajananthan, R. (2023). Cyberbullying among university students: A study on Sri Lankan universities. *Journal of Business Studies*, *10*(2), 74 - 95.
- Nalaka, S., & Diunugala, H. (2020). Factors associating with social media related crime victimization: Evidence from the undergraduates at a public university in Sri Lanka. *International Journal of Cyber Criminology*, *14*(1), 174 - 184.
- Oksanen, A., & Keipi, T. (2013). Young people as victims of crime on the Internet: A population-based study in Finland. *Vulnerable Children and Youth Studies*, *8*(4), 298 - 309. https://doi.org/10.1080/17450128.2012.752119
- Outlaw, M., Ruback, B., & Britt, C. (2002). Repeat and multiple victimizations: The role of individual and

contextual factors. *Violence and Victims*, *17*(2), 187 - 204.

- Prabhashini, I.G.N. (2025). A study on the relationship between education level and criminality in Sri Lanka (with special reference to prison statistics of Sri Lanka 2019 - 2023). *International Journal of Research and Innovation in Social Science*, *9*(1), 966 - 970. https://dx.doi.org/10.47772/IJRISS.2025.9010079
- van Wilsem, J. (2013). 'Bought it, but never got it': Assessing risk factors for online consumer fraud victimization. *European Sociological Review*, *29*(2), 168 - 178. https://doi.org/10.1093/esr/jcr053

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

The Moderating Role of the Employment Sector on the Relationship Between Culture of Support Policies and Work-Life Balance in the Western Province, Sri Lanka

Mathotage RR[1], Jothirathne PAAU[2]
[1,2]*Department of Industrial Management, Faculty of Applied Sciences, Wayamba University of Sri Lanka*
raveenamathota@gmail.com[1], anoja@wyb.ac.lk[2]

ABSTRACT

Work-life balance has become an important concern as employees increasingly manage work responsibilities alongside family and personal demands. Although organizational support mechanisms can facilitate work-life balance, limited empirical evidence has examined whether the influence of a supportive workplace culture differs between public- and private-sector employees in Sri Lanka. This study addressed this gap by quantitatively examining the influence of Culture of Support Policies on employees' Work-Life Balance and investigating whether this relationship was moderated by employees' employment sector. The target population included employees in public and private-sector organizations in Sri Lanka, with the study focusing on the Western Province due to its concentration of public and private organizations. Data were collected using a structured questionnaire, yielding 602 responses, of which 510 were retained after screening. A pilot study was conducted before the main survey, and the constructs demonstrated satisfactory internal consistency. Data were analyzed using JASP through Structural Equation Modeling and multi-group Structural Equation Modeling. The results indicated that Culture of Support Policies were significantly and positively related to Work-Life Balance. The relationship was significant in both the private sector and the public sector. Although the estimated coefficient was higher in the private sector, the difference between the sector-specific structural paths was not statistically significant. Therefore, the findings did not provide sufficient statistical evidence that Employment Sector moderated the relationship between Culture of Support Policies and Work-Life Balance. The findings highlighted the relevance of supportive workplace practices across both employment sectors.

KEYWORDS: Culture of Support Policies, Private Sector, Public Sector, Structural Equation Modeling, Work-Life Balance

1 INTRODUCTION

Work-Life Balance (WLB) has become an important concern as employees increasingly manage work, family, and personal responsibilities. Organizations have introduced Family-Friendly Policies (FFPs) to support employees in managing these competing demands. Among these, Culture of Support Policies promote a supportive work environment through organizational values, managerial behaviours, and workplace norms that recognize employees' family responsibilities. Previous research indicated that organizational support can reduce work-family interference and enhance WLB (Nassani et al., 2024).

Despite considerable attention to WLB and organizational support, several research gaps remain. First, limited attention has been given to Culture of Support Policies as a distinct dimension of FFPs (Allen, 2001). Second, limited evidence examines whether its relationship with WLB differs between public and private sector employees, particularly in developing countries (Feeney & Stritch, 2019). Third, Sri Lankan studies have largely focused on

factors affecting WLB within individual organizations or specific employee groups, providing limited comparative evidence on sectoral differences (Maduwansha, 2024).

Therefore, this study examines the influence of Culture of Support Policies on employees' WLB and investigates the moderating role of Employment Sector among employees in Sri Lanka's Western Province. By addressing these gaps, the study contributes sector-specific evidence to the limited Sri Lankan literature. The objectives of the study are:

1. To examine the influence of Culture of Support Policies on employees' WLB.
2. To examine the moderating role of Employment Sector in the relationship between Culture of Support Policies and employees' WLB.

2 LITERATURE REVIEW

Work-Life Balance (WLB) refers to an individual's ability to effectively manage work and personal responsibilities while maintaining satisfactory functioning across different life domains (Clark, 2000). A Culture of Support Policies encompasses organizational values, managerial behaviours, and workplace norms that help employees manage family responsibilities alongside work demands (Kossek et al., 2014). Such support can facilitate WLB by creating a workplace environment in which employees perceive greater organizational and managerial support for managing competing responsibilities (Allen, 2001). Empirical evidence generally supports this positive relationship, with studies highlighting the roles of organizational support and supportive workplace cultures in facilitating WLB (Feeney & Stritch, 2019; Fan et al., 2021; Heras et al., 2021). Sri Lankan evidence is also consistent with this pattern, as work-life support culture, organizational support, and family support have been identified as important factors associated with WLB (Jayawardana & Weerasinghe, 2020; Maduwansha, 2024).

However, limited evidence has examined whether the influence of a Culture of Support Policies on WLB.Therefore, the following hypothesis is proposed:

H1: Culture of Support Policies will have a positive impact on employees' Work-Life Balance.

However, the extent to which organizational support influences work-life balance may vary across employment sectors. Public and private sector organizations can differ in their organizational practices, flexibility, managerial support, and implementation of family-supportive policies. Therefore, based on the identified gap , the following hypothesis is proposed:

H2: Work-Life Balance differs significantly between private and public sector employees.

3 METHODOLOGY

A quantitative, cross-sectional survey design was employed to collect primary data from employees in public- and private-sector organizations through a structured questionnaire available in English and Sinhala. The questionnaire measured Culture of Support Policies and WLB using established measurement items adapted from previous studies. A pilot study was conducted before the main survey to assess the clarity and

reliability of the instrument. The target population comprised employees working in public and private sector organizations in Sri Lanka, while the study was geographically limited to the Western Province. Convenience sampling was used to recruit participants based on their accessibility and willingness to participate.

4 DATA COLLECTION & ANALYSIS

A total of 602 responses were obtained. After data cleaning, 510 valid responses were retained for the final analysis. The reported Cronbach's alpha values for the constructs exceeded 0.70, indicating satisfactory internal consistency (Hair et al., 2019). Descriptive statistics and reliability analyses were conducted first, followed by Structural Equation Modeling (SEM) to examine the relationship between Culture of Support Policies and WLB. Model fit was assessed using commonly reported fit indices. Multi-group SEM was used to estimate the structural relationship separately for public and private sector employees.

5 RESULTS & DISCUSSION

The demographic profile of the 510 respondents was examined. The final sample consisted of 266 women (52.2%) and 244 men (47.8%). Of the respondents, 296 (58.0%) were employed in the private sector and 214 (42.0%) in the public sector. The respondents represented several industries, including manufacturing (30.4%), other industries (16.7%), education (12.7%), healthcare (11.6%), and other services (28.6%). Respondents with more than 11 years of work experience constituted the largest group (38.0%), followed by those with 3-5 years 25.9%.

The Structural Equation Modeling analysis based on 510 valid responses demonstrated a good model fit, with $\chi^2 = 44.32$, $df = 13$, $CFI = 0.980$, $TLI = 0.968$, $NFI = 0.972$, $RMSEA = 0.069$, and $SRMR = 0.035$ (Hair et al., 2019). The results indicated that Culture of Support Policies significantly predicted WLB ($\beta = 0.511$, $z = 10.80$, $p < 0.001$, 95% CI). The AVE values of 0.557 for Culture of Support Policies and 0.708 for WLB indicated satisfactory convergent validity. Therefore, H1 was supported. The measurement and structural results are presented in Figure 1.

A multi-group SEM was conducted to examine whether the structural relationship between Culture of Support Policies and WLB differed between private and public-sector employees. The unconstrained model, in which the structural regression path was freely estimated across the two sectors, demonstrated an acceptable model fit ($\chi^2 = 87.45$, $df = 49$, $CFI = 0.989$, $TLI = 0.990$, $RMSEA = 0.056$, $SRMR = 0.059$). Culture of Support Policies significantly predicted WLB in both the private sector ($\beta = 0.553$, $z = 7.423$, $p < 0.001$, 95% CI and the public sector ($\beta = 0.468$, $z = 5.027$, $p < 0.001$, 95% CI. The model explained 23.43% of the variance in WLB among private-sector employees and 17.98% among public-sector employees.

To formally assess whether the structural relationship differed between the two sectors, the unconstrained model was compared with a constrained model in which the Culture of Support Policies and WLB regression coefficient was

constrained to be equal across sectors. The constrained model yielded $\chi^2 = 90.70$ with 50 degrees of freedom, compared with $\chi^2 = 87.45$ with 49 degrees of freedom for the unconstrained model. Thus, the models differed by $\Delta\chi^2 = 3.25$ and $\Delta df = 1$. Based on the standard chi-square difference calculation, this difference was not statistically significant ($p \approx 0.071$). Accordingly, although the estimated structural coefficient was higher in the private sector than in the public sector, the difference between the sector-specific path coefficients did not reach the 5% significance level. Therefore, H2 was not supported as a statistically significant moderation effect. The findings indicated that Culture of Support Policies was positively associated with WLB, consistent with previous research highlighting the importance of organizational and supervisory support (Wadsworth & Owens, 2007). Although the estimated association was stronger among private-sector employees ($\beta = 0.553$) than public-sector employees ($\beta = 0.468$), the difference between the sector-specific paths was not statistically significant ($\Delta\chi^2 = 3.25$, $\Delta df = 1$, $p \approx 0.071$). Thus, the findings indicated a positive relationship between Culture of Support Policies and WLB across both sectors but did not provide sufficient evidence that employment sector significantly moderated the strength of this relationship.

6. CONCLUSION

This study investigated the relationship between Culture of Support Policies and employees' WLB and the moderating role of employment sector. The first objective examined the influence of Culture of Support Policies on employees' WLB and found a significant positive relationship, supporting H1. The second objective examined whether Employment Sector moderated this relationship. Although Culture of Support Policies significantly predicted WLB in both the private and public sectors, the difference between the sector-specific structural paths was not statistically significant. Therefore, H2 was not supported. Overall, the findings indicate that Culture of Support Policies is positively associated with WLB across both sectors, but the strength of this relationship did not significantly differ by Employment Sector.

This study was limited to employees in Sri Lanka's Western Province. Its cross-sectional design and reliance on self-reported data also limited causal interpretation. In addition, organizational and occupational differences were not examined in detail. Future research could

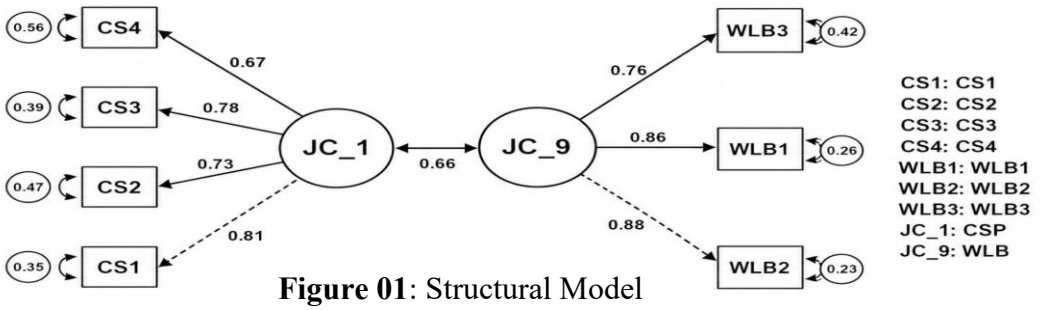


Figure 01: Structural Model

include employees from multiple provinces and adopt longitudinal designs. Furthermore, could examine factors such as supervisor support, organizational commitment, job demands, job autonomy, and organizational climate to provide a broader understanding of WLB.

REFERENCES

- Allen, T. D. (2001). Family-supportive work environments: The role of organizational perceptions. *Journal of Vocational Behavior, 58*(3), 414-435.https://doi.org/10.1006/jvbe.2000.1774
- Clark, S. C. (2000). Work/family border theory: A new theory of work/family balance. *Human Relations, 53*(6), 747–770.https://doi.org/10.1177/0018726700536001
- Fan, Y., Potočnik, K., & Chaudhry, S. (2021). A process-oriented, multilevel, multidimensional conceptual framework of work-life balance support: A multidisciplinary systematic literature review and future research agenda. *International Journal of Management Reviews, 23*(4), 486-515. https://doi.org/10.1111/ijmr.12254
- Feeney, M. K., & Stritch, J. M. (2019). Family-friendly policies, gender, and work-life balance in the public sector. *Review of Public Personnel Administration, 39*(3), 422-448. https://doi.org/10.1177/0734371X17733789
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage.
- Heras, M. L., Rofcanin, Y., Bal, P. M., Stollberger, J., Kim, S., & Mayer, M. C. J. (2021). Family-supportive organisational culture, work–family balance satisfaction and government effectiveness: Evidence from four countries. *Human Resource Management Journal, 31*(1), 1-22 https://doi.org/10.1111/1748-8583.12317
- Jayawardana, A., & Weerasinghe, T. D. (2020). Flex-work and work-life balance: Effect of role conflicts and work-life support organizational culture. *Sri Lankan Journal of Management, 24*(2), 49-76 https://doi.org/10.33939/SLJM.24.02.03.2019
- Kossek, E. E., Hammer, L. B., Kelly, E. L., & Moen, P. (2014). Designing work, family & health organizational change initiatives. *Organizational Dynamics, 43*(1), 53-63. https://doi.org/10.1016/j.orgdyn.2013.10.007
- Maduwansha, N. A. N. J. (2024). Factors affecting the work-life balance of academic staff of public universities in Sri Lanka. *International Journal of Social Statistics, 1*(2). https://doi.org/10.31357/ijss.v1i02.8273
- Nassani, A. A., Badshah, W., Grigorescu, A., Cozorici, A. N., Yousaf, Z., & Zhan, X. (2024). Participatory leadership and supportive organisational culture: Panacea for job satisfaction: the regulatory role of work-life balance. *Heliyon, 10*(16), e36043. https://doi.org/10.1016/j.heliyon.2024.e36043

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Online Hotel Review Sentiment and Booking Intention: A Moderated Study Among Sri Lankan Consumers

Prabhash KPK[1], Habaragoda BS[2]
[1,2]*Department of Industrial Management, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*
kpprabhashk@gmail.com[1], sandareka@wyb.ac.lk[2]

ABSTRACT

Online reviews are widely seen as important for hotel booking decisions, yet previous research has examined this role either through computational sentiment analysis or consumer surveys using self reported perceptions and researcher written review stimuli. This study addresses this gap by analyzing the sentiment of 26,676 authentic Booking.com reviews using VADER and examining its association with booking intention among 77 past hotel bookers in Sri Lanka. Respondents rated five general perceptions of online reviews: usefulness, positive sensitivity, negative sensitivity, volume perception, and timeliness perception. They were then shown a review representing a specific sentiment level and rated their booking intention. Moderated regression showed that review sentiment was significantly and positively related to booking intention. Review volume perception significantly moderated this relationship, while timeliness perception showed only a marginal pattern that was not robust. Usefulness, positive sensitivity, and negative sensitivity did not significantly moderate the relationship, although the small sample limited power. These exploratory findings suggest that consumers' general perceptions of the review ecosystem may matter more for responses to review sentiment than engagement with individual review content, providing implications for hospitality practice and a refinement of the Elaboration Likelihood Model in the context of authentic online reviews. Limitations include the small cross-sectional sample; future research should use larger samples and multiple review exposures.

KEYWORDS: Booking Intention, eWOM, Online Reviews, Sentiment Analysis

1 INTRODUCTION

Review valence is consistently found to be one of the strongest predictors of purchase intention across contexts and cultures (Qiu & Zhang, 2024). Lexicon- and rule-based sentiment analysis tools such as VADER (Hutto & Gilbert, 2014) are now routinely applied to hotel review corpora to extract sentiment (Ounacer et al., 2023; Nariman, 2025), but these tools are typically used as managerial diagnostics rather than as inputs into models of individual consumer decisions. This gap is sharper still in the Sri Lankan context. Jayathilaka et al. (2020) reported that online reviews play a meaningful role in inbound travellers' accommodation decisions at Sri Lankan boutique hotels, and Sally (2022) applied lexicon-based sentiment analysis to Sri Lankan user-generated content. Yet no identified study has paired an objectively scored sentiment measure with a moderated model of booking intention among Sri Lankan consumers, leaving it unclear which consumer perceptions of reviews, if any, condition the link between sentiment and booking intention. The present study therefore examines whether a VADER sentiment score predicts hotel booking intention among Sri Lankan consumers, and whether this relationship is moderated by five consumer perceptions, grouped under the Elaboration Likelihood Model into central-route (usefulness, positive

and negative sensitivity) and peripheral-route (volume, timeliness) perceptions.

2 LITERATURE REVIEW

User-generated content is positively related to hotel bookings (Ye et al., 2011), positively framed reviews with numerical ratings raise booking intentions and trust (Sparks & Browning, 2011), and review valence is among the strongest predictors of purchase intention in a meta-analysis of 156 studies (Qiu & Zhang, 2024). In Sri Lanka specifically, Jayathilaka et al. (2020) reported that online reviews played an important role in inbound travellers' hotel decisions. Sentiment-scoring research has advanced from lexicon-based tools (Ounacer et al., 2023; Nariman, 2025) to aspect-based BERT approaches (Nadeem et al., 2025). Review usefulness, grounded in the Information Adoption Model (Sussman & Siegal, 2003), reflects argument quality and source credibility, with information quality a key antecedent of trust (Filieri et al., 2015); Park et al. (2007) found quality matters more for highly involved consumers while quantity matters more for less involved ones. Positive and negative review sensitivity relate to negativity bias, the theorized tendency for negative information to weigh more heavily than positive information (Rozin & Royzman, 2001). Review volume and timeliness perception function primarily as peripheral, heuristic cues; Zhao et al. (2015) found volume a significant predictor of booking-platform intention, while Almutairi (2024) found it non-significant among Saudi bookers. The Elaboration Likelihood Model (Petty & Cacioppo, 1986) distinguishes central-route processing, careful evaluation of content, from peripheral-route processing, reliance on surface cues, and predicts that central-route perceptions should more strongly condition persuasive effects. The Stimulus-Organism-Response framework (Mehrabian & Russell, 1974) complements this by treating sentiment as the stimulus, the five perceptions as the organism, and booking intention as the response. Table 1 lists the hypotheses.

3 METHODOLOGY

Research philosophy. The study is positivist, treating its variables as objectively measurable. *Research approach.* A deductive approach was used, testing the hypotheses in Table 1 against data.

Research method. This study used a quantitative, cross sectional research design. It combined sentiment analysis of real hotel reviews with a survey of hotel customers. *Research strategy.* A survey strategy was used, with each respondent reading one real review within the questionnaire.

Data collection instruments and source. A total of 26,676 hotel reviews were

Table 1: Summary of hypotheses

Hyp.	Statement	Path
H1	Sentiment is positively related to booking intention.	Direct effect
H2a	Usefulness moderates the sentiment-intention link.	Central route
H2b	Positive sensitivity moderates that link.	Central route
H2c	Negative sensitivity moderates that link.	Central route
H3a	Volume perception moderates that link.	Peripheral route
H3b	Timeliness perception moderates that link.	Peripheral route

collected from Booking.com and analyzed using VADER (Valence Aware Dictionary and sEntiment Reasoner), a lexicon and rule based sentiment analysis tool implemented in Python with pandas (Hutto & Gilbert, 2014). VADER checks the words in each review against a prepared sentiment dictionary and also considers factors such as negation, word intensity, punctuation, and capitalization. It then gives each review a sentiment score from −1 (very negative) to +1 (very positive). This method was selected because it is transparent and has been used with hotel reviews (Nariman, 2025) and Sri Lankan user generated content (Sally, 2022). VADER gives each review a single score, so mixed, sarcastic, or translated reviews may be misclassified. These were not recoded by hand, and the scores were not validated against respondents' own perception of sentiment.

The survey had two parts. In Part A, participants answered questions about their general views of online hotel reviews (Q5 to Q21). In Part B, each participant was shown one review randomly selected from the 100 selected reviews. They then immediately rated their intention to book that hotel (Q22 to Q25).

This order was used to separate participants' general views about online reviews from their reaction to a specific review. The study therefore examined whether people's general views about online reviews were related to how strongly the sentiment of a specific review was associated with their booking intention.

Population. The target population was Sri Lankan consumers who had booked a hotel before, as they have first-hand experience of reviews and booking.

Sampling technique and sample. from the 26,676 reviews, 100 were selected by stratified sampling to cover the full sentiment range and offset the usual positive bias of online reviews. Each respondent was shown one of these reviews at random, so not all 100 were seen. Sentiment was the only review-level variable in the models; length, topic, hotel, and star rating were not controlled. Respondents were recruited by purposive sampling, online and in person, because first-hand booking experience was required and no list of past hotel bookers exists. The 77 valid responses form a small sample that limits power and generalizability.

Data analysis. The data were first tested for reliability using Cronbach's alpha. Moderated regression was then used to test the relationships. Each of the five perceptions was tested separately as a moderator because the sample size was too small to include all interaction terms in one model. A model with all five moderators needs eleven predictors, about seven cases each, far below the 10 to 15 usually recommended, so separate models were the main test and a combined model served as a check.

4 DATA COLLECTION AND ANALYSIS

Of the 77 valid responses, 67.5% were female and 64.9% were aged 25-34, with 66.2% holding undergraduate qualifications. Respondents came from 13 districts, mainly Galle (29.9%) and Gampaha (19.5%). Seven exact duplicate responses, likely caused by accidental resubmission, were retained in the primary analysis and examined separately in a robustness check. Cronbach's alpha was computed for each of the six multi-

Table 2: Reliability analysis results

Construct	α	Assessment
Booking intention	0.907	Excellent
Timeliness perception	0.868	Good
Volume perception	0.861	Good
Usefulness	0.857	Good
Positive sensitivity	0.752	Acceptable
Negative sensitivity	0.647	Below threshold

item constructs (Table 2). Five met the conventional 0.70 threshold; negative review sensitivity did not, with two items (skepticism toward excessive positivity, and an absolute rather than graded negativity rule) showing weaker item-total correlations. Removing either item did not raise alpha above 0.70, suggesting conceptual heterogeneity. Because low reliability adds measurement error that weakens interaction effects, results for this construct are interpreted with caution.

5 RESULTS AND DISCUSSION

5.1 Direct Effect of Sentiment on Booking Intention (H1)

A linear regression tested whether review sentiment predicts booking intention. As sentiment moved from the most negative possible score (-1) to the most positive (+1), predicted booking intention increased by 0.93 points on the 1-5 scale, from approximately 2.81 to 3.74. This relationship was statistically significant, $b = 0.466$, $p = .008$, $R^2 = .091$. The fitted equation was Booking Intention = 3.276 + 0.466 × Sentiment. Although this single predictor explains a modest 9.1% of the variance in booking intention, a proportion that should not be overstated, the relationship is statistically significant and indicates a genuine, non-random association between sentiment and booking decisions. H1 was therefore supported.

Table 3: Moderation analysis results

Hyp.	Moderator	p	f^2	Result
H2a	Usefulness	.694	.002	Not supported
H2b	Positive sensitivity	.659	.003	Not supported
H2c	Negative sensitivity	.910	.000	Not supported
H3a	Volume perception	.029	.068	Supported (small)
H3b	Timeliness perception	.076	.045	Marginal (suggestive)

5.2 Moderation Analysis (H2a-H3b)

Each of the five perceptions was tested as a moderator in a separate model, given that the sample size could not reliably support eleven simultaneous predictors. Results are summarized in Table 3. The three central-route perceptions, usefulness, positive sensitivity, and negative sensitivity, showed no significant moderating effect (all $p > .65$, $f^2 < .01$).

In contrast, volume perception significantly strengthened the sentiment-intention relationship, a small effect by Cohen's (1988) benchmark ($b = 0.594$, $p = .029$, $f^2 = .068$): for respondents low in volume perception, sentiment barely moved booking intention, whereas for respondents high in volume perception, sentiment had a noticeably stronger effect. Timeliness perception showed a similar but weaker, marginal and suggestive pattern ($p = .076$, $f^2 = .045$) that stops short of a confirmed effect. This pattern directly reverses the prediction of the Elaboration Likelihood Model, under which central-route perceptions were expected to condition sentiment's effect more strongly than peripheral cues. Instead, general trust in the review ecosystem, reflected in whether a hotel has enough and sufficiently recent reviews, appears to matter more than how carefully a

consumer engages with any single review's content.

5.3 Robustness and Multicollinearity Checks

When all five moderators were entered into a single combined model, the volume and timeliness interactions both lost statistical significance ($p = .125$ and $p = .388$ respectively), despite volume perception's direct effect remaining significant ($p = .038$). This reflects multicollinearity: volume and timeliness perception correlated at $r = .786$, with Variance Inflation Factors of 3.90 and 3.37, both above the conventional caution threshold of 3, indicating the two constructs may partly capture a shared underlying dimension. Removing seven exact duplicate responses left the H1 and H3a effects stable ($p = .007$ and $p = .040$ respectively), while the marginal H3b effect weakened further ($p = .142$), indicating that only the sentiment and volume-perception findings are robust.

5.4 Final Predictive Model

Combining the two statistically supported effects yields Booking Intention = 3.253 + 0.403 × Sentiment + 0.563 × Volume_c + 0.594 × (Sentiment × Volume_c), where Volume_c is volume perception mean-centered. This model explained 25.5% of the variance in booking intention ($R^2 = .255$), substantially higher than the sentiment-only model ($R^2 = .091$), confirming that volume perception meaningfully improves prediction of booking intention beyond sentiment alone.

6 CONCLUSION

This study found that VADER derived review sentiment was significantly and positively associated with hotel booking intention among Sri Lankan consumers, suggesting that authentic review content is related to booking decisions. Of the five perceptions tested as moderators, only review volume perception significantly moderated this relationship, with a small effect, although it lost significance in the combined model; timeliness showed only a marginal effect that did not survive robustness testing. The three central route perceptions showed no significant moderating effects, contrary to the pattern expected under the Elaboration Likelihood Model, although with 77 respondents small effects cannot be ruled out. Theoretically, the study bridges sentiment analysis and consumer behavior research, addresses a gap in the Sri Lankan literature, and offers tentative evidence that volume perception matters for responses to sentiment. Practically, hotel managers and online travel agencies may benefit from building review volume and managing sentiment, though findings are exploratory. The study is limited by its small sample ($N = 77$), cross-sectional design, single platform, single review exposure, low reliability of the negative sensitivity scale, unvalidated VADER scores, lack of attention checks, and uncontrolled review features such as length, topic, and hotel. Future research should use larger, more diverse samples, attention checks, controls for review-specific features, a refined negative sensitivity measure, manipulation checks, and multiple review exposures analyzed with approaches such as multilevel models.

REFERENCES

- Almutairi, O. (2024). Implication of online consumer reviews on hotel bookings in the Kingdom of Saudi Arabia. International Journal of Business and Management, 19(2), 15.

- Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum Associates.
- Filieri, R., Alguezaui, S., & McLeay, F. (2015). Why do travelers trust TripAdvisor? Antecedents of trust towards consumer-generated media and its influence on recommendation adoption and word of mouth. Tourism Management, 51, 174-185.
- Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. Proceedings of the International AAAI Conference on Web and Social Media, 8(1), 216-225.
- Jayathilaka, R., Dharmasena, T., Rezahi, N., & Haththotuwegama, S. (2020). The impact of online reviews on inbound travellers' decision making. Quality & Quantity, 54(3), 1005-1021.
- Mehrabian, A., & Russell, J. A. (1974). An approach to environmental psychology. MIT Press.
- Nadeem, A., Missen, M. M. S., Al Reshan, M. S., Memon, M. A., Asiri, Y., Nizamani, M. A., Alsulami, M., & Shaikh, A. (2025). Resolving ambiguity in natural language for enhancement of aspect-based sentiment analysis of hotel reviews. PeerJ Computer Science, 11, e2635.
- Nariman, D. (2025). Sentiment analysis of hotel reviews using lexicon-based methods: A comparative study of VADER and TextBlob. In L. Barolli (Ed.), Advances on Broad-Band Wireless Computing, Communication and Applications. Springer.
- Ounacer, S., Mhamdi, D., Ardchir, S., Daif, A., & Azzouazi, M. (2023). Customer sentiment analysis in hotel reviews through natural language processing techniques. International Journal of Advanced Computer Science and Applications, 14(1), 569-578.
- Park, D.-H., Lee, J., & Han, I. (2007). The effect of on-line consumer reviews on consumer purchasing intention: The moderating role of involvement. International Journal of Electronic Commerce, 11(4), 125-148.
- Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. Advances in Experimental Social Psychology, 19, 123-205.
- Qiu, K., & Zhang, L. (2024). How online reviews affect purchase intention: A meta-analysis across contextual and cultural factors. Data and Information Management, 8(2), 100058.
- Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. Personality and Social Psychology Review, 5(4), 296-320.
- Sally, S. (2022). Sentiment analysis on YouTube smart phone unboxing video reviews in Sri Lanka. International Journal of Research - GRANTHAALAYAH, 10(11), 53-63.
- Sparks, B. A., & Browning, V. (2011). The impact of online reviews on hotel booking intentions and perception of trust. Tourism Management, 32(6), 1310-1323.
- Sussman, S. W., & Siegal, W. S. (2003). Informational influence in organizations: An integrated approach to knowledge adoption. Information Systems Research, 14(1), 47-65.
- Ye, Q., Law, R., Gu, B., & Chen, W. (2011). The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings. Computers in Human Behavior, 27(2), 634-639.
- Zhao, X. R., Wang, L., Guo, X., & Law, R. (2015). The influence of online reviews to online hotel booking intentions. International Journal of Contemporary Hospitality Management, 27(6), 1343-1364.

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

The Impact of Lean Manufacturing Practices on Operational Performance: Evidence from the Sri Lankan Beverage Industry

Athapattu ARNC[1], Jothirathne PAAU[2]
[1,2]*Department of Industrial Management, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*
nirashaathapattu736@gmail.com[1], anoja@wyb.ac.lk[2]

ABSTRACT

Lean manufacturing aims to achieve smooth production flow by eliminating waste, helping firms meet quality, cost, delivery, and productivity goals. Sri Lankan beverage manufacturers face inefficiencies, high costs, excess inventory, low productivity, quality issues, and delivery delays, yet integrated evidence on Lean's multi-dimensional impact in this sector remains limited. This study examines the extent of Lean adoption in the Sri Lankan beverage industry and tests both direct and indirect relationships between Lean Manufacturing Practices and Operational Performance with Production Flow Optimization as a mediator. Data was collected from 81 employees in production and operations roles via a structured questionnaire, capturing five Lean constructs: Employee-Oriented Practices, Total Quality Management, Total Productive Maintenance, Just-in-Time, and 5S. Hypotheses were evaluated using PLS-SEM in SmartPLS, following preliminary analysis in SPSS. The results show that Lean manufacturing practices significantly and positively affect both production flow optimization and operational performance, and that production flow optimization partially mediates this relationship, supporting Lean's role in enhancing performance both directly and through improved production flow. Lean practices were adopted above the scale midpoint, with 5S and employee-oriented practices most consistent and JIT lowest; the findings rest on a convenience sample and a cross-sectional design and should therefore be interpreted with caution.

KEYWORDS: Beverage Industry, Lean Manufacturing, Operational Performance, Production Flow Optimization

1 INTRODUCTION

Mass production has driven manufacturers toward continuous quality and productivity improvement, yet waste in materials, time, and resources remains a persistent challenge. Lean manufacturing addresses this through practices such as Employee-Oriented Practices (EOP), Total Quality Management (TQM), Total Productive Maintenance (TPM), Just-in-Time (JIT), and 5S, which together reduce costs and inventory while improving quality, productivity, and delivery. Lean is especially relevant to beverage manufacturing, where high production volumes, interdependent processes, and unpredictable demand create significant waste. Although Sri Lankan beverage companies are increasingly adopting Lean practices, limited research examines their effect on production flow and operational performance. Prior studies elsewhere show Lean improves performance often through an intermediate mechanism, yet whether production flow optimization mediates this relationship remains unexplored locally.

This study aims to examine the effect of lean manufacturing practices on operational performance and on production flow optimization, the effect of production flow optimization on

operational performance, and whether production flow optimization mediates the relationship between lean manufacturing practices and operational performance in the Sri Lankan beverage industry.

2 LITERATURE REVIEW

Lean manufacturing, rooted in the Toyota Production System (TPS), has evolved from a set of discrete tools into a comprehensive philosophy centered on waste elimination and continuous improvement. Bhamu and Sangwan (2014) note that lean is most effective as an integrated bundle of practices, 5S, JIT, TPM, and TQM, rather than isolated initiatives. This study draws on TPS and the Resource-Based View, which frames lean-derived capabilities as strategic resources for competitive advantage.

Empirical studies confirm positive links between lean practices and operational performance. Chavez et al. (2013) show lean implementation enhances quality, inventory control, and productivity. Operational performance is assessed across five dimensions, cost, delivery, inventory, quality, and productivity, while production flow optimization, measured through cycle time, supply coordination, and layout design, is a key linking mechanism (Afy-Shararah & Rich, 2018), with relevance confirmed in food and beverage settings (Khusaini et al., 2014). Recent evidence suggests lean effects operate through mediators such as quality systems (Budianto et al., 2023), though not production flow.

Within Sri Lanka, empirical evidence remains notably scarce**.** De Alwis Seneviratne et al. (2021) found JIT among the most widely adopted practices across 342 manufacturing SMEs, improving customer satisfaction and waste reduction. Wickramasinghe and Wickramasinghe (2017), using 1,189 respondents from the apparel sector, found lean practices significantly enhance manufacturing performance. Most directly relevant, Silva and Warnapura (2021) found lean practices improve business performance through operational performance as a partial mediator, in Sri Lankan food manufacturing. A recent apparel study (Abhayagunarathna et al., 2024) found that continuous improvement, leadership support, and supplier involvement improved operational performance whereas waste elimination did not.

Despite this, three critical gaps remain unaddressed locally**.** No Sri Lankan study examines the beverage subsector, despite its distinct continuous-batch and hygiene demands; none tests production flow optimization, distinct from operational performance itself, as a mediating mechanism and none treats all five lean practices as a single integrated bundle.

3 METHODOLOGY

This study adopted a quantitative research design to examine the effect of lean manufacturing practices on operational performance in the Sri Lankan beverage industry, with production flow optimization as a mediating variable. A quantitative approach was chosen because it enables objective, numerical testing of hypothesized relationships among measurable variables, and a cross-sectional survey was feasible within the study's time and access constraints.

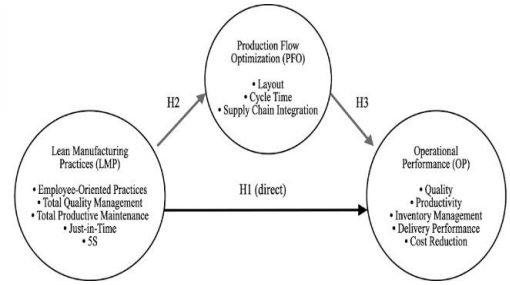


Figure 1: Conceptual model

The study proposes four hypotheses. H1: Lean manufacturing practices have a significant positive impact on operational performance. H2: Lean manufacturing practices have a significant positive impact on production flow optimization. H3: Production flow optimization has a significant positive impact on operational performance. H4: Production flow optimization mediates the relationship between Lean manufacturing practices and operational performance. Figure 1 presents the conceptual model.

The population includes employees, supervisors, executives, and managers involved in production and operations at Sri Lankan beverage manufacturing companies. The convenience sampling method was used due to accessibility constraints, as no complete sampling frame of beverage production staff exists. The 81 valid questionnaires collected exceed the minimum of 51 needed to detect the smallest structural path ($\beta = 0.350$) at 5% significance and 80% power (Kock & Hadaya, 2018).

4 DATA COLLECTION & ANALYSIS

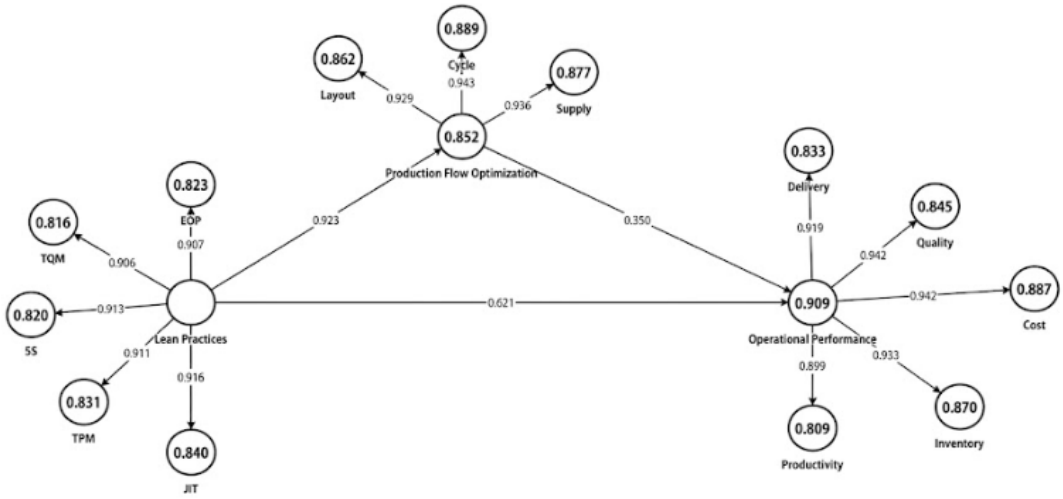
Primary data were collected via a structured questionnaire using validated measurement scales, comprising 39 items across all three constructs, rated on a five-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree), which limits respondent burden. Questionnaires were distributed directly to eligible respondents in beverage manufacturing organizations.

Data were analyzed in two stages. IBM SPSS Statistics was first used for preliminary analysis (reliability, descriptive statistics, normality, correlation). SmartPLS was then used to estimate the measurement model (indicator reliability, internal consistency, convergent validity) and the structural model (path coefficients, R^2, bootstrapped significance and mediation) via Partial Least Squares Structural Equation Modeling (PLS-SEM). PLS-SEM was chosen because it suits prediction-oriented mediation models with modest samples (Hair et al., 2019); bootstrapping needs no distributional assumptions.

5 RESULTS AND DISCUSSION

All four hypotheses were supported. The descriptive statistics of all first-order constructs and the Pearson correlation coefficients between them are summarized below. All five Lean manufacturing practices recorded means above 3.85 on the five-point scale, indicating a relatively high level of Lean practice in Sri Lankan beverage manufacturing firms; 5S (mean = 4.037) and EOP (mean = 3.996) were the most consistently implemented practices.

All correlations among the study variables were positive and significant at the 0.01 level, providing initial support for the proposed relationships prior to structural testing; correlations between Lean practices and cost, quality outcomes were particularly strong, suggesting a potentially important role for TQM, TPM and EOP type practices in reducing costs and improving quality.



.

Figure 2: Path model

Table 1: Results of the hypothesis tests

Hypothesis	Relationship	β	t-value	p-value	Decision
H1	Lean Manufacturing Practices → Operational Performance	0.621	5.592	<0.001	Supported
H2	Lean Manufacturing Practices → Production Flow Optimization	0.923	58.432	<0.001	Supported
H3	Production Flow Optimization → Operational Performance	0.35	3.252	<0.001	Supported
H4	LMP → PFO → OP	0.3636	5.601	<0.001	Supported

According to Fig. 2, all path coefficients are positive and significant at the 1% significance level, and Table 1 gives the bootstrapping results for all hypothesized direct and indirect effects. Lean manufacturing practices had a significant direct effect on operational performance and an even stronger effect on production flow optimization, which in turn significantly affected operational performance. Because both the direct and the specific indirect paths are significant, production flow optimization partially mediates the relationship between Lean manufacturing practices and operational performance.

R^2 values of 0.25, 0.50 and 0.75 indicate weak, moderate and substantial explanatory power (Hair et al., 2019). The observed R^2 values were 0.855 for production flow optimization and 0.904 for operational performance, both substantial. Collinearity was not a concern (VIF 1.395–3.545, below 5.0).

6 CONCLUSION

All five lean practices (EOP, TQM, TPM, JIT, 5S) were adopted above the scale midpoint, with 5S and EOP most consistent. All four hypotheses were supported: lean practices affected production flow optimization (β = 0.923) more strongly than operational performance directly (β = 0.621), and production flow optimization partially mediated this relationship, consistent with Silva and Warnapura's (2021) partial-mediation finding in Sri Lankan food manufacturing, though this study

identifies production flow, specifically, as the mediating mechanism. JIT's comparatively lower adoption (mean = 3.856), unlike De Alwis Seneviratne et al.'s (2021) SME findings, may reflect changeover constraints in process industries (Maalouf & Zaduminska, 2019), although this was not tested. The study is limited by its convenience sample, cross-sectional design, and reliance on self-reported data. Beverage manufacturers should prioritize production-flow investment, layout, cycle time, and supply coordination, and consider adapted JIT approaches where changeover constraints are confirmed. Future research should adopt longitudinal designs, test the JIT explanation directly, disaggregate operational performance by dimension, and extend this model to other Sri Lankan process industries.

REFERENCES

- Abhayagunarathna, W. D. G. G., Adihetti, I., & Arachchi, V. S. (2024). How lean management drives operational excellence in Sri Lanka's apparel sector: A pathway to competitive advantage. *International Journal of Contemporary Business Research, 3*(1), 23–45. https://doi.org/10.4038/ijcbr.v3i1.3
- Afy-Shararah, M., & Rich, N. (2018). Operations flow effectiveness: A systems approach to measuring flow performance. *International Journal of Operations & Production Management, 38*(11), 2096–2123. https://doi.org/10.1108/IJOPM-09-2016-0575
- Bhamu, J., & Sangwan, K. S. (2014). Lean manufacturing: Literature review and research issues. *International Journal of Operations & Production Management, 34*(7), 876–940. https://doi.org/10.1108/IJOPM-08-2012-0315
- Budianto, B., Feri, Z. O., Nurlaila, Q., Sitorus, H., & Suryatman, T. H. (2023). Quality system as a mediating variable of the relationship between lean manufacturing and operational performance in the food industry. *SINERGI, 27*(2), 201–210. https://doi.org/10.22441/sinergi.2023.2.008
- Chavez, R., Gimenez, C., Fynes, B., Wiengarten, F., & Yu, W. (2013). Internal lean practices and operational performance: The contingency perspective of industry clockspeed. *International Journal of Operations & Production Management, 33*(5), 562–588. https://doi.org/10.1108/01443571311322724
- De Alwis Seneviratne, R. Y. H., Dissanayake, K., Premaratna, S. P., Melegoda, N., Ranwala, S., & Fernando, A. (2021). Lean practices and benefits: Study of Sri Lankan small and medium enterprises. *Journal of Business Studies, 8*(Special Issue), 1–21. https://doi.org/10.4038/jbs.v8i0.65
- Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of partial least squares structural equation modeling. *European Business Review, 31*(1), 2–24. https://doi.org/10.1108/EBR-11-2018-0203
- Khusaini, N. S., Jaffar, A., & Yusoff, N. (2014). A survey on lean manufacturing practices in Malaysian food and beverages industry. *Applied Mechanics and Materials, 564*, 632–637. https://doi.org/10.4028/www.scientific.net/AMM.564.632
- Kock, N., & Hadaya, P. (2018). Minimum sample size estimation in PLS-SEM: The inverse square root and gamma-exponential methods. *Information Systems Journal, 28*(1), 227–261. https://doi.org/10.1111/isj.12131
- Maalouf, M., & Zaduminska, M. (2019). A case study of VSM & SMED in the food processing industry. *Management & Production Engineering Review,10*(2),60–68. https://doi.org/10.24425/mper.2019.129569
- Silva, C. W. C., & Warnapura, H. H. S. C. (2021). Impact of lean manufacturing practices on operational and business performance: Evidence from Sri Lankan food industry. *Sri Lanka Journal of Social Sciences and Humanities, 1*(2), 69–76. https://doi.org/10.4038/sljssh.v1i2.39
- Wickramasinghe, G. L. D., & Wickramasinghe, V. (2017). Implementation of lean production practices and manufacturing performance: The role of lean duration. *Journal of Manufacturing Technology Management, 28*(4), 531–550. https://doi.org/10.1108/JMTM-08-2016-0112

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Association between the Marketing Mix and Sri Lankan Audience Engagement with Sinhala Cinema

Madhuranga WADS [1], Wattegama EJ[2]
[1,2] *Department of Industrial Management, Faculty of Applied Sciences, Wayamba University of Sri Lanka Kuliyapitiya, Sri Lanka*
sulakshanamadhuranga@gmail.com[1], erandiw@wyb.ac.lk[2]

ABSTRACT

There is a constant fall in the number of viewers visiting cinema halls for Sinhala films in Sri Lanka over the last decade particularly among the youth who are moving towards foreign films and online streaming services. Reviewed literature has shown that there are few studies which have focused on movie marketing through the traditional marketing mix approach of Product, Price, Place and Promotion in the South Asian indigenous cinema environment. This research aims at examining the effect of marketing mix variables namely Film Content and Quality, Ticket Pricing and Value Perception, Cinema Hall Experience and Accessibility and Movie Promotion and Communication on Sri Lankan Audience Engagement who are ranging from 15 to 40 years of their age. Using a quantitative cross-sectional design, a bilingual (Sinhala and English) structured questionnaire has used for collecting data among 220 cinemagoers who have been selected purposively. Reliability analysis indicates high internal consistency for all the variables. After stepwise multiple regression analysis, it is found that Ticket Pricing and Value Perception and Cinema Hall Experience and Accessibility as significant predictors explaining 23.7% of the variance in audience. The significance correlation between Cinema Hall Experience and Accessibility and Movie Promotion and Communication (r = 0.898) indicates a serious multicollinearity and discriminant-validity concern: these two scales may not be empirically distinct, and this must be resolved before firm managerial conclusions are drawn about them.

KEYWORDS: Audience Engagement, Marketing Mix, Sinhala Cinema

1 INTRODUCTION

Sinhala cinema in Sri Lanka faces declining theatre attendance, which is often described as most pronounced among younger audiences. Press reports indicate that annual cinema attendance fell from about 17 million in 2001 to 7.2 million in 2009 (Sunday Times, 2010), and an industry representative reported that only about 100 of 187 cinemas were operating in 2021 and that Sinhala film audiences had fallen by 58.35% by 2019 (Daily Mirror, n.d.). These figures come from press reports rather than systematic audience surveys, and age-specific attendance data could not be identified. Despite the cultural legacy of Sinhala cinema, leisure consumption has shifted towards imported foreign films, streaming services and social media content. The industry has responded with modern multiplex facilities, digital projection and increased production, yet attendance has not recovered, and film promotion may not be resonating with the values of modern consumers (Mohammadian & Sezavar Habibi, 2012).

Two gaps motivate this study. The contextual gap is that existing 4Ps studies of cinema audiences come mainly from Western, East Asian or Middle Eastern settings (Prag & Casavant, 1994; Mohammadian & Sezavar Habibi, 2012; Ülker Demirel et al., 2018), with little evidence from Sinhala or wider South Asian cinema, while the cultural-industry marketing mix literature remains largely

conceptual (Shahhosseini & Tohidy Ardahaey, 2011). The empirical and theoretical gap is that these studies treat Product, Price, Place and Promotion as separate predictors without testing whether their measures are empirically distinct (discriminant validity) or how the four dimensions jointly relate to audience engagement. The present study therefore examines the association between Film Content and Quality (Product), Ticket Pricing and Value Perception (Price), Cinema Hall Experience and Accessibility (Place), and Movie Promotion and Communication (Promotion) and the Audience Engagement of Sri Lankan viewers aged 15–40 years, assessing the predictive value of each dimension.

2 LITERATURE REVIEW

For theatrical film consumption, marketing mix corresponds to: film content and quality (artistry, actors, photography, genre), ticket pricing and value perception (affordable cost, value of price / performance, promotions), cinema hall experience, and accessibility (comfort of auditorium, acoustics/quality of vision, cleanliness, safety and physical location), and movie promotion and communications (trailers, promotion in social media, reviews, star publicity) (Litman, 1983; Mohammadian & Habibi, 2012; Saygin, 2025).

2.1 Film Content and Quality (Product)

The film itself is the experiential product. Early research (Litman, 1983; Ravid, 1999) showed that star appeal, plot and production scale are associated with lower perceived risk and with box-office success. Box-office revenue is an aggregate behavioural outcome, whereas Audience Engagement in this study is a self-reported construct; box-office evidence is therefore used only as indirect support that film attributes shape audience demand, and these attributes are expected to be positively associated with engagement. In Nigerian cinema, culturally relevant content has also been linked to audience engagement (Lima Foehn, 2025).

2.2 Ticket Pricing and Value Perception (Price)

Pricing of cultural services signals value as well as monetary cost (Prag & Casavant, 1994), and perceived value for money may be associated with higher attendance. The evidence is mixed, however Mohammadian and Sezavar Habibi (2012) reported that price was the one marketing mix variable that was not influential among Iranian cinema audiences, whereas perceived value, customer experience and service quality were found to affect loyalty among Cinema XXI customers in Bandung (Mahmudah & Iskamto, 2024).

2.3 Place and Promotion: Distinct Constructs or Overlapping Measures?

Physical attributes of cinema halls (acoustics, screen projection and ergonomic seating) fall under Place (Management and Planning Organization of the Country, 2002; Palmer, 2005), although comfort and ambience are sometimes treated as physical evidence in the extended services mix (Palmer, 2005). Contemporary film promotion relies on trailers, social media and electronic word of mouth (Liu, 2006; Saygin, 2025), and promotion, cast and distribution diversity have been linked to purchase intention in Istanbul (Ülker Demirel et al., 2018). Place and Promotion are therefore treated here as theoretically distinct constructs: one concerns the physical venue and access,

the other communication about films. Whether respondents' scores on the two scales are empirically distinct is a measurement question that earlier studies assumed rather than tested; it is examined here through correlation and reliability-based checks rather than assumed.

2.4 Conceptual Framework and Hypotheses

Based on established theoretical foundations, the conceptual model posits direct relationships between the four marketing mix elements and those are positively associated with Audience Engagement;

H1: Film Content and Quality is positively associated with audience engagement with Sinhala cinema.

H2: Ticket Pricing and Value Perception is positively associated with audience engagement with Sinhala cinema.

H3: Cinema Hall Experience and Accessibility is positively associated with audience engagement with Sinhala cinema.

H4: Movie Promotion and Communication is positively associated with audience engagement with Sinhala cinema.

3 METHODOLOGY

In order to achieve the study objective a quantitative, deductive, and cross-sectional survey approach was followed. The target population comprised Sri Lankans aged 15–40 years with an active or intermittent involvement in theatrical films. In the absence of a sampling frame of national cinemagoers, a convenience sampling strategy was used at tertiary-level institutions, corporate locations, and social media cinemagoing groups (as used by Kumar, 2024). This recruitment route is likely to over-represent students, younger respondents and relatively active cinemagoers, so the findings describe this recruited sample and cannot be generalised to the wider Sri Lankan cinema audience.

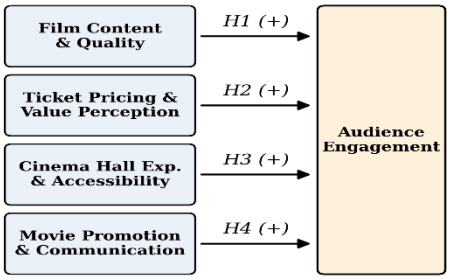


Figure 1: Research Model

Cochran's formula for an infinite population gives a minimum of 384 respondents for estimating a proportion at a 95% confidence level and ±5% margin of error. This criterion was not met: data screening left N = 220 valid responses, so estimates are less precise than the Cochran target implies. Adequacy was therefore judged against regression-specific guidance (Green, 1991), which recommends $N \geq 50 + 8k$ (= 82) to test the overall model and $N \geq 104 + k$ (= 108) to test individual predictors, with k = 4 predictors. Both are satisfied, but the smaller sample limits power to detect small effects.Primary data were gathered through a structured, self-administered Google Form in English and Sinhala. It included 22 items on a 5-point Likert scale from 1 ("Strongly Disagree") to 5 ("Strongly Agree"): Film Content and Quality (6 items), Ticket Pricing and Value Perception (3), Cinema Hall Experience and Accessibility (4), Movie Promotion and Communication (6) and Audience Engagement (4), together with demographic questions. The 22 items were developed by the researcher on the basis of the cinema marketing-mix literature and the study objectives; they

were not adopted from a single existing validated scale. The questionnaire was prepared in English and translated into Sinhala by the researcher. A second, independent Sinhala translation was produced by a peer, and the two versions were compared and reconciled into a single final Sinhala version to improve translation accuracy. Formal back-translation and expert linguistic validation were not carried out, so full linguistic equivalence between the two versions cannot be assumed.

The pilot study (n = 43, separate from the final sample of N = 220) was used to check item clarity and estimate preliminary reliability. Some items were removed from each construct, leaving the 22 items used in the final survey. Reliability alone does not establish instrument validity, and no factor-analytic validation was performed. Careless responders were identified using the intra-individual response standard deviation (Response_SD across 22 items); four straight-liners (Response_SD = 0.00) were found and retained in the analysis

Data were analysed in SPSS. Descriptive statistics summarised the constructs, and Pearson product-moment correlations examined the relationships between constructs and screened for multicollinearity. Stepwise multiple regression was used as an exploratory procedure to identify the most parsimonious set of predictors of Audience Engagement. Because stepwise selection is data-driven and can yield unstable models, a theory-driven simultaneous regression including all four hypothesised predictors is also reported as a robustness check, with VIF values for both models. Residual diagnostics comprised a scatterplot of standardised residuals against predicted values and a Normal P-P plot. A hypothesis was treated as supported only if its coefficient was significant ($p < 0.05$) in the hypothesised positive direction; a significant coefficient in the opposite direction was treated as contradicting the hypothesis.

4 RESULTS

In the final sample (N = 220), the distribution according to gender was equal: 51.8% male (n = 114) and 48.2% female (n = 106). The distribution according to age was the following: 70.5% (n = 155) of 20-30 years old; 23.6% (n = 52) of 30-40 years old; 4.1% (n = 9) of 40-50 years old; 1.8% (n = 4) above 50 years. Thirteen respondents (5.9%) were therefore older than the stated 15–40 target range and were retained in the analysis, which is a departure from the eligibility criterion and a limitation. The sample is predominantly young (94.1% aged 20–40) but, being recruited by convenience, is a student-heavy sample of active cinemagoers rather than a representative sample of Sri Lankan audiences. Occupation wise, students constituted 54.1% (n = 119) and employed people 38.6% (n = 85). The distribution according to moviegoers' frequency: 45.9% (n = 101) monthly or even more; 35.9% (n = 79) sometimes; 17.7% (n = 39) rarely, which confirms that the sample is weighted towards active cinemagoers.

Cronbach's Alpha coefficients were calculated for the pilot study (n = 43) and, separately, for the final sample (N = 220). In the pilot, Ticket Pricing and Value Perception ($\alpha = 0.669$) and Audience Engagement ($\alpha = 0.680$) fell slightly below the 0.70 benchmark (Nunnally, 1978), which may reflect the

Table 1: Measurement Scale Reliability Statistics

Construct	Items	Pilot α (n=43)	Final α (N=220)
Film Content & Quality (FCQ)	6	0.746	0.901
Ticket Pricing & Value (TPV) Perception	3	0.669	0.853
Cinema Hall Experience (CHE)	4	0.793	0.944
Movie Promotion & Communication (MPC)	6	0.830	0.955
Audience Engagement (AE)	4	0.680	0.887

small pilot sample and short scales (3 and 4 items). In the final sample (Table 1), all constructs exceeded 0.70 (0.853–0.955). Alpha indicates internal consistency only and does not establish construct validity; values above 0.95, as for Movie Promotion and Communication, may also signal item redundancy.

The audience Engagement achieved the highest mean score (M = 4.15, SD =0.85), closely followed by Film Content and Quality (M = 4.14, SD = 0.85). Ticket Pricing and Value Perception showed moderate agreement (M = 3.85, SD = 0.97), while Cinema Hall Experience (M = 3.40, SD = 1.27) and Movie Promotion (M = 3.27, SD = 1.20) scored lower, accompanied by higher variance. In the analysis of Pearson correlation (Table 2), Ticket Pricing and Value Perception showed the highest bivariate positive correlation with Audience Engagement ($r = 0.441$, $p < 0.01$). The second strongest positive correlation was revealed between Film Content and Quality ($r = 0.260$, $p < 0.01$). However, Cinema Hall Experience ($r = -0.405$, $p < 0.01$) and Movie Promotion ($r = -0.328$, $p < 0.01$) showed strong negative correlation with Engagement. At the same time, the extremely high positive correlation between Cinema Hall Experience and Movie Promotion was found ($r = 0.898$, $p < 0.01$), which indicates strong multicollinearity of Place and Promotion variables. The stepwise multiple regression procedure with Audience Engagement as dependent variable was performed. In the first step, Ticket Pricing and Value Perception variable was included ($R = 0.441$, $R^2 = 0.194$, $F = 52.552$, $p < 0.001$). In Step 2, Cinema Hal Experience and Accessibility were included into the model significantly improving its predictive power ($R = 0.487$, $R^2 = 0.237$, Adjusted $R^2 = 0.230$, $\Delta R^2 = 0.043$, $F(2, 217) = 33.718$, $p < 0.001$).

Film Content and Quality and Movie Promotion and Communication variables were not included in the final equation because they do not add any value when Place and Price are considered.

Variance Inflation Factors (VIF = 1.362) remained well below conservative thresholds which was demonstrating that multicollinearity was resolved within the retained two predictor model.

5 DISCUSSION

The results yield valuable empirical insight regarding theatrical cinema consumption by Sri Lankan youth. First,

Table 2: Pearson Correlation Matrix (N = 220)

Constru	AE	FCO	TPV	CH	MP
AE	1.00				
FCO	0.26	1.00			
TPV	0.44	0.70	1.00		
CHE	-	-	-	1.00	
MPC	-	-	-	0.89	1.00

Ticket Pricing and Value Perception proved to be the main positively contributing factor towards audience engagement ($\beta = 0.316$, $p < 0.001$). This fact confirms the tenets of service marketing theory (Palmer, 2005; Mohammadian & Habibi, 2012) stating that audiences assess attendance at the movies on the basis of "value for money" ratio, rather than simple costs. Young viewers perceive their experience as valuable enough in terms of art and emotions in relation to ticket costs. Conversely, Cinema Hall Experience and Accessibility was a significant negative predictor ($\beta = -0.242$, $p = 0.001$), contrary to the hypothesised positive direction and to cinema service-experience findings elsewhere (e.g., Mahmudah & Iskamto, 2024). This result should be interpreted cautiously. First, Cinema Hall Experience and Movie Promotion are almost perfectly correlated ($r = 0.898$), and in the simultaneous model Movie Promotion reverses sign while Cinema Hall Experience becomes more negative, a pattern consistent with a suppression-type or unstable-estimation effect, although no formal suppression analysis was conducted. Second, one possible interpretation, not tested by these data, is that culturally devoted viewers of Sinhala films attend traditional single-screen cinemas with older infrastructure, whereas less-engaged viewers use multiplexes and expect better facilities; future studies could test this by recording the type of cinema attended.

Film Content and Quality was not retained, although it correlated positively with engagement ($r = 0.260$). In the simultaneous model its unique coefficient was small and non-significant ($\beta = -0.10$, $p = .224$) and its VIF was moderate (2.01), so its exclusion mainly reflects limited incremental explanatory power beyond Ticket Pricing and Value Perception, with which it shares substantial variance ($r = 0.708$), rather than severe collinearity. It is possible that audiences judge film quality and ticket value together, but this was not tested.

Finally, the very high correlation between Place and Promotion ($r = 0.898$) is the most important caveat of this study. It may indicate that respondents evaluate the cinema experience and its promotion as a combined impression, but it may equally reflect measurement overlap or insufficient discriminant validity between the two scales, for instance because items refer to similar facilities and communication cues. The present data cannot distinguish between these explanations; factor-analytic evidence on the item-level data is needed before concluding that Place and Promotion are perceived as a single construct.

6 CONCLUSIONS & RECOMMENDATIONS

Within this convenience sample, perceived ticket value was the strongest statistical predictor of audience engagement with Sinhala films. The traditional approach that treats Product, Price, Place and Promotion as independent variables did not hold up cleanly in these data: the very high correlation between the Place and Promotion scales suggests that they overlap or measure related aspects of the cinema experience, and their discriminant validity should be tested before they are treated as separate marketing levers. All findings are associations and not evidence of causal effects.

Film producers and directors: Film Content and Quality was positively correlated with engagement ($r = 0.260$)

but was not a significant unique predictor, so this study provides no direct evidence that quality improvements alone raise engagement. Because quality and ticket value were strongly associated (r = 0.708), producers may nevertheless wish to ensure that quality and informative trailers support audiences' perception of value; this is a managerial suggestion rather than a tested finding.

Film Corporation of Sri Lanka: Create national value-pricing strategies that will include discounts for students, flexible matinee ticket pricing, and subscription tickets aimed at engaging youth audiences in both urban and provincial circuits.

Cinema hall operators: Because the Place coefficient was negative and unstable, this study does not support infrastructure investment (for example 4K projection or acoustic upgrades) as a way of raising engagement; any such investment is a broader managerial consideration not demonstrated by these data. Operators may wish to trial in-theatre and social media check-in promotions for young audiences and to gather qualitative evidence on what audiences mean by "cinema hall experience" before major investment.

Marketing agencies: Allocate marketing budgets from traditional print/poster promotions to more interactive digital marketing and social media influencer promotion as well as via eWOM channels.

This study has several limitations. It used a convenience sample recruited through tertiary institutions, corporate locations and social media groups, which is dominated by students and active cinemagoers (5.9% of respondents were outside the 15–40 target range), and its N = 220 is below the Cochran target of 384.

Table 3: Final Stepwise Regression Coefficients

Model Parameter	B	SE	β	t	Sig.	VIF
Constant	3.627	0.345		10.51	0.000	
Ticket Pricing & Value	0.277	0.061	0.316	4.569	0.000	1.362
Cinema Experience	-0.161	0.046	-0.242	-3.491	0.001	1.362

Table 4: Summary of Hypothesis Testing Outcomes

Hypothesis	Empirical Result	Decision
H1: FCQ & Engagement	Excluded (β = −0.10, p = .224, simultaneous model)	Not Supported
H2: TPV & Engagement	β = 0.316, t = 4.569, p < 0.001	Supported
H3: CHE & Engagement	β = -0.242, t = -3.491, p = 0.001	Not Supported
H4: MPC & Engagement	Excluded (β = +0.19, p = .151, simultaneous model)	Not Supported

The cross-sectional design supports associations only, and the results may not generalise to other age groups or to Tamil- and English-language cinema. Cronbach's alpha does not establish construct validity, and the overlap between Place and Promotion (r = 0.898) remains unresolved. Residual analysis showed heteroscedasticity, so significance tests are approximate. Further studies should validate the scales with exploratory/confirmatory factor analysis, use heteroscedasticity-robust estimation and longitudinal or SEM designs, and consider electronic word-of-mouth as a mediator.

REFERENCES

- Daily Mirror. (n.d.). *Cinema: A local industry that is not an industry*. https://dailymirror.lk/insight/Cinema-A-

local-industry-that-is-not-an-industry/374-218466

- Ali Shahhosseini, F. T., & Farzan, F. (2011). Marketing mix practices in the cultural industry: The film sector. *The Canadian Center of Science and Education Journal*, *4*(1), 120 - 128.
- Green, S. B. (1991). How many subjects does it take to do a regression analysis? *Multivariate Behavioral Research*, *26*(3), 499-510.
- Kheireddine, M. H., & Saleh, M. (2025). The impact of marketing mix on customer satisfaction: An empirical investigation. *International Journal of Management Studies*, *12*(1), 45 - 58.
- Kumar, S. (2024). Impact of marketing strategies on consumer buying behaviour with specific reference to movies as an experiential medium. *International Research Journal on Advanced Engineering and Management*, *8*(2), 112 - 124.
- Lima, O. P., & Foehn, M. R. (2025). Cultural relevance & audience engagement in domestic film industries.*Journal of Media & Entertainment Economics*, *19*(1), 88-104.
- Litman, B. R. (1983). Predicting success of theatrical movies: An empirical study. *Journal of Popular Culture*, *16*(4), 159-175.
- Liu, Y. (2006). Word of mouth for movies: Its dynamics & impact on box office revenue. *Journal of Marketing*, *70*(3), 74-89.
- Mohammadian, M., & Habibi, F. (2012). The impact of marketing mix on attracting audiences to the cinema. *Management Science Letters*, *2*(3), 997-1006.
- Nunnally, J. C. (1978). *Psychometric Theory* (2nd ed.). New York: McGraw-Hill.
- Palmer, A. (2005). *Principles of Services Marketing* (4th ed.). London: McGraw-Hill Education.
- Prag, J., & Casavant, J. (1994). An empirical study of the determinants of revenues and marketing expenditures in the motion picture industry. *Journal of Cultural Economics*, *18*(3), 217 -235.
- Ravid, S. A. (1999). Information, blockbusters, and stars: A study of the film industry. *The Journal of Business*, *72(*4), 463 - 492.
- Saygin, E. P. (2025). Movie trailers as an integrated marketing communication tool in contemporary cinema. *OPUS Journal of Society Research*, *15*(2), 205-219.
- Standards Organization. (2002). *Cinema Design Criteria & Acoustic Regulations*. National Planning Department Report.

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Determinants of Work-Life Balance: A Case Study of the Furniture Manufacturing Sector in Sri Lanka

Dasanayaka TPNN[1], Deegahawature MMDR[2]
[1,2] *Department of Industrial Management, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*
nipuninavodya111@gmail.com[1], dharsana@wyb.ac.lk[2]

ABSTRACT

Balancing work and life (WLB) is an important aspect of employee wellbeing and productivity, but it has been under-researched in developing nations manufacturing companies. This study investigates the organizational aspects associated with WLB of 268 employees of two furniture-manufacturing companies in Sri Lanka. Hypotheses were tested for the relationship between the Work-Family Border perspective and the Work-Family Enrichment perspective, with organizational demand, coworker support, job autonomy and job satisfaction. After reverse coding the negatively worded items, simple linear regression was used to test the four hypotheses. All four were supported: job autonomy ($\beta = 0.820$, $R^2 = 0.720$, $p < 0.001$), job satisfaction ($\beta = 0.619$, $R^2 = 0.346$, $p = 0.005$) and coworker support ($\beta = 0.449$, $R^2 = 0.239$, $p < 0.001$) were positive predictors, while lower organizational demand was associated with higher WLB ($\beta = 0.042$, $R^2 = 0.053$, $p < 0.001$). When all four variables (Demand, Coworker Support, Autonomy, and Job Satisfaction) are included in the same model ($R^2 = 0.338$, $F(4, 263) = 33.49$, $p < 0.001$), job satisfaction is no longer significant. The results expand knowledge of WLB theory to a labor-intensive manufacturing setting and inform focused and resource efficient interventions.

KEYWORDS: Coworker Support, Furniture Manufacturing , Job Autonomy, Organizational Demand, Work-Life Balance

1 INTRODUCTION

Work-life balance (WLB) is the balance between the time, cognition and emotion an individual allocates to work and to personal or family life (Geetha, 2023). Organizational conditions, namely the demands placed on employees and the resources available to manage them, are key determinants of WLB (Bjärntoft, (2020)), (Domagalska-Grędys, 2025). In furniture manufacturing, assembly-line work, varied shifts and a production-oriented culture may erode workers' balance (Ganiyu, 2018).

Resources such as coworker support, job autonomy and job satisfaction may offset these demands, but their relative importance in labor-intensive settings is unclear. Sri Lankan manufacturing workers face long hours, pressure and limited flexibility (Imawa, 2022), yet WLB research has concentrated on developed economies (Zychová, 2024), (Geetha, 2023) and, in Sri Lanka, on service and white-collar jobs (Kajenthiran, 2019), (Chimote, 2013). Organizational determinants, particularly coworker support, are therefore less studied than individual ones.

This study aims to identify the organizational factors influencing WLB among furniture manufacturing employees in Sri Lanka by analyzing the relationships of organizational demand, coworker support, job autonomy and job satisfaction with WLB, and to suggest organizational-level interventions.

2 LITERATURE REVIEW

Organizational factors are the structural, cultural and relational conditions that shape employees' opportunities for WLB. They operate on the demand side (workload, performance pressure) and the resource side (peer relationships, job design, job satisfaction).

Four are examined: organizational demand, coworker support, job autonomy and job satisfaction.

2.1 Organizational Demand

Organizational demand refers to the workload, time pressure, role ambiguity and performance expectations placed on employees. High demand is consistently associated with work-family conflict, lower job satisfaction and adverse health effects, and is most harmful when decision latitude is low. In manufacturing, production targets, shift work and machine-paced lines are typically heavy and limit flexibility in working hours and family responsibilities (Ganiyu, 2018). Competing and volatile demands add stress and complicate workload management (Gaim, 2018). In a Polish manufacturing and services sample, organizational demand was among the strongest predictors of work-life imbalance after controlling for individual factors. Hence:

H1: There is a significant relationship between organizational demand and work-life balance.

2.2. Coworker Support

Coworker support is the support, encouragement and social interaction workers receive from colleagues, and it buffers the negative impact of work demands. It is negatively related to role conflict and positively related to WLB and wellbeing. In manufacturing, where work depends on collaboration as much as on formal policy, peer relationships such as task coverage and shift swapping shape workers' sense of balance. Hence:

H2: Coworker support has a significant positive relationship with work-life balance.

2.3 Job Autonomy

Discretion over how, when and by what means work is done is a strong and reliable predictor of motivation, wellbeing and WLB. Autonomy is linked to self-efficacy (Saragih, 2011), to performance and involvement, especially among intrinsically motivated employees (Dysvik, 2011), and to psychological wellbeing (Clausen, 2022) and satisfaction in technical occupations (Zychová, 2024) it promotes positive attitudes and reduces stress. Perceived control over schedule and method predicted WLB among office workers with flexible arrangements (Bjärntoft, (2020)). In manufacturing, where assembly-line processes and managerial control limit independence, perceived autonomy may strongly affect WLB, consistent with evidence that quality of work life predicts manufacturing performance (Layer, 2009). Hence:

H3: Job autonomy has a significant positive effect on work-life balance.

2.4 Job Satisfaction

Job satisfaction is employees' overall evaluative response to their work and working context, including pay, conditions, relationships and the work itself (Judge, 2020). It correlates with productivity, absenteeism and intention to leave (Aziri, 2011), and cultural factors shape its experience in Asian work settings (Zhu, 2013). Satisfaction and WLB appear mutually interdependent in the Sri Lankan context (Chimote, 2013), and in manufacturing satisfaction is both a determinant and a result of WLB (Domagalska-Grędys, 2025). Where intrinsic rewards are lacking, recognition, promotion and positive relationships may matter more. Hence:

H4: Job satisfaction has a significant positive impact on work-life balance.

3 METHODOLOGY

The design of this study is quantitative, deductive and positivist for testing hypothesis statistically. The data was obtained efficiently and with little interviewer and social-desirability bias using a self-administered structured questionnaire. The management

conducted a survey of employees of two furniture-manufacturing companies (Company A and Company B), who were assured anonymity and confidentiality. The adapted instrument was based on literature and was rated on a 5-point Likert scale that included measures related to organizational demand, coworker support, job autonomy, job satisfaction and WLB. The number of valid responses (268) is more than adequate for regression with four predictors (10-20 per predictor). To analyze data, descriptive statistics, Cronbach's alpha, Shapiro-Wilk tests and Q-Q plot, Variance Inflation Factor and Tolerance, Pearson correlation were used, in R (version 4.3.x). Simple linear regression (β, t, p, R^2) was subsequently used to test each hypothesis followed by a multiple regression with all four predictors entered together (WLB = β0 + β1 Organizational Demand + β2 Coworker Support + β3 Job Autonomy + β4 Job Satisfaction + ε).

4 RESULTS AND DISCUSSION

Descriptive statistics (n = 268) show the highest mean for coworker support (4.25, range 1.67–5.00), followed by job autonomy (3.79, range 2.33–4.83), job satisfaction (3.71, range 2.00–4.80) and organizational demand (3.39, range 2.33–4.44). The mean WLB score was 2.91 (range 2.00–5.00, median 3.00) which was slightly below the scale midpoint. Reverse coding resulted in acceptable to excellent Cronbach's alphas (α = 0.94 for WLB; 0.74–0.88 for other scales). All variables were not normally distributed (Shapiro-Wilk), but were large enough for the Central Limit Theorem to allow for parametric regression (n = 268). The VIF values were close to 1.0 (between 1.12 and 1.19), which means there was not a concern of multicollinearity. Table 1 shows four simple regressions (H1–H4), all of which were confirmed.

Items of the demand scale were reverse coded, meaning that a higher score represents a lower level of demand, and

Table 1: Summary of Hypothesis Testing Results for Organizational Factors (Source: Author compiled). Note: organizational demand is reverse coded, so the positive β indicates that lower demand is associated with higher WLB.

H	Predictor	β	R^2	p	Result
H1	Organizational demand	0.042	0.053	<0.001	Negative Relationship
H2	Coworker support	0.449	0.239	<0.001	Positive Relationship
H3	Job autonomy	0.820	0.720	<0.001	Positive Relationship
H4	Job satisfaction	0.619	0.346	0.005	Positive Relationship

the positive coefficient for the organizational demand (H1: β = 0.042, R^2 = 0.053, $p < 0.001$) suggests that lower demands are associated with higher WLB. In the individual models, job autonomy had the largest coefficient (H3: β = 0.820, R^2 = 0.720, $p < 0.001$), followed by job satisfaction (H4: β = 0.619, R^2 = 0.346, p = 0.005) and coworker support (H2: β = 0.449, R^2 = 0.239, $p < 0.001$), showing that discretion over work, satisfaction and shop-floor solidarity are key resources for WLB. In the multiple regression (Table 2), WLB = −0.148 + 0.334 (Organizational Demand) + 0.323 (Coworker Support) + 0.127 (Job Autonomy) + 0.018 (Job Satisfaction). The model was significant ($F(4, 263) = 33.49$, $p < 0.001$) and explained 33.75% of the variance in WLB (R^2 = 0.338, adjusted R^2 = 0.327; residual standard error = 0.340). The majority of respondents' residuals were near zero (median = 0.023) with a longer positive tail (−0.928 to 1.673), indicating that some reported a significantly higher rating for WLB than that predicted.

Three predictors contributed significantly. Organizational demand (β = 0.334, SE = 0.060, t = 5.57, $p < 0.001$) and coworker support (β = 0.323, SE = 0.044,

Table 2: Multiple Linear Regression Predicting Work-Life Balance, n = 268 (Source: Author compiled). Note: β = unstandardized coefficient; $R^2 = 0.338$, adjusted $R^2 = 0.327$, $F(4, 263) = 33.49$, $p < 0.001$

Predictor	β	SE	t	p
(Intercept)	−0.148	0.297	−0.498	0.619
Organizational demand	0.334	0.060	5.570	<0.001
Coworker support	0.323	0.044	7.419	<0.001
Job autonomy	0.127	0.048	2.659	0.008
Job satisfaction	0.018	0.049	0.376	0.707

t = 7.42, p < 0.001) had the largest coefficients: a one-point increase on either scale (for demand, a higher reverse-coded score, i.e., lower demand) was associated with about a third of a point higher WLB. Job autonomy had a smaller significant effect (β = 0.127, SE = 0.048, t = 2.66, p = 0.008); job satisfaction was not significant (β = 0.018, SE = 0.049, t = 0.38, p = 0.707). H1–H3 therefore remained supported, but H4 did not once the other factors were controlled, suggesting that satisfaction does not explain WLB beyond demand, support and autonomy. Because the coefficients are unstandardized, relative sizes should be read with caution, and about two-thirds of the variance remains unexplained, pointing to other, including individual-level, determinants.

5 CONCLUSION AND RECOMMENDATIONS

Limitations include the cross-sectional design, two companies in one sub-sector, and the focus on organizational determinants. Future research should use longitudinal designs, add individual-level determinants, and cover other labor-intensive sub-sectors and developing countries.

The limitations are the cross-sectional design, two companies in one sub-sector and the focus on organizational determinants. The association of a higher reverse coded score (lower demand) with future scale was approximately 1/3 point higher WLB. Job autonomy had a smaller significant effect (β = 0.127, SE = 0.048, t = 2.66, p = 0.008); job satisfaction was not significant (β = 0.018, SE = 0.049, t = 0.38, p = 0.707). Thus, when the other factors were controlled, H1–H3 remained supported, but not H4, indicating that satisfaction is not sufficient to account for WLB beyond demand, support and autonomy. Relative sizes should be interpreted with caution because the coefficients are unstandardized, and approximately two-thirds of the variance is not accounted for, indicating that there are other factors, such as individual level factors, involved.

REFERENCES

- Aziri, B. (2011). Job satisfaction: A literature review. Management Research and Practice. 77-86. https://d1wqtxts1xzle7.cloudfront.net/51841101/f7-libre.pdf
- Bakker, A. B. ((2007)). The Job Demands-Resources model: State of the art. *Journal of Managerial Psychology*, 22(3), 309-328. https://www.emerald.com/jmp/article-abstract/22/3/309/236386
- Bjärntoft, S. ((2020)). Organizational and job-related determinants of work-life balance: A systematic review. *International Journal of Environmental Research and Public Health*, 17(4), 1418. https://www.mdpi.com/1660-4601/17/4/1418
- Chimote, N. K. (2013). Work-life balance among IT and manufacturing employees: A comparative study. . *Prabandhan: Indian Journal of Management*, 6(1), 34-42. https://ir.ucc.edu.gh/server/api/core/bitstreams/7f26c6cd-bcc5-4d6f-9a64-c12597341b10/content
- Clausen, T. (2022). Job autonomy and psychological well-being: A longitudinal study. Work & Stress. 36(1), 45-63.

https://www.tandfonline.com/doi/full/10.1080/1359432X.2021.1972973

- Domagalska-Grędys, M. (2025). Organizational support and employee work-life balance in production environments. Sustainability,. 17(3), 210. https://ejournal.45mataram.ac.id/index.php/economina/article/view/2683
- Dysvik, A. (2011). Intrinsic motivation as a moderator on the relationship between perceived job autonomy and work performance. *European Journal of Work and Organizational Psychology*, 20(3), 367-387. https://www.tandfonline.com/doi/abs/10.1080/13594321003590630
- Gaim, M. (2018). Competing demands and organizational performance. *A review. Journal of Business Research*, 88(2), 210-220. https://onlinelibrary.wiley.com/doi/full/10.1002/bse.2906
- Ganiyu, I. F. (2018). Work-life balance challenges in labour-intensive manufacturing industries. *Journal of Industrial Relations*, 60(2), 245-262. https://scholar.google.com/scholar?cluster=9252669477870101392&hl=en&as_sdt=0,5
- Geetha, R. (2023). A systematic literature review of work-life balance research in manufacturing sectors. Employee Relations. 45(1), 55-77.
- Imawa, K. (2022). Working conditions and employee wellbeing in Sri Lankan manufacturing firms. *South Asian Journal of Management*, 29(1), 112-129.
- Jayasundara, J. M. ((2023)). Challenges to work-life balance among Sri Lankan factory workers. *Sri Lanka Journal of Management Studies*, 5(2), 40-58.
- Judge, T. A. (2020). Job satisfaction: A conceptual review. *Journal of Organizational Behavior*, 41(3), 197-211.
- Kajenthiran, K. (2019). Individual determinants of work-life balance: Evidence from developing countries. *Asian Journal of Business Research*, 9(1), 1-18.
- Karasek, R. A. (1979). Job demands, job decision latitude, and mental strain. Administrative Science Quarterly. 24(2), 285-308. https://www.jstor.org/stable/2392498
- Layer, J. K. (2009). Quality of work life and perceived organizational performance in manufacturing. Human Factors and Ergonomics in Manufacturing. 19(5), 413-441.
- Morgeson, F. P. (2005). The work design questionnaire: Developing and validating a comprehensive measure. *Journal of Applied Psychology*, 91(6), 1321-1339.
- Rahim, N. B. (2020). Supervisor and family support as moderators of work-life balance and wellbeing. *Journal of Family Studies*, 26(2), 165-181.
- Saragih, S. (2011). The effects of job autonomy on work outcomes: Self efficacy as an intervening variable. *International Research Journal of Business Studies*, 4(3), 203-215.
- Shukla, A. (2016). Coworker support and its influence on work-life balance.. *The Journal of Business Perspective*, 20(3), 191-201.
- Zhu, Y. (2013). Job satisfaction in the Asian context: Cultural and universal determinants. . *Asia Pacific Journal of Management*, 30(4), 1101-1120.
- Zychová, K. (2024). Job autonomy and satisfaction in technical occupations. Central European Business Review. 13(1), 22-40.

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Impact of Social Influence on Investment Decisions: The Mediating Role of Financial Literacy Among Employees in Sri Lanka

Madhuwanthi MPDP[1], Karunarathne EACP[2]

[1,2]*Department of Industrial Management, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*

poornimadhuwanthi@gmail.com[1], chami@wyb.ac.lk[2]

ABSTRACT

Investment decisions are shaped not only by financial considerations but also by social and behavioural influences. The increasing availability of investment-related information through peers, social media, and financial experts highlights the importance of understanding how these influences affect investment decisions.This study investigates the impact of social influence on investment decisions among employees in Sri Lanka, with particular attention to the mediating role of financial literacy. This research focuses on three dimensions of social influence;peer pressure, social media influence, and expert opinion, while considering financial literacy as a mediator. A structured questionnaire was used to collect data from 260 employees aged 20 years and above from the public, private,self-employed and other employment sectors in Sri Lanka. Descriptive statistics, reliability analysis, correlation analysis, linear regression, and bootstrap mediation analysis were conducted using SPSS software to examine the relationships among the variables. The findings indicate that social influence significantly affects investment decisions and financial literacy, while financial literacy significantly influences investment decisions. Furthermore, financial literacy partially mediates the relationship between social influence and investment decisions.The findings highlight the importance of financial literacy in supporting informed investment decisions among employees in Sri Lanka.

KEYWORDS: Financial Literacy, Investment Decisions, Social Influence

1 INTRODUCTION

Investment is an important aspect of personal financial management, enabling individuals to achieve long-term financial goals. With the rapid development of digital technology, employees have access to investment-related information through peers, social media platforms, financial advisors, and investment communities. Social influence can shape investment choices through peer pressure, social media influence, and expert opinions (Dang, 2024).Previous studies indicate that social media and peer influence may encourage trend-following and herd behaviour (Junaidi & Nurhidayah, 2023; Rachakonda et al., 2024), while expert recommendations can provide information and reduce uncertainty (Joseph et al., 2025).

Financial literacy enables individuals to understand financial information, evaluate investment alternatives, and make informed decisions. Studies have shown that financially literate individuals are confident and independent in investment decisions (Peiris, 2021; Darwish, 2025). However, limited research has examined employees in Sri Lanka and the mediating role of financial literacy. Tennekoon et al. (2021) and Kumari (2020) found that financial literacy significantly influences investment decisions in Sri Lanka. However, limited research has examined the mediating role of financial literacy in the relationship

between social influence and investment decisions among Sri Lankan employees.. Therefore, this study investigates the impact of social influence on investment decisions among employees in Sri Lanka while examining the mediating role of financial literacy.

2 LITERATURE REVIEW

2.1 Social Influence

Social influence refers to the extent to which individuals' beliefs and behaviours are shaped by others in their social environment (Dang, 2024). Ali-Rind et al. (2023) highlighted the importance of peer effects in financial decision-making, particularly through herding and imitation behaviours. Similarly, Dang (2024) identified peer pressure, social media influence, and expert opinions as factors affecting investment decisions. These findings demonstrate the importance of social influence in understanding individuals' investment choices and financial decision-making behaviour.

2.2 Peer Pressure

Peer pressure occurs when individuals are encouraged to make investment choices similar to those of their friends or colleagues. It can contribute to herd behaviour, where investors imitate others' decisions to gain social approval or avoid missing opportunities (Bajaj & Kaur, 2025). Such behaviour may influence employees to adopt investment strategies without adequately evaluating risks and returns (Dang, 2024).

2.3 Social Media Influence

Social media has become an important source of investment information and advice. Platforms such as Facebook, YouTube, and LinkedIn enable individuals to access investment-related content and observe market trends (Joseph et al., 2025; Rachakonda et al., 2024). Frequent exposure to such content may influence investment decisions and encourage investors to follow trending information rather than conduct independent analysis (Rachakonda et al., 2024).

2.4 Expert Opinion Influence

Expert opinions from financial analysts, advisors, and commentators can influence investment decisions by providing information, increasing confidence, and reducing uncertainty (Joseph et al., 2025). However, excessive reliance on expert recommendations may limit investors' independent evaluation of investment alternatives.

2.5 Financial Literacy

Financial literacy refers to the knowledge and skills required to understand financial products, evaluate risks, and make informed financial decisions (Peiris, 2021). Higher financial literacy enables individuals to critically evaluate investment information and reduces their vulnerability to external influences (Bajaj & Kaur, 2025).

2.6 Investment Decisions

Investment decision-making involves selecting investment alternatives based on expected returns, risk tolerance, and personal financial objectives. Behavioral finance suggests that investment decisions can be influenced by social and emotional factors, including herding behaviour (Bajaj & Kaur, 2025; Rachakonda et al., 2024).

Most studies explored social influence and financial literacy among students or general investors in foreign contexts. However, this study focuses on employees in Sri Lanka and tests

financial literacy as a mediating mechanism. Thus, this study addresses gaps in both population and methodology by targeting a specific employee group in an emerging economy.

3 METHODOLOGY

3.1 Conceptual Framework

Social influence is conceptualized as a multidimensional independent variable comprising peer pressure, social media influence, and expert opinion. These dimensions are considered components of the overall social influence construct, which is examined in relation to investment decisions and financial literacy. The framework retains financial literacy as the mediating variable between social influence and investment decisions.

3.2 Hypothesis Development

Four hypotheses were tested to examine the direct relationships among the study variables and the mediating role of financial literacy:

H1: Social influence has a significant impact on investment decisions

H2: Social influence has a significant impact on Financial Literacy

H3: Financial Literacy has a significant impact on investment decisions

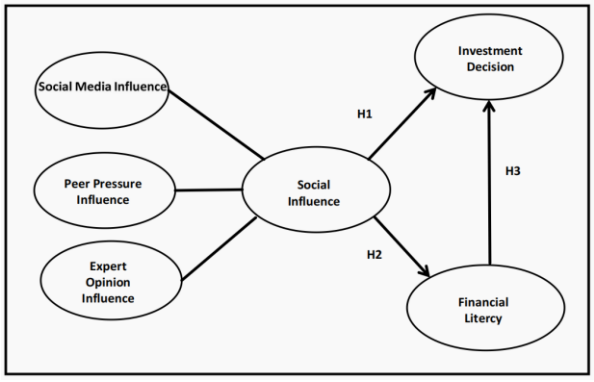


Figure 01: Conceptual Framework

H4: Financial Literacy mediates the relationship between Social Influence and Investment Decision.

3.3 Data Collection

The data were collected using a structured questionnaire from employees aged 20 years and above in Sri Lanka, including public-sector employees, private-sector employees, self-employed individuals, and in other sectors. Data collection was conducted through an online survey and the survey was distributed using social media (Facebook, WhatsApp) and LinkedIn.

3.4 Data Analysis

The collected data were coded, screened, and analysed using IBM SPSS Statistics. A total of 260 valid responses were used for the analysis. Reliability of the measurement scales was assessed using Cronbach's alpha, while descriptive statistics were used to summarise the study variables. Factor analysis was conducted to assess the suitability of the data, followed by Pearson correlation analysis to examine the relationships among Social Influence, Financial Literacy, and Investment Decision. Regression analysis was then performed to test the direct relationships proposed in H1, H2, and H3. Finally, bootstrap mediation analysis was conducted to examine the mediating role of Financial Literacy in the relationship between Social Influence and Investment Decision (H4). The mediation analysis used bootstrapping with 95% confidence intervals.

4 RESULTS AND DISCUSSION

All measurement scales demonstrated acceptable internal consistency, with Cronbach's alpha values ranging from

Table 1: Reliability and Descriptive Statistics

Variable	Cronbach's α	M	SD
Social media Influence	0.783	2.519	0.674
Peer pressure Influence	0.776	3.138	0.612
Expert opinion Influence	0.820	3.737	0.671
Financial Literacy	0.951	3.831	0.634
Investment Decision	0.857	3.893	0.471

Table 2: Regression Analysis

Hypothesis	Relationship	R	R^2	F	p
H1	Social Influence→ Investment Decision	0.659	0.434	197.960	0.0
H2	Social Influence→ Financial Literacy	0.551	0.304	112.748	0.0
H3	Financial Literacy→ Investment Decision	0.735	0.540	302.353	0.0

0.776 to 0.951. Descriptive statistics are presented in Table 1.

Overall, the results indicate relatively high levels of Financial Literacy and Investment Decision among the respondents.

The regression results indicate that Social Influence significantly predicted Investment Decision, explaining 43.4% of its variance ($R^2 = 0.434$, $F = 197.960$, $p < 0.001$). Therefore, H1 was supported. Social Influence also significantly predicted Financial Literacy, explaining 30.4% of its variance ($R^2 = 0.304$, $F = 112.748$, $p < 0.001$), supporting H2. Financial Literacy significantly predicted Investment Decision and explained 54.0% of its variance ($R^2 = 0.540$, $F = 302.353$, $p < 0.001$), supporting H3.

These findings suggest that social factors are associated with employees' investment decisions, while financial literacy is also strongly related to investment decision-making

Table 3: Correlation Between Variables

Variable	Financial Literacy	Investment Decision	Social Influence
Financial Literacy	1.000	0.735	0.551
Investment Decision	0.735	1.000	0.659
Social Influence	0.551	0.659	1.000

To examine whether Financial Literacy mediates the relationship between Social Influence and Investment Decision, a bootstrap mediation analysis was conducted.

The indirect effect of Social Influence on Investment Decision through Financial Literacy was 0.2734, with a 95% bootstrap confidence interval of [0.1149, 0.4766]. Since the confidence interval did not include zero, the indirect effect was significant. The direct effect of Social Influence on Investment Decision remained significant after Financial Literacy was included in the model ($\beta = 0.336$, $p < 0.001$). Therefore, Financial Literacy partially mediates the relationship between Social Influence and Investment Decision, supporting H4. The findings indicate that financial literacy represents an important pathway through which social influence is associated with investment decision-making, while social influence also has a direct effect on investment decisions.

5 CONCLUSION

This study examined the impact of social influence on investment decisions among employees in Sri Lanka and investigated the mediating role of

Table 4: Mediation Analysis

Hypothesis	Total Effect	Direct Effect	Indirect Effect	CI(95%)
H4	0.6124	0.3390	0.2734	[0.1149,0.4766

financial literacy. The findings revealed that social influence significantly affects investment decisions and financial literacy, while financial literacy also has a significant positive effect on investment decisions. Furthermore, financial literacy partially mediates the relationship between social influence and investment decisions, indicating that it strengthens employees' ability to evaluate socially influenced investment information. The study highlights the importance of developing financial literacy among employees to support informed and independent investment decisions. These findings provide useful implications for financial education and investment decision-making in the Sri Lankan context.

Future research could use different samples and larger populations to improve the reliability of findings. Longer study periods and additional behavioral factors could also provide deeper insights into the relationship between social influence, financial literacy, and investment decisions.

REFERENCES

- Bajaj, S., & Kaur, B. (2025). Herding in the modern financial landscape: An empirical exploration of behavioural finance dynamics. International Journal of Science and Management Studies, 8(3), 259–264. https://doi.org/10.51386/25815946/ijsms-v8i3p122
- Dang, P. G. (2024). The impact of social influences on investment decision-making: A multi-method analysis of Vietnamese investors. Journal of Economics, Finance and Management Studies, 7(12), 6978–6989.https://doi.org/10.47191/jefms/v7-i12-05
- Darwish, F. S. (2025). Financial literacy and investment decision: An empirical study from the Palestine Stock Exchange. Frontiers in Behavioral Economics, 4. https://doi.org/10.3389/frbhe.2025.1444022
- Junaidi, J., & Nurhidayah, N. (2023). Social media impact on trading behavior: An examination among Indonesian young adult investors with capital market literacy as a mediator. JEMA: Jurnal Ilmiah Bidang Akuntansi Dan Manajemen, 20(1), 136–155. https://doi.org/10.31106/jema.v20i1.19687
- Kumari, D. A. T. (2020). The impact of financial literacy on investment decisions: With special reference to undergraduates in Western Province, Sri Lanka. Asian Journal of Contemporary Education, 4(2)110–126. https://doi.org/10.18488/journal.137.2020.42.110.126
- Peiris, T. U. I. (2021). Effect of financial literacy on individual savings behavior The mediation role of intention to saving. European Journal of Business & Management Research, 6(5), 94–99. https://doi.org/10.24018/ejbmr.2021.6.5.1064
- Rachakonda, N., Kishore Kumar, C., Balivada, S. P., Mahesh, K., & Paparao, A. (2024). The influence of social media on investment decisions: A behavioural finance perspective in Hyderabad City. Journal of Information Systems Engineering & Management, 2025(28s).
- Rind, A. A., Boubaker, S., & Jarjir, S. L. (2023). Peer effects in financial economics: A literature survey. Research in International Business & Finance, 64, Article 101873. https://doi.org/10.1016/j.ribaf.2022.101873
- Tennekoon, S. T. M. S., & Liyanage, C. (2021). Impact of financial literacy levels among Sri Lankan investors on investment choices. South Asian Journal of Finance, 1(2), 123–147. https://doi.org/10.4038/sajf.v1i2.35
- Utami, E. M., Yuliani, R., & Pesakovic, G. (2025). Financial knowledge and social influence on Generation Z intention to invest: The mediating role of financial attitude and literacy. Media Ekonomi Dan Manajemen, 40, 121–147.

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Effect of Sustainable Development Goals on Tourism Growth: Evidence from Asian Developed Economies

Jayathunga PDR[1], Deegahawature MMDR[2]
[1,2]*Department of Industrial Management, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*
desinijayathunga@gmail.com[1], dharsana@wyb.ac.lk[2]

ABSTRACT

The achievement of Sustainable Development Goals (SDGs) is increasingly important for promoting sustainable tourism development; however, limited empirical evidence examines how SDG-related economic, social, and environmental dimensions influence tourism growth in developed Asian economies. Therefore, this study examines the impact of selected SDG-related indicators on tourism growth in developed Asian economies over 1998–2020, using secondary data. GDP represents the economic dimension; HDI and unemployment represent the social dimension; while CO_2 emissions and renewable energy consumption represent the environmental dimension. Using panel econometric techniques, the Westerlund cointegration test confirmed a long-run relationship among the variables, followed by the ARDL-PMG approach to estimate short- and long-run effects. GDP positively affects tourism growth in the short run, while CO_2 emissions have a significant positive long-run effect. Unemployment negatively affects tourism growth in both the short and long run, while HDI is significant only in the long run, whereas renewable energy consumption is significant only in the short run. The Dumitrescu–Hurlin test indicates bidirectional causality between GDP and tourism growth, supporting both the Tourism-Led Growth and Growth-Led Tourism hypotheses.

KEYWORDS: Panel Econometric Analysis, Sustainable Development Dimensions, Sustainable Development Goals, Tourism Growth

1 INTRODUCTION

Tourism is a major global sector with significant economic, social, and environmental impacts. Asia and the Pacific account for nearly 30% of global tourism receipts (WTO, 2021), while tourism is closely linked to the achievement of the United Nations' 17 Sustainable Development Goals (SDGs) (UNWTO, 2016). Empirical studies have also identified significant relationships between tourism, economic growth, and sustainability dimensions (Wu & Wu, 2019; Wani et al., 2024). Developed Asian economies provide an important context for this study due to their mature tourism markets, high human development, and advanced economic and infrastructure conditions. However, limited evidence examines the combined effects of SDG-related economic, social, and environmental dimensions on tourism growth in these economies, particularly in terms of short-run, long-run, and causal relationships.

Therefore, this study examines the effects of selected SDG-related indicators on tourism growth in developed Asian economies during 1998–2020. GDP represents the economic dimension; HDI and unemployment represent the social dimension; and CO_2 emissions and renewable energy consumption represent the environmental dimension. Panel econometric techniques are employed to examine their short- and long-run effects and causal relationships.

2 LITERATURE REVIEW

Sustainable development integrates economic, social, and environmental dimensions, while the Sustainable Development Goals (SDGs) provide a global framework for achieving sustainable development. Tourism is closely linked to these dimensions through its contribution to economic activity, employment, human development, and environmental outcomes (UNWTO, 2016).

International tourism receipts are widely used to measure tourism growth (Wani et al., 2024; Wu & Wu, 2019). GDP represents the economic dimension and is associated with the Tourism-Led Growth (TLG) and Growth-Led Tourism (GLT) hypotheses. Empirical studies have identified significant relationships between tourism and economic growth (Wu & Wu, 2019; Wani et al., 2024). CO_2 emissions and renewable energy consumption represent the environmental dimension, reflecting the relationship between tourism, energy use, and environmental quality (Paramati et al., 2017). HDI and unemployment represent the social dimension, reflecting human development and employment conditions that can influence tourism performance (Nunkoo et al., 2019; Wani et al., 2024). The variables were selected based on their extensive use in prior studies, enabling comparison of results. However, limited studies have jointly examined these economic, social, and environmental SDG-related dimensions and their short-run, long-run, and causal effects on tourism growth in developed Asian economies. This study addresses this gap using panel econometric techniques.

3 METHODOLOGY

This study examines the short- and long-run effects of selected economic, social, and environmental SDG-related indicators on tourism growth in five developed Asian economies-Singapore, Japan, South Korea, Israel, and Hong Kong-using annual data for 1998–2020, based on data availability.

Tourism growth is measured by international tourism receipts (TR) in current US dollars. The economic dimension is represented by GDP (current US$). The social dimension is represented by the Human Development Index (HDI) and unemployment rate (UNEMP) (% of the labour force). The environmental dimension is represented by CO_2 emissions per capita (CO_2) (metric tons) and renewable energy consumption (REN) (% of total final energy consumption). Data were obtained from the World Bank, International Labour Organization (ILO), International Energy Agency (IEA), and United Nations Development Programme (UNDP), according to the respective indicators.

The study hypothesizes that GDP, HDI, and REN positively affect tourism growth, while CO_2 emissions and unemployment negatively affect tourism growth in both the short and long run. The variables were transformed into logarithmic form for econometric analysis.

4 RESULTS AND DISCUSSION

Descriptive statistics were first conducted to summarize the characteristics of the variables. The correlation analysis revealed significant relationships among the study variables. Tourism receipts (TR) were positively

Table 1: Descriptive Statistics

Variable	Mean	Std. Dev.	Skewness	Kurtosis
TR	22.87	0.89	-0.04	2.32
GDP	26.47	1.75	-0.01	2.34
CO2Emit	2.14	0.24	-0.37	2.24
HDI	-0.12	0.04	-0.58	2.64
UNEMP	1.59	0.44	0.99	3.27
REN	0.49	1.45	-0.55	2.09

correlated with GDP ($r = .653$, $p < .001$), HDI ($r = .648$, $p < .001$), and CO_2 emissions ($r = .216$, $p < .01$), while negative correlations were observed with unemployment ($r = -.634$, $p < .001$) and renewable energy consumption (REN) ($r = -.434$, $p < .001$). GDP was positively associated with CO_2 emissions ($r = .541$, $p < .001$) and HDI ($r = .455$, $p < .001$), but negatively associated with unemployment ($r = -.463$, $p < .001$). CO_2 emissions were negatively correlated with unemployment ($r = -.316$, $p < .001$) and positively correlated with REN ($r = .190$, $p < .01$), whereas HDI was negatively related to unemployment ($r = -.381$, $p < .001$). Unemployment and REN also showed a positive correlation ($r = .361$, $p < .001$). Overall, the correlations indicate meaningful associations among tourism, economic, social, and environmental dimensions, with no exceptionally high correlation among the explanatory variables.

Multicollinearity results (Table 2) indicate no severe multicollinearity issue among variables.

Pesaran et al. (2004) CD test in Table 3 rejects the null hypothesis, confirming significant cross-sectional dependence among countries.

Table 2: Multicollinearity Results

Variable	VIF	1/VIF
GDP	1.97	0.5071
UNEMP	1.70	0.5889
CO_2 Emit	1.63	0.6149
HDI	1.41	0.7114
REN	1.32	0.7595

Table 3: CD Test Results

Variable	CD-test	P-value
TR	12.127	0.000
GDP	11.735	0.000
UNEMP	6.630	0.000
HDI	14.853	0.000
CO_2Emit	1.923	0.055
REN	3.687	0.000

The Pesaran–Yamagata test confirmed significant slope heterogeneity, with a Delta statistic of 7.57 and an adjusted Delta statistic of 9.07, both with p values below .001, rejecting slope homogeneity across countries.

Accordingly, second-generation unit root tests- CIPS, CADF, and Herwartz and Siedenburg (2008)-were employed.

As shown in Table 6, the variables are non-stationary at level but become stationary at first difference, confirming that all variables are I (1). Based on the Westerlund (2007) cointegration test, cointegration is confirmed at the 5% significance level (Table 4)

Therefore, the ARDL-PMG approach is appropriate for examining the long-run and short-run relationships among the variables (Table 5)

Table 4: Westerlund Cointegration Test

Statistic	Z-value	Robust P-value
Gt	-2.652	0.020
Ga	2.158	0.160
Pt	-1.661	0.040
Pa	1.207	0.100

Table 6: Stationarity Test Results

Variable	CIPS		CADF		Herwartz & Siedenburg	
	Level	First Difference	Level	First Difference	Level	First Difference
TR	−1.497	−3.800***	−1.977	−2.398*	0.5518	−2.4719***
GDP	−2.763	−3.741***	−1.571	−2.950***	0.433	−2.3950***
HDI	−2.140	−3.931***	−2.836***	−3.272***	0.0043	−1.9567**
REN	−2.008	−3.748***	−1.881	−2.736**	0.5323	−1.8376**
UNEMP	−1.888	−2.886***	−2.255	−2.625**	−0.5894	−2.0598**
CO_2Emit	−2.200	−5.257***	−1.516	−2.710**	0.7461	−1.6310*

Note. * p < .10, **p < .05, and ***p < .01

Table 5: PMG Test Result.

Short- Run Coefficients (SR)	
ECT(t-1)	-0.396**
Δ GDP	1.235**
Δ CO_2 Emit	0.456
Δ HDI	1.416
Δ UNEMP	-0.628*
Δ REN	0.177*
Long-Run Coefficients	
GDP	-0.339
CO_2 Emit	4.464***
HDI	8.882***
UNEMP	-1.075***
REN	-0.143

The Error Correction Term (ECT) is negative and significant (-0.396), confirming the existence of long-run equilibrium, with 39.6% of deviations from the equilibrium corrected annually. Therefore, when the variables deviate from their long-run equilibrium.

GDP has a positive and significant short-run effect but no significant long-run effect, with a 1% increase in GDP raising tourism receipts by 1.235%, possibly reflecting the characteristics of mature tourism markets (Eugenio-Martin et al., 2004).

CO_2 emissions have a positive and significant long-run effect, with a 1% increase associated with a 4.464% increase in tourism receipts, indicating the link between tourism expansion and carbon-intensive activities (Wangzhou et al., 2022). Renewable energy consumption (REN) has a weakly significant positive short-run effect, with a 1% increase associated with a 0.17% increase in tourism receipts.

HDI has a positive and significant long-run effect, with a 1% increase associated with an 8.882% increase in tourism receipts, highlighting the role of human development in tourism competitiveness (Goffi et al., 2019). Unemployment negatively affects tourism receipts in both the long and short run, with a 1% increase reducing tourism receipts by approximately 1.075% and 0.62%, respectively.

The Dumitrescu–Hurlin causality (2012) test results in the Table 7 indicates bidirectional causality between tourism receipts and economic sustainability, while the other sustainability dimensions exhibit unidirectional causality. These findings support both the Tourism-Led Growth and Growth-Led Tourism hypotheses, indicating a mutually reinforcing relationship between tourism and economic development).

Table 7: Causality Test Resul

Direction	Statistics	Causality
TR → GDP	2.6212**	Exists
GDP → TR	4.2618***	Exists
TR → HDI	0.2072	Does not exist
HDI → TR	3.3099***	Exists
TR → REN	3.8951***	Exists
REN → TR	−1.4413	Does not exist
TR → CO_2 Emit	1.5484	Does not exist
CO_2 Emit → TR	3.5845***	Exists
TR → UNEMP	5.5840***	Exists
UNEMP → TR	1.1901	Does not exist

Note. * p < .10, **p < .05, and ***p < .01.

5 CONCLUSION

According to the findings of the study, the impact of Sustainable Development Goals (SDGs) on tourism growth in developed Asian countries varies depending on the time horizon considered, particularly between the short run and the long run. Therefore, policymakers should take these time-dependent effects into consideration when formulating and implementing policies related to SDGs and tourism development, recognising that the outcomes of such policies may vary over time. Furthermore, future research could extend the scope of this study by examining whether the relationship between SDGs and tourism growth differs across Asian economies at different levels of development.

Limitations, including geographical scope, sample size, and the number of variables considered, suggest opportunities for future studies

REFERENCES

- Dumitrescu, E.-I., & Hurlin, C. (2012). Testing for Granger non-causality in heterogeneous panels. Economic Modelling, 29(4), 1450–1460.
- Eugenio-Martín, J. L., Morales, N. M., & Scarpa, R. (2004). Tourism and economic growth in Latin American countries: A panel data approach. SSRN Electronic Journal.
- Goffi, G., Cucculelli, M., & Masiero, L. (2019). Fostering tourism destination competitiveness in developing countries: The role of sustainability. Journal of Cleaner Production, 209, 101–115.
- Herwartz, H., & Siedenburg, F. (2008). Homogeneous panel unit root tests under cross sectional dependence: A Monte Carlo comparison. Computational Statistics & Data Analysis, 52(10), 5115–5129.
- Nunkoo, R., Seetanah, B., Jaffur, Z. R. K., & Ramkissoon, H. (2019). Tourism and economic growth: A meta-regression analysis. Journal of Travel Research.
- Paramati, S. R., Shahbaz, M., & Alam, M. S. (2017). Does tourism degrade environmental quality? A comparative study of Eastern and Western European Union. Transportation Research Part D: Transport and Environment, 50, 1–13.
- Pesaran, m. h., schuermann, T., & Weiner, s. m. (2004). modeling regional interdependencies using a global error-correcting macroeconometric model. Journal of Business & Economic Statistics, 22(2), 129–162.
- Pesaran, M. H., & Yamagata, T. (2008). Testing slope homogeneity in large panels. Journal of Econometrics, 142(1), 50–93.
- United Nations Development Programme. (2022). Human Development Report 2021/2022: Uncertain Times, Unsettled Lives: Shaping Our Future in a Transforming World. United Nations Development Programme.
- UNWTO. (2016). Tourism and the Sustainable Development Goals. World Tourism Organization (UNWTO).
- Wangzhou, K., Wen, J. J., Wang, Z., Wang, H., Hao, C., & Andlib, Z. (2022). *Revealing the nexus between tourism*

development and CO_2 emissions in Asia: Does asymmetry matter? Environmental Science and Pollution Research, 29(52), 79016–79024.

- Wani, M. J. G., Loganathan, N., & Mujalli, A. (2024). The impact of sustainable development goals (SDGs) on tourism growth: Empirical evidence from G-7 countries. Cogent Social Sciences, 10(1), 2397535.
- Westerlund, J. (2007). Testing for error correction in panel data. Oxford Bulletin of Economics and Statistics, 69(6), 709–748.
- World Tourism Organization. (2021). International Tourism Highlights, 2021 Edition. World Tourism Organization (UNWTO).
- Wu, T.-P., & Wu, H.-C. (2019). Tourism and economic growth in Asia: A bootstrap multivariate panel Granger causality. International Journal of Tourism Research, 21(1), 87–96.

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

How Consumers Value Traditional Drawings in Sri Lanka: The Role of Art-Related Characteristics in Perceived Value

Ariyawansha GVADD[1], Karunarathne EACP[2]

[1,2]*Department of Industrial Management, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*

dinudula2020@gmail.com[1], *chami@wyb.ac.lk*[2]

ABSTRACT

Traditional drawings continue to hold value as a form of artistic expression, reflecting creativity, craftsmanship, cultural meaning, and individual artistic skills. However, limited attention has been given to understanding the factors that influence consumers' perceived value of traditional drawings, particularly the role of individual art-related characteristics. Therefore, this study aimed to examine the factors influencing customers' perceived value of traditional drawings, focusing on personality traits, art interest, art knowledge, and art involvement. The study also examined whether perceived value differed across selected demographic characteristics. A quantitative research approach was adopted, and data were collected from 321 respondents using a structured questionnaire. The collected data were analyzed using descriptive statistics, reliability and validity analysis, exploratory factor analysis, correlation analysis, multiple regression analysis, an independent-samples t-test, and one-way ANOVA. The findings indicated that art knowledge and art involvement had significant positive effects on the perceived value of traditional drawings, whereas personality traits and art interest did not. Among the examined factors, art knowledge had the strongest influence on customer-perceived value. The regression model was statistically significant and explained 10.1% of the variance in perceived value. The findings suggest that consumers' evaluation of traditional drawings is more strongly associated with their art-related knowledge and involvement than with general personality traits or art interest. The study provides useful insights for artists, galleries, and art-related businesses in developing strategies to enhance the perceived value of traditional drawings. Future research may examine additional factors such as authenticity, aesthetic preferences, artistic quality, emotional attachment, artists' reputations, and cultural influences.

KEYWORDS: Art Interest, Art Involvement, Art Knowledge, Customer Perceived Value, Personality Traits, Traditional Drawing

1 INTRODUCTION

Art is a major form of human expression that communicates creativity, culture, identity, and emotion. Traditional drawings continue to hold value among consumers due to their craftsmanship, tangible nature, and perceived authenticity. Customer perceived value (CPV) refers to consumers' overall evaluation of a product based on its perceived benefits and costs (Sweeney & Soutar, 2001). In the context of art, these benefits may include aesthetic, emotional, social, and symbolic value.

Previous studies suggest that art-related evaluations are influenced by individual characteristics. Art knowledge enhances the interpretation of artistic features and aesthetic judgments (Leder et al., 2004), while art involvement encourages attention and emotional engagement (Holbrook & Hirschman, 1982; Zaichkowsky, 1994). Art interest also influences responses to visual art (Hüttl-Maack, 2018), while personality characteristics, particularly openness to experience, are associated with aesthetic

preferences (Afhami & Mohammadi-Zarghan, 2018; Fayn et al., 2018).

However, limited research has examined how these factors collectively influence CPV of traditional drawings. Existing studies have largely focused on aesthetic preferences, art appreciation, and purchase intentions rather than CPV. Accordingly, this study was designed to examines the influence of art interest, art knowledge, art involvement, and personality traits on CPV of traditional drawings. Further, study targeted to investigate differences across demographic groups, and provide recommendations for artists and other relevant stakeholders.

2 LITERATURE REVIEW

CPV refers to the consumer's overall evaluation of the benefits received from a product relative to the sacrifices involved. Sweeney and Soutar (2001) established CPV as a multidimensional construct encompassing functional, emotional, and social value dimensions. In art consumption, value can extend beyond utilitarian benefits to include aesthetic experience, authenticity, emotional meaning, and symbolic significance. Accordingly, consumers may assign different levels of value to an artwork even when its physical characteristics are similar.

Art interest represents an individual's level of attraction, curiosity, and engagement toward art. Hüttl-Maack (2018) found that consumers' art interest can influence responses to visual art, including value-for-money perceptions and willingness to pay. Afhami and Mohammadi-Zarghan (2018) also reported associations between art interest, personality characteristics, and aesthetic judgment styles. These findings suggest that consumers who are more interested in art may attend more closely to artistic qualities and therefore perceive greater value.

On the other hand, art knowledge reflects an individual's ability to understand artistic techniques, styles, concepts, and contextual meanings. The model of aesthetic appreciation proposed by Leder et al. (2004) emphasizes the role of knowledge and cognitive processing in aesthetic evaluation. Greater knowledge may allow consumers to recognize technical skill, composition, symbolism, and craftsmanship that less knowledgeable viewers may overlook. Consequently, art knowledge is expected to positively contribute to the perceived value of traditional drawings.

Involvement towards art refers to the degree of personal relevance and psychological engagement an individual has with art. Experiential consumption research emphasizes that involvement can increase attention, feelings, and personal meaning during consumption (Holbrook & Hirschman, 1982). Zaichkowsky (1994) similarly conceptualized involvement as the perceived relevance of an object to an individual's needs and values. In traditional drawing contexts, greater involvement may increase appreciation for the craftsmanship and effort invested in creating the work.

Personality traits, final factor considered in this study, implies how individuals respond emotionally and cognitively to aesthetic stimuli. Research on aesthetic experience has frequently identified openness to experience as an important predictor of interest in novelty, complexity, and artistic experiences (Fayn et al., 2018; Pelowski et al., 2017).

Afhami and Mohammadi-Zarghan (2018) also identified relationships among Big Five characteristics, aesthetic judgment styles, and art interest. Thus, personality traits may influence how consumers interpret and value traditional drawings.

3 METHODOLOGY

A quantitative, explanatory, cross-sectional research design was employed. The target population consisted of individuals with an interest in or exposure to art-related activities. Because a complete sampling frame was unavailable, convenience sampling was used. A total of 321 usable responses were collected using a structured questionnaire with five-point Likert-scale statements ranging from strongly disagree to strongly agree. CPV was measured using six items adapted from Sweeney and Soutar (2001). Art interest was measured using two items adapted from Cotter et al. (2023), art knowledge using nine items informed by An et al. (2021), and art involvement using four involvement items adapted from Zaichkowsky (1994). Personality traits was assessed using the Ten-Item Personality Inventory (Gosling et al., 2003), a brief measure of the Big Five dimensions. The study analyzes personality as a single composite variable rather than examining each dimension separately

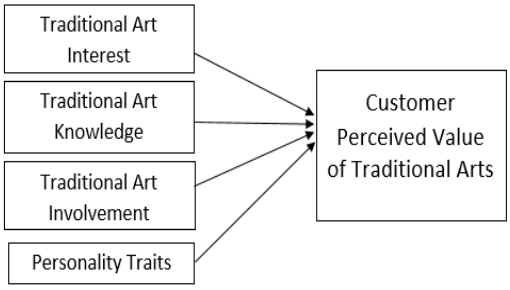


Figure 1: Conceptual Framework

Data were analyzed using SPSS software and descriptive statistics were used to summarize the sample profile. Cronbach's alpha assessed internal consistency, while exploratory factor analysis (EFA) using principal component analysis with Varimax rotation examined the factor structure. Pearson correlation was used to examine bivariate relationships, and multiple linear regression was employed to determine the unique effects of the four predictors on CPV. Independent-samples t-test and one-way ANOVA were used to examine differences in CPV across gender, age, education, and income.

4 RESULTS AND DISCUSSION

4.1 Respondent Profile and Descriptive Statistics

The sample consisted of 154 females (47.98%) and 167 males (52.02%). The largest age group was 25-34 years (45.48%), followed by 18-24 years (41.74%). Respondents with no income accounted for 28.97% of the sample, followed by 25.55% with earning LKR 50,000–100,000 and 22.43% earning below LKR 50,000. The mean CPV for traditional drawings was 3.74 (SD = 0.29), indicating a generally positive evaluation. The mean values for art interest, art knowledge, and art involvement were 3.59, 3.32, and 3.52, respectively.

Table 1: Descriptive Statistics of Study Variables

Construct	N	Mean	SD
Personality Traits	321	3.0991	.55258
Art Interest	321	3.5903	.58731
Art Knowledge	321	3.3205	.38853
Art Involvement	321	3.5234	.55361
CPV	321	3.7440	.29490

Table 2: Reliability and Exploratory Factor Analysis Summary

Construct / Test	Items	Cronbach's α
Personality Traits	10	.711
Art Interest	2	.701
Art Knowledge	9	.699
Art Involvement	4	.793
CPV	6	.695
KMO	—	.809
Bartlett's test	—	$p < .001$

4.2 Reliability and Validity Analysis

Reliability analysis indicated generally acceptable to marginal internal consistency across the constructs. Art involvement showed the highest reliability, while CPV and art knowledge were slightly below the commonly used benchmark, but were very close to .70.

EFA produced a five-factor solution explaining 43.276% of the total variance. The KMO value of 0.809 indicated good sampling adequacy, and Bartlett's Test of Sphericity was significant ($\chi^2 = 2389.213$, $df = 465$, $p < .001$), confirming that the correlation matrix was suitable for factor analysis. The results provided reasonable evidence of construct structure for subsequent analysis.

4.3 Correlation Analysis

Pearson correlation analysis showed that CPV had significant positive relationships with art knowledge ($r = .287$, $p < .001$), art involvement ($r = .274$, $p < .001$), and art interest ($r = .107$, $p = .028$). Personality traits showed a weak negative and non-significant correlation with CPV ($r = -.052$, $p = .177$). These results indicate that knowledge and involvement were more strongly associated with CPV than interest or personality traits. Because correlations do not control for the simultaneous influence of other predictors, multiple regression was conducted.

4.4 Multiple Regression Analysis

The multiple regression model was statistically significant, $F(4,316) = 8.854$, $p < .001$. The model produced $R = .317$, $R^2 = .101$, and adjusted $R^2 = .089$, indicating that the four predictors jointly explained 10.1% of the variance in CPV (8.9% after adjustment). Art knowledge was the strongest significant predictor ($\beta = .189$, $p = .005$), followed by art involvement ($\beta = .159$, $p = .016$). Art interest ($\beta = .036$, $p = .507$) and personality traits ($\beta = .017$, $p = .756$) were not statistically significant.

4.5 Customer Perceived Value and Demographics

Art knowledge and art involvement had significant positive effects on CPV, supporting Leder et al. (2004) and Holbrook and Hirschman (1982), who highlighted the roles of knowledge and experiential engagement in aesthetic evaluation. These findings are also consistent with Hernando and Campo (2017) and Spee et al. (2022).

Although art interest was positively correlated with CPV, it had no significant unique effect in the regression model. Similarly, personality traits were non-significant, suggesting that art-specific knowledge and involvement may be more

Table 3: Multiple Regression Results for CPV

Predictor	β	p-value	VIF
Personality Traits	.017	.756	1.053
Art Interest	.036	.507	1.061
Art Knowledge	.189	.005	1.603
Art Involvement	.159	.016	1.529
R^2	.101	—	—
Adjusted R^2	.089	—	—
F(4,316)	8.854	< .001	—

influential than general interest and personality characteristics (Hüttl-Maack, 2018; Afhami & Mohammadi-Zarghan, 2018; Fayn et al., 2018).

The regression model was significant, explaining 10.1% of CPV variance (adjusted R^2=8.9%). Its modest explanatory power may reflect other unexamined factors, including authenticity, artistic quality, emotional value, and artist reputation (Pelowski et al., 2017; Spee et al., 2022).

Regarding demographics, males reported slightly higher CPV (M = 3.78) than females (M = 3.71), t(292.979) = −1.979, p = .049. No significant differences were found across age (p = .654), education (p = .403), or income (p = .601). However, the income result requires caution due to the significant Levene's test (p = .046). Overall, art-related characteristics appeared more informative than most demographic variables in explaining CPV.

5 CONCLUSION AND RECOMMENDATIONS

This study examined the factors influencing CPV of traditional drawings. Art knowledge and art involvement were significant positive predictors, with art knowledge showing the strongest effect, while art interest and personality traits were non-significant. The regression model explained 10.1% of the variance in CPV. Gender differences were significant, whereas age, education, and income showed no significant differences.

Based on these findings, artists and galleries should enhance consumer knowledge and involvement through exhibitions, workshops, artist demonstrations, and information about artistic techniques and craftsmanship. Marketing strategies should emphasize the creative process, originality, and artistic background of traditional drawings.

Future research should consider additional factors such as authenticity, artistic quality, emotional attachment, price perception, and artist reputation. Larger and more diverse samples, qualitative approaches, and separate analyses of the Big Five personality dimensions could further strengthen understanding of CPV.

REFERENCES

- Afhami, R., & Mohammadi-Zarghan, S. (2018). The Big Five, aesthetic judgment styles, and art interest. *Europe's Journal of Psychology*, *14*(4), 764–775. https://doi.org/ 10.5964/ejop.v14i4.1479
- An, D., Jeong, B., & Youn, N. (2021). Effects of art consumption on consumer well-being. *The Journal of Consumer Affairs*, *56*(2), 685–702. https://doi.org/10.1111/joca.12429
- Cotter, K.N., Rodriguez-Boerwinkle, R.M., Christensen, A.P., Fekete, A., Smith, J.K., Smith, L.F., & Silvia, P.J. (2023). Updating the Aesthetic Fluency Scale: Revised long and short forms for research in the psychology of the arts. *PLOS ONE*, *18*(2), e0281547. https://doi.org/10.1371/journal.pone.0281547
- Fayn, K., Silvia, P., Erbas, Y., Tiliopoulos, N., & Kuppens, P. (2018). Nuanced aesthetic emotions: Emotion differentiation is related to knowledge of the arts and curiosity. *Cognition and Emotion*, *32*(3). https://doi.org/10.1080/02699931.2017.1322554
- Gosling, S.D., Rentfrow, P.J., & Swann, W.B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*, 504–528.

https://doi.org/10.1016/S0092-6566(03)00046-1

- Holbrook, M.B., & Hirschman, E.C. (1982). The experiential aspects of consumption: Consumer fantasies, feelings, and fun. *Journal of Consumer Research*, *9*(2), 132–140. https://doi.org/10.1086/208906
- Hüttl-Maack, V. (2018). Visual art in advertising: New insights on the role of consumers' art interest and its interplay with the hedonic value of the advertised product. *Journal of Product & Brand Management*, *27*(3), 262–276. https://doi.org/10.1108/ JPBM-02-2017-1424
- Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology*, *95*(4), 489–508. https://doi.org/10.1348/000712604236981 1
- Pelowski, M., Gerger, G., Chetouani, Y., Markey, S., & Leder, H. (2017). But is it really art? The classification of images as "art"/"not art" and correlation with appraisal and viewer interpersonal differences. *Frontiers in Psychology*, *8*, 1729. https://doi.org/10.3389/fpsyg.2017.01729
- Spee, M., Pelowski, M., Arato, J., Mikuni, J., Tran, S., Eisenegger, C., & Leder, H. (2022). Social reputation influences on liking and willingness-to-pay for artworks. *PLOS ONE*, *17*(4), e0266020. https://doi.org/10.1371/journal.pone.0266020
- Sweeney, J.C., & Soutar, G.N. (2001). Consumer perceived value: The development of a multiple item scale. *Journal of Retailing*, *77*(2), 203–220. https://doi.org/10.1016/S0022-4359(01)00041-0
- Zaichkowsky, J.L. (1994). The personal involvement inventory: Reduction, revision, and application to advertising. *Journal of Advertising*, *23*(4), 59–70. https://doi.org/10.1080/00913367.1994.10673459

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Geographic Variation in Category-Level Spending Across Outlet Locations in Sri Lankan Grocery Retailing

Perera VPM[1], Prabodanie RAR[2]
[1,2]*Department of Industrial Management, Faculty of Applied Sciences, Wayamba University Sri Lanka, Kuliyapitiya, Sri Lanka*
pasanmanodya20@gmail.com[1], ranga@wyb.ac.lk[2]

ABSTRACT

Geographic variation in consumer expenditure is an important consideration in retail analytics, particularly when evaluating category-level spending patterns across multiple outlets. This study investigates geographic differences in fresh-product, dry-goods, and luxury-product expenditure across grocery retail outlet locations in Sri Lanka. It also evaluates outlet-level expenditure concentration using Pareto/ABC classification. A quantitative observational approach was adopted using a publicly available retail dataset. Descriptive and inferential techniques, including analysis of variance, Kruskal–Wallis tests, effect-size estimation, random-effects variance decomposition, and outlet-level bootstrapping, were employed. The findings indicate detectable differences in expenditure across outlets, although outlet identity accounts for a relatively small proportion of overall spending variation. Outlet-associated variation is comparatively greater for luxury and fresh products than for dry goods. Pareto/ABC analysis indicates relatively dispersed expenditure contributions rather than concentration among a small number of outlets. The study provides a methodological foundation for interpreting geographic expenditure patterns while recognizing the importance of location specific inventory-management and distribution policies empirically validated with sales data.

KEYWORDS: Customer Segmentation, Geographic Expenditure Variation, Grocery Retailing, Pareto/ABC Classification, Retail Analytics

1 INTRODUCTION

Understanding geographic variation in category-level expenditure is important for retail analytics and inventory planning in multi-outlet grocery retailing. Effective inventory management requires retailers to maintain appropriate stock levels at theire distribution centers and retail outlets to reduce stockouts and to ensure product availability in response to customer demand. When inventory and distribution policies are applied uniformly across a retail chain, these approaches may overlook variations in customer expenditure patterns across outlet locations, potentially limiting their ability to refelect local market characteristics.Geographic differences in expenditure may arise from variations in customer demographics, purchasing power, lifestyle preferences,accessibility, and local market conditions. Consequently,expenditure patterns for fresh products,dry goods,and luxury products may differ across outlets within the same retail chain.Examining these differences can provide insights into the spatial heterogeneity of consumer spending and help improve inventory management and product distribution strategies to ensure availabilty of right products at right locations. Nevertheless, observed expenditure differences should not automatically be interpreted as evidence of differences in underlying demand or as justification for specific inventory-management interventions. Against this background,this study investigates expenditure on fresh products,dry products,luxury products across 20 grocery retail outlet locations in Sri Lanka.Specifically,it evaluates the magnitude and variation of outlet-

associated expenduture differences and examines expenditure concentration using Pareto/ABC classification. By considering both category-level spending patterns and the distribution of expenditure.The findings are intended to contribute to retail analytics and inventry classification research by clarifying the extent to which expenditure patterns vary across locations,while distinguishing statistical evidence of such variation from its potential managerial implications.

2 LITERATURE REVIEW

Geographic variation in retail expenditure has been examined in studies emphasizing differences in consumer purchasing behaviour and store-level market conditions. Fox et al. (2004) demonstrated variations in shopping and spending across retail formats, while Hoch et al. (1995) identified differences in price responsiveness associated with store characteristics. Montgomery (1997) further highlighted the relevance of store-specific information in retail decision-making. However, these studies provide limited evidence regarding the magnitude of category-level expenditure differences across Sri Lankan grocery outlet locations. Research on cross-category purchasing relationships (Russell & Petersen, 2000) and retail assortment planning (Mantrala et al., 2009) also emphasizes the importance of examining product categories separately. Customer segmentation introduces additional methodological challenges. Although clustering can identify groups with similar purchasing characteristics, using expenditure-derived segments to explain the same expenditure variables risks circular inference (Kriegeskorte et al., 2009). Consequently, pre-existing customer classifications require validation before being treated as independent explanatory factors. Furthermore, statistical significance in large datasets does not necessarily indicate substantial practical effects (Lin et al., 2013). The hierarchical structure of observations within retail outlets also requires consideration of between-outlet and within-outlet variation (Raudenbush & Bryk, 2002). Pareto/ABC classification provides a framework for examining relative expenditure contributions; however, outlet-level expenditure rankings should not be interpreted as direct measures of inventory criticality (Flores & Whybark, 1987). Within Sri Lanka, existing research on grocery store choice (Vithanage et al., 2023) provides relevant retail context but does not establish the magnitude of category-specific outlet effects. Accordingly, this study addresses the gap by integrating expenditure comparisons, effect-size estimation, cluster-validity diagnostics, and descriptive Pareto/ABC analysis to evaluate geographic spending variation while avoiding unsupported causal and operational conclusions.

3 METHODOLOGY

3.1 Research Design and Data

To study the geographic variations in category-level spending, this study used a publicly available grocery retail dataset contains customer expenditure on fresh, dry, luxury products at retail outlets located across different cities in Sri Lanka. The analytical sample comprised 774,033 valid observations across 20 outlet locations, with expenditure recorded in Sri Lankan rupees for fresh products, dry goods, and luxury products. Outlet location served as the categorical explanatory variable, the records were treated as pooled cross-sectional data, and their hierarchical structure and limitations in data provenance are acknowledged.

3.2 Statistical Analysis

Statistical analyses were conducted using Stata. Expenditure differences across outlets were examined using one-

way ANOVA, heteroskedasticity-robust tests, Kruskal–Wallis tests, and effect-size measures (η^2 and ω^2). Intraclass correlation coefficients were estimated using random-effects generalized least squares models to assess between-outlet variation. Category-specific outlet effects were compared using differences in η^2, with 95% confidence intervals obtained through 2,000 outlet-level bootstrap replications. Spending-derived outlet tiers were excluded from confirmatory testing to avoid circularity. Chi-square tests, Cramér's V, and within-group diagnostics assessed associations between cluster membership and outlet location. Finally, Pareto/ABC analysis evaluated expenditure concentration using 70/90 and 80/95 cumulative classification thresholds.

4 RESULTS AND DISCUSSION

4.1 *Expenditure Variation Across Outlets*

Conventional analysis of variance (ANOVA) identified statistically significant differences in fresh-product, dry-goods, and luxury-product expenditure across the 20 outlet locations ($p < 0.001$). Heteroskedasticity-robust tests and Kruskal–Wallis analyses produced consistent evidence of between-outlet differences. However, effect-size estimates indicated that outlet identity accounted for only a small proportion of total expenditure variation, ranging from 1.37% for dry goods to 4.44% for luxury products (Table 1). Intraclass correlation coefficients similarly indicated that most expenditure variability occurred within rather than between outlets. These findings were consistent with H1, although conventional significance tests require cautious interpretation because the observations were nested within only 20 outlets.

Table 1. Summary of Outlet-Level Expenditure Effects and Bootstrap Comparisons

Category	ANOVA F	η^2	ICC
Luxury sales	1,894.49	44	0.0468
Fresh sales	1,409.03	33	0.0352
Dry goods sales	564.53	37	0.0144

Outlet-level bootstrap comparisons (2,000 replications)

Table 2. Bootstrap Comparison of Outlet-Associated Expenditure Variance Across Product Categories

Comparison	Difference in η^2	95% percentile CI
Luxury−Dry goods	0.0308	[0.0221, 0.0326]
Fresh−Dry goods	0.0196	[0.0144, 0.0206]

ANOVA tests used F (19, 774013), with $p < 0.001$ for all categories. ICC = intraclass correlation coefficient; CI = confidence interval. Bootstrap confidence intervals were estimated by resampling the 20 outlets.

4.2 Category-Specific Outlet Effects

The proportion of expenditure variance associated with outlet location was compared across fresh products, dry goods, and luxury products using eta-squared (η^2) estimates. Outlet-level bootstrap analysis with 2,000 replications indicated greater outlet-associated variation for luxury and fresh products than for dry goods, with both 95% confidence intervals excluding zero (Table 1). These findings supported H2, indicating that the magnitude of outlet-related expenditure variation differed across product categories. However, the observed differences were relatively small, suggesting that outlet location explained only a limited proportion of overall expenditure variability.

4.3 Pareto/ABC Classification

Outlet-level Pareto/ABC analysis indicated relatively dispersed expenditure contributions across the 20 locations. Under the 70% and 90% cumulative thresholds, Class A comprised 12 outlets for luxury and fresh products and 13 for dry goods, representing 60%–65% of all outlets. The alternative 80% and 95% thresholds classified 15 outlets as Class A in every category, demonstrating sensitivity to classification criteria. These findings indicate that recorded expenditure was not concentrated among a small number of outlets. Therefore, the resulting ABC groups describe relative expenditure contributions rather than independently establishing inventory criticality or optimal replenishment policies.

5 CONCLUSION AND RECOMMENDATIONS

This study examined geographic variation in category-level expenditure across 20 grocery retail outlet locations in Sri Lanka. The findings revealed statistically significant expenditure differences across outlets, although outlet location explained only a small proportion of overall spending variation. Outlet-associated differences were comparatively greater for luxury and fresh products than for dry goods, while most expenditure variability occurred within outlets. Pareto/ABC classification indicated relatively dispersed expenditure contributions rather than concentration among a few outlets. These findings highlight the importance of considering effect sizes, hierarchical data structures, and segmentation validity when interpreting retail expenditure patterns. Retailers are encouraged to evaluate outlet-specific inventory and delivery planning, validate existing customer clusters, and use Pareto/ABC classification cautiously. However, expenditure patterns alone cannot establish optimal inventory policies. Further research incorporating inventory performance, demand variability, and outlet characteristics is necessary to determine whether geographic information can improve retail planning and operational decisions.

REFERENCES

- Flores, B. E., & Whybark, D. C. (1987). Implementing multiple criteria ABC analysis. Journal of Operations Management, **7**(1–2), 79–85. https://doi.org/10.1016/0272-6963(87)90008-8
- Fox, E. J., Montgomery, A. L., & Lodish, L. M. (2004). Consumer shopping and spending across retail formats. The Journal of Business, **77**(S2), S25–S60. https://doi.org/10.1086/381518
- Hoch, S. J., Kim, B.-D., Montgomery, A. L., & Rossi, P. E. (1995). Determinants of store-level price elasticity. Journal of Marketing Research, 32(1), 17–29. https://doi.org/10.1177/002224379503200104
- Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F., & Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. Nature Neuroscience, 12(5), 535–540. https://doi.org/10.1038/nn.2303
- Lin, M., Lucas, H. C., & Shmueli, G. (2013). Too big to fail: Large samples and the p-value problem. Information Systems Research, 24(4), 906–917. https://doi.org/10.1287/isre.2013.0480
- Mantrala, M. K., Levy, M., Kahn, B. E., Fox, E. J., Gaidarev, P., Dankworth, B., & Shah, D. (2009). Why is assortment planning so difficult for retailers? A framework and research agenda. Journal of Retailing, 85(1), 71–83. https://doi.org/10.1016/j.jretai.2008.11.006
- Montgomery, A. L. (1997). Creating micro-marketing pricing strategies using supermarket scanner data. Marketing Science, 16(4), 315–337. https://doi.org/10.1287/mksc.16.4.315
- Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods

(2nd ed.). SAGE Publications. https://www.sagepub.com/shop/buy-a-book/hierarchical-linear-models-2-9230?page=1

- Russell, G. J., & Petersen, A. (2000). Analysis of cross category dependence in market basket selection. Journal of Retailing, 76(3), 367–392. https://doi.org/10.1016/S0022-4359(00)00030-0
- Vithanage, C., Wattage, P., Kariyawasam, S., Wilson, C., & Khanal, U. (2023). Choice modeling of grocery shopping behavior: A Sri Lankan case study. Cities, 143, Article 104610. https://doi.org/10.1016/j.cities.2023.104610

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Crafting An Ethical Business Climate: An Entrepreneurial Value Perspective on Product Harm Crisis Management

Sabarjah S[1], Herath HMSP[2]
[1]*Department of Commerce and Financial Management, Faculty of Commerce and Management Studies, University of Kelaniya, Sri Lanka*
[2]*Department of Banking and Finance, Faculty of Business Studies and Finance, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*
s.sabarjahvictory@gmail.com[1], shan@wyb.ac.lk[2]

ABSTRACT

This conceptual study examines how an ethical business climate could be crafted from an entrepreneurial perspective, with special attention to product-harm crises, corporate social responsibility (CSR), entrepreneur ethics, and entrepreneurial values. The study addresses a gap in predominantly consumer-oriented product-harm crisis research by focusing on the entrepreneur as an important source of ethical decision-making. A systematic review of the literature was performed with academic papers sourced from databases such as Web of Science, Science Direct and Emerald. Search terms combined included CSR, Entrepreneur Ethics, Product-Harm Crisis, Ethical Decision Making, and Entrepreneur Values. Using content analysis, it allows to combine theoretical arguments with empirical data to establish relationships between constructs. The review supports a sequential mechanism in which entrepreneurial values shape CSR orientation, CSR orientation influences ethical product and brand decisions, and repeated ethical decisions contribute to an ethical business climate. The proposed framework further links ethical climate with product-harm crisis preparedness and management. The study integrates value–attitude–behavior and upper-echelons perspectives with CSR, ethical decision-making, and product-harm crisis literature, shifting attention from consumers' reactions toward entrepreneurs' role in building ethical organizational conditions. Finally, the study concludes that the proposed framework is conceptual and it requires empirical testing and validation via a future study.

KEYWORDS: Corporate Social Responsibility, Entrepreneur Ethics, Entrepreneurial Values, Ethical Decision-Making, Product-Harm Crisis

1 INTRODUCTION

Consumers' purchasing decisions are influenced by factors such as brand credibility and trustworthiness, alongside price and performance. Product-harm crises, stemming from defective products, can significantly harm consumers and damage brand equity. Existing research mainly focuses on consumer perceptions and corporate communication, neglecting the impact of entrepreneurs' values on ethical practices. This study addresses the gap by exploring how entrepreneurial values shape the corporate social responsibility (CSR) orientation and crisis preparedness. It integrates the Value-Attitude-Behaviour (VAB) hierarchy with Upper Echelons Theory to propose a framework linking individual values to organizational ethical practices in response to product-harm crises.

2 LITERATURE REVIEW

2.1 Business Ethics and Corporate Social Responsibility

Business ethics concerns the application of moral principles to business decisions and stakeholder responsibilities (Carroll, 1991). CSR extends these obligations beyond economic and legal requirements to encompass social well-

being and stakeholder accountability (Dahlsrud, 2006). While shareholder theory prioritizes profit within ethical and legal boundaries (Friedman, 1970). The Stakeholder Theory emphasizes the responsibilities towards groups affected by organizational decisions (Freeman, 1984; Donaldson & Preston, 1995). Carroll's (1991) framework further identifies economic, legal, ethical, and philanthropic responsibilities. The entrepreneurs play an important role in crisis management and responsible business practices (Wattegama et al., 2013). Beyond reputation management, genuine CSR requires product safety, transparent disclosure, corrective action, and fair stakeholder treatment (Klein & Dawar, 2004).

2.2 Entrepreneurial Values and Ethical Decision-making

Entrepreneurial values are enduring beliefs that guide decision-making and organizational practices beyond profit maximization (Rokeach, 1973; Schwartz, 1999; Wattegama et al., 2013). The VAB theory explains how personal values shape attitudes and subsequent behaviours, suggesting that entrepreneurs' beliefs about responsibility, justice, and stakeholder welfare may influence CSR orientation and ethical product decisions (Homer & Kahle, 1988). Complementarily, the Upper Echelons Theory proposes that organizational outcomes reflect senior decision-makers' characteristics, experiences, and values (Hambrick, 2007). Empirical research further links executive values with CSR-related decisions, highlighting their potential influence on organizational policies and ethical practices (Chin et al., 2013).

2.3 *Product*-Harm Crises and Ethical Business Climate

A product-harm crisis occurs when a product causes or is perceived to pose risks to consumers, potentially damaging brand reputation, consumer trust, and organizational performance. Situational Crisis Communication Theory (SCCT) explains how crisis characteristics and perceived organizational responsibility influence crisis responses (Coombs, 2007). The Ethical Climate Theory highlights how shared organizational norms shape ethical decision-making, transparency, and corrective (Victor & Cullen, 1988). Previous studies have primarily examined consumer responses, including responsibility attribution, purchase intentions, negative word-of-mouth, and brand evaluations (Dawar, 1998; Lin et al., 201). However, integrating entrepreneurial values with organizational ethical practices and crisis preparedness remains important. Combining these perspectives enables a broader understanding of product-harm crisis management at individual, organizational, and stakeholder levels.

2.4 Conceptual and Theoretical Synthesis

This study integrates entrepreneurship, CSR, business ethics, and crisis management to examine how entrepreneurial values may influence product-harm crisis preparedness. The framework comprises five dimensions: entrepreneurial values, CSR orientation, ethical product and brand decisions, ethical business climate, and crisis preparedness. The VAB paradigm and Upper Echelons Theory explain how personal values may influence organizational decisions (Homer & Kahle, 1988). The Stakeholder Theory connects these values to CSR orientation

(Freeman, 1984; Choongo et al., 2019). The Ethical Climate Theory and SCCT explain how organizational ethical practices and crisis responses may strengthen preparedness (Victor & Cullen, 1988).

3 METHODOLOGY

The present study has reviewed the literature to investigate the interconnections among entrepreneurial values, CSR, ethical decision-making, and product-harm crisis management. Utilizing databases, such as, Web of Science and Science Direct, the present study has focused on key terms associated with these concepts and examined the role of entrepreneurs in ethical decision-making. The content analysis highlighted recurring theories and empirical findings, although the study does not assert causality due to reliance on secondary data. The propositions developed aim to guide future empirical research, emphasizing the intricate relationships between entrepreneurial values and ethical practices in CSR and crisis management.

4 RESULTS AND DISCUSSION

4.1 Entrepreneurial Values and CSR Orientation

Entrepreneurial literature emphasizes that personal values shape how entrepreneurs interpret situations, establish priorities, and make decisions. These values influence organizational practices concerning employee well-being, customer satisfaction, organizational culture, and product quality. The study also suggests that entrepreneurs' ethical and social responsibility values are associated with their CSR orientation. Entrepreneurs who prioritize ethical conduct and stakeholder welfare may demonstrate a stronger commitment to CSR, consistent with Stakeholder Theory, which emphasizes considering the interests of groups affected by organizational decisions (Freeman, 1984). Furthermore, the VAB hierarchy explains how personal values shape attitudes that subsequently guide behaviour (Feng et al., 2014). Accordingly, the present study suggests the following proposition;

Proposition 1: Entrepreneurial values direct firm's orientation toward CSR.

4.2 CSR Orientation and Ethical Product / Brand Decisions

CSR orientation may encourage entrepreneurs to consider stakeholder welfare and societal impacts alongside price and performance when making product and brand decisions. Research links CSR associations with consumer evaluations and brand responses, highlighting the importance of ethical product attributes in responsible consumption (Peloza et al., 2013). Accordingly, the present study suggests the following proposition;

Proposition 2: A stronger CSR orientation leads to the consideration of ethical attributes in product and brand decisions.

4.3 Ethical Decisions and the Development of an Ethical Business Climate

A single policy statement is not likely to create ethical climate. It is more likely to be reinforced through repeated managerial decisions, routines and signals about what the organization deems acceptable. When entrepreneurs make these decisions a regular part of their organizations' expectations, it is because they prioritize product safety, honest talk, stakeholder interests, and accountability.

Thus, entrepreneurial values may be translated into organizational climate through repeated ethical decision making (Hunt et al., 1989). Accordingly, the study has suggested the following proposition;

Proposition 3: Repeated ethical product and brand decisions are positively associated with the development of an ethical business climate.

4.4 Ethical Business Climate and Product-Harm Crisis Preparedness

An ethical business climate fosters crisis preparedness by promoting transparency, responsibility, and accountability. Companies with such values are better positioned to identify risks, communicate effectively, and implement corrective actions during product-harm crises. While an ethical climate does not prevent such crises, it enhances the quality of organizational responses. (Daniel & Coombs, 2006; Coombs, 2007). Accordingly, the study has suggested the following proposition;

Proposition 4: A stronger ethical business climate is positively associated with product-harm crisis preparedness and responsible crisis management.

4.5 Proposed Conceptual Mechanism

The mechanism connects the individual and organizational levels of analysis. The entrepreneurial values represent the entrepreneur-level starting point; CSR orientation captures an attitudinal or strategic orientation; ethical product and brand decisions represent observable organizational choices; ethical climate reflects the cumulative organizational environment; and crisis preparedness represents the downstream organizational capability. The model therefore extends consumer-centered product-harm crisis research by introducing an upstream explanation of how ethical organizational conditions may be constructed before a crisis occurs.

5 CONCLUSION

After analyzing the content that are available in the literature the present study has suggested the propositions for creating an ethical business climate based on entrepreneurial values. Accordingly, those argue that the values influence CSR orientation, which in turn shapes ethical product and brand decisions, leading to a cycle of ethical practices that foster an ethical business environment. This environment is essential for effective management of product-harm crises. The study has integrated various theoretical perspectives and suggested that the firms should focus on the values and decision-making orientations of entrepreneurs and managers, alongside their formal practices. The future researchers should empirically validate those propositions while considering factors like organizational size and cultural contexts.

REFERENCES

- Carroll, A. (1991). The pyramid of corporate social responsibility: toward the moral management of organizational stakeholders. *Journal of Business Horizons*, 39 - 48.
- Chin, M., Hambrick, D.C., & Trevino, L.K. (2013). Political Ideologies of CEOs: The Influence of Executives' Values on Corporate Social Responsibility. *Administrative Science Quarterl,* 197 - 232.
- Choongo, P., Paas, L.J., , Masurel, E., van Burg, E, & Lungu, J. (2019). Entrepreneurs' personal values and CSR orientations: evidence from SMEs in Zambia. *Journal of Small Business and Enterprise Development*, 545 - 570.
- Coombs, W. (2007). Protecting Organization Reputations During a Crisis: The Development and Application of

Situational Crisis Communication Theory. *Corporate Reputation Review*, 163 - 176.

- Coombs, W.T. (2007). Ongoing crisis communication: Planning, managing, and responding. *Sage*.
- Dahlsrud, A. (2006). How corporate social responsibility is defined: An analysis of 37 definitions. *Corporate Social Responsibility and Environmental Management, 15*, 1-13.
- Daniel, L., & Coombs, W.T. (2006). How should a company respond to a product harm crisis? The role of corporate reputation and consumer-based cues. *Business Horizons*, 379-385.https://doi.org/doi:10.1016/j.bushor.2006
- Dawar, N. (1998). Product-harm crises and the signaling ability of brands. *International Studies of Management & Organization*, 109 -119
- Donaldson, T., & Preston, L.E. (1995). The stakeholder theory of the corporation: Concepts, evidence, and implications. *Academy of Management Review*, 65 - 91.
- Feng, J.J., Wattegama, W.G.E.J., Ishaq, M., & Ping, Q. (2014). The Mechanism of Entrepreneurs' decisionmaking in the Spillover Effects of Industry Unethical Behavior: A Theoretical Reflection under the Chinese Cultural Context. *International Journal of Advances in Management Science. 3*(3), 97 - 104
- Freeman, R.E. (1984). Strategic management: A stakeholder approach.
- Friedman, M. (1970). The social responsibility of business is to increase its Profits. 122-126.
- Hambrick, D.C. (2007). Upper echelons theory: An update. . *Academy of management review,*, 334 - 343.
- Homer, P.M., & Kahle, L.R. (1988). A structural equation test of the value-attitude-behavior hierarchy. *Journal of Personality and Social Psychology*, 638 - 646.
- Hunt, S., Wood, V.R., & Chonko, L.B. (1989). Corporate Ethical Values and Organizational Commitment in Marketing. *Journal of Marketing*, 79 - 90.
- Klein, J., & Dawar, N. (2004). Corporate social responsibility and consumers' attributions and brand evaluations in a product–harm crisis. *International Journal of Research in Marketing*.
- Lin, Chieh-Peng., Chen, Shwu-Chuan, Chiu, & Chou-Kang. (2011). Understanding purchase intention during product-harm crises: Moderating effects of perceived corporate ability and corporate social responsibility. *Journal of Business Ethics*, *102*(3), 455 - 471
- Peloza, J., White, K., & Shang, J. (2013). Good and Guilt-free; The role of self accountability in influencing preferences for products with ethical attitudes. *Journal of Marketing*, 104 -119.
- Rokeach, M. (1973). The Nature of Human Values. *New York, United States: The Free Press.*
- Schwartz, S. (1999). Cultural value differences: some implications for work. *Applied Psychology: An International Review*, 23 - 48.
- Victor, B., & Cullen, J. B. (1988). The organizational bases of ethical work climates. *Administrative science quarterly*, 101 - 125.
- Wattegama, W.G.E.J., Feng, J., & Qing, P. (2013). An Entrepreneur's Value Perspective in making an Ethical Business Decision Under the Context of Product-harm Crises Management: A Research Model. In: ICBM Committee (Eds.) Resilience Amid Turbulence: Revisiting Ethics. Proceedings of Tenth International Conference on Business Management; 10 December, Gangodawila, Nugegoda, Sri Lanka: Faculty of Management Studies and Commerce, University of Sri Jayewardenepura: pp. 135 - 150, ISSN 2235 - 9761

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Predictors of Generation Z's Purchase Intention for Green Packaged Products in Sri Lanka: The Mediating Role of Perceived Value

Rasangani NWGS[1], Wattegama EJ[2]
[1,2] *Department of Industrial Management, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*
rasanganishashini1@gmail.com[1], erandiw@wyb.ac.lk[2]

ABSTRACT

Environmental awareness has led firms to adopt green packaging, yet little is known about what predicts Generation Z's purchase intention for green packaged products in Sri Lanka, or whether perceived value explains how Theory of Planned Behavior (TPB) constructs relate to it. This study examines the environmental concern, attitude, social influence, and perceived behavioral control as predictors of purchase intention, with perceived value as a mediator. Data from 208 Generation Z consumers who were selected purposively and were collected via online surveys, have analyzed with SPSS and PROCESS Model 4. The results have shown that environmental concern, attitude, and social influence positively associate with purchase intention whereas the perceived behavioral control was not. In jointly estimated models, perceived value partially mediated the attitude association (a×b = 0.095, 95% CI [0.030, 0.195]) and fully mediated the perceived behavioral control association (a×b = 0.122, 95% CI [0.027, 0.231]), but not those of environmental concern or social influence. The eco-labeling that communicates functional and environmental benefits may strengthen the consumers' purchase intention for green packaged products.

KEYWORDS: Generation Z, Green Packaging, Perceived Value, Purchase Intention

1 INTRODUCTION

Consumers are increasingly attentive to the environmental consequences of what was bought by them (ElHaffar et al., 2020). The Generation Z comprises the individuals who were born between 1995 and 2012 (Jayatissa, 2023) and these consumers are often portrayed as attentive to the environmental impact of products, including packaging. The green packaging can be understood as degradable, reusable, and non-toxic materials and production processes (Singh & Pandey, 2018). While interest in green consumption is growing globally, research on the determinants of green packaging purchase intention among Generation Z in Sri Lanka remains limited. The literature on green consumption has mainly focused on the direct relationship between green purchase intention (PI) and variables, such as, Environmental Concern (EC), Attitude towards Green Packaging (AT), Social Influence (SI), and Perceived Behavioral Control (PB) without considering the possibility of mediation of Perceived Value (PV) in the process, mainly in the context of Sri Lanka. The PV was included as a mediating variable in the relationship between EC, AT, SI, PB, and PI in the present study. The first objective of the study was to test the direct effects of EC, AT, SI, and PB on PI. The second objective was to test the indirect effects of EC, AT, SI, and PB on PI through PV. Furthermore, it was aimed at testing whether PV mediates the relationship between the aforementioned four independent variables and the dependent variable of PI. It was expected that the results of this study will contribute

to the development of Value-Based Theory of Planned Behavior (TPB).

2 LITERATURE REVIEW

The TPB suggests that behavioral intention is influenced by attitude, subjective norm, and perceived behavioral control Ajzen (1991). It has been widely applied in examining green consumption behavior Paul et al. (2016). Environmental concern includes the awareness of environmental issues among consumers and their action towards solving these issues (Xiao et al., 2013). Previous studies have indicated that there exist positive correlations between environmental concern and green product evaluation and purchase intention (For example, Erdem, 2021). Nevertheless, according to some scholars, it is not necessary for purchasing the products. Green Attitude is an indicator of how favorable or unfavorable the consumers consider eco-packaging to be. It is considered a strong determinant of green purchase intention (Paul et al., 2016). Social Influence represents the subjective norm within the TPB framework and indicates consumers' perception of pressure to consume green due to family, friends, and social media (Ajzen, 1991). Moreover, Perceived Behavioral Control denotes the consumers' perceptions about easy availability of green products (Ajzen, 1991). The consumers who feel that green products are more environmentally and functionally beneficial have stronger purchase intentions (Chen & Chang, 2012; Konuk, 2018). PV has also become the key mediator which is linking environmental attitude to purchase action instead of directly affecting purchase intention (Chen and Chang, 2012). Following the above literature, it is hypothesized that EC, AT, SI, and PB will each positively affect the PI (Hypothesis 1 –4), while PV will mediate the effect of each of these four variables on PI (Hypothesis 5 – 8).

3 METHODOLOGY

A quantitative and explanatory approach was used in the present study to examine both the direct and mediating relationships between EC, AT, SI, PB, PV, and PI. The target population was included the Generation Z consumers in Sri Lanka (who were born between 1995 and 2012) who had awareness or knowledge about green packaged products. With reference to the number of responses required according to Morgan's table when the population size is not known, the sample size for the present study was estimated to be 400. However, only 208 usable questionnaires could be obtained by means of a non-probability purposive sampling technique using an online self-administered survey, and it was enough for conducting multiple regressions and mediation analyses. All the variables (EC, AT, SI, PB, PV, PI) were assessed with four questions based on reliable scales using a 5-point Likert Scale. A pilot study was conducted with 35 respondents who were selected conveniently revealed satisfactory to high internal consistency (Cronbach’s α were ranged between 0.621 - 0.867). The data were screened for missing values and inconsistencies before analysis. Descriptive and correlation analyses summarized the respondents’ characteristics and bi-variate relationships. Total effects of EC, AT, SI, and PB on PI were tested using multiple linear regression in SPSS.

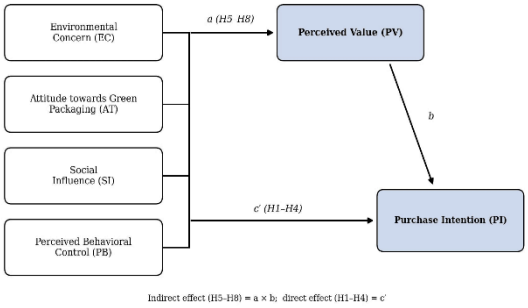


Figure 1: Research Model

Multicollinearity was assessed using tolerance and Variance Inflation Factor (VIF) statistics. Mediation was tested using Hayes PROCESS macro v4.3.1. Model 4, with 5,000 bootstrap resamples and 95% percentile confidence intervals; an indirect effect was considered significant when its confidence interval excluded zero.

Based on the research model, as shown in Figure 1, the following hypotheses have been established.

H_1: EC is positively associated with PI.
H_2: AT is positively associated with PI.
H_3: SI is positively associated with PI.
H_4: PB is positively associated with PI.
H_5: PV mediates the EC–PI relationship.
H_6: PV mediates the AT–PI relationship.
H_7: PV mediates the SI–PI relationship.
H_8: PV mediates the PB–PI relationship.

4 RESULTS AND DISCUSSION

The majority of the 208 respondents belonged to the Colombo (21.2%) and Gampaha (14.4%) districts and had a monthly income of between LKR 50,000 and LKR 100,000 (61.5%). All six constructs had high mean values (M= 4.30 - 4.47), implying that all the constructs were positively perceived by the consumers; attitude towards green packaging was perceived most highly (M = 4.47), while perceived behavioral control had the lowest mean value (M = 4.30). The correlation coefficient for all the pairs was positive and moderate to very strong (r = 0.572 - 0.805), with tolerance values of 0.268–0.400 and VIF values of 2.50–3.73, indicating no serious multicollinearity. The regression model significantly predicted Purchase Intention, explaining 60.6% of its variance ($R^2 = 0.606$, adjusted $R^2 = 0.598$, $F(4, 203) = 78.081$, $p < 0.001$). As shown in Table 1, Environmental Concern ($\beta = 0.225$, $p = 0.007$), Attitude towards Green Packaging ($\beta = 0.320$, $p < 0.001$), and Social Influence ($\beta = 0.216$, $p = 0.012$) had significant positive direct effects on Purchase Intention, supporting the hypotheses, H_1 – H_3. Attitude was the strongest direct predictor, consistent with the TPB in Ajzen's (1991) and prior findings that attitude is the leading determinant of green purchase intention among young consumers as stated by Paul et al. (2016). Perceived Behavioral Control did not significantly predict Purchase Intention ($\beta = 0.115$, $p = 0.136$), and H_4 was not supported, suggesting that for Sri Lankan Generation Z, ease of access alone does not drive purchase intention unless it also generates perceived value.

Mediation (as shown in Table 1); PV was positively associated with PI in every model ($b = 0.282$, $p < .001$). Indirect effects were significant for AT ($a \times b = 0.095$, 95% CI [0.030, 0.195]) and PB (0.122, [0.027, 0.231]) but not for EC (0.018, [−0.028, 0.083]) or SI (0.034, [−0.032, 0.150]), supporting H_6 and H_8 but not H_5 and H_7. Mediation was partial for AT ($c' = 0.225$, $p = 0.002$) and indirect-only for PB ($c' = -0.007$, $p = 0.931$), so PB relates to PI only through PV; EC and SI related to PI directly ($c' = 0.207$, p=.01,

Table 1: Multiple Regression of Purchase Intention on the Four Predictors (Standardized Coefficients)

Path	β	p	Decision
EC → PI	0.225	0.007	Supported
AT → PI	0.320	0.001	Supported
SI → PI	0.216	0.012	Supported
PB → PI	0.115	0.136	Not Supp.

c′ = 0.182, p = 0.030). These are associations consistent with mediation in a cross-sectional design, not the causal evidence. PV thus complements the TPB constructs for AT and PB, but the findings do not establish a "value-based" TPB.

5 CONCLUSION

This study examined predictors of Generation Z's green packaged product purchase intention in Sri Lanka, with perceived value as a mediator. In a sample of 208 Generation Z consumers, the largest share of whom (21.2%) lived in Colombo, five out of eight hypotheses (H_1, H_2, H_3, H_6, and H_8) were supported. AT and SI were associated with PI, and PV partially mediated the AT association and fully mediated the PB association, but not EC or SI. As the managerial implications of the findings the marketers could encourage positive attitudes through information, eco-labeling, and certification, with suitable pricing and availability (not measured here), and policymakers could consider eco-labeling regulation and incentives where these implications require testing. Limitations include the cross-sectional design, self-reported data (possible common method bias), and intention rather than behavior (as stated by ElHaffar et al., 2020).

Table 2: Total (c), direct (c′), and indirect (a×b, via PV) effects on PI (standardized & other predictors as covariates)

Path	a	c′	a×b	95% CI	c	Type
EC→ PI	.065	.207	.018	[−.028, .083]	.225	None
AT→ PI	.336	.225	.095	[.030, .195]	.320	Partial
SI→P I	.120	.182	.034	[−.032, .150]	.216	None
PB→ PI	.433	−.007	.122	[.027, .231]	.115	Indirect only

REFERENCES

- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179 - 211. doi:10.1016/0749-59 78(91)90020-T
- Chen, Y., & Chang, C. (2012). Enhance green purchase intentions: The roles of green perceived value, green perceived risk, and green trust. *Management Decision, 50*(3), 502 - 520. doi:10.1 108 /00251741211216250
- ElHaffar, G., Durif, F., & Dubé, L. (2020). Towards closing the attitude-intention-behavior gap in green consumption: A narrative review of the literature and an overview of future research directions. *Journal of Cleaner Production, 275*. doi:10.1016/j.jclepro.2020.122556
- Erdem, E. (2021). Exploring the factors affecting purchase intention towards green packaged products of Indian consumers. *E3S Web of Conferences* (p. 01057). EDP Sciences. doi:10.1051/e3sconf/20212510 1057
- Jayatissa, K. (2023). Generation Z – A New Lifeline: A Systematic Literature Review. *Sri Lanka Journal of Social Sciences and Humanities, 3*(2), 179 - 186. doi:10.4038/sljssh.v3i2.110
- Konuk, F. (2018). The role of store image, perceived quality, trust and perceived value in predicting consumers' purchase intentions towards organic private label

food. *Journal of Retailing and Consumer Services, 48*, 304 - 310. doi:10.1016/j.jretconser.2018.04.011

- Paul, J., Modi, A., & Patel, J. (2016). Predicting green product consumption using theory of planned behavior and reasoned action. *Journal of Retailing and Consumer Services, 29*, 123 - 134. doi:10.1016/j.jretco nser.2015.11.006
- Singh, G., & Pandey, N. (2018). The determinants of green packaging that influence buyers' willingness to pay a price premium. *Australasian Marketing Journal, 26*(3), 221 - 230. doi:10.1016/j.ausmj.2018. 06.001
- Xiao, C., Dunlap, R., & Hong, D. (2013). The nature and bases of environmental concern among Chinese citizens. *Social Science Quarterly, 94*(3), 672 - 690. doi:10.1111/j.1540-6237.2012.00883.x

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research –2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Phase-Specific Modeling of Dengue Transmission Dynamics and Temporal Surges in the Gampaha Medical Officer of Health Area, Sri Lanka

Fernando MWJJ[1], Dharmawardane PMN[2]
[1,2]*Department of Mathematical Sciences, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*
jjfernando772@gmail.com[1], *maheshd@wyb.ac.lk*[2]

ABSTRACT

Dengue fever is a very dangerous virus infection found in many tropical and subtropical areas around the world. In Sri Lanka, it is a major public health problem in the Western Province. The main aim of this study is to characterize and retrospectively evaluate the temporal dynamics of dengue transmission in the Gampaha Medical Officer of Health (MOH) area using a phase-specific Susceptible-Infected-Recovered-Deceased (SIRD) compartmental model, with estimation of transmission intensity and comparison of epidemic trajectories. This SIRD model was fitted to daily clinical data provided by Medical Office of Health officers in the Gampaha MOH area from December 14,2023 to May 31,2026. The temporal trajectory of dengue transmission was divided into eight distinct phases demarcated by seasonal monsoonal transitions and rapid surges in active cases. Model parameters and the basic reproduction number derived via the Next Generation Matrix approach were optimized through non-linear least-squares fitting. In-sample goodness of fit was evaluated using R^2, Root Mean squared Error (RMSE) and Mean Absolute Error (MAE). The basic reproduction number remained above the epidemiological threshold across all phases, reflecting ongoing transmission potential, while continuous endemicity was confirmed by active cases remaining non-zero throughout the monitoring period. The maximum spreading intensity was observed during Phase 7. Cyclic active case oscillations in 2026 were explained by a second-order Fourier series model with linear trend (R^2=79.44%). Multi-year tracking confirmed that 2026 epidemic trajectory accelerated significantly, reaching 400 cumulative cases by the end of May faster than in 2024 and 2025. These findings provide mathematical basis for anticipating changes in dengue transmission and supporting vector-control planning and healthcare resource allocation.

KEYWORDS: Basic reproduction number, Epidemic trajectory, Longitudinal trajectory, Next generation matrix, SIRD compartmental model

1 INTRODUCTION

Dengue remains a major public health concern in tropical and subtropical regions, including Sri Lanka, where recurring seasonal outbreaks place a considerable burden on the healthcare system. The disease is mainly transmitted by female Aedes aegypti and Aedes albopictus mosquitoes (Fernando et al., 2024). Gampaha District, located in the Western Province, is one of the country's areas and consistently reports a high number of dengue cases (Epidemiology Unit, 2024). Seasonal rainfall and other climatic conditions can create suitable mosquito breeding environments, contributing to changes in dengue transmission (Fernando et al., 2024). Although routine surveillance provides information on reported cases, simple case counts do not fully describe changes in transmission dynamics. In addition, detailed host-vector models can be difficult to apply because continuous information on mosquito populations and breeding sites is not always available

(Nishiura, 2006). Mathematical compartmental models provide an alternative approach for describing disease transmission using available case data. Therefore, this study uses a multi-phase SIRD model based on seasonal rainfall and other climate conditions and develops a complementary Fourier harmonic model with linear trend to analyze dengue transmission in the Gampaha MOH area over a 900-day period from 2023 to 2026.

2 LITERATURE REVIEW

Nishiura (2006) highlighted the importance of mathematical modelling in understanding infectious disease transmission and estimating key epidemiological measures such as the basic reproduction number (R_0). The study emphasized that R_0 estimation methods should balance theoretical assumptions with the availability and quality of clinical data. It also noted that recurring epidemic patterns can be influenced by interactions between pathogen characteristics and environmental factors.

Calafiore et al. (2020) demonstrated the usefulness of compartmental SIRD models for representing changing epidemic dynamics. Their work showed that allowing model parameters to vary across different time periods can improve the representation of distinct epidemic waves while maintaining a relatively simple mathematical structure. Although developed for COVID-19, this framework provided the scientific basis for our study, serving as a practical proxy to model dengue dynamics when fine-scale mosquito data are unavailable.

Fernando et al. (2024**)** examined long-term dengue patterns in Sri Lanka using data covering two decades from 2000 to 2020. Their findings showed that dengue has remained endemic in the country, with major epidemic increases observed during periods such as 2017 and 2019.

Karasinghe et al. (2024) investigated seasonal ARIMA models for dengue forecasting in the Colombo region. Their results demonstrated that time-series forecasting can provide early indications of changes in dengue incidence, potentially supporting public health authorities in planning vector-control activities before increases in clinical cases occur.

Mathematical epidemiology uses deterministic compartmental ODE models to evaluate transmission pathways and threshold dynamics (Martcheva, 2015). Modeling host-pathogen interactions through structured compartments enables the systematic derivation of R_0 and parameter estimation, offering actionable insights for surveillance and intervention design. Paz-Bailey et al. (2025) discussed Wolbacia interventions against dengue.

3 METHODOLOGY

The host population across health states is formulated using a four-compartment ordinary differential equation system under a closed population.

The governing system is (Alsaeed et al., 2020):

$$\frac{dS}{dt} = -\frac{\beta SI}{N},$$
$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I - \mu I, \quad (1)$$

$$\frac{dR}{dt} = \gamma I,$$
$$\frac{dD}{dt} = \mu I,$$

where *S*, *I*, *R* and *D* denote the number of susceptible, active infected, recovered and deceased individuals respectively and the parameters are defined as follows:

- β – Transmission rate
- γ – Recovery rate
- μ - Mortality rate

The basic reproduction number (R_0) was mathematically derived from the SIRD model using the Next Generation Matrix (NGM) method. It represents the average number of new infections generated by one infectious individual in a fully susceptible population (Diekmann et al., 1990; Van den Driessche & Watmough, 2002).

Right hand side of the second differential equation in system (1) can be separated into the rate of new infections (2) and the rate of removal (3):

$$F = \frac{\beta SI}{N}, \quad (2)$$
$$V = (\gamma + \mu)I. \quad (3)$$

At the Disease-Free Equilibrium (DFE), *S*=*N* and *I*=0. Accordingly, the corresponding Jacobian terms at the DFE are:

$$f = \frac{\partial F}{\partial l}\Big|_{S=N} = \beta, \quad (4)$$

$$v = \frac{\partial V}{\partial I} \Big|_{S=N} = \gamma + \mu. \quad (5)$$

The Next Generation Matrix can be obtained as follows:

$$K = fv^{-1}. \quad (6)$$

Since the model has one infected compartment, *K* is a 1×1 matrix. (Martcheva, 2015).

$$R_0 = \rho(fv^{-1}) = \frac{\beta}{\gamma + \mu}, \quad (7)$$

where ρ is the spectral radius.

A value of $R_0 > 1$ denotes that dengue transmission can expand in a susceptible population, where as $R_0 < 1$ denotes that infections will decline towards disease extinction. In an endemic setting, sustained transmission is corroborated when active infection persists continuously above zero over time. $R_0 = 1$ denotes that the outbreak as controlled.

4 DATA COLLECTION AND ANALYSIS

The dengue data used in this study were obtained from the official records of the Medical Officer of Health (MOH) Office in Gampaha, Sri Lanka**.** The data included daily numbers of infected cases, recoveries, and dengue-related deaths over a period of 900 days, from December 14, 2023 to May 31, 2026, as shown in Figure 1.

To capture shifts in transmission dynamics, the 900-day dataset was divided into eight epidemic phases based on seasonal monsoonal transitions combined with observed rapid surges and

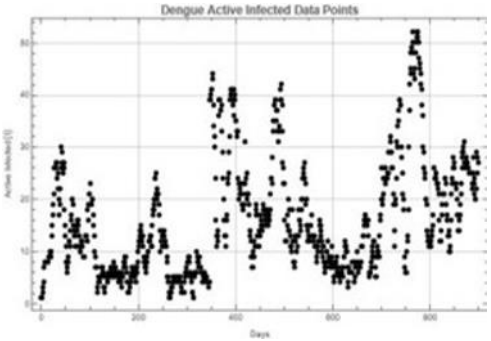


Figure 1: Active infected cases

Table 1: Description of the eight phases.

Phase	Date	Primary Justification and seasonal context
1	Dec,2023-Apr,2024	Northeast Monsoon to First Inter-Monsoon transition
2	May,2024-Jun,2024	Southwest Monsoon onset
3	Jul,2024-Sep,2024	Southwest Monsoon period
4	Oct,2024-Dec,2024	Second Inter-Monsoon to Northeast Monsoon transition
5	Dec,2024-Sep,2025	Northeast Monsoon through First Inter-Monsoon and Southwest Monsoon
6	Sep,2025-Dec,2025	Southwest Monsoon to second Inter-Monsoon and North east Monsoon transition
7	Jan,2026-Feb,2026	Northeast Monsoon period
8	Feb,2026 -May,2026	Northeast Monsoon to first inter-Monsoon and Southwest Monsoon onset

trajectory inflection points in daily active inflections. Each wave was analyzed separately using a mathematical tool, where the SIRD model was fitted to the observed daily active infected cases using the NonlinearModelFit function. The model parameters β, γ, and μ were estimated for each wave to capture variations in dengue transmission dynamics over time as shown in Table 1. Active infected cases were defined as the period a person remains actively sick from the date they become infected until the date they recover or die. This was calculating using individual infected dates and recovery dates from the record.

5 RESULTS AND DISCUSSION

The SIRD model was applied separately to the eight epidemic waves identified from the 900-day dengue dataset. For each wave, the model parameters β, γ, and μ were estimated, and the corresponding basic reproduction number (R_0) was calculated. The results obtained for all eight phases are presented in Table 2.

Table 2: Estimated SIRD Transmission Parameters (β, γ) and Basic Reproduction Numbers (R_0) Across the Eight Epidemic Phases.

Phase	β	γ	R_0
1	0.2344	0.0774	3.0284
2	0.1827	0.0845	2.1621
3	0.2725	0.1054	2.5853
4	0.4010	0.0959	4.1814
5	0.3421	0.0644	5.3121
6	0.1500	0.1199	1.2510
7	0.4024	0.0539	7.4243
8	0.1500	0.1146	1.3089

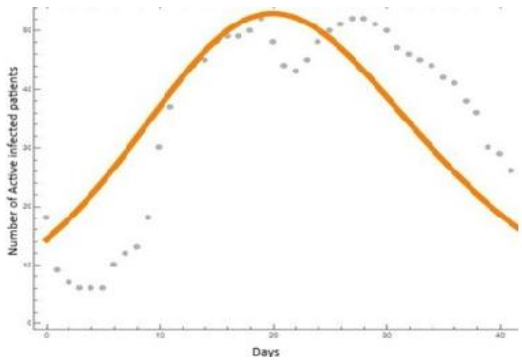


Figure 2: Optimized SIRD model fit for phase 7 active dengue cases

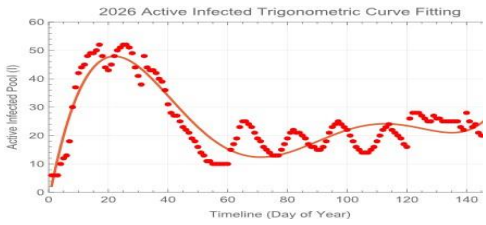


Figure 3: Active infected trigonometric curve.

Initial conditions as in beginning, set $S_0 = N$ with I_0 initialized from recorded cases. For phases 2-8, the final values of each phase served as the initial conditions for the next to ensure dynamic continuity across all 900 days. The dengue mortality rate found between 0.0001 and 0.0005, this say that in this area dying is minimized.

As shown in Table 2, the basic reproduction number (R_0) remained strictly greater than 1.0 across every single phase over the entire 900-day evaluation. This critical finding indicates that dengue transmission was never fully eliminated in the Gampaha MOH area, confirming that the region experiences continuous and sustainable transmission potential across all eight phases. Among all phases, the highest transmission intensity occurred during Phase 7 (Jan 01, 2026 – Feb 20, 2026), recording a peak basic reproduction number of $R_0 = 7.4243$. The calibrated SIRD model showed high in sample goodness-of-fit and calibration

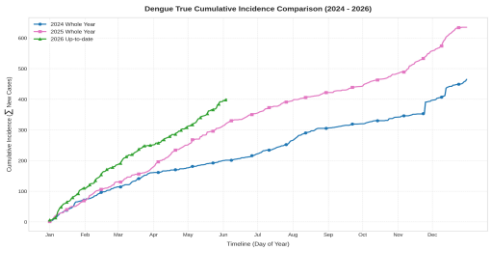


Figure 4: 2024, 2025 and 2026 cumulative dengue cases in Gampaha MOH area.

accuracy for Phase 7, achieving an R^2 value of 0.9915 alongside low error scores (RMSE= 2.5185, MAE = 2.0513) (Chai & Draxler, 2014). The model fit for this high-intensity period is illustrated in Figure 2.

According to Table 1, Phase 7 shows a period in which clinical data showed a major expansion of dengue cases in the area. The rapid climb to peak caseloads confirmed that 2026 experienced an accelerated transmission surge compared to previous years shown in Figure 4.

To capture the detailed daily oscillations and cyclical behavior during the first five months of 2026, a complementary second-order Fourier series model with a linear trend was developed as shown in Figure 3 and the corresponding model is given as follows:

$$\begin{aligned} I(t) &= -1383.53 + 16.908t \\ &- 190.91 \sin\left(\frac{2\pi t}{365}\right) \\ &+ 1430.54 \cos\left(\frac{2\pi t}{365}\right) \\ &- 236.89 \sin\left(\frac{4\pi t}{365}\right) \\ &- 50.048 \cos\left(\frac{4\pi t}{365}\right). \end{aligned} \quad (8)$$

Here $I(t)$ represents the daily active dengue cases and t denotes the day of the year. The fitted trigonometric model achieved an Adjusted R^2 of approximately 0.7944 (79.44%), confirming that nearly 80% of the daily case variations in early 2026 were successfully captured. As displayed in Figure 3, this harmonic curve. Effectively tracks the rapid drop following the January peak, the mid-year endemic fluctuations, and the upward pre-monsoonal climb in late May.

6 CONCLUSION

This study used a data-driven SIRD model and a second-order Fourier model to study dengue transmission in the Gampaha MOH area over a 900-day period from December 14, 2023 to May 31, 2026. The results showed that the R_0 was greater than 1 in all eight phases. This indicates that dengue transmission remained active throughout the study period. The highest transmission level was observed in Phase 7.

The second-order Fourier model with linear trend given in (8) also provided a useful representation of the dengue pattern in 2026, explaining approximately 79.44% of the variation in daily cases. The estimated mortality rate remained very low across the phases. Overall, the results show that dengue transmission in the Gampaha MOH area is persistent and varies over time. The proposed phase calibrated modeling framework accurately tracks historical transmission dynamics and temporal surges, providing valuable quantitative insights to assist health authorities in evaluating intervention optimizing resource allocation. According to that as a future enhancement this research good foundation to add rainfall factors like rainfall, temperature in the model. In addition, mosquito population data can be included as a separate compartment to develop a host-vector model.

REFERENCES

- Alsaeed, N., Siddiqui, M., & Alqaissi, E. (2020). An agent-based simulation of the SIRD model of COVID-19 spread. *International Journal of Biology and Biomedical Engineering.* 14, 210-217.
- Calafiore, G. C., Novara, C. & Possieri, C. (2020). A time-varying SIRD model for the COVID-19 contagion in Italy. Annual reviews in control, 50, 361-372.
- Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature, Geosci. Model Dev., 7, 1247–1250
- Diekmann, O., Heesterbeek, J. A. P., & Metz, J. A. J. (1990). On the definition and the computation of the basic reproduction ratio R₀ in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, 28(4), 365–382.
- Epidemiology Unit. (2024). Sri Lanka Weekly Epidemiological Report. Epidemiology Unit Sri Lanka.
- Fernando, L., Wickramasinghe, V. P., Kalra, P., Kastner, R. & Gallagher, E. (2024). Epidemiological burden of dengue in Sri Lanka: A systematic review of literature from 2000-2020. IJID regions, 13, 100436.
- Karasinghe, N., Peiris, S., Jayathilaka, R., & Dharmasena, T. (2024). Forecasting weekly dengue incidence in Sri Lanka: Modified autoregressive integrated moving average modeling approach. PLoS One, 19(3), e0299953.
- Martcheva, M. (2015). *An Introduction to Mathematical Epidemiology*, Springer.
- Nishiura, H. (2006). Mathematical and statistical analyses of the spread of dengue, Dengue Bulletin, 30, 51-67
- Paz-Bailey, G., Jernigan, D. B., Laserson, K., Zielinski-Gutierrez, E., & Petersen, L. (2025). New solutions against the dengue global threat: Opportunities for *Wolbachia* interventions. *International Journal of Infectious Diseases, 157*, Article 107923.
- Van den Driessche, P. & Watmough, J. (2002). Reproduction numbers and sub threshold endemic equilibria for compartmental models of disease transmission. Mathematical Biosciences, 180(1–2), 29–48.

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research –2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Developing an RSA-AES Hybrid Encryption Framework for Secure Student Results: A Case Study of the Faculty of Applied Sciences, Wayamba University of Sri Lanka

Shreen MT[1], Dahanayaka SD[2]
[1,2]*Department of Mathematical Sciences, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*
shreenthasleem@gmail.com[1], shalikadahanayaka@wyb.ac.lk[2]

ABSTRACT

This paper focuses on developing and analyzing a hybrid encryption model, combining Rivest-Shamir-Adleman (RSA) and Advanced Encryption Standard (AES), with an additional password-authentication mechanism, to secure students' examination results in the Faculty of Applied Sciences, Wayamba University of Sri Lanka (FAS-WUSL). Two problems were identified in the existing result-publication process: stored data were not encrypted, and the non-confidential student registration number was used as the sole means of access. The proposed system, built using Python and MySQL, assigns unique RSA-2048 keys to all students. Examination results for each semester are encrypted using AES-256 in EAX, an authenticated-encryption with associated data (AEAD) mode, while the AES session key is encrypted using the student's RSA public key. Access is granted only after authentication with the registration number and a unique hash generated from an individually salted SHA-256 password. Performance testing showed that AES-256 encryption and decryption of a semester record averaged 0.47 ms and 0.23 ms, respectively, across 50 repeated trials. Repeated file-system access for key retrieval contributed substantially to hybrid-pipeline latency compared with operation using pre-loaded keys. Comparison of RSA key-generation time across 1024-, 2048-, and 4096-bit keys showed that RSA-2048 required substantially less key-generation time than RSA-4096 under the experimental conditions, while a comprehensive security-strength evaluation was outside the study's scope. A targeted authentication test confirmed that a valid registration number paired with an incorrect password was rejected, showing that registration-number knowledge alone is insufficient to access another student's results.

KEYWORDS: AES-256, Authentication, Hybrid encryption, RSA-2048, Student records

1 INTRODUCTION

The use of digital systems to manage university examination results has become common practice; however, digital publication of results raises security concerns due to the risk of unauthorized access. At FAS-WUSL, results were commonly published as scanned or photocopied result sheets without an authentication process. A specific security vulnerability arose because students' registration numbers were not confidential but served as unique identifiers. Consequently, using a registration number alone to authenticate a result request could allow a student to access another student's results.

Results stored in plaintext could also be exposed following a database security breach. Encryption and authentication address different security requirements: encryption protects confidentiality, while authentication verifies the identity of the requester. This study therefore developed and evaluated a hybrid RSA-AES encryption scheme supported by independent salted-hash password authentication for securing student examination results at FAS-WUSL. Specifically, the study aimed to encrypt semester results using AES-256, protect the

AES session keys using RSA-2048, implement salted SHA-256 password authentication, simulate the academic environment across eight degree programmes using Grade Point Average (GPA) and Cumulative Grade Point Average (CGPA), and evaluate the security and performance of the developed system.

2 LITERATURE REVIEW

AES is a symmetric cipher standardized by the National Institute of Standards and Technology (NIST), offering resistance to cryptanalysis and efficient processing (National Institute of Standards and Technology, 2001). Since AES uses a shared key, RSA can be used for secure key distribution. Fujisaki et al. (2001) demonstrated that RSA with Optimal Asymmetric Encryption Padding (RSA-OAEP) provides resistance to adaptive chosen-ciphertext attacks and is suitable for session-key encryption. Dworkin (2007) described Galois/Counter Mode (GCM) as an authenticated-encryption mode that uses an authentication tag to detect unauthorized modifications.

EAX is an authenticated-encryption with associated data (AEAD) mode providing confidentiality and authentication through an authentication tag. Bellare et al. (2004) introduced EAX as a mode combining encryption and message authentication. In this study, EAX was selected for AES-256 encryption of examination-result records. RSA key length also affects the balance between security and computational cost, with larger keys requiring greater computational resources (Barker & Roginsky, 2019).

Akter et al. (2023) demonstrated the use of hybrid RSA-AES encryption to combine AES efficiency with RSA key-distribution security. FERPA requires educational records to be protected from unauthorized access, highlighting the need for both encryption and authentication (U.S. Department of Education, 2024). Nawusu et al. (2024) similarly identified security concerns in student academic record systems, but did not specifically examine hybrid RSA-AES protection for examination results.

Provos and Mazières (1999) emphasized the use of salted password hashing. In this prototype, individually salted SHA-256 was used for password verification; however, modern guidance recommends adaptive or memory-hard password-hashing schemes for production systems (Turan et al., 2010; National Institute of Standards and Technology, 2025). Cryptographic key management is also essential because compromise of RSA private keys could expose encrypted session keys (Barker, 2020).

Although previous studies provide foundations for encryption, key exchange, and academic-record security, limited attention has been given to the vulnerability of using a non-confidential registration number as the sole access credential for student results. This study therefore investigates a hybrid RSA/AES system with independent password authentication to address this access-control weakness.

3 METHODOLOGY

The study used a quantitative experimental design based on the FAS-WUSL case study. Students were authenticated using individually salted SHA-256 password hashes. Each semester result was encrypted using AES-256 in EAX mode, while the AES session key was encrypted using the student's RSA-2048

public key with RSA-OAEP. The encrypted result, session key, nonce, and authentication tag were stored in MySQL.

The system was developed using Python, PyCryptodome, hashlib, secrets, MySQL, and Streamlit. RSA keys were tested both through persistent storage and when pre-loaded into memory to distinguish file-system access time from cryptographic processing. Production deployment would require secure private-key storage, access control, key rotation, backup, and recovery (Barker, 2020).

Experiments were conducted on a Dell Inspiron 3501 laptop with an 11th Gen Intel® Core™ i5-1135G7 processor, 8 GB RAM, and Windows 11 Home 64-bit. The software environment used Python 3.13.11, PyCryptodome 3.23.0, MySQL, and Streamlit 1.58.0.

Performance was evaluated using RSA key-generation, RSA-only, AES-only, and hybrid encryption/decryption times. RSA key generation was measured once for 1024-, 2048-, and 4096-bit keys, while RSA-only, AES-only, and hybrid tests were repeated 50 times using perf_counter() and averaged. Hybrid tests used both disk-loaded and pre-loaded keys. Data-size impact was tested using 1, 5, 10, and 20 KB records. The simulation included 30 students and 220 encrypted semester records.

4 RESULTS AND DISCUSSION

Table 1 shows RSA key - generation time across the three tested key sizes. Generation time increased sharply with key length, reaching 22,651.10 ms at 4096 bits RSA-2048 was selected for the prototype based on its measured key-generation time and the adopted key-size and guidance (Barker & Roginsky, 2019).

Table 1: RSA Key Generation Time

Key Size	Generation Time (ms)
1024-bit	61.24
2048-bit	260.56
4096-bit	22,651.10

A comprehensive cryptographic security-strength evaluation was outside the study's scope. AES-256 in EAX mode encrypted and decrypted a semester record in under one millisecond on average (0.47 ms and 0.23 ms). Table 2 compares hybrid performance with RSA keys read from disk and pre-loaded into memory. Decryption time decreased from 48.15 ms to 4.32 ms when keys were cached, showing that most of the latency was caused by repeated file-system access rather than cryptographic processing.

For the 5–20 KB records tested, encryption time remained approximately 2.2–2.3 ms. All authentication trials using a valid registration number with an incorrect password were rejected, confirming that the registration number alone was insufficient for access. Because encrypted rather than plaintext results were stored, a database-only compromise would not directly expose results in plaintext; however, production security would also depend on RSA private -key protection, application credentials, database access controls, and key management. Caching RSA key material in server memory could reduce decryption latency compared with repeated disk retrieval.

The study was limited by its use of simulated data, a single development machine, and the absence of concurrent network testing. Additionally, salted SHA-256 was used for password verification; although salting helps protect against

Table 2: Hybrid RSA-AES Performance

Condition	Encrypt (ms)	Decrypt (ms)
Disk I/O per call	1.77	48.15
Keys pre-loaded	0.87	4.32

precomputed attacks, SHA-256 is not designed specifically for password storage. Future work should evaluate an adaptive or memory-hard password-hashing scheme such as Argon2id or appropriately configured PBKDF2.

5 CONCLUSION

This study developed and evaluated a student-specific hybrid RSA-AES encryption prototype with independent password authentication for simulated FAS-WUSL examination results. Under the experimental conditions, RSA-2048 was selected based on its lower key-generation time than RSA-4096, while AES-256 in EAX mode achieved sub-millisecond encryption and decryption times. Pre-loading RSA keys into memory substantially reduced hybrid decryption latency compared with repeated disk retrieval. Authentication testing showed that a registration number alone was insufficient to pass the implemented authentication check. The study was limited to simulated data and a single development environment; therefore, future work should investigate concurrent access, secure key management, modern password-hashing schemes, and web-based deployment.

REFERENCES

- Akter, R., Khan, M. A. R., Rahman, F., Soheli, S. J., and Suha, N. J. (2023). RSA and AES Based Hybrid Encryption Technique for Enhancing Data Security in Cloud Computing. International Journal of Computational and Applied Mathematics & Computer Science, 3, 60–71. https://doi.org/10.37394/232028.2023.3.8
- Barker, E. (2020). *Recommendation for Key Management: Part 1: General* (NIST Special Publication 800-57 Part 1 Revision 5). National Institute of Standards and Technology.
- Barker, E., and Roginsky, A. (2019). *Transitioning the use of cryptographic algorithms and key lengths* (NIST Special Publication 800-131A Revision 2). National Institute of Standards and Technology.
- Bellare, M., Rogaway, P., and Wagner, D. (2004). The EAX mode of operation. In Fast Software Encryption 2004 (pp. 389–407). Springer. https://doi.org/10.1007/978-3-540-25937-4_25
- Dworkin, M. J. (2007). Recommendation for block cipher modes of operation: Galois/Counter Mode (GCM) and GMAC (NIST Special Publication 800-38D). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-38D
- Fujisaki, E., Okamoto, T., Pointcheval, D., and Stern, J. (2001). RSA-OAEP is secure under the RSA assumption. In J. Kilian (Ed.), Advances in Cryptology - CRYPTO 2001 (pp. 260-274). Springer. https://doi.org/10.1007/3-540-44647-8_16
- National Institute of Standards and Technology. (2001). Advanced Encryption Standard (AES) (Federal Information Processing Standards Publication 197). U.S. Department of Commerce. https://doi.org/10.6028/NIST.FIPS.197
- National Institute of Standards and Technology. (2025). *Digital identity guidelines: Authentication and lifecycle management* (NIST Special Publication 800-63B-4). National Institute of Standards and Technology.
- Nawusu, Y. A.-W., Danaa, A. A. A., Mumin, D., and Ismail, S. (2024). Student academic record systems and their security issues. *Asian Journal of Research in Computer Science, 17*(5), 190–200. https://doi.org/10.9734/ajrcos/2024/v17i5448
- Provos, N., and Mazieres, D. (1999). A future-adaptable password scheme. In Proceedings of the 1999 USENIX Annual

Technical Conference (pp. 81-91). USENIX Association.

- Turan, M. S., Barker, E., Burr, W., and Chen, L. (2010). Recommendation for password-based key derivation: Part 1: Storage applications (NIST Special Publication 800-132). National Institute of Standards and Technology.
- U.S. Department of Education. (2024). Family Educational Rights and Privacy Act (FERPA). https://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research –2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Comparative Study of Analytical and Numerical Methods for Some Trigonometric Integrals

Tharushika AHP[1], De Silva TMM[2]
[1,2]*Department of Mathematics, Faculty of Science, University of Kelaniya, Kelaniya, Sri Lanka*
ahpt261@kln.ac.lk[1], mihiris@kln.ac.lk[2]

ABSTRACT

The evaluation of trigonometric integrals can be challenging when conventional calculus techniques become ineffective due to oscillatory behaviour, singularities, or complex rational structures. This study presents a comparative framework of analytical and numerical strategies for evaluating trigonometric integrals, focusing on analytical methods, Cauchy's Residue Theorem (CRT), the Logarithmic Residue Theorem (LRT), and numerical approaches, including Composite Simpson's rules, and Gaussian quadrature. The analytical methodology transforms trigonometric integrals into contour integrals over the unit circle, followed by systematic identification of poles, residue computation, and contour evaluation. Numerical methods implemented in MATLAB are employed to obtain approximations and assess computational errors against exact residue-based solutions. Results demonstrate that the effectiveness of each method depends on the structural complexity of the integrand and the required accuracy. Residue-based methods provide exact and theoretically rigorous solutions, while Gaussian quadrature generally achieves higher numerical accuracy than Composite Simpson's rules, particularly for oscillatory and computationally demanding integrals. Furthermore, LRT offers computational advantages for integrals involving logarithmic singularities and multiple zeros or poles. For the six examples considered, the study concludes that residue-based and numerical methods are complementary, and the results guide the choice of an integration strategy based on the integrand's structure and the required accuracy.

KEYWORDS: Cauchy's Residue Theorem, Gaussian Quadrature Rule, Logarithmic Residue Theorem, Simpson's Rule, Trigonometric Integrals

1 INTRODUCTION

Trigonometric integrals of the form

$$\int_0^{2\pi} R(\sin\theta, \cos\theta)d\theta, \qquad (1)$$

where R is a rational function, appear in signal processing, quantum mechanics, electrical engineering, and Fourier analysis (Brown & Churchill, 2009). Evaluating them directly with basic calculus is difficult because the functions may oscillate rapidly, have singularities, or exhibit complex rational structures. Complex analysis provides an alternative: Cauchy's Residue Theorem turns the realintegral into a contour integral in the complex plane (Ren, 2023; Chen et al., 2023). The Logarithmic Residue Theorem extends CRT by relating the contour integral of f'/f to the numbers of zeros and poles of a meromorphic function f (Miao, 2024). Complementarily, numerical methods such as composite Simpson's rules and Gaussian quadrature provide general-purpose strategies that give approximations rather than exact values (Atkinson, 1989; Masud et al., 2024). While extensive literature exists on the individual application of residue theorems or numerical quadrature to trigonometric integrals, in the literature reviewed, we did not find a study that has systematically approached the same trigonometric integral using Cauchy's

Residue Theorem, the Logarithmic Residue Theorem, and several numerical methods and then measures the numerical error against the exact residue results. This study addresses this gap by presenting six worked examples and conducting a direct, side-by-side comparison of analytical and numerical methods. The scope is limited to these examples: Example $1-3$ are integrals over $[0,2\pi]$ with rational integrands and simple poles, Example 4 and 5 are contour integrals of meromorphic functions over $|z|=1$, and Example 6 is an improper oscillatory integral. The main objectives of the study are to:

(1) Use the residue theorem to calculate trigonometric integrals and conduct a comparison with numerical integration methods.
(2) Investigate the Logarithmic Residue Theorem for integrals with logarithmic singularities and analyse its comparative advantages or limitations relative to Cauchy's Residue Theorem.
(3) Conduct a systematic performance analysis of both analytical and numerical methods and synthesize the findings into a practical framework for selecting the optimal integration strategy based on integral complexity and precision requirements.

2 LITERATURE REVIEW

2.1 Residue Theorems

The Residue Theorem follows from the Cauchy–Goursat theorem (Orloff, 2018) and the Laurent series; Qin (2023) proved it by excluding each singular point with a small circle in a multiply connected region. It states the following (Brown & Churchill, 2014).

Theorem (Residue Theorem). Let D be a simply connected domain and C a positively oriented simple closed contour lying in D. If f is analytic on and inside C except at finitely many singular points $z_1, \dots, z_n$ inside C, then

$$\oint_C f(z)\,dz = 2\pi i \sum_{k=1}^{n} \operatorname{Res}(f, z_k). \qquad (2)$$

Ren (2023) uses this result with $z = e^{i\theta}$ to evaluate trigonometric integrals, and Chen et al. (2023) computed residues by partial fractions and by power-series expansion. Huang (2023) evaluated $\int_0^{2\pi} \frac{\cos\theta}{5-\cos\theta} d\theta = \pi/6$ in this way (Example 2 below).

For a function f that is meromorphic inside C and analytic and non-zero on C, the LRT states that (Miao, 2024)

$$\frac{1}{2\pi i}\oint_C \frac{f'(z)}{f(z)} dz = N - P, \qquad (3)$$

where N and P are the numbers of zeros and poles of f inside C, counted with multiplicity. Miao (2024) extended this result by inserting a function φ that is analytic on and inside C. If f has zeros a_k of order m_k and poles b_j of order n_j inside C, then

$$\frac{1}{2\pi i}\oint_C \varphi(z)\frac{f'(z)}{f(z)} dz = \sum_{k=1}^{p} m_k \varphi(a_k) - \sum_{j=1}^{q} n_j \varphi(b_j). \qquad (4)$$

Numerical integration techniques serve as essential complementary tools when analytical methods are impractical. Masud et al. (2024) compared classical quadrature rules on smooth test

integrands. They reported that Gaussian quadrature usually gives the highest accuracy for a given number of function evaluations, that Simpson's $1/3$ and $3/8$ rules, which use quadratic and cubic interpolation, are less accurate but simpler, and that Gaussian quadrature is less effective near singularities. Atkinson (1989) and Davis & Rabinowitz (1984) gave the theory of Gaussian quadrature; Davis and Rabinowitz also discussed the fast convergence of equally spaced rules for smooth periodic integrands integrated over a full period, which is relevant to Examples $1-4$.

3 METHODOLOGY

3.1 Analytical framework

- Complex substitution: Transform the real trigonometric integral into a complex contour integral using $z = e^{i\theta}$, converting $d\theta$ to $dz\, iz$ (Chen et al., 2023).
- Pole identification: Locate and classify singularities (poles) of the transformed inte grated within the unit circle $|z| = 1$ (Chen et al., 2023).
- Residue calculation:

– Simple poles:

$$\text{Res}(f, z_0) = \lim_{z \to z_0} (z - z_0)\, f(z). \quad (5)$$

– Logarithmic Residue theorem (Miao, 2024): $g'(z)\,/g(z)$ at $z = b$ counts zeros/poles.

$$\text{Res}\left(\frac{g'}{g}, b\right) = \begin{cases} n, (\text{if } b \text{ is a zero of multiplicity } n) \\ -m, (\text{if } b \text{ is a pole of multiplicity } m). \end{cases}$$

- Contour integration (Ren, 2023; Huang, 2023):

– Apply the Residue theorem to obtain exact solutions:

$$\oint_C f(z)dz = 2\pi i \sum \text{Res}\,(f, z_k). \quad (6)$$

– Logarithmic Residue theorem: for integrals that can be expressed as

$$\oint_C \frac{f'(z)}{f(z)}\, dz \quad (7)$$

relate to the results back to the $f(z)$ within C's zeros/poles.

3.2 Numerical framework

All numerical values were computed in MATLAB R2024a (The MathWorks Inc., 2024). Examples $1-3$ were integrated in their original real form over $[0, 2\pi]$. The contour integrals of Examples 4 and 5 were integrated along the parametrization $z = e^{i\theta},\ dz = ie^{i\theta}d\theta,$ so that $\oint_C f(z)dz = \int_0^{2\pi} f(e^{i\theta}) i\, e^{i\theta} d\theta.$ Example 6 was integrated over $[0, \infty)$.

Composite Simpson's rules (Masud et al., 2024**):**

$1/3$ Rule: $\int_a^b f(x)\, dx \approx$

$$\frac{h}{3}\left[f(x_0) + 4\sum_{\text{odd}} f(x_i) + 2\sum_{\text{even}} f(x_i) + f(x_n)\right], \quad (8)$$

$3/8$ Rule: $\int_a^b f(x)\, dx \approx$

$$\frac{3h}{8}\left[f(x_0) + 3\sum_{i=1}^{n-1} f(x_i) + 2\sum_{i=1}^{n-2} f(x_i) + f(x_n)\right]. \quad (9)$$

Gaussian Quadrature (Atkinson, 1989; Masud et al., 2024): Gaussian quadrature uses optimally selected nodes and weights for accurate numerical integration with fewer function evaluations. Gauss-Kronrod quadrature extends this approach by providing an error estimate through comparison of two approximations. In this study, MATLAB's built-in integral function, which uses adaptive Gauss–Kronrod quadrature, was employed to evaluate the integrals with specified error tolerances.

Table 1: Absolute error of each numerical integration method

	Composite Simpson's 1/3	Composite Simpson's 3/8	Gauss-Kronrod quadrature
Example 01	1.55×10^{-15}	1.55×10^{-15}	0
Example 02	7.77×10^{-16}	7.77×10^{-16}	3.44×10^{-10}
Example 03	4.02×10^{-17}	4.02×10^{-17}	7.90×10^{-11}
Example 04	2.06×10^{-16}	2.06×10^{-16}	6.33×10^{-17}
Example 05	–	–	0
Example 06	–	–	9.12×10^{-5}

3.3 Comparative Analysis

Absolute error is computed as

$$\text{Abs. Err.} = |\text{Approx.} - \text{Exact}| \quad (10)$$

for each test integral. This allows a quantitative comparison of the accuracy of each numerical method against the residue calculus.

4 RESULTS AND DISCUSSION

Example 01: $\int_0^{2\pi} \frac{d\theta}{\sin\theta + \cos\theta + 9}$. Use $z = e^{i\theta}$. The integral turns into a contour integral around the unit circle. Only the pole $z = \frac{-9+\sqrt{79}}{1-i}$ lies inside (Figure 1), and its residue gives the exact value. Exact answer: $2\pi/\sqrt{79}$.

Example 02: $\int_0^{2\pi} \frac{\cos\theta}{5 - 3\cos\theta} d\theta$. Inside the unit circle (Figure 2), we have poles at $z = 0$ and $z = {}^1/_3$. Sum their residues. Exact answer: $\pi/6$ (Huang, 2023).

Example 03: $\int_0^{2\pi} \frac{\cos\theta}{2 + \sin\theta} d\theta$. The simple poles $z = 0$ and $z = (\sqrt{3} - 2)i$ lie inside the unit circle and $z = -(\sqrt{3} + 2)i$ lies outside (Figure 3). The residues

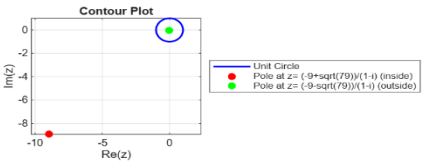
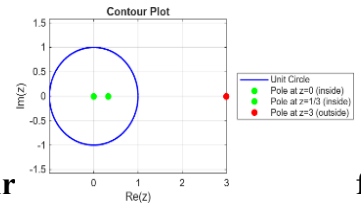


Figure 1: Contour visualization for Example 1



Figur for Example 2

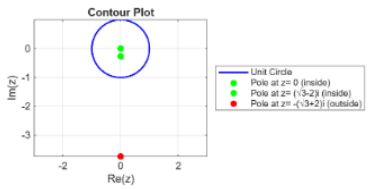


Figure 3: Contour visualization for Example 3

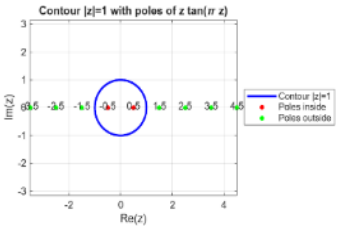


Figure 4: Contour visualization for Example 4

are −1 and +1, which have equal magnitude and opposite sign, so exact answer: 0.

Example 04: $\int_{|z|=1} z\tan(\pi z)\, dz$. Notice $\tan(\pi z) = \frac{\sin(\pi z)}{\cos(\pi z)}$. Here, LRT works nicely. The only singularities come from $\cos(\pi z) = 0$ at $z = \pm 1/2$, and those inside the unit circle (Figure 4), Add up the residues. Exact answer: 0.

Example 05: $\int_{|z|=1} z^n \tan(\pi z)\, dz$. The zeros at $z = \pm 1/2$ gives residues that depend on whether n is even or odd (Miao, 2024). Exact answer: 0 when n is odd; $-4i(1/2)^n$ when n is even.

Example 06: $\int_0^{\infty} \frac{x \sin x}{x^2+3} dx$. This is an improper integral with an oscillatory term. Extend to the complex plane and close the contour with a large semicircle in the upper half-plane. The integrand $\frac{ze^{i\pi}}{z^2+3}$ has a simple pole at $z = \sqrt{3}i$. Exact answer: $\frac{\pi}{2}e^{-\sqrt{3}}$.

4.1 Error Analysis

The accuracy of the numerical integration methods was evaluated by comparing their errors, as presented in Table 1. These results demonstrate that the relative accuracy of the numerical methods depends on the test integral and the integration method employed.

5 CONCLUSION

This study evaluated six trigonometric integrals using residue-based analytical methods and numerical techniques. CRT and LRT provided exact analytical solutions, while numerical methods produced accurate approximations whose performance depended on the integrand's characteristics and the numerical scheme used. Both CRT and LRT yield exact solutions for trigonometric integrals, but their effectiveness depends on the integrand's structure: CRT is particularly useful for integrals with isolated poles and offers clear geometric interpretation, while LRT can simplify problems involving complex zeros, poles, and their multiplicities. The two residue theorems serve as complementary analytical tools, with the choice of the method determined by the integrand's characteristics. Overall, the structure and analytical properties of the integrand should guide the selection of an appropriate evaluation method. For rational trigonometric integrals satisfying the conditions for residue-based analysis, analytical methods provide exact results and additional structural insight, while numerical methods remain useful for verification and for cases where analytical evaluation is difficult or unavailable. Although demonstrated these observed patterns provide a basis for extending the investigation to a broader class of trigonometric integrals in future work.

REFERENCES

- Atkinson, K. E. (1989). *An Introduction to Numerical Analysis*, 2nd ed. New York: John Wiley & Sons.
- Brown, J. W. and Churchill, R. V. (2009). *Complex Variables and Applications*. New York: McGraw-Hill.
- Chen, Y., Lu, Y. and Wang, Z. (2023). Residue theorem and its application. *Highlights in Science, Engineering and Technology*,*38*,290-299. https://doi.org/10.54097/hset.v38i.5819
- Davis, P. J. and Rabinowitz, P. (1984). *Methods of Numerical Integration*, 2nd ed. Orlando, FL: Academic Press.
- Huang, Y. (2023). Application and extension of residue theorem. *Highlights in Science, Engineering and Technology*, *72*, 971-975. https://doi.org/10.54097/p2mvpv42
- Masud, M., Shimi, F. N. and Gope, R. C. (2024). Numerical integration techniques: A comprehensive review. *International Journal of Innovative Science and Research Technology*, 2744-2755.
- Miao, S. (2024). Residue theorem and logarithmic residue theorem. *Science and Technology of Engineering, Chemistry and Environmental Protection*, *1*. https://doi.org/10.61173/xctcqz13
- Orloff, J. (2018). *Complex Variables with Applications*. Massachusetts Institute of Technology, LibreTexts.

- Qin, Z. (2023). Proof and application of Cauchy's residue theorem. *Highlights in Science, Engineering and Technology*, *72*, 873-878. https://doi.org/10.54097/zfg4j539
- Ren, Z. (2023). Application and analysis of the residue theorem. *Highlights in Science, Engineering and Technology*, *38*,643-646. https://doi.org/10.54097/hset.v38i.5898
- The MathWorks Inc. (2024). *MATLAB, Version 24.1 (R2024a)*. Natick, Massachusetts: The MathWorks Inc.

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research –2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Examination Scheduling Using Integer Linear Programming: A Case Study of the Faculty of Applied Sciences, Wayamba University of Sri Lanka

Hashini AI[1], Dahanayaka SD[2]

[1,2]*Department of Mathematical Sciences, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka.*

imalkahashini20@gmail.com[1], shalikadahanayaka@wyb.ac.lk[2]

ABSTRACT

Examination timetabling is a complex scheduling problem because examinations must be arranged across available days, sessions, and venues while meeting academic and institutional requirements. At the Faculty of Applied Sciences (FAS), Wayamba University of Sri Lanka (WUSL), examination timetables are prepared manually, requiring considerable time and effort due to student conflicts, limited examination venues, laboratory requirements, and departmental restrictions. In this study, an Integer Linear Programming (ILP) model was developed to prepare feasible examination timetables by considering theory and practical examinations, project presentations, different types of student conflicts, venue capacities, laboratory availability, and departmental requirements. Existing examination timetabling models are generally designed for specific institutional requirements and do not directly address the combination of student conflict categories and resource restrictions considered in the FAS context. The main objective was to reduce the total number of examination days while discouraging relaxed conflict pairs from being scheduled on the same day through a small penalty. The model was implemented in MATLAB and tested using examination data from two academic semesters. The examination period was reduced from 16 to 14 days in Semester 1 and from 20 to 19 days in Semester 2 while satisfying the required scheduling constraints. The optimality gaps were 0% and 0.26%, with solution times of 31.25 and 32.52 seconds, respectively. These results demonstrated that the proposed ILP model could provide a practical decision-support approach for improving manual examination scheduling at the FAS.

KEYWORDS: Examination timetabling, Integer Linear Programming, Scheduling, Student conflicts

1 INTRODUCTION

Examination timetabling is an important administrative activity in universities that involves assigning examinations to available days, sessions, and venues while satisfying academic and institutional requirements. At the FAS, WUSL students follow General, Joint Major, and Special degree programs with different subject combinations. Some students also sit examinations for previous-year subjects while continuing at their current academic level, resulting in various types of student conflicts. Consequently, manual timetable preparation requires considerable time and effort to manage these conflicts while considering available venues, laboratories, and departmental requirements. Although various examination timetabling approaches had been proposed, institutional requirements vary considerably. This study developed an ILP model to generate examination timetables for FAS while considering student conflicts, examination types, and resource limitations.

2 LITERATURE REVIEW

Examination timetabling had been addressed using several approaches, including graph-based methods, evolutionary algorithms, and mathematical optimization. Bania and Duarah (2018) focused on representing examination conflicts through graph coloring by representing examinations as vertices and conflicts as edges, whereas the present study required a formulation that could incorporate conflict, resource, practical, and departmental restrictions simultaneously.

Mohammed et al. (2022) used a genetic algorithm to search for timetables by evaluating conflicts and constraint violations. Such evolutionary approaches can handle complex search spaces, but their solutions depend on the design of the algorithm and evaluation process. Compared with graph-based and evolutionary approaches, ILP was selected in this study because the required scheduling rules could be represented explicitly as mathematical constraints and incorporated within a single optimization model.

Laisupannawong and Sumetthapiwat (2024) developed an ILP model to minimize examination days and improve room utilization. Al-Hawari et al. (2020) used a three-phase ILP approach for time-slot assignment, examination-day scheduling, and hall allocation, while Aslan et al. (2017) applied binary integer programming to examination-time and classroom allocation. These studies show that ILP can incorporate multiple scheduling requirements within a structured optimization framework.

Although previous studies demonstrated the usefulness of ILP and other optimization approaches, their formulations were developed for specific institutional requirements. The combination of different student conflict categories, practical examinations, project presentations, and resource limitations in the FAS context was not directly addressed in these studies. This gap motivated the development of a faculty-specific ILP model for the examination scheduling requirements of the FAS, WUSL.

3 METHODOLOGY

The required examination data were collected from the FAS, WUSL. The data used in this study were obtained from the first semester examination of the academic year 2023/2024 and the second semester examination of the academic year 2022/2023. The main examination venues were the Library Hall (LH), Main Hall (MH) and Lecture room 09 (LR 09). Additional lecture rooms and laboratories were used where required according to examination type and capacity. The ILP model was implemented in MATLAB R2026a using the intlinprog optimization solver with the HiGHS solver to obtain the examination timetable.

3.1 Conflict Matrix Construction and Conflict Classification

A Conflict Matrix (CM) was constructed using the examination registration data by comparing the student lists of every pair of examinations. The elements of the conflict matrix were defined as:

$CM_{i,j} = 1$, if examinations i and j share at least one common student; 0 otherwise.

To reflect the different scheduling requirements of students, the identified conflicts were classified into three categories. Strict conflicts were defined as

conflicts where two examinations shared at least one proper student and therefore could not be scheduled on the same day. Medium conflicts were defined as conflicts where a proper student in one examination wrote a previous-year subject in the other. Relaxed conflicts were defined as conflicts where the common students wrote previous-year subjects in both examinations.

3.2 ILP Model Formulation

The ILP model was constructed by defining the sets, indices, parameters, and decision variables required for examination scheduling.

3.2.1 Sets and indices

Table 1 presents the sets and indices used in the proposed ILP model.

Table 1: Sets and Indices

Symbol	Description
$\boldsymbol{E}$	Set of all examinations, indexed by e
$\boldsymbol{E_{th}}$	Subset of theory examinations,
$\boldsymbol{E_{pr}}$	Subset of practical examinations,
$\boldsymbol{E_{pp}}$	Subset of project presentation examinations,
$\boldsymbol{D}$	Set of available scheduling days, $D = \{1,2, \dots, D_{max}\}$, indexed by d
$\boldsymbol{S}$	Set of daily sessions, $S = \{1,2\}$ (1 = Morning 9:00am, 2 = Afternoon 1:00pm)
$\boldsymbol{CP_{strict}}$	Set of strict conflict pairs (Proper + Proper)
$\boldsymbol{CP_{medium}}$	Set of medium conflict pairs (Proper + Repeat)
$\boldsymbol{CP_{relaxed}}$	Set of relaxed conflict pairs (Repeat + Repeat)
$\boldsymbol{K}$	Index set over $CP_{relaxed}$, indexed by k
$\boldsymbol{dept(e)}$	Department of examination e
$\boldsymbol{P}$	Set of presentations, indexed by k
$\boldsymbol{V_k}$	Venue option(s) available to presentation k
$\boldsymbol{dept_{pres}(k)}$	Department of presentation k

3.2.2 Parameters

The parameters used in the model are listed in Table 2.

Table 2: Parameters used

Parameter	Description
$\boldsymbol{sc_e}$	Number of students enrolled in examination e
$\boldsymbol{R_1, R_2, R_3}$	Capacity of Hall 1(LH), Hall 2(MH), Hall 3(LR 09)
$\boldsymbol{C_{total}}$	Combined hall capacity per session
$\boldsymbol{M}$	Big-M constant for day-usage linking ($M = \boldsymbol{E} * \boldsymbol{S}$)
$\boldsymbol{\beta}$	Penalty weight for repeat-student same-day scheduling = 0.05

3.2.3 Decision Variables

Table 3 summarizes the decision variables defined for the proposed model.

Table 3: Decision Variables

Variable	Definition
$\boldsymbol{x_{e,d,s}}$	1 if exam e is held on day d, session s; 0 otherwise
$\boldsymbol{y_d}$	1 if at least one examination occupies day d; 0 otherwise
$\boldsymbol{p_{k,d}}$	1 if both examinations of relaxed pair k on day d; 0 otherwise
$\boldsymbol{q_{k,d,r}}$	1 if presentation k is held on day d, venue r; 0 otherwise

3.2.4 Objective Function

The primary objective was to minimize the total number of examination days. A secondary penalty was included to discourage relaxed conflict pairs from occurring on the same day:

$$\min Z = \sum_{d \in D} y_d + \beta \sum_{k \in K} \sum_{d \in D} p_{k,d}. \quad (1)$$

The penalty weight β was set to 0.05 to give a much smaller importance to same-day relaxed conflicts than to reducing examination days. One additional examination day increases the objective by 1.0, whereas one same-day relaxed conflict adds only 0.05. Therefore, $\beta = 0.05$ was used to discourage relaxed conflicts without giving them the same importance as examination-day reduction. A full test of different β values was not carried and is suggested as future work.

Subject to:

Each examination was assigned to the required day and session.

$$\sum_{d \in D} \sum_{s \in S} x_{e,d,s} = 1, \forall\, e \in E_{\text{th}}, \quad (2)$$

$$\sum_{d \in D} \sum_{s \in S} x_{e,d,s} = 2 \times \text{days_needed}_{\text{e}}, \quad \forall\, e \in E_{\text{pr}}, \quad (3)$$

Practical examinations were scheduled across the required consecutive days and both daily sessions.

$$x_{e,d,1} = x_{e,d,2}, \forall\, e \in E_{\text{pr}},\ d \in D, \quad (4)$$

$$\sum_{d=1}^{D_{\max}-1} z_{e,d} = 1,\ \forall \text{e} \in E_{\text{pr}}: \text{days_needed}_{\text{e}} > 1, \quad (5)$$

$$x_{e,d,1} = z_{e,d} + z_{e,d-1}, \forall\, e \in E_{\text{pr}}: \text{days_needed}_{\text{e}} > 1,\ d \in \text{D}. \quad (6)$$

Examinations sharing proper students were not scheduled on the same day.

$$u_{e_1,d} + u_{e_2,d} \leq 1, \quad \forall\, (e_1, e_2) \in CP_{\text{strict}},\ d \in D. \quad (7)$$

Examinations involving both proper & repeat and repeat & repeat students were not scheduled in the same session.

$$x_{e_1,d,s} + x_{e_2,d,s} \leq 1, \quad \forall (e_1, e_2) \in CP_{\text{relaxed}}, d \in D,\ s \in S. \quad (8)$$

$$x_{e_1,d,s} + x_{e_2,d,s} \leq 1, \quad \forall (e_1, e_2) \in CP_{\text{medium}}, d \in D,\ s \in S. \quad (9)$$

The total enrollment assigned to a session did not exceed the available hall capacity.

$$\sum_{e \in E_{th}} s\, c_e \cdot x_{e,d,s} \leq C_{\text{total}}, \quad \forall d \in,\ s \in S. \quad (10)$$

Additional constraints were incorporated to maintain practical examination continuity, prevent simultaneous allocation of the same laboratory, enforce departmental practical restrictions, assign each project presentation to an available day, session and venue, prevent presentation venue conflicts, separate same-department presentations and theory examinations, and link examination assignments with the number of days used.

4 RESULTS & DISCUSSION

Table 4 summarizes the results obtained from the proposed model.

Table 4: Summary of Model results

Performance Measure	Semester 1	Semester 2
Total examinations	64	70
Theory exams	53	47
Practical exams	11	10
Presentations	0	13
Decision variables	3,601	3,554
Total Constraints	31,243	31,306
Manual timetable	16 days	20 days
ILP timetable	14 days	19 days
Days Reduced	2 days	1 day
Objective value	14.00	19.05
Optimality gap	0%	0.26%
Solver time	31.25 seconds	32.52 seconds

The ILP model reduced the examination period in both semesters. For Semester the number of examination days was reduced from 16 to 14, giving a 12.5% reduction. For Semester 2, it was reduced from 20 to 19 days, giving a 5% reduction. The reduction in examination days was achieved by allowing the model to assign examinations efficiently across the available daily sessions and venue capacity while satisfying the defined hard scheduling constraints. The smaller reduction in Semester 2 may be associated with its different examination structure, particularly the inclusion of 13 project presentations and the additional scheduling requirements associated with them. The generated timetables satisfied the defined scheduling requirements, including student conflict restrictions, session restrictions for students writing previous-year subjects, practical examination requirements, laboratory availability, hall capacity, departmental restrictions, and presentation requirements. A representative portion of the generated examination timetable for semester 1 is presented in Table 5.

Table 5: A portion of resulted timetable for semester 1

Day	Course	Start	End	Venue	Count
1	CMIS 1123	9:00am	11:30am	LH(160)+ MH(52)	212
	IMGT 4123	9:00am	11:30am	LR09	36
	MMOD 3124	9:00am	12:00pm	MH	48
	CMIS 2123	1:00pm	3:30pm	LH(160)+MH(32)	192
	CMIS 3134	1:00pm	4:00pm	MH	42
2	CMIS[P] 2123	8:30am	10:30am	CMIS Lab	48
		10:45am	12:45pm	CMIS Lab	48
		1:15pm	3:15pm	CMIS Lab	48
		3:30pm	5:30pm	CMIS Lab	48
	ELTN 4122	9:00am	11:00am	LH	16
	IMGT 3142	9:00am	11:00am	LH	32
	STAT 4124	9:00am	12:00pm	LH	39
3	ELTN 4114	9:00am	12:00pm	MH	27
	MATH 2114	9:00am	12:00pm	LH(160)+MH(31)	191
	MATH 3114	9:00am	12:00pm	MH	9
	CMIS 3114	1:00pm	4:00pm	LH	124
	IMGT 4133	1:00pm	3:00pm	LH	36

A small penalty of $\beta = 0.05$ was included in the objective function to discourage undesirable same-day relaxed conflicts. Therefore, the Semester 2 objective value was 19.05 while the timetable used 19 examination days.

The proposed ILP model was tested on a computer running Windows 11 (64-bit) with 8 GB of RAM**.** The model was implemented in MATLAB R2026a using the intlinprog solver. Semester 1 contained 3,601 decision variables and 31,243 constraints and was solved in 31.25 seconds. Semester 2 contained 3,554 decision variables and 31,306 constraints and was solved in 32.52 seconds. A 0% optimality gap was obtained for Semester 1, while a 0.26% optimality gap was obtained for Semester 2. Semester 1 had an objective value and dual bound of 14.00, while Semester 2 had an objective value of 19.05 and a dual bound of 19.00. Thus, the Semester 2 solution was very close to the best bound identified by the solver.

5 CONCLUSION

This study developed and evaluated an ILP based examination timetabling model specifically for the FAS, WUSL, using real examination data from two academic semesters. The model reduced the examination period from 16 to 14 days in Semester 1 and from 20 to 19 days in Semester 2 while satisfying the defined scheduling requirements. The results showed that the ILP approach provided a practical method for improving the examination scheduling process within a reasonable computational time. However, the model did not directly consider examination spacing, so some students may still have examinations on consecutive days. Future work could consider examination spacing to achieve a more balanced timetable and test different values of the penalty weight β.

6 REFERENCES

- Al-Hawari, F., Al-Ashi, M., Abawi, F., & Alouneh, S. (2020). A practical three-phase ILP approach for solving the examination timetabling problem. *International Transactions in Operational Research*, *27*(2), 924-944. https://doi.org/10.1111/itor.12471
- Aslan, E., Şimşek, T., & Karkacıer, A. (2017). A binary integer programming model for exam scheduling problem with several departments. *Bilgi Ekonomisi ve Yönetimi Dergisi, 12*(2), 169-175.
- Bania, R. K., & Duarah, P. (2018). Exam time table scheduling using graph coloring approach. *International Journal of Computer Sciences and Engineering*, *6*(5), 84-93.
- Laisupannawong, T., & Sumetthapiwat, S. (2024). An integer linear programming model for the examination timetabling problem: A case study of a University in Thailand. *Advances in Operations Research*, *2024*(1), Article 6636563. https://doi.org/10.1155/2024/6636563
- Mohammed, M. F., Abobaker, R. A., Aljalil, A. M., & Maatuk, A. M. (2022, July). Examinations timetabling system based on a genetic algorithm. *2022 International Conference on Engineering & MIS (ICEMIS)* (pp. 1-6). IEEE.

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Box–Jenkins Modeling and Forecasting of Monthly Solar Radiation : A Case Study in Sri Lanka

Imalsha RD[1], Abeygunawardana RAB[2]
[1,2]*Department of Statistics, Faculty of Science, University of Colombo, Sri Lanka.*
dilkiimalsha01@gmail.com[1] , rab_abey@stat.cmb.ac.lk[2]

ABSTRACT

This study aims to develop and evaluate a seasonal Box–Jenkins time-series model for forecasting monthly solar radiation in Sri Lanka using observations from 2014 to 2023. Monthly mean solar radiation data from 23 observation stations were obtained from the NASA POWER, while relevant meteorological data were obtained from the Department of Census and Statistics, Sri Lanka. Stationarity was assessed using the Augmented Dickey Fuller (ADF) and Kwiatkowski Phillips Schmidt Shin (KPSS) tests, while the autocorrelation function (ACF) and partial autocorrelation function (PACF) were examined to identify temporal and seasonal dependence. Seasonal autoregressive integrated moving average (ARIMA) models were evaluated using the Akaike information criterion (AIC), corrected Akaike information criterion (AICc), Bayesian information criterion (BIC) and residual diagnostics. Data from 2014–2022 were used for model estimation, while 2023 was retained as an independent test period. The seasonal ARIMA$(1,0,1)(0,1,1)_{12}$ model adequately captured the temporal and seasonal behaviour of monthly solar radiation (AIC = 90.32, BIC = 101.05), with independent residuals (Ljung–Box p = 0.6758). For 2023 out-of-sample forecasting, the model achieved an Root Mean Square Error (RMSE) of 0.3353, Mean Absolute Error (MAE) of 0.2711, and Mean Absolute Percentage Error (MAPE) of 5.21%, indicating reasonably accurate short-term forecasts. These findings confirm that the monthly solar radiation series contains a clear seasonal pattern that can be effectively captured using seasonal ARIMA modelling.

KEYWORDS: Augmented Dickey Fuller, Box–Jenkins, Kwiatkowski Phillips Schmidt Shin, Solar radiation

1 INTRODUCTION

Solar radiation is the primary energy input to the Earth's surface and plays an important role solar energy generation. The availability of solar radiation varies considerably over time due to seasonal and atmospheric influences. Consequently, reliable prediction of solar radiation is important for solar- energy resource assessment and operational decision-making. Time-series forecasting methods provide an approach for modelling temporal dependence within observations without requiring information on explanatory variables. The Box–Jenkins methodology provides a systematic framework for time-series model identification, estimation, diagnostic checking and forecasting. Seasonal ARIMA models extend this framework to time series exhibiting periodic patterns, making them particularly suitable for monthly environmental observations.

In Sri Lanka, solar radiation exhibits temporal variability associated with seasonal changes in atmospheric conditions. Therefore, developing an appropriate statistical forecasting model can provide useful information for understanding and predicting future monthly solar radiation patterns. This study aims to develop and evaluate a Box–Jenkins time-series model for monthly solar radiation using observations from 2014 to 2023. The specific objectives are to assess the stationarity and temporal dependence of the series, identify an

appropriate seasonal ARIMA model, evaluate its adequacy using residual diagnostics, and assess its forecasting performance using an out-of-sample test period.

2 LITERATURE REVIEW

Solar radiation is strongly associated by seasonal and meteorological conditions, making accurate forecasting important for solar energy planning. Previous studies have shown that rainfall and monsoon patterns are associated with variations in solar radiation. Teong et al. (2017) found that higher rainfall during monsoon periods was generally associated with lower solar radiation, while Sendanayake (2015) demonstrated that rainfall and sunshine duration could be used to predict solar radiation in Sri Lanka. Similarly, Falayi (2013) identified cloud cover, rainfall, and relative humidity as important factors associate with solar radiation.

Several studies have also applied statistical and machine learning methods for solar radiation prediction. Karasu et al. (2017) showed that Gaussian Process Regression provided highly accurate solar radiation predictions, while Huang et al. (2021) reported strong performance of machine learning models, including XGBoost and Random Forest. Voyant et al. (2017) reviewed various machine-learning methods and highlighted their effectiveness in modelling nonlinear relationships in solar radiation. More recently, Abisoye et al. (2024) demonstrated the potential of machine-learning techniques for improving solar radiation prediction. However, these approaches mainly focus on meteorological predictors and machine learning techniques.

Reikard (2009) demonstrated that ARIMA model can provide effective solar-radiation forecasts, because they can identify systematic temporal dependence in solar radiation. Alsharif et al. (2019) applied ARIMA models to daily and monthly solar radiation forecasting, demonstrating the usefulness of time-series models for capturing temporal patterns. Shadab et al. (2019) used single-point monthly solar-insolation data from JMI University, New Delhi, India, for 1984–2017 and applied the Box–Jenkins approach to develop a multiplicative seasonal ARIMA model. They identified ARIMA$(1,0,1)(0,1,1)_{12}$ as an appropriate model for monthly solar radiation forecasting. Similarly, Chodakowska et al. (2023) evaluated ARIMA models for solar-radiation forecasting using monthly and hourly data from Jordan and Poland.

Although several international studies have concluded the effectiveness of ARIMA and SARIMA models for solar-radiation forecasting, comparatively limited attention has been given to monthly solar-radiation forecasting for Sri Lanka using the Box–Jenkins approach. Therefore, aim of this study is addresses the gap by developing appropriate seasonal time-series model for monthly solar radiation using Box–Jenkins approach in Sri Lanka.

3 METHODOLOGY

3.1 Data

Monthly solar radiation observations covering the period from 2014 to 2023 were used in the analysis and data are collected from Department of Census and Statistics, Sri Lanka, NASA POWER

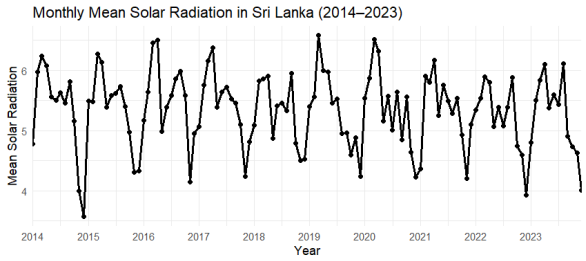


Figure 1: Monthly Mean Solar radiation in Sri Lanka (2014-2023)

Data Access Viewer. The observations were aggregated by year and month to obtain the monthly mean solar radiation. The resulting series consisted of 120 monthly observations. The monthly observations were represented in Fig. 1 as a time series with a seasonal frequency of 12.

3.2 Box-Jenkins Model

The Box–Jenkins methodology was applied through the stages of identification, estimation, diagnostic checking and forecasting. Initially, the monthly solar radiation series was examined graphically to identify temporal variation and possible seasonal behaviour. Seasonal decomposition was also performed to investigate the underlying components of the series.

Stationarity was assessed using the Augmented Dickey Fuller (ADF) and KPSS tests. The autocorrelation function (ACF) and partial autocorrelation function (PACF) were subsequently examined to identify potential autoregressive and moving-average structures. Seasonal ARIMA models were estimated using the forecast package in R. Model selection was based primarily on AIC, AICc and BIC, together with residual diagnostics.

The adequacy of the selected model was assessed using residual analysis and the Ljung–Box test. A satisfactory model was expected to produce residuals with no significant remaining serial correlation. The final model was then used to generate future monthly solar-radiation forecasts.

3.3 Forecast Validation

To assess predictive performance, the observations from 2014–2022 were used as the training period, while the 12 observations from 2023 were retained as an independent test period. Forecast accuracy was evaluated using the mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE).

4 RESULTS AND DISCUSSION

Solar radiation showed a mean value of 5.35 kWh/m²/day with low variability (SD = 0.69), ranging from 3.03 to 6.90 kWh/m²/day and approximately normally distributed. Cloud amount, humidity, and wind speed showed moderate variability with nearly symmetric distributions, while the temperature variables exhibited relatively low variation, with negative skewness indicating the influence of lower temperature observations and positive kurtosis suggesting slightly peaked distributions. Rainfall exhibited the highest variability (SD = 153.69 mm) and a positively skewed distribution (skewness = 1.48), indicating occasional heavy rainfall events. These meteorological conditions provide useful information for interpreting variations in solar radiation, as higher cloud amount, humidity, and rainfall are generally associated with atmospheric conditions that can reduce the amount of solar radiation reaching the surface, whereas temperature and wind speed may influence atmospheric conditions

indirectly. However, these meteorological variables were not included as explanatory variables in the Box–Jenkins model, which focused specifically on the temporal and seasonal behavior of monthly solar radiation.

The stationarity of the series was assessed using the Augmented Dickey–Fuller (ADF) and KPSS tests, Both the ADF ($p = 0.01$) and KPSS ($p = 0.10$) tests indicated that the series was stationary.

The ACF plot in Fig. 2 and PACF plot in Fig. 3 examined to identify the temporal and seasonal dependence structure of the monthly solar radiation series. Based on the Box–Jenkins model-selection procedure, the selected model for the complete 2014–2023 series was seasonal ARIMA$(1,0,1)(0,1,1)_{12}$, with AIC = 90.32, AICc = 90.71, and BIC = 101.05.

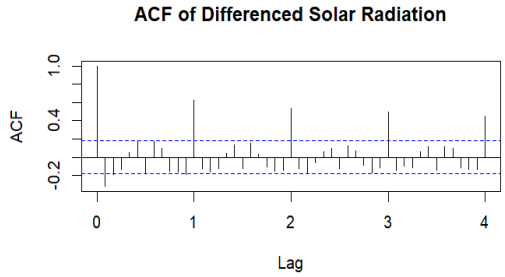


Figure 2: ACF plot of the differenced monthly solar radiation

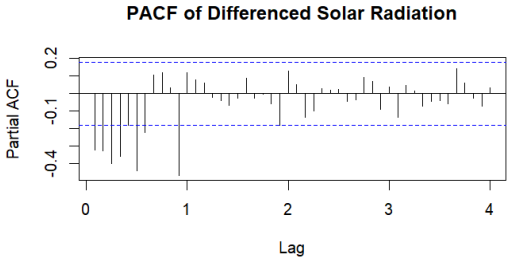


Figure 3: PACF plot of the differenced monthly solar radiation

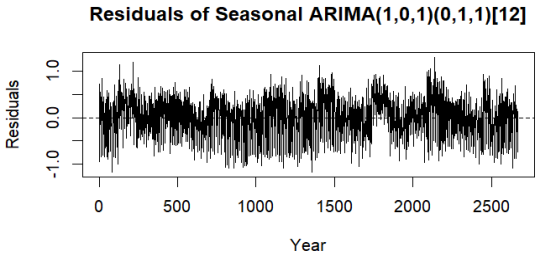


Figure 4: Residual plot of the seasonal ARIMA$(1,0,1)(0,1,1)_{12}$ model

The model included significant temporal and seasonal components, capturing the annual variation in monthly solar radiation. Residual diagnostics in Fig. 4 indicated model adequacy. The Ljung–Box test produced a p-value of 0.6758, indicating no significant residual autocorrelation.

For validation, the model fitted using 2014–2022 data was used to forecast the 12 months of 2023. The model achieved an RMSE of 0.3353, MAE of 0.2711, MAPE of 5.21%. These results indicate that the Box–Jenkins approach provided reasonably accurate forecasts and effectively captured the temporal and seasonal behavior of monthly solar radiation.

5 CONCLUSION

This study demonstrated that the Box–Jenkins approach can effectively model and forecast monthly solar radiation in Sri Lanka. The selected SARIMA$(1,0,1)(0,1,1)_{12}$ model adequately captured the temporal and seasonal behavior of the series and produced satisfactory forecasting accuracy for the independent 2023 test period. The findings provide useful evidence for applying seasonal time-series forecasting to Sri Lankan solar radiation data and can support short-term

solar energy assessment and planning. Overall, the study contributes empirical evidence on the applicability of the Box–Jenkins approach to monthly solar-radiation forecasting in a multi-station Sri Lankan context.

REFERENCES

- Abisoye, B. O., Sun, Y., & Wang, Z. (2024). Machine learning forecasting model for solar energy radiation. *International Journal of Computer Theory and Engineering, 16*(2), 66–75.
- Alsharif, M. H., Younes, M. K., & Kim, J. (2019). Time series ARIMA model for prediction of daily and monthly average global solar radiation: The case study of Seoul, South Korea. *Symmetry, 11*(2), 240.
- Chodakowska, E., Nazarko, J., Nazarko, L., Rabayah, H. S., Abendeh, R. M., & Alawneh, R. (2023). ARIMA models in solar radiation forecasting in different geographic locations. *Energies, 16*(13), 5029.
- Department of Census and Statistics. (2025). *Statistical data sheet 2025*. Department of Census and Statistics, Sri Lanka.
- Falayi, E. O. (2013). The impact of cloud cover, relative humidity, temperature and rainfall on solar radiation in Nigeria. *Energy and Power, 3*(6), 119–127.
- Huang, L., Kang, J., Wan, M., Fang, L., Zhang, C., & Zeng, Z. (2021). Solar radiation prediction using different machine learning algorithms and implications for extreme climate events. *Frontiers in Earth Science, 9*, 596860.
- Karasu, S., Altan, A., Saraç, Z., & Hacıoğlu, R. (2017). Prediction of solar radiation based on machine learning methods. *The Journal of Cognitive Systems, 2*(1), 16–20.
- NASA POWER. (2026). *NASA Prediction of Worldwide Energy Resources (POWER): Data access viewer*. NASA Langley Research Center.
- Reikard, G. (2009). Predicting solar radiation at high resolutions: A comparison of time series forecasts. *Solar Energy, 83*(3), 342–349.
- Sendanayake, S., Miguntanna, N., & Jayasinghe, M. T. R. (2015). Predicting solar radiation for tropical islands from rainfall data. *Journal of Urban and Environmental Engineering, 9*(2), 109–118.
- Shadab, A., Said, S., & Ahmad, S. (2019). Box–Jenkins multiplicative ARIMA modeling for prediction of solar radiation: A case study. *International Journal of Energy and Water Resources, 3*, 305–318.
- Teong, K. V., Sukarno, K., Chang, J. H. W., Chee, F. P., Ho, C. M., & Dayou, J. (2017). The monsoon effect on rainfall and solar radiation in Kota Kinabalu. *International Conference on Science and Natural Resources*.
- Voyant, C., Notton, G., Kalogirou, S. A., Nivet, M. L., Paoli, C., Motte, F., & Fouilloy, A. (2017). Machine learning methods for solar radiation forecasting: A review. *Renewable Energy, 105*, 569–582.

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Modeling Hospital-Based Diabetes Mortality in Sri Lanka

Wijekoon WMRT[1], Dharmawardane PMN[2], Dissanayaka DMDK[3]

[1,2]*Department of Mathematical Sciences, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*

[3]*Department of Physical Sciences, Faculty of Applied Sciences, Rajarata University of Sri Lanka, Mihintale, Sri Lanka*

wijekoonrashmi@gmail.com[1] , maheshd@wyb.ac.lk[2], dilukdissanayaka12345@gmail.com[3]

ABSTRACT

Diabetes Mellitus (DM) is a public health concern that leads to many illnesses and deaths. In Sri Lanka, it is a common chronic disease. This study aimed to estimate the overall hospital-based mortality rate due to DM and assess the associations of diabetes category, hospital type, and Regional Directorate of Health Services (RDHS) divisions with mortality, providing baseline evidence on diabetes-related mortality within the Sri Lankan healthcare context. The analysis was based on 1,036 hospital records from 523 institutions across 26 RDHS divisions. The dataset comprised 67,607 live discharges and 546 deaths due to DM, resulting in an overall hospital-based mortality proportion of 0.8%. Descriptive analysis was conducted to examine mortality patterns across diabetes categories, hospital types, and RDHS divisions. Binomial logistic regression models were fitted using R to assess the associations between mortality and the selected explanatory variables. Mortality varied considerably across hospital settings and geographical regions. The final model incorporating diabetes category, hospital type, and RDHS division demonstrated improved model fit compared with the intercept-only model (AIC = 1027.6). Diabetes category and hospital type showed significant associations with hospital-recorded mortality, while significant regional differences were also observed after adjusting for the other factors. The findings highlight variations in hospital-based diabetes mortality across healthcare settings and regions in Sri Lanka. These results provide a useful baseline for future studies using cross-sectional data and for developing mathematical models of diabetes-related mortality in Sri Lanka.

KEYWORDS: Binomial Logistic Regression, Diabetes Category, Diabetes Mellitus, Hospital-Based Mortality Rate

1 INTRODUCTION

Diabetes Mellitus (DM) is an important public health problem worldwide. The number of people living with diabetes has increased rapidly, and it contributes to mortality rates through complications of DM. Sri Lanka has also experienced a considerable burden of DM (World Health Organization: WHO, 2024). According to Rannan-Eliya et al. (2023), the crude prevalence of diabetes among adults was 23% in 2018/2019, demonstrating the importance of DM as a national health concern in Sri Lanka. Hospital mortality provides useful information about severe outcomes among patients. However, mortality may vary according to diabetes category, diabetes type, hospital characteristics and geographical location. Therefore, modeling DM can help to identify significant factors associated with differences in mortality. This study provides quantitative evidence for understanding patterns of mortality across different diabetes categories, hospital type and Regional Directorate of Health Services (RDHS) division. Moreover, it also investigates hospital-based mortality in Sri Lanka using the hospital data and a binomial logistic regression framework for identifying factors associated with

hospital mortality and supports healthcare planning in Sri Lanka.

2 LITERATURE REVIEW

This literature review summarises the existing evidence on diabetes-related mortality, associated factors, variation across healthcare settings, and also statistical approaches used to analyse mortality. This provides the background for examining hospital-based diabetes mortality in Sri Lanka.

DM is a chronic metabolic disease caused due to the significant changes in blood glucose. The World Health Organization identifies diabetes as a chronic disease damages blood vessels and increases the risk of cardiovascular disease, kidney failure, neurological complications and other serious health outcomes. According to Diabetesatlas.org (2025) by 2050, approximately 1.9 million adults aged 20–79 years in Sri Lanka are projected to have diabetes.

Rannan-Eliya et al. (2023) used the Sri Lanka Health and Ageing Survey (SLHAS) 2018/2019 data to estimate the prevalence of diabetes among Sri Lankan adults was 23.0% with an age-standardized prevalence of 21.8%. The study also identified substantial variation in diabetes prevalence according to age, sex, ethnicity, socioeconomic status, and geographical area. This study identified Sri Lanka as a global hotspot for diabetes, having the highest diabetes prevalence in South Asia. Therefore, these findings show the importance of analyzing DM population and healthcare data specific to Sri Lanka.

A prospective population based study reported that individuals with diabetes experienced more hospital admissions and had higher all-cause mortality than individuals without diabetes. This also suggests that identifying factors associated with mortality among hospitalized DM patients can provide useful information for healthcare planning (Johansson et al., 2024).

Munasinghe et al. (2025) conducted a retrospective population based study using Sri Lankan mortality data collected from 2004 to 2020. Age-adjusted mortality rates were calculated, and trends were examined according to disease category, age, and sex. The most important finding of this study was that the non-communicable disease (NCD) mortality increased significantly in Sri Lanka over the study period. In addition, women had higher total mortality while men had higher hospital mortality for most NCD categories.

Although previous studies have identified diabetes prevalence, mortality and DM patterns, there remains value in analyzing hospital-based diabetes mortality according to diabetes types, hospital characteristics and geographical divisions. The study addresses this gap using hospital records and a binomial logistic regression framework.

3 METHODOLOGY

3.1 Study design

This study uses hospital-based data on diabetes related outcomes in Sri Lanka. The outcome was defined using the number of deaths and live discharges.

3.2 Statistical Model

The outcome consists of two mutually exclusive outcomes as deaths and live discharges. The data were modeled using binomial logistic regression model. For the ith observation, representing a specific hospital-level

record in the dataset, the total number of hospital outcomes was calculated as,
$N_i = D_i + L_i$ where, D_i represents the number of deaths and L_i represents the number of live discharges for observation i, respectively.

Then the observed mortality rate (p_i) was calculated as,

$$p_i = \frac{D_i}{N_i} \times 100\% \quad (1)$$

A logistic regression model can be represented as,

$$\text{logit}(p_i) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \quad (2)$$

where, β_0 is the intercept and β_1, β_2 and β_3 are the regression coefficients corresponding to the selected predictor variables x_1, x_2, and x_3.

The coefficients were interpreted using odds ratios. An odds ratio greater than 1 indicates higher odds of death, whereas an odds ratio below 1 indicates lower odds of death.

The final model was selected by comparing several logistic regression models with different combinations of explanatory variables. The models were compared based on their Akaike Information Criterion (AIC) values, and the model with the lowest AIC was selected as the final model.

3.3 Data Collection and Analysis

The dataset consisted of 1036 observations and 523 unique health care institutions distributed across 26 RDHS divisions including 67,607 live discharges and 546 deaths, where the total number of recorded hospital outcomes was 68,153. The data used in this study were obtained from statistics on diabetic patients recorded and published in 2024 by the Medical Statistics Unit of the Ministry of Health, Sri Lanka.

According to the International Classification of Diseases (ICD), diabetes cases were categorized as follows: Type 1 DM (E10), Type 2 DM (E11), Malnutrition-related DM (E12), other specified DM (E13) and Unspecified DM (E14). According to hospital-based records, the number of hospitals that have each of the aforementioned diabetes categories is listed in Table 1.

Hospitals were classified into three categories based on the number of available patient beds. Type A, Type B, and Type C consist of more than 100, 50 to100, and less than 50 patient beds respectively.

For descriptive analysis, mortality rate was calculated according to the diabetes categories, hospital classification and RDHS division. Then several binomial logistic models given below were tested.

Model 0: Intercept only
Model 1: Diabetes category
Model 2: Diabetes category + Hospital type
Model 3: Diabetes category + Hospital type + RDHS division

The model comparisons were done using AIC and statistical significance of predictors. The overall analysis was done using R software.

4 RESULTS AND DISCUSSION

4.1 Model selection

In Table 2, the reduction in AIC from Model 0 to Model 3 indicates improved model fit after incorporating hospital type and geographical variation.

Table 1: Distribution of hospital-based Records by Diabetes Category

Diabetes category	No. of hospitals
E10	285
E11	364
E12	20
E13	61
E14	306

Although the model including hospital classification produced an even smaller AIC, it generated extremely large standard errors and warnings concerning fitted probabilities of 0 or 1.

This indicates sparse-data problems caused by some hospital-classification categories having very few observations or no deaths. Similarly, the multilevel model incorporating institution and RDHS random effects produced a convergence warning. Therefore, according to the current analysis and (2) binomial logistic regression model containing diabetes category, hospital type and RDHS division is the most appropriate model to report as the final model (3):

$$\begin{aligned}\text{logit}(p_i) &= \beta_0 + \beta_1(\text{Diabetes Category}) \\ &\quad + \beta_2(\text{Hospital Type}) \\ &\quad + \beta_3(\text{RDHS Division})\end{aligned} \tag{3}$$

4.2 Mortality according to diabetes category

The mortality rates vary between diabetes categories as in Table 3.

The largest number of deaths occurred in E14, with 271 deaths, followed by E11 with 209 deaths. However, the highest mortality proportion was observed for E10 followed by E11.

Table 2: Logistic models produced

Model	Predictors	AIC
Model 0	Intercept only	1584.3
Model 1	Diabetes Category	1582.6
Model 2	Diabetes Category + Hospital type	1168.0
Model 3	Diabetes Category + Hospital type + RDHS division	1027.6

4.3 Mortality according to hospital type

Hospital type showed a strong association with mortality as the hospital types such as Type A, Type B, and Type C are significant in the final model (3).

Table 3: Hospital-based mortality according to diabetes category

Diabetes category	Overall deaths (Island wide)	Total hospital outcomes	Mortality (%)
E10	62	6,473	0.96
E11	209	23,846	0.88
E12	0	83	0.00
E13	4	1,059	0.38
E14	271	36,692	0.74

Table 4: Estimated coefficients and *p*-values for hospital type in the final model

Hospital Type	Estimate	*p*-value
Type A	-1.0384	<0.001
Type B	-1.7329	<0.001
Type C	-3.1629	<0.001

In Table 4, negative estimated coefficients indicate lower estimated odds of death, and all the types are statistically significant at 5% level of significance as the p-values of all the types are less than 0.05. The highest mortality rate among the hospital classifications was observed in teaching hospitals (2.42%), followed by National hospitals (1.67%) and District General hospitals (0.77%). These differences highlight the differences in patient severity and the complexity of cases treated at different hospital levels corresponding to their own resources.

4.4 Mortality according to regional variation

Model 2 performance increased after adding RDHS divisions. Several RDHS divisions showed significant differences. Among the RDHS divisions, Colombo recorded the highest observed mortality proportion at approximately 2.35%, followed by Galle at 1.83%, Gampaha at 1.20%, Puttalam at 1.15% and Kurunegala at 1.08%. These geographical changes reflect the differences in hospital capacity, and population characteristics.

4.5 Overall mortality

The analysis included a total of 67607 live discharges and 546 deaths were recorded. Then form (1), the overall hospital-based mortality rate can be obtained as follows:

$$\frac{546}{68153} \times 100\% \approx 0.8\%$$

This indicates that approximately 8 deaths occurred per 1000 recorded hospital outcomes in the study dataset.

5 CONCLUSION

This study focuses on diabetes mortality in Sri Lanka using hospital records and a binomial logistic regression approach. The analysis included 1,036 observations from 523 hospitals in 26 RDHS divisions. In total there were 67,607 patients discharged alive and 546 deaths giving overall hospital-based mortality rate of 0.8%. The descriptive findings demonstrated considerable variation in hospital-recorded mortality across diabetes categories, hospital types, and RDHS divisions. E10 had the highest mortality proportion among the diabetes categories while the largest number of deaths were recorded from E14. The logistic regression analysis indicated that hospital type was strongly associated with mortality, while the inclusion of RDHS division further improved model fit. The final model incorporating diabetes category, hospital type and RDHS division provided a more suitable framework for explaining differences in hospital-based mortality. The findings provide useful evidence regarding patterns of diabetes-related hospital mortality in Sri Lanka. Future research should incorporate individual-level clinical data, additional patient characteristics, and observations collected over multiple time periods. Such data would enable more comprehensive multilevel and longitudinal modelling of diabetes-related outcomes and provide stronger evidence for healthcare planning and resource allocation in Sri Lanka. Furthermore, the overall mortality estimate obtained in this study may provide an empirical parameter for the development and calibration of future ordinary differential equation (ODE) models of diabetes-related mortality and disease progression.

REFERENCES

- Diabetesatlas.org (2025, March 20). *Sri Lanka*. https://diabetesatlas.org/data -by-location/country/sri-lanka/
- Johansson, M., Åkesson, A., Nilsson, P. M., & Melander, O. (2024). Longitudinal assessment of the impact of prevalent diabetes on hospital admissions and mortality in the general population: A prospective population-based study with 19 years of follow-up. *BMC Public Health*, *24*(1), 2948.
- Munasinghe, H., Dissanayaka, P., Jayasundara, M., Piyasena, M. P., Sivaprasad, S., & Nugawela, M. D. (2025). Trends in mortality from diabetes and other non-communicable diseases among Sri Lankan adults: A retrospective population-based study, 2004-2020/2021. *Journal of Clinical Medicine*, *14*(23), 8568.
- Rannan-Eliya, R. P., Wijemunige, N., Perera, P., Kapuge, Y., Gunawardana, N., Sigera, C., Jayatissa, R., Herath, H. M. M., Gamage, A., Weerawardena, N., Sivagnanam, I., Dalpatadu, S., Samarage, S., Samarakoon, U., Samaranayake, N., Pullenayegam, C., Perera, B., & SLHAS Collaborators. (2023). Prevalence of diabetes and pre-diabetes in Sri Lanka: A new global hotspot-estimates from the Sri Lanka Health and ageing survey 2018/2019. *BMJ Open Diabetes Research & Care*, *11*(1)
- World Health Organization. (2024, November 14). *Diabetes*. World Health Organization.

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Predicting Daily Coconut Retail Prices in Major Markets of Sri Lanka

Wanigasekara WADS[1], Appuhamy PADAN[2]
[1,2]*Department of Mathematical Sciences, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka.*
dswanigasekara57@gmail.com[1] , asankanidarshani@wyb.ac.lk[2]

ABSTRACT

Coconut retail prices vary considerably across markets in Sri Lanka due to differences in transportation costs, intermediaries, supply and demand, and market characteristics. Limited understanding of these market-specific price variations hinders price management and informed decision-making among stakeholders. This study aims to identify the best predictive model for coconut retail prices in two major Sri Lankan markets: the Dambulla Economic Centre and Pettah Market. Daily retail and wholesale price data from 1 January 2022 to 31 October 2025 were obtained from the Central Bank of Sri Lanka. Wholesale price, lagged prices, weekly volatility, spread, and month were used as predictors. The study developed linear (Autoregressive Integrated Moving Average (ARIMA) and ARIMA with exogenous variables (ARIMAX), machine learning (Random Forest and Extreme Gradient Boosting (XGB)), and hybrid models, which were evaluated and compared based on their predictive performance using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2) to identify the best model for each market. Retail prices in both markets increased over the study period, with Pettah Market recording significantly higher retail prices than Dambulla. Standalone machine learning models outperformed other models in both markets, with Random Forest performing best for Dambulla (RMSE = 2.26, MAE = 1.56, MAPE = 1.02%, R^2 = 98.15%) and XGB for Pettah (RMSE = 8.32, MAE = 5.10, MAPE = 2.40%, R^2 = 88.50%). The results provide early insights into future price disparities, enabling authorities to take action to reduce market-level price variations.

KEYWORDS: Coconut, Random Forest, Retail Price, XGB

1 INTRODUCTION

Coconut (*Cocos nucifera L.*) is one of the most important agricultural commodities in Sri Lanka, contributing substantially to domestic consumption and export earnings. Coconut prices are influenced by supply shortages, demand, climatic conditions, transportation costs, market demand, intermediary involvement, and supply chain conditions (Bandara & Kumari, 2020). The lack of market integration and information poses challenges for stakeholders in maintaining reasonable price differentials across markets and making informed decisions. Moreover, these price variations create uncertainty for retailers, traders, and consumers in making day-to-day buying and selling decisions.

The two major Sri Lankan markets, the Dambulla Economic Centre and Pettah Main Market, have distinct market structures. The Dambulla market primarily functions as a supply-driven agricultural distribution center, where retail prices are closely linked to wholesale prices. In contrast, the Pettah market is predominantly an urban, demand-driven market characterized by longer supply chains, greater intermediary involvement, and higher transportation costs. These structural differences may lead to variations in price transmission, price margins, and volatility between the

two markets. Limited knowledge of market-specific price dynamics may contribute to persistent price disparities. Therefore, market-specific prediction models are needed rather than a single general model for both markets. To address this research gap, this study aims to develop market-specific daily coconut retail price prediction models for the selected markets using linear, non-linear, and hybrid approaches.

2 LITERATURE REVIEW

Existing studies have explored various methodologies for agricultural price forecasting. Abeysekara and Waidyarathne (2020) found that the Autoregressive Integrated Moving Average (ARIMA) model performed well in forecasting coconut prices in Sri Lanka and other coconut-producing countries. For instance, Ahalya et al. (2023) applied an ARIMA (1,1,1) models to forecast monthly coconut-price of Coimbatore and Polachchi markets in Tamil Nadu, India. However, Dilrukshi et al. (2022) highlighted the suitability of Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models for capturing price volatility. Prasert et al. (2021) demonstrated that ARIMA with Exogenous Variables (ARIMAX) models can improve forecasting accuracy by incorporating the effects of climatic variability. Machine learning approaches, such as Random Forest (RF) and Extreme Gradient Boosting (XGB), can capture complex and nonlinear relationships among variables without requiring strict distributional assumptions (Krishna, 2026). Sarpong et al. (2023) found that a hybrid framework combining ARIMA and a neural network significantly outperformed standalone models when capturing nonlinear patterns in price data. Despite these advances, most previous studies, particularly those conducted in Sri Lanka, have primarily focused on monthly, regional, or national-level data for coconut price forecasting, with limited attention given to market-specific price variations. Therefore, this study aims to address this research gap.

3 METHODOLOGY

3.1 Data Collection and Analysis

Daily retail and wholesale price data for the two selected markets, spanning from January 2022 to October 2025, were collected from the Central Bank of Sri Lanka. The retail price per coconut (R_t) at each market was considered as the response variable, and the daily wholesale price (W_t), lagged prices (W_{t-1} & R_{t-1}), price spread ($P_Spread = R_t - W_t$), spread ratio (S_Ratio), weekly volatility of wholesale price (Std (W_t)) and retail price (Std (R_t)), and Month were considered as the predictors for the models developed.

Data preprocessing techniques were applied to ensure data quality. These included missing-value treatment, data cleaning, encoding of categorical variables, and data splitting, with 80% data for training and 20% for testing. Missing values were imputed using linear interpolation to preserve the continuity of the daily data (Noor et al., 2015). The Variance Inflation Factor (VIF) was used to assess multicollinearity among the predictor variables, and highly correlated variables were removed accordingly. Initially, exploratory data analysis was conducted to examine the behavior and patterns of coconut prices across the two markets.

3.2 Model Development

The model development comprised three main modeling architectures: linear, nonlinear, and hybrid approaches. These approaches were independently applied to each market based on the observed characteristics and temporal behavior of retail coconut prices over the study period. For the linear component, ARIMA and ARIMAX models were developed. The Box–Jenkins methodology was adopted for ARIMA model development, encompassing model identification, parameter estimation, diagnostic checking, and forecasting. The most appropriate model among the candidate linear models was selected by considering both residual diagnostic tests and forecasting performance based on Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). Moreover, machine learning models, namely RF and XGB, were developed for each market to capture complex nonlinear relationships among the variables. RF is an ensemble learning method that combines multiple decision trees to capture complex nonlinear patterns in price behavior. XGB is a sequential ensemble learning algorithm that iteratively improves predictions using gradient boosting, making it suitable for capturing irregular and volatile patterns in time-series data. Next, hybrid models were developed using the residuals obtained from a linear model (either ARIMA or ARIMAX) as the target variable for the machine learning models. The predictors identified in the best-performing nonlinear model were then used as input variables for the same machine learning model identified for each market to capture the unexplained nonlinear patterns in the residuals of a linear model. The prediction accuracy of both non-linear and hybrid models for each market was evaluated using RMSE, MAE, MAPE, and the coefficient of determination (R^2).

4 RESULTS AND DISCUSSION

The analysis revealed considerable variation in daily coconut retail and wholesale prices across the two markets, with Pettah Market recording significantly higher prices. During the period of study, retail prices ranged from Rs. 59 to Rs. 190 in Dambulla and from Rs. 70 to Rs. 250 in Pettah, with average prices of Rs. 97 and Rs. 136, respectively. Moreover, the Pettah market exhibited a significantly larger gap between the mean retail price (Rs. 136) and the wholesale price (Rs. 112) compared with the Dambulla market.

Fig. 1 illustrates the daily retail prices in both markets during the study period. Accordingly, the Dambulla market exhibited an upward trend with irregular fluctuations. In contrast, the Pettah market showed relatively stable behavior during the middle of the period, followed by a sharp increase in prices accompanied by some fluctuations.

Table 1 presents a comparison of the prediction performance of the models developed for daily retail prices in the Dambulla market. As shown in Table 1, the RF model outperformed the other models for this market, achieving the lowest RMSE (Rs. 2.26), MAE (Rs. 1.56), and MAPE (1.02%), and the highest R^2 (98.15%). XGB ranked second, with an RMSE of Rs. 3.98, MAE of Rs. 2.81, and R^2 of 94.28%.

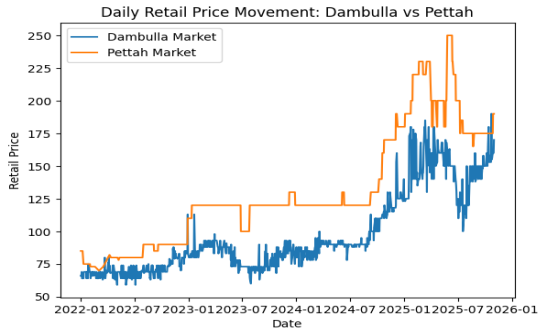


Figure 1: Daily Coconut Retail Prices of Dambulla and Pettah Market

The hybrid models underperformed both standalone machine learning models. Since both standalone linear models failed to satisfy some diagnostic assumptions, they were excluded from the model comparison and considered unsuitable for capturing the complex dynamics of daily coconut retail prices in the Dambulla market.

Table 2 illustrates a comparison of the predictive performance of the models developed for Pettah Market.

According to Table 2, XGB achieved the best predictive performance, with the lowest RMSE of Rs. 8.32, MAE of Rs. 5.10, MAPE of 2.40%, and the highest R^2 value of 89%.

Table 1: Comparison of Predictive Performance of Models for Coconut Retail Prices in the Dambulla Market

Model	MAE	RMSE	MAPE (%)	R^2
RF	1.56	2.26	1.02	0.98
XGB	2.81	3.98	1.8	0.94
ARIMAX-RF	6.11	8.61	4.33	0.73
ARIMA-RF	9.97	14.36	7.27	0.26

Table 2: Comparison of Predictive Performance of Models for Coconut Retail Prices in the Pettah Market

Model	MAE	RMSE	MAPE (%)	R^2
XGB	5.10	8.32	2.40	0.89
RF	8.27	10.80	4.16	0.81
ARIMA-XGB	20.52	23.42	10.58	0.11
ARIMAX -XGB	37.05	40.94	20.17	-1.72

The RF model ranked second in terms of predictive performance, achieving an RMSE of Rs. 10.80, MAE of Rs. 8.27, MAPE of 4.16%, and R^2 of 80.63%. In contrast, both hybrid models demonstrated poor predictive performance. The ARIMA-XGB hybrid recorded an RMSE of Rs. 23.42 and an R^2 of 11%. The ARIMAX-XGB hybrid exhibited the weakest overall performance, with a negative R^2 value, indicates that the hybrid model performs worse than the training-mean baseline on the test set, suggesting poor generalization to unseen data and possible overfitting or model misspecification.

Moreover, both standalone linear models were excluded from the model comparison because they failed to satisfy certain diagnostic assumptions and were therefore considered unsuitable for modeling the complex dynamics of daily coconut retail prices in the Pettah Market. The poor performance of the hybrid models may also be explained by the noisy residuals produced by the linear models, which may have adversely affected the forecasting performance of the hybrid approach.

5 CONCLUSION

The study developed and compared market-specific daily coconut price

forecasting models for the Dambulla Economic Centre and Pettah Main Market in Sri Lanka. The study revealed meaningful market disparities. Daily coconut retail and wholesale prices in the Pettah market were significantly higher than those in the Dambulla market during the study period, consistent with its urban retail market structure and longer supply chain. Overall, the standalone machine learning models demonstrated superior predictive performance in predicting daily coconut retail prices across both markets. Random Forest emerged as the best-performing model for the Dambulla market, achieving the lowest RMSE (Rs. 2.26), MAE (Rs. 1.56), and MAPE (1.02%), while attaining the highest R^2 (98.15%). For the Pettah market, XGB outperformed the competing models, recording an RMSE of Rs. 8.32, MAE of Rs. 5.10, MAPE of 2.40%, and R^2 of 88.50%. The results suggest that incorporating classical linear models into the hybrid framework may have introduced residual noise that limited forecasting accuracy. Consequently, standalone machine learning models appeared to be more effective in capturing the complex and volatile dynamics of agricultural commodity prices. These market-specific predictions provide early insights into price disparities, support wholesale–retail spread monitoring, and help stakeholders make informed decisions to reduce market price variation.

REFERENCES

- Abeysekara, M. G. D., & Waidyarathne, K. P. (2020). The coconut industry: A review of price forecasting modelling in major coconut producing countries. *Coconut Research & Development Journal*, *36*, 17–26.
- Ahalya, S. P., Murugananthi, D., Rohini, A., Parimala Devi, R., & Kalpana, M. (2023). Study on predicting the price of coconut in Tamil Nadu market. *Asian Journal of Agricultural Extension, Economics & Sociology*, *41*(10), 149–158.
- Bandara, J. M. A. K., & Kumari, D. A. T. (2020). Factors affecting coconut price determinants in Sri Lanka: An integrative review. *Applied Economics and Business*, *4*(2), 61–71.
- Dilrukshi, K. B. M. T., Jayalath, K. V. V. N., Dissanayake, S. N., & Perera, S. M. S. D. (2022). Retail price volatility: The case of coconut in Sri Lanka. *Sri Lankan Journal of Agricultural Economics*, *23*(1).
- Krishna, P. A., Narayana, G. V., Kotha, S. K., & Pattnayak, D. (2026). Machine learning based agricultural price forecasting for major food crops in India using environmental and economic factors. *Biology and Life Sciences Forum*, *54*(1), Article 7.
- Noor, N. M., Al Bakri Abdullah, M. M., Yahaya, A. S., & Ramli, N. A. (2015). Comparison of linear interpolation method and mean method to replace the miss ing values in environmental dataset. *Materials Science Forum*, *803*, 278–281.
- Prasert, T., & Rungreunganun, V. (2021). Thai coconut price forecasting using ARIMA model. *International Journal of Advanced Research in Engineering and Technology*, *12*(1), 950–961.
- Sarpong, S. R. M. N. Y., Sokkalingam, R., Othman, M., Azad, A. S., Syahrantau, G., & Arifin, Z. (2023). Intelligent hybrid ARIMA-NARNET time series model to forecast coconut price. *IEEE Access*, *11*, 48568–48577.

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research - 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

A Hybrid Deep Learning Model for Forecasting of Stock Market Prices in Colombo Exchange

De Silva RSA[1], Lakshmi MPDS[2]
[1,2] *Department of Mathematical Sciences, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*
sadeepaarosh21@gmail.com[1] , saroja@wyb.ac.lk[2]

ABSTRACT

The prediction of financial stock market data becomes more complicated because of the non-linear, volatile nature of it. But forecasting of this financial data is very important for financial decision making and risk management. Prediction difficulties of this type of data could be successfully overcome with machine learning techniques. Deep learning is one of the most applicable techniques in stock price prediction. Deep learning models such as Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) are mostly used in predicting stock prices. None of study has been done using the combination of three such models in a one single hybrid architecture for forecasting of stock market prices in Colombo Stock Exchange. The performance of standalone deep learning models and hybrid architectures are compared in this study for predicting the closing stock price at Colombo Stock Exchange. After the data preprocessing, all possible combinations of RNN, CNN, LSTM and GRU comprising up to three combined architectures, were tested and the model performance was evaluated through the use of the Root Mean Squared Error, Mean Absolute Error and Mean Absolute Percentage Error. Among the single models LSTM performed better with a Mean Absolute Error (MAE) value of 0.011070 but have high prediction errors. In two model architectures CNN-LSTM model performed well with a MAE value of 0.1185 When consider about the three model hybrid architectures CNN-GRU-LSTM architecture performed better with a MAE value of 0.1150. This shows that hybrid models outperform standalone models.

KEYWORDS: Deep learning, Forecasting, Hybrid architectures, Stock prices

1 INTRODUCTION

The stock market is a volatile market and forecasting stock prices is a difficult but a very crucial task (He et al., 2023). It is also an important enabler for economic development as it raises capital for companies and enables investors to be part of the economic growth of the nation. A variety of factors such as supply and demand, company's performance, market sentiment, government policies, investor confidence and international events may affect the stock prices (Team, 2026).

Traditional forecasting methods struggle to identify different complex patterns in stock market data, but machine learning and deep learning techniques have the ability to identify those patterns. When consider about hybrid models, those models are more capable of understanding nonlinear patterns effectively than using standalone deep learning models. Correct stock price prediction helps investors, financial analysts and companies make informed stock market investments, minimize financial risks and optimize investment portfolios. In this study a newly implemented hybrid deep learning models were introduced to forecast the daily stock market prices in Colombo Stock Exchange market.

2 LITERATURE REVIEW

Mondal et al. (2014) investigated Autoregressive Integrated Moving Average (ARIMA) for the stock market forecasting. In that study, they have discovered that ARIMA can predict more than 85 percent accurately on relatively linear and stationary data. But the model did not perform well in the face of nonlinear and highly volatile stock market conditions. The use of Artificial Intelligence (AI) and Machine Learning (ML) has provided more advanced methods when it comes to financial stock market forecasting (Ferdowsy, 2024). Yunita, et al. (2025) compared recurrent neural network architectures for time-series forecasting. They studied the performance of RNN, LSTM and GRU models. Results showed that LSTM and GRU were much more effective than the traditional RNN architectures, because of their capability of avoiding vanishing gradient problem and retaining information for longer periods. Even the performance of single deep learning models has been successful, recent studies show that there is still potential to improve forecasting performance with hybrid deep learning architectures (Adesina & Obokoh, 2025). A typical problem with stock data is missing data because of holidays, The study by Melina et al. (2024) explored higher-order methods for dealing with missing data in financial time-series data. Adesina and Obokoh (2025) showed that the hybrid model of Seasonal Autoregressive Integrated Moving Average (SARIMA)-LSTM had the best forecasting performance among all the models they tested. Researchers consistently find that deep learning models perform better than traditional methods when handling nonlinear and volatile financial datasets. Additionally, hybrid architectures have produced promising results by merging the strengths of various forecasting models (Zaheer et al., 2023).

The novelty of our study is identified two hybrid deep learning architectures for Colombo stock price exchange forecasting, with the aim of improving prediction accuracy and stability. The proposed approach provides useful insights for investors and financial analysts in making informed decisions and may also support policymakers in understanding stock price trends. Furthermore, our study contributes to the limited body of research on advanced deep learning based financial forecasting in the Sri Lankan stock market.

3 METHODOLOGY

3.1 Data Collection

The data in this study is historical daily stock price data from 2009 to 2023 in Colombo Stock Exchange (CSE). There are about 5027 daily observations of stock market prices. In that data set the targeted variable was the closing stock price, which shows a very high non-linear volatile nature. Therefore, this data set is more suitable for my study.

3.2 Data Preprocessing

Stock market data contains missing values, which can be caused by non-trading days. The missing of observations can cause the data to be non-sequential and impact on the learning ability of deep learning models. The method used for filling missing values was the concave interpolation method and this method was implemented based on the methodology

suggested by Melina et al., (2024). The concave function is given in equation (1).

$$P_t = \frac{P_{t-1}+P_{t+1}}{2} \quad (1)$$

where, P_t is estimated value at time t, P_{t-1} is previously available observation and P_{t+1} is available observation.

The outliers were detected using Inter Quartile Range (IQR) method to reduce the impact of extreme values.

The Min–Max normalization (Equation (2)) was used for the normalizing of the data set,

$$X' = \frac{X-X_{min}}{X_{max}-X_{min}} \quad (2)$$

where, X' is the normalized value, X is the original value, $Xmin$ is the minimum value of the feature and $Xmax$ is the maximum value of the feature.

A 60-day lookback window was used where the stock price data for each of the past 60 consecutive trading days were use to predict the closing stock price of the next trading day. Then the data set was split into the ratio of 80:20 with 80% of the observations allocated to the training set and 20% to the test set.

3.3 Model Training

All the selected Standalone, two model and three model hybrid architectures were trained with a standard training procedure to facilitate the comparison of those architectures. The implementation was done using the Python programming language with the aid of the TensorFlow and Keras deep learning frameworks. Adam optimizer was used to compile those hybrid models and early Stopping was used in the training process to prevent overfitting and to achieve higher generalization. Every model was trained at most 75 epochs, using a batch size of 32. Three evaluation metrics were used to evaluate the models. They are Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and the Mean Absolute Percentage Error (MAPE) Which are all useful indicators of how well the forecasting models are performing.

4 RESULTS AND DISCUSSION

The developed standalone, two model hybrid and three model hybrid deep learning architectures were trained and evaluated using the same preprocessed dataset and identical model parameters. The performances of the above used models were evaluated using RMSE, MAE and MAPE. Table 1 represent the forecasting performance of the three-model hybrid deep learning architectures. This results clearly demonstrate that the developed models were capable of learning the temporal characteristics of the past stock price data. But differences in forecasting accuracy were observed among the various architectures, which shows that the architecture you select significantly affect the prediction performance. Among these developed architectures, the CNN-GRU-LSTM model achieved the best overall performance as the combination of CNN, GRU, and LSTM enabled the model to learn both local feature representations and long-short-term sequential dependencies simultaneously, resulting in highly accurate stock price predictions. which has a 0.1150 MAE value followed with a 1.0927 MAPE value. But when we considered the models such as GRU-LSTM-CNN and LSTM-GRU-CNN, they were not able to provide satisfactory results. These two Models recorded relatively large prediction errors including MAPE values with 12.8959 and 13.2877

Table 1: Performance comparison of three model hybrid deep learning architectures

Model	MAE	RMSE	MAPE (%)
CNN-LSTM-GRU	0.2689	0.3653	2.5676
CNN-GRU-LSTM	0.1150	0.1716	1.0927
GRU-CNN-LSTM	0.2601	0.3567	2.5148
GRU-LSTM-CNN	1.4176	1.6017	12.8959
LSTM-CNN-GRU	0.2244	0.3202	2.1698
LSTM-GRU-CNN	1.4693	1.6596	13.2877

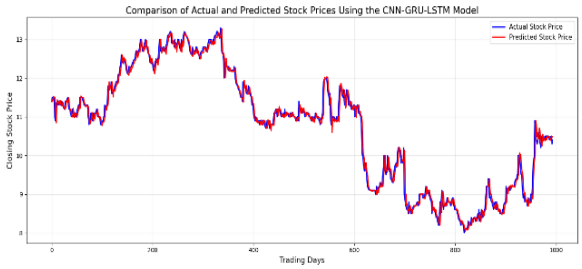


Figure 1:Comparison of actual and predicted stock prices

respectively, which indicate that the arrangement of layers within a hybrid architecture has a significant impact on forecasting performance. This improvement can be attributed to the complementary learning capabilities of different neural network architectures. The CNNs effectively extract informative local features, whereas LSTM and GRU networks capture sequential dependencies over different time horizons. By integrating these capabilities into a single architecture, that three model hybrid architectures were able to capture the nonlinear and dynamic behavior of stock market data more effectively than standalone networks. Fig.1 shows the relevant plot for the predicted values of the CNN-GRU-LSTM architecture. The and predicted stock prices further support the findings of this study. The best performing model produced prediction curve that closely followed the actual stock price movements throughout the testing period, with little deviations during the period. This close agreement between the predicted and the actual values confirms the ability of the proposed hybrid deep learning models to provide reliable forecasts for the stock prices.

5 CONCLUSION

This study investigated the use of deep learning and hybrid deep learning architectures for forecasting the closing stock price in the Colombo Exchange stock market. The results show that deep learning models were able to effectively learn the temporal patterns in past stock price data, but their forecasting performance varies considerably depending on the architecture used. Out of the evaluated models, the CNN-GRU-LSTM architecture showed the best overall forecasting performance, with MAE of 0.1150. The close agreement between the actual and predicted prices during the testing period supports the effectiveness of this architecture. The findings suggest that combining CNN, GRU and LSTM allows different characteristics of stock price data to be learned within the same model. The architecture benefits from the feature

extraction capability of CNN and the ability of GRU and LSTM to learn sequential patterns from past observations. In triple hybrid models performance varied significantly when the order of the layers was changed. While CNN-GRU-LSTM performed highly effectively, GRU-LSTM-CNN and LSTM-GRU-CNN produced relatively high error values. This indicates that the sequence in which different neural network components are arranged is an important consideration when developing hybrid forecasting models. As the final conclusion this shows that the hybrid models are performing better than the standalone models and also the sequence of arranging the neural network components effect the accurate prediction capability of a hybrid model.

REFERENCES

- Adesina, O. S., & Obokoh, L. O. (2025). A hybrid framework of deep learning and traditional time series models for exchange rate prediction. Scientific African, 29, e02818.
- Ferdowsy, F. (2024, March 7). Deep Learning-Based Time Series Prediction Techniques. Erepo.uef.fi; Itä-Suomen yliopisto.
- He, M., Wang, Y., Zeng, Q., & Zhang, Y. (2023). Forecasting aggregate stock market volatility with industry volatilities: The role of spillover index. Research in International Business and Finance, 65, 101983–101983.
- Melina, M., Sukono, Napitupulu, H., & Mohamed, N. (2024). Investment risk forecasting model using extreme value theory approach combined with machine learning. AIMS Mathematics, 9(11), 33314–33352.
- Mondal, P., Shit, L., & Goswami, S. (2014). Study of Effectiveness of Time Series Modeling (Arima) in Forecasting Stock Prices. International Journal of Computer Science, Engineering and Applications, 4(2), 13–29.
- Team, C. (2026, March 31). 7 Key Factors Affecting Stock Market Prices: What Drives Market Movements? CapMint.
- Yunita, A., Pratama, M. I., Almuzakki, M. Z., Ramadhan, H., Akhir, E. A. P., Firdausiah Mansur, A. B., & Basori, A. H. (2025). Performance analysis of neural network architectures for time series forecasting: A comparative study of RNN, LSTM, GRU, and hybrid models. MethodsX, 15, 103462.
- Zaheer, S., Anjum, N., Hussain, S., Algarni, A. D., Iqbal, J., Bourouis, S., & Ullah, S. S. (2023). A Multi Parameter Forecasting for Stock Time Series Data Using LSTM and Deep Learning Model. Mathematics, 11(3), 590.

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research - 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

An Empirical Investigation of the Short-Run and Long-Run Dynamics of Electricity Consumption and Industrial Output in Sri Lanka: A VECM Approach

Anjalee WAH[1], Aponsu GMLM[2]
[1,2]*Department of Mathematical Sciences, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*
harshikaanjali836@gmail.com[1], lakshana@wyb.ac.lk[2]

ABSTRACT

This study examines the short-run and long-run relationship between electricity consumption and industrial output in Sri Lanka using monthly data from January 2000 to December 2024. Given recurring electricity supply constraints, dependence on imported fuels, and periodic power shortages, understanding the electricity–industrial output nexus is important for energy and industrial policy. Data on electricity consumption, industrial output, exchange rates, interest rates, and the Colombo Consumer Price Index (CCPI) were obtained from the Central Bank of Sri Lanka, the Department of Census and Statistics, and the Ceylon Electricity Board. The study employs the Johansen cointegration approach and a Vector Error Correction Model (VECM), incorporating exchange rate, interest rate, and inflation controls. Unit root tests confirm that all variables are integrated of order one, while the Johansen trace test identifies one cointegrating relationship. The long-run estimates indicate that the exchange rate is positively associated with industrial output, whereas inflation has a significant negative relationship; electricity consumption has no statistically significant long-run effect. The error correction term is negative and significant (-0.0501; p=0.0031), indicating that approximately 5.01% of disequilibrium is corrected monthly, with full adjustment taking about 20 months. Wald tests confirm significant short-run dynamics, while Granger causality tests indicate unidirectional causality from electricity consumption and interest rates to industrial output. Residual diagnostics confirm the absence of serial correlation and structural instability, though non-normality is detected, plausibly reflecting the 2022 economic crisis. Overall, the findings highlight the importance of reliable electricity supply, exchange rate stability, and coordinated monetary and industrial policies in supporting Sri Lanka's industrial development.

KEYWORDS: Cointegration, Electricity Consumption, Granger Causality, Industrial Output, Vector Error Correction Model

1 INTRODUCTION

Electricity is a fundamental driver of industrial activity in developing economies. In Sri Lanka, manufacturing and export-oriented industries depend heavily on stable and affordable electricity supply, yet the energy landscape is characterized by heavy reliance on imported fossil fuels, a hydropower-dominated generation mix vulnerable to seasonal variability, and periodic supply shortages. The 2022 economic and energy crisis, which produced rolling blackouts of up to 13 hours per day, illustrated how supply-side electricity shocks can disrupt industrial output.

Although the energy–growth nexus has been studied extensively for Sri Lanka, most existing work uses aggregate Gross Domestic Product rather than industrial output, obscuring sector-specific dynamics, and few studies embed the relationship within a multivariate framework accounting for exchange rate, interest rate, and inflation effects. This study addresses these gaps using a Vector

Error Correction Model (VECM) that jointly incorporates electricity consumption, industrial output, and three macroeconomic controls, using monthly data covering January 2000 to December 2024 (300 observations) obtained from the Central Bank of Sri Lanka, the Department of Census and Statistics of Sri Lanka, and the Ceylon Electricity Board.

The specific objectives are to: (1) examine the long-run equilibrium relationship between electricity consumption and industrial output using the Johansen cointegration approach; (2) estimate the short-run dynamics and the speed of adjustment through the Vector Error Correction Model; (3) determine the direction of causality using Granger causality tests; and (4) assess the dynamic response of industrial output to shocks using impulse response functions and variance decomposition. The findings are intended to inform energy infrastructure planning and industrial policy in Sri Lanka.

2 LITERATURE REVIEW

The relationship between energy consumption and economic growth has been studied since Kraft and Kraft (1978), who found that Gross National Product Granger-caused energy consumption in the United States rather than the reverse. Subsequent studies report four broad patterns: unidirectional causality from energy to output, unidirectional causality from output to energy, bidirectional causality, and no significant relationship, with results varying by country, sample period, and methodology.

For Sri Lanka, Morimoto and Hope (2004) documented a bidirectional relationship between electricity supply and economic growth, while Kiriella et al. (2020) reviewed the literature and concluded that the direction of causality is highly sensitive to variable choice and sample period. Asghar (2008) applied Johansen cointegration and Granger causality within a Vector Error Correction Model across five South Asian countries and found that, for Sri Lanka, causality ran from Gross Domestic Product to electricity consumption rather than the reverse, using aggregate output rather than industrial output.

Evidence from other developing economies is mixed: Ghosh (2002) and Jumbe (2004) reported bidirectional causality for India and Malawi, Adom (2011) and Aktas and Yilmaz (2008) found causality running from output to electricity for Ghana and Turkey, and Shiu and Lam (2004) found electricity consumption Granger-caused Gross Domestic Product in China, consistent with energy availability acting as a binding constraint in rapidly industrializing economies, whereas Cheng (1995) found no significant link for the more service-oriented United States economy.

More recent evidence reinforces both the importance and the sector-specific nature of the electricity–output link. Hassan and Kankanamge (2021) combined multivariate Johansen (1988) cointegration with the Toda–Yamamoto causality test for Sri Lanka and found that electricity consumption and economic growth are cointegrated, with causality running from growth to electricity consumption, again using aggregate output rather than industrial output. At the industry level, Onatunji (2022) examined electricity consumption and industrial output in ECOWAS countries over 1971–2017 and found unidirectional causality

from electricity consumption to industrial output in Benin, Burkina Faso, Gambia, Ghana, Nigeria, and Sierra Leone, but no causal link in several other countries, highlighting that results are country-specific.

A recurring limitation in the Sri Lankan literature is reliance on aggregate Gross Domestic Product, which conflates manufacturing, services, and agriculture and obscures the sector-specific role of electricity, even though manufacturing is typically the most electricity-intensive and supply-sensitive sector. Few studies embed a full set of macroeconomic controls: exchange rate depreciation raises import costs while improving export competitiveness, and interest rates affect the cost of capital for industrial investment, both interacting with electricity access in shaping industrial expansion. This study addresses these gaps by focusing on industrial output as a distinct dependent variable within a multivariate Vector Error Correction Model that also controls for the exchange rate, interest rate, and inflation.

3 METHODOLOGY

Monthly data covering January 2000 to December 2024 (300 observations) were used for five variables: electricity consumption, industrial output (the average across low-, medium-, and high-technology industries), the exchange rate, the interest rate, and the Colombo Consumer Price Index as a measure of inflation, sourced from the Central Bank of Sri Lanka, the Department of Census and Statistics of Sri Lanka, and the Ceylon Electricity Board. All variables were transformed into natural logarithms before analysis.

The empirical strategy proceeds in four steps. First, the Augmented Dickey–Fuller and Phillips–Perron tests establish the order of integration of each series, following the Augmented Dickey–Fuller test equation;

$$\Delta Y_t = \delta + \gamma_t + \varphi Y_{t-1} + \sum \theta_i \Delta Y_{t-i} + u_t \quad (1)$$

where Y_t is the natural log of the series, φ is the parameter tested against a unit root, and the lag order is chosen using the Schwarz Information Criterion. Second, the optimal Vector Autoregression lag length is selected using the Akaike, Schwarz, and Hannan–Quinn information criteria. Third, the Johansen (1988) maximum likelihood procedure tests for cointegration using the Trace and Maximum Eigenvalue statistics. Fourth, if cointegration is confirmed, the Vector Autoregression is reformulated as a VECM:

$$\Delta X_t = \phi + \alpha\beta' X_{t-1} + \sum \Gamma_i \Delta X_{t-i} + \varepsilon_t \quad (2)$$

where $\beta' X_{t-1}$ is the cointegrating (long-run equilibrium) relationship and α is the vector of adjustment speeds.

The joint short-run significance of each variable is assessed using the Wald test, and predictive causality is examined using pairwise Granger causality tests. Impulse response functions, based on a Cholesky decomposition ordered by electricity, industrial production, exchange rate, interest rate, and CCPI, trace the response of industrial production to a one-standard-deviation shock in each variable over a twelve-month horizon. Forecast error variance decomposition quantifies each variable's contribution to the variance of industrial production's forecast error. Model adequacy is

assessed using the Breusch–Godfrey serial correlation and Autoregressive Conditional Heteroskedasticity (ARCH) Lagrange Multiplier tests, correlogram Q-statistics, the normality test, and the Cumulative Sum (CUSUM) and CUSUM of Squares stability tests.

4 RESULTS AND DISCUSSION

4.1 Unit Root and Cointegration Results

Table 1 shows Augmented Dickey–Fuller unit root test results (Phillips–Perron tests gave the same result but aren't shown). For all five variables, the p-value is above 0.05 at the level (non-stationary) but below 0.05 after first-differencing (stationary) meaning every variable is integrated of order one, I(1). That satisfies the requirement for running the Johansen cointegration test.

The Schwarz and Hannan–Quinn information criteria select a lag length of two for the underlying Vector Autoregression, so the Vector Error Correction Model is estimated with one lag in first differences. The Johansen Trace statistic for the null hypothesis of no cointegration is 70.268, exceeding the 5% critical value of 69.819 ($p = 0.046$), and therefore indicates at least one cointegrating relationship.

The Maximum Eigenvalue test, however, does not reject the null at the conventional 5% significance level. This difference reflects the distinct hypotheses tested by the two statistics and suggests that the evidence on the cointegrating rank is not unequivocal. As a robustness check, the Johansen procedure was re-estimated under alternative lag specifications, with the results remaining consistent with a cointegrating rank of r=1. Accordingly, the subsequent analysis adopts r=1, while acknowledging the mixed evidence across the two Johansen statistics (Johansen, 1988).

4.2 Long-Run Equation

The normalized long-run equation for industrial production is:

$$L_IP = -0.025\, L_ELEC + 0.087\, L_EXR + 0.035\, L_INT - 0.113\, L_CCPI + 4.789 \quad (3)$$

Only the exchange rate ($t = 2.351$) and the CCPI ($t = -2.510$) are statistically significant in the long run; electricity ($t = -0.621$) and the interest rate ($t = 1.565$) are not. A weaker rupee is associated with higher long-run industrial output, plausibly through improved export competitiveness that offsets higher import costs, while sustained inflation erodes industrial output, plausibly through rising input costs.

4.3 Short-Run Dynamics

Table 2 summarizes the Wald test results for the joint significance of each variable's short-run coefficients. The error correction term is negative and significant ($\alpha = -0.050$, $t = -2.99$, $p = 0.003$), implying that approximately 5% of any deviation from long-run equilibrium is corrected each month, with full adjustment taking roughly 20 months.

Table 1: Augmented Dickey–Fuller Unit Root Test Results

Variable	Level	1st diff.	Order
Electricity	0.6120	0.0004	I(1)
Ind. production	0.2497	0.0000	I(1)
Exchange rate	0.5280	0.0000	I(1)
Interest rate	0.0658	0.0000	I(1)
Colombo CPI	0.2843	0.0000	I(1)

Table 2: Wald Test for Joint Short-Run Coefficient Significance

Variable	Chi-sq.	p	Result
Ind. production	8.017	0.0046	Sig.
Electricity	10.203	0.0014	Sig.
Exchange rate	4.039	0.0445	Sig.
Interest rate	4.429	0.0353	Sig.
Colombo CPI	1.554	0.2125	Not sig.

A 1% increase in the lagged monthly change in electricity consumption is associated with an approximate 0.021% increase in the current monthly change in industrial production.

4.4 Causality, Impulse Responses, and Variance Decomposition

Table 3 shows the pairwise Granger causality results. Electricity consumption Granger-causes industrial production with no evidence of reverse causality, and the interest rate Granger-causes industrial production; the exchange rate and inflation show no significant causal relationship in either direction.

Impulse responses show that an electricity shock raises industrial production, peaking near period two before gradually declining, while an exchange rate shock is initially negative and an interest rate shock is positive and persistent.

Variance decomposition shows industrial production is mostly explained by its own shocks (94.4% at month 12), with interest rate, inflation, and electricity contributing small, growing shares; the exchange rate contributes negligible variance.

Table 3: Pairwise Granger Causality Results

X → Y (H_0: no Granger-cause)	F	p	H_0
ELEC → IP	6.225	0.0023	Rej
IP → ELEC	1.274	0.2812	NR
INT → IP	3.150	0.0443	Rej
IP → INT	1.935	0.1463	NR
EXR → IP	1.541	0.2158	NR
CCPI → IP	2.269	0.1052	NR

Note. IP = industrial production; ELEC = electricity; EXR = exchange rate; INT = interest rate; CCPI = Colombo CPI; "Rej" = reject H0 (5%); "NR" = fail to reject.

4.5 Diagnostics Checking

The Breusch–Godfrey serial correlation test ($F = 1.472$, $p = 0.231$) and the ARCH Lagrange Multiplier test ($F = 0.222$, $p = 0.638$) fail to reject their null hypotheses, confirming the absence of serial correlation and conditional heteroskedasticity. The CUSUM test also confirms parameter stability. The Jarque–Bera test rejects normality ($p < 0.001$), reflecting outliers linked to the 2022 economic crisis; this is common in macro-financial series and does not affect the consistency of the coefficient estimates, though it affects the exactness of finite-sample inference.

5 CONCLUSION

This study examined the long-run and short-run dynamics between electricity consumption and industrial output in Sri Lanka over January 2000 to December 2024 using a five-variable Vector Error Correction Model. All variables were integrated of order one, and the Johansen Trace test identified one cointegrating

relationship. In the long run, only the exchange rate and inflation were statistically significant; electricity consumption was not. However, the error correction term was negative and significant, with about 5% of disequilibrium corrected monthly, and electricity had a significant positive short-run effect on industrial production, confirmed by the Wald test and by unidirectional Granger causality running from electricity to industrial output. Impulse response and variance decomposition results reinforced electricity's meaningful, if secondary, role relative to industrial output's own inertia.

These findings suggest that reliable electricity supply matters most for the short-run performance of Sri Lankan industry, while exchange rate and price-level conditions dominate the long run. Minimizing short-term supply disruptions is therefore likely to yield the most immediate industrial output gains, alongside exchange rate stability and inflation control as complementary policy levers.

REFERENCES

- Aktas, C., and Yilmaz, V. (2008). Causality between electricity consumption and economic growth in Turkey. *Journal of Economic and Social Research, 10*(1), 45–60.
- Asghar, Z. (2008). Energy–GDP relationship: A causal analysis for the five countries of South Asia. *Applied Econometrics and International Development, 8*(1), 167–180.
- Cheng, B. S. (1995). An investigation of cointegration and causality between energy consumption and economic growth. *Journal of Energy and Development, 21*(1), 73–84.
- Ghosh, S. (2002). Electricity consumption and economic growth in India. *Energy Policy, 30*(2), 125–129.
- Hassan, Y. A., and Kankanamge, A. (2021). Electricity consumption and economic growth in the presence of structural breaks: Evidence from Sri Lanka. *Journal of Smart Economic Growth*, *6*(1), 1–17.
- Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control, 12*(2), 231–254.
- Jumbe, C. B. L. (2004). Cointegration and causality between electricity consumption and GDP: Empirical evidence from Malawi. *Energy Economics, 26*(1), 61–68.
- Kiriella, K. K. C. S., Peiris, W. L. T., Samarakoon, W. H. A., Samarasinghe, K. T. A. B., Wijayapala, W. D. A. S., and Dias, M. P. (2020). The relationship between economic growth and electricity consumption in Sri Lanka. *Journal of Economics, Business and Management, 8*(3), 1–6.
- Kraft, J., and Kraft, A. (1978). On the relationship between energy and GNP. *Journal of Energy and Development, 3*(2), 401–403.
- Morimoto, R., and Hope, C. (2004). The impact of electricity supply on economic growth in Sri Lanka. *Energy Economics, 26*(1), 77–85.
- Onatunji, O. G. (2022). Electricity consumption and industrial output: Fresh evidence from economic community of West African states (ECOWAS). *Journal of Economic and Administrative Sciences*, *41*(1), 381–398.
- Shiu, A., and Lam, P.-L. (2004). Electricity consumption and economic growth in China. *Energy Policy, 32*(1), 47–54.

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research -2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Battery Remaining Useful Life Estimation in Electric Vehicles Using Multiple Linear Regression and Random Forest Models

Palansooriya PHAI [1], Francisco GS[2]
[1,2]*Department of Mathematical Sciences, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*
anuupalansooriya0206@gmail.com[1] , gfrancisco@wyb.ac.lk[2]

ABSTRACT

Electric vehicles (EVs) are considered a promising solution for sustainable transformation. However, one of the major challenges associated with EVs is battery degradation, which can reduce energy storage capacity, driving range, and overall vehicle performance over time. Therefore, accurate estimation of battery Remaining Useful Life (RUL) is essential for effective battery management and maintenance. To address this challenge, Multiple Linear Regression (MLR) and Random Forest (RF), a Machine Learning (ML) algorithm, were employed to estimate the RUL of EV batteries. The Hawaii Natural Energy Institute (HNEI) dataset was used to evaluate and compare the performance of these models. The RF model was trained and evaluated across three train-test split ratios using Coefficient of Determination (R^2), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) to identify the optimal train-test split ratio. The MLR model was developed in R software using a Variance Inflation Factor (VIF) based multicollinearity reduction procedure for the 80:20 split. The results showed that the RF model achieved the best performance, with the lowest prediction error (MAE=13.8218 cycles, RMSE=21.6124 cycles, MAPE=11.8529%) and the highest R^2 value (0.9953). Subsequently, an Explainable Artificial Intelligence (XAI) analysis was conducted on the best-performing model using Shapley Additive Explanations (SHAP) plot. XAI analysis employing SHAP plot was subsequently conducted on the best-performing model. The analysis consistently identified Time at 4.15V, Discharge Time, and Time Constant Current as the most influential predictors of battery RUL, with these findings further corroborated by impurity based and SHAP based feature importance measures.

KEYWORDS: Electric Vehicles, Multiple Linear Regression, Random Forest, Remaining Useful Life

1 INTRODUCTION

Elecric Vehicles (Evs) are becoming increasingly popular, due to their potential to reduce greenhouse gas emissions, improve urban air quality, reduce dependency on fossil fuels, improve energy efficiency, and reduce noise pollution. A key challenge in battery management lies in accurately estimating the Remaining Useful Life (RUL), which provides an essential measure of how long a battery can function reliably. RUL is a well-liked research area in the field of electric battery research, as it has clear significance for assessing the safety of batteries. According to a widely accepted generalization, a battery's lifespan degrades over time when its remaining capacity reaches 70-80% of its initial value. This degradation pattern is very complex and non-linear. Without accurate RUL predictions, systems are defenseless against unplanned breakdowns, security vulnerabilities, and inefficient maintenance practices. These challenges hinder the development of sustainable energy systems and drive the demand for more reliable and interpretable solutions.

The objectives of this research are as follows:

- Develop and evaluate the Multiple Linear Regression (MLR) and Random Forest (RF) models for predicting the RUL of lithium-ion (Li-ion) batteries.
- Find the best split ratio and best-performing model.
- Identify the most influential battery characteristics affecting battery RUL using Explainable Artificial Intelligence (XAI) techniques.

2 LITERATURE REVIEW

Recent studies have demonstrated the effectiveness of various Machine Learning (ML) techniques for battery RUL prediction. Extreme Gradient Boosting (XGBoost) has been identified as an effective algorithm, achieving near-perfect prediction performance in Li-ion battery RUL estimation (Karthick et al., 2024). Similarly, RF has demonstrated strong predictive performance, with improved evaluation metrics and lower execution time, highlighting its suitability for accurate RUL prediction (Sekhar et al., 2023). Shiblee and Laaksonen (2025) showed that the Particle Swarm Optimization Algorithm (PSO) is used to optimize different parameters of XGBoost to ensure precise and reliable predictions of State of Health (SOH) and RUL for Li-ion batteries.

In addition to prediction accuracy, model interpretability has received increasing attention. Several models have been developed for RUL prediction and classification, with Shapley Additive Explanations (SHAP) based analysis used to interpret model predictions and a Graphical User Interface (GUI) implemented to support practical battery RUL prediction (Paneru et al., 2024).

Ensemble based approaches have also shown promising results, with RF combined with Extra Trees Regressor achieving the best performance based on Mean Squared Error (MSE) (Onyenagubo et al., 2025). Samanta and Williamson (2023) have identified RF and Gradient Boosting (GB) as consistently strong performers, while also noting that linear regression still holds value as a simple, interpretable baseline model.

Other studies have explored hybrid modelling approaches, where Support Vector Regression (SVR) was used to predict the low-frequency component representing the main battery degradation trend, while a Long Short-Term Memory (LSTM) network optimized using the Sparrow Search Algorithm (SSA) was employed to predict the high-frequency component associated with capacity regeneration (Fan et al., 2025). Tran et al. (2025) demonstrated that separating the stages of the ML pipeline preprocessing, feature engineering, training, and evaluation helps prevent data leakage and makes results more reproducible.

3 METHODOLOGY

3.1 Dataset Description

The Hawaii Natural Energy Institute (HNEI) evaluated 14 NMC-LCO 18 650 batteries with a nominal capacity of 2.8 Ah after being cycled over a thousand times at a temperature of 25°C with a CC-CV rate of C/2 and a discharge rate of 1.5C. There were 8 independent variables and the RUL target variable, with 15064 data points. The independent variables were cycle index(F1), discharge time (s)(F2), time at 4.14 V (s)(F3), time constant current (s)(F4), decrement 3.6-3.4 V (s)(F5), maximum voltage discharge (V)(F6), minimum voltage

charge (V)(F7), and charging time (s)(F8).

3.2 Dataset Preprocessing

The battery RUL prediction was carried out in a step-by-step process. After data collection and exploratory data understanding, the dataset was cleaned by handling missing values and removing duplicate records. Exploratory data analysis (EDA) was then carried out using descriptive statistics and correlation analysis (heatmap), followed by feature selection. Outliers were detected and handled using Interquartile Range (IQR) method. The cleaned dataset was divided into training and testing subsets using three train-test split ratios (80:20, 70:30, and 60:40), and normalization was applied to the training data. MLR was developed in R as the statistical approach, and RF was developed in Jupyter Notebook (Python) as the ML approach, chosen for its demonstrated effectiveness in battery RUL prediction in most previous studies. Model performance on the test data was assessed using R^2, RMSE, MAE, and MAPE. The optimal split ratio was the one that yielded the highest R^2 and the lowest RMSE, MAE, and MAPE. The two models were then compared under that ratio to identify the best-performing approach, to which XAI was then applied to the selected model before deployment as a web application.

4 RESULTS AND DISCUSSION

4.1 Descriptive Statistics

Descriptive statistics, including the mean, standard deviation (SD) and quartiles, were computed for all input features and the target variable (RUL) to summarize their central tendency, spread and distribution. The dataset contains 15,064 observations with no missing values. Cycle index ranges from 1 to 1,134 (mean = 556.16, SD = 322.38), and the target variable RUL ranges from 0 to 1,133 cycles (mean = 554.19, median = 551), indicating an even spread across the battery life span. Voltage variables are stable, with maximum discharge voltage averaging 3.9082 V (SD = 0.0910) and minimum charge voltage averaging 3.5779 V (SD = 0.1237).

4.2 Analysis of Cycle Index Variable

The distributions of RUL and Cycle Index are nearly symmetric and evenly spread in Fig.1. According to Fig. 2, Cycle Index and RUL have a perfect negative correlation of -1.00, indicating a complete linear inverse relationship between the two variables. Cycle Index is a counter variable among all independent variables, and it does not indicate the actual physical or chemical condition of the battery. Therefore, Cycle Index was removed from the dataset. After the outlier removal process, the dataset contained 14421 observations with 7 independent variables.

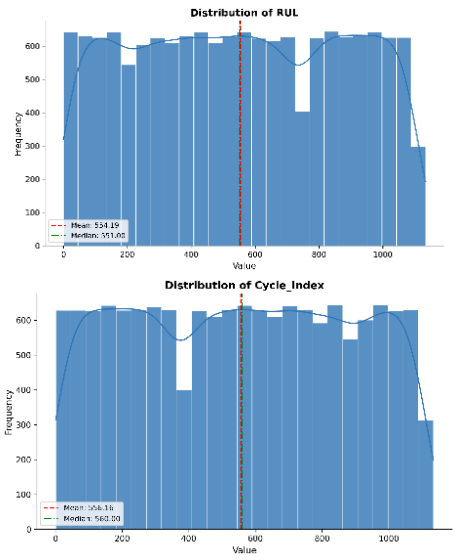


Figure 1: Distribution of Cycle Index and RUL

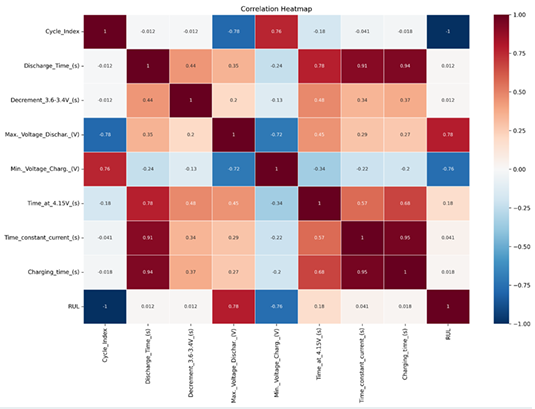


Figure 2: Correlation Heatmap

4.2. Multiple Linear Regression Model

The iterative VIF-based reduction procedure was applied across five model versions, as summarized in Table 1. The final MLR model was expressed as Equation 1.

$$RUL = -2.786 \times 10^{-16} + (6.177 \times 10^{-1} \times Max\ Voltage\ Discharge) - (7.442 \times 10^{-2} \times Min\ Voltage\ Charge) + (3.324 \times 10^{-1} \times Charging\ Time) \quad (1)$$

A Multiple R^2 value of 0.9554 was obtained. The assumptions of normality, linearity, homoscedasticity, and independence of residuals were satisfied**.**

Table 1: VIF Values across Model Iterations

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
F2	117.22	71.72	—	—	—
F5	32.35	25.31	24.45	19.56	—
F6	16.81	16.77	16.03	12.93	**5.10**
F7	4.09	3.56	3.56	3.37	**3.06**
F3	233.44	—	—	—	—
F4	91.21	66.87	30.28	—	—
F8	7.01	6.97	6.44	4.93	**3.53**

4.3. Analyze Multiple Linear Regression and Machine Learning Results

Table 2 reports on the metric values across ML and MLR models. Based on the results in Table 2, RF under the 80:20 split was confirmed as the best-performing model, as it has the highest R^2 and lower MAE, RMSE, and MAPE values. The R^2 value of 0.9953 obtained for RF is high. Therefore, it was verified that all metrics in Table 2 were computed exclusively on the held-out 20% test set, using predictions generated by a model trained only on the remaining 80% of the data. R^2 was calculated using Equation 2.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (2)$$

where y_i is the observed RUL, $\hat{y}_i$ is the predicted RUL, $\bar{y}$ is the mean observed RUL, and n is the number of test observations.

4.4. SHAP Global Feature Importance (Bar Plot)

To supplement impurity-based importance with a model-agnostic attribution measure, SHAP global feature importance was computed as the mean absolute SHAP value across all test samples, as shown in Fig. 3. Time at 4.15V (s) recorded the highest mean |SHAP| value of 156.24, followed by Discharge Time (s) at 97.07 and Time constant current (s) at 38.93. The SHAP-based ranking was consistent with the RF impurity scores, reinforcing the reliability of the feature selection and confirming that Time at 4.15V (s), Discharge Time (s), Time constant current (s) had the strongest average influence on predicted battery RUL across the test set.

Table 2: Comparative Model Performance Metrics under the 80:20 Train-Test Split

Model	R^2	MAE	RMSE	MAPE (%)
RF	0.9953	13.8218	21.6124	11.8529
MLR	0.9554	52.04	65.72	46.29

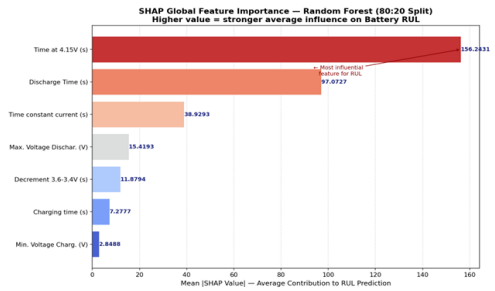


Figure 3: SHAP Global Feature Importance - Random Forest(80:20 Split)

5 CONCLUSION

This study sought to address a significant challenge in EV battery management, namely the accurate prediction of the RUL of Li-ion batteries. The RF model achieved superior predictive performance using the 80:20 train-test split, explaining 99.53% of the variation in battery RUL, with an MAE of 13.82 cycles and a MAPE of 11.85%. These findings demonstrate that RF can effectively capture the complex and non-linear degradation patterns of EV batteries. The application of XAI techniques further identified Time at 4.15 V, Discharge Time, and Time Constant Current as the most influential predictors of battery RUL.

The present study has several limitations that should be acknowledged. The best-performing model could be deployed in a prototype web-based application via ONNX Runtime to enable real time RUL prediction. The dataset was generated under controlled laboratory cycling conditions and therefore may not fully represent the variability associated with real-world driving conditions. In addition, the dataset was limited to 14 batteries of a single battery chemistry, which may restrict the generalizability of the findings to other battery chemistries and operating conditions. Future studies could address these limitations by incorporating larger and more diverse datasets, including batteries with different chemistries and data collected under real-world operating conditions. Further investigation using additional ML models and advanced hyperparameter optimization techniques could also provide a broader assessment of battery RUL prediction performance. Also, other XAI methods such as Partial Dependence Plots and Individual Conditional Expectation plots could be explored and compared with SHAP to provide a more comprehensive understanding of model behavior.

REFERENCES

- Fan, Y., Lin, Z., Wang, F., & Zhang, J. (2025). A hybrid approach for lithium-ion battery remaining useful life prediction using signal decomposition and machine learning. *Scientific Reports*, 15(1), 8161.
- Karthick, K., Ravivarman, S., & Priyanka, R. (2024). Optimizing electric vehicle battery life: a machine learning approach for sustainable transportation. *World Electric Vehicle Journal*, 15(2), 60.
- Onyenagubo, C., Ismail, Y., Belu, R., & Lacy, F. (2025). Forecasting the Remaining Useful Life of Lithium-Ion Batteries Using Machine Learning Models-A Web-Based Application. *Algorithms*, 18(6), 303.
- Paneru, B., Thapa, B., Mainali, D. P., Paneru, B., & Shah, K. B. (2024). Remaining useful life prediction for batteries utilizing an explainable AI approach with a predictive application for decision-making. *arXiv preprint arXiv:2409.17931*.

- Samanta, A., & Williamson, S. (2023). Machine learning-based remaining useful life prediction techniques for lithium-ion battery management systems: A comprehensive review. *IEEJ Journal of Industry Applications*, 12(4), 563-574.
- Sekhar, J. C., Domathoti, B., & Santibanez Gonzalez, E. D. (2023). Prediction of battery remaining useful life using machine learning algorithms. *Sustainability*, 15(21), 15283.
- Shiblee, M. F. H., & Laaksonen, H. (2025). Battery-Insight-PSO: A machine learning model for accurate prediction of state of health and remaining useful life in lithium-ion batteries. *Future Batteries*, 100114.
- Tran, K., Le, T., Huynh, B., Trinh, H. C., & Nguyen, V. R. (2025). End-to-End Framework for Predicting the Remaining Useful Life of Lithium-Ion Batteries. *arXiv preprint arXiv:2505.16664*.

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research-2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

A Comparative Study of Black-Scholes and Binomial Tree Model and Monte Carlo Simulation Method for Pricing European Call Options on Netflix Stock

Ekanayaka EMDJ[1], Karunananda GACM[2]
[1,2]*Department of Mathematical Sciences, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*
jalendra12014@gmail.com[1] , chamly@wyb.ac.lk[2]

ABSTRACT

Accurate valuation of options is essential for pricing derivatives and financial risk management. European option pricing involves models such as the Black-Scholes Model (BSM), Binomial Tree Model (BTM), and Monte Carlo (MC) simulation method, each with different assumptions and computational procedures. These differences can lead to different pricing results when implemented with market data. This study evaluates and compares the pricing performance of these three approaches for European call options on Netflix stock by examining their theoretical prices against observed market prices across multiple valuation dates and strike prices. Historical stock prices and option chain data were analyzed over 20 trading days from 20.04.2026 to 15.05.2026, using multiple strike prices with a fixed expiry date. All three methods tended to overestimate observed market prices, while the BTM performed more accurately, with the lowest values for Mean Absolute Error (MAE) of 0.8964, Root Mean Squared Error (RMSE) of 0.9401, and Mean Absolute Percentage Error (MAPE) of 40.0256%. The pairwise Wilcoxon signed-rank tests indicated statistically significant differences between all pairs of model errors. The moneyness analysis further showed that pricing performance varied across option categories, with the BSM recording a lower MAE of 0.9982 for In-the-Money(ITM) options. Overall, the results indicate that the BSM achieved the lowest overall pricing errors for the Netflix European call options examined, while the findings also demonstrate differences in model performance across moneyness categories. The study contributes an empirical comparison of the three pricing approaches using multiple valuation dates, strike prices, statistical error measures, and moneyness categories.

KEYWORDS: Binomial Tree Model, Black-Scholes Model, European Call Options, Monte Carlo Simulation, Option Pricing

1 INTRODUCTION

Financial derivatives play a key role in today's financial markets, since their usage allows investors to manage their portfolios, reduce risks, and predict future price movements. Options have become one of the most important financial instruments in the financial market.

Call option: gives the right to the holder to purchase the underlying asset at a certain price within a certain period of time or at the end of the period.

As option prices impact the choice of strategies applied in option trading and overall portfolio performance, various mathematical models have been proposed for estimating the theoretical value of an option. Models such as the Black-Scholes Model (BSM), Binomial Tree Model (BTM), and Monte Carlo (MC) simulation have become benchmarks in the field of option pricing theory. This study aims to compare the pricing accuracy of these three different option pricing approaches using Netflix, Inc. (NFLX) Call Options and identify the option pricing method that demonstrates the highest pricing accuracy and consistency when estimating NFLX call option market prices.

2 LITERATURE REVIEW

Recent studies showed mixed findings regarding the performance of these models. Sidharth (2025) compared the BTM and BSM for European call options using NASDAQ data. About 200,000 binomial steps were needed to converge to the BSM price, highlighting its computational efficiency; however, MC simulation and different maturities were not considered. Li (2024) compared the BSM and BTM using Kweichow Moutai data, while Cheng (2024) evaluated all three methods using NASDAQ data. Neither study compared estimated prices with actual market prices. Tambingon et al. (2019) compared NFLX call and put prices with BSM values, finding both call and put options were underpriced, but did not consider alternative methods. Longstaff et al. (2001) introduced the Least Squares Monte Carlo (LSM) method for American option pricing, achieving accuracy comparable to finite-difference methods. However, it was not compared with analytical or lattice-based methods for European options. Few studies have evaluated these models using standardized accuracy measures and non-parametric tests against market prices, particularly for volatile technology stocks such as Netflix. Therefore, this study compares the BSM, BTM, and MC methods for pricing Netflix European call options.

3 METHODOLOGY

3.1 Data

Historical daily closing prices for NFLX shares were collected from the Nasdaq Market website from 06.03.2026 to 15.05.2026. Netflix call option prices were obtained for 20 trading days from 04.20.2026 to 15.05.2026. The market mid-price is calculated as the average of the bid and ask prices. The risk-free interest rates were obtaind form U.S. Department of the Treasury website.

3.2 Assumptions

For non-dividend-paying stocks such as NFLX, the early exercise of call options is generally not optimal; therefore, American call option prices are expected to be very close to their European call option prices.

3.3 Black-Scholes Model

The BSM was developed by Black and Scholes(1973), provides a theoretical estimate of European option values under the assumption of constant volatility, zero transaction costs, and the normal distribution of the underlying stock prices. Black-Scholes European call option pricing formula is:

$$C_{BSM} = S_0 N(d_1) - K e^{-rT} N(d_2) \quad (1)$$

where,

$$d_1 = \frac{ln\left(\frac{S_0}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}} \quad (2)$$

$$d_2 = d_1 - \sigma\sqrt{T} \quad (3)$$

Here, C_{BSM} is European call option price, S_0 is current stock price, K is strike price, σ is annualized volatility, r is risk–free interest rate, T is time to maturity, and $N(d)$ is the Cumulative Distribution Function(CDF) for the standard normal distribution.

3.4 Binomial Tree Model

The BTM proposed by Cox, Ross, and Rubinstein (1979). A simplified 2-step binomial price three is shown in Fig. 1.

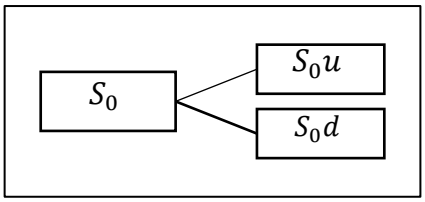


Figure 1: Simplified 2-step binomial price tree

It assumes that at each time step, the stock price can move either up or down creating a recombining tree of prices.

$$u = e^{\sigma\sqrt{\Delta t}} \quad (4)$$

$$d = e^{-\sigma\sqrt{\Delta t}} \quad (5)$$

where u and d are up and down movement factors respectively, σ is stock price volatility, and Δt is the time step length. The risk-neutral probability of an upward price movement is given by:

$$p = \frac{e^{r\Delta t} - d}{u - d} \quad (6)$$

where r is the risk free interest rate and p is the risk neutral probability of an upward movement. At maturity, the European call option payoff is

$$C_T = max(S_T - K, 0) \quad (7)$$

where S_T is stock price at maturity and K is strike price. The option value is then obtaind through backward induction as

$$C_{BTM} = e^{-r\Delta t}[pC_u + (1-p)C_d] \quad (8)$$

where C_u and C_d denote the option values following upward and downward movements, respectively, and C_{BTM} is the resulting binomial tree option value.

3.5 Monte Carlo Simulation

MC Simulation is a stochastic numerical technique that simulates a large number of potential future paths of the stock price to estimate derivative values.

The terminal stock price for simulation path i is:

$$S_{T,i} = S_0\, exp\left[\left(r - \frac{1}{2}\sigma^2\right)T + \sigma\sqrt{T}Z_i\right] \quad (9)$$

The simulated payoff and MC option price are:

$$Payoff_i = max(S_{T,i} - K, 0) \quad (10)$$

$$C_{MC} = e^{-rT}\frac{1}{M}\sum_{i=1}^{M} Payoff_i \quad (11)$$

where S_0 is initial stock price, $S_{T,i}$ is terminal stock price for path i, K is strike price, r is risk-free interest rate, σ is annualized volatility, T is time to maturity, Z_i is a standard normal random variable, and M is the number of simulated paths.

3.6 Statistical Analysis

The Friedman test was applied to assess whether there were significant differences in the daily MAE among the three option pricing methods, followed by pairwise Wilcoxon Signed-Rank tests identify differences in the daily MAE between individual model paris. Hypotheses of the Friedman Test are:

H_0: The three models have the same distribution of daily MAE values.

H_1: At least one model has a different distribution of daily MAE values.

3.7 Moneyness Classification.

$$Moneyness = \frac{S}{K} \quad (12)$$

Options were classified using Stock price (S) and strike price (K) as In-the-Money (ITM) when $S > K$, At-the -Money (ATM) when $S \approx K$ and Out-of-the-Money (OTM) when $S < K$.

4 RESULTS AND DISCUSSION

4.1 Stock Price Movements

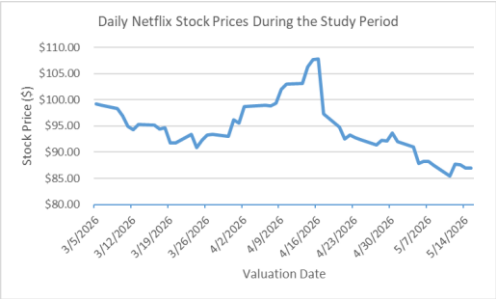


Figure 2: Netflix daily closing prices over the study period

Fig. 2 presents the closing price for NFLX between 06.03.2026 and 15.05.2026. During early April, an upward trend in the price can be observed, reaching its peak of around $108.00 in mid-April.

4.2 Pricing Accuracy Comparison

As shown in Table 2, the BSM recorded the lowest error values among the three methods, with MAE of 0.8964, RMSE of 0.9401, and MAPE of 40.0256%, respectively. The difference between the BSM and BTM were only 0.0077 in MAE and 0.0079 in RMSE, while difference between the BSM and MC were 0.0131 and 0.0138, respectively, indicating broadly similar overall pricing performance over the study period.

Table 1: Descriptive Statistics of Daily MAE Values

Pricing Method	BSM	BTM	MC
Mean	1.0117	1.1355	1.0245
Median	1.0125	1.1616	1.0271
SD	0.3009	0.2455	0.3034
Min	0.5138	0.6648	0.5255
Max	1.7345	1.5431	1.7491

Table 2: Average Pricing Accuracy Metrics Across All Contracts

Pricing Method	MAE	RMSE	MAPE
BSM	0.8964	0.9401	40.0256
BTM	0.9041	0.9480	40.3896
MC	0.9095	0.9539	40.5939

Table 1 indicates that the BTM has a higher average MAE of 1.1355, but its lower Standard Deviation (SD) of 0.2455 makes it much more consistent, indicating day-to-day stability compared with the alternative methods.

4.3 Statistical Comparison

The Friedman test results show a statistically significant difference among the MAE of three option pricing methods ($p < 0.05$) according to the Table 3. In Table 4, pairwise Wilcoxon signed-rank test results show statistically significant differences in MAE between each pair of option pricing methods.

Table 3: Friedman test results

Statistic	Value
Chi-Square Statistic	32.4
Degrees of Freedom	2
p-value	9.214e-08

Table 4: Pairwise Wilcoxon signed-rank test results

Comparison	Wilcoxon V Statistic	p-value
BSM vs BTM	17	0.0003948
BSM vs MC	1	0.0001113
BTM vs MC	192	0.0004826

Table 5: Average Pricing Errors by Moneyness Category

Moneyness Category	BSM MAE	BTM MAE	MC MAE
ITM	0.9982	1.0010	1.0125
ATM	1.1855	1.1917	1.1985
OTM	1.1694	1.1719	1.1816

4.4 Performance Across Moneyness Categories

To further evaluate model performance, the average pricing errors were examined according to moneyness categories as shown in Table 5.

The BSM produced the lowest average MAE in each moneyness category, although the differences among the three models were relatively small. All three models recorded the highest average pricing errors for ATM options, while comparatively lower errors were observed for ITM options.

5 CONCLUSION

In conclusion, while all three models provide reliable estimates for European call option pricing, the BSM produced the most accurate and consistent valuations for the NFLX call option data examined in this study. However, given the marginal differences in pricing errors, the BTM and MC simulation remain viable alternatives when different market conditions or option characteristics are involved. Practically, these findings rovide useful guidance for investors, traders, financial analysts, and organizations managing derivative instruments. The BSM, with lower pricing errors and simpler computations, is appropriate for valuing European call options similar to those studied. This study has several limitations. The evaluation was performed on European call options for a single underlying asset - NFLX limiting generalizability to other assets. The analysis utilized historical volatility, assuming constant volatility over the valuation period, whereas in practice volatility varies with time. Additionally, the analysis covered a specific time period (April-May 2026) and may not represent model performance during high-volatility markets or financial crises. Future studies could extend the analysis to longer periods and high-volatility market conditions, while incorporating time-varying volatility to further evaluate model performance under different market environments.

REFERENCES

- Cheng, Z. (2024). Pricing European Call Options with Visualization Based on the Binomial Model, Monte-Carlo Simulation, and Classical Black-Scholes Model. *Highlights in Science, Engineering and Technology*, *88*, 311-317.
- Li, J. (2024, May). Comparison of Option Pricing Based on Black-Scholes and Binomial Tree: Evidence from Moutai's Share Price. In *9th International Conference on Financial Innovation and Economic Development (ICFIED 2024)* (pp. 342-351). Atlantis Press.
- Longstaff, F. A., & Schwartz, E. S. (2001). Valuing American options by simulation: A simple least-squares approach. The review of financial studies, 14(1), 113-147.
- Sidharth, V. (2025). A Comparison of the Binomial Asset Pricing Model and the Black-Scholes Formula in European Call Option Pricing. *Available at SSRN 5137136*.
- Tambingon, D. A., Titaley, J., & Manurung, T. (2019). Black-Scholes Model in Determining European Option Prices on Netflix, Inc. *d\'Cartesian: Jurnal Matematika dan Aplikasi*, *8*(2), 80-85.

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Assessment of Factors Affecting Soil Organic Carbon Content in Sri Lankan Tea-Growing Soils

Wijethunga WMYN[1], Karunananda GACM[2]
[1,2]*Department of Mathematical Sciences, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka.*
yasithaw20@gmail.com[1], chamly@wyb.ac.lk[2]

ABSTRACT

Soil Organic Carbon (SOC) is a key indicator of soil fertility and sustainable tea production, yet estate-level information on the factors that control it is rarely available to plantation managers in Sri Lanka. Existing Sri Lankan studies have mostly focused on individual estates, selected districts, or specific management systems, so the combined influence of elevation and soil chemical properties across major tea-growing regions remains insufficiently documented. This study therefore examined the relationships between SOC and elevation and soil chemical variables using data from 196 tea estates covering 24 agro-ecological regions. Data were analysed in R 4.3.0 using correlation, simple and multiple linear regression, one-way analysis of variance with Tukey post hoc test, and Random Forest regression. Mean SOC was 2.69% (SD = 1.10%) and increased from 1.59% in Low Country to 2.09% in Mid Country and 3.29% in Up Country estates. Elevation and nitrogen individually explained 61.97% and 60.81% of SOC variance, respectively. The final stepwise model (elevation, nitrogen, water pH, and calcium chloride pH) had an adjusted R^2 of 0.6895 and a root mean square error (RMSE) of 0.6138, while the Random Forest model gave an apparent R^2 of 0.7049 and RMSE of 0.5968. These measures reflect in-sample fit only, as no independent test set was used. The study provides a multi-variable, estate-level assessment of SOC in Sri Lankan tea soils and identifies elevation and nitrogen as its principal correlates, supporting more targeted soil management.

KEYWORDS: Elevation, Multiple linear regression, Nitrogen, Random Forest, Soil organic carbon

1 INTRODUCTION

Tea production is an already existing part of the Sri Lankan agricultural industry. There are three main elevations at which tea is grown: the Up Country (>1200 m), the Mid Country (600-1200 m), and the Low Country (<600 m). These elevations have their own climate and specific soil-forming factors. The Soil Organic Carbon (SOC) is a constituent of soil fertility as it provides the soil structure, soil moisture, nutrients recycling, and carbon sequestration (Brady & Weil, 2016; Das & Biswas, 2025).The SOC in tea soils depends on the elevation, nutrient status, and land management (Jobbágy & Jackson, 2000; Dandeniya et al., 2022). Those people who manage tea soils without knowing the content of SOC use inorganic fertilizers excessively increasing the expenses and accelerating the process of soil degradation. In Sri Lanka, there have been studies concerning the relation between SOC and the plantation management, as well as the change of land use (Wanniarachchi et al., 2012; Rotawewa & Lokupitiya, 2023).

This study addresses that gap, aiming to (i) describe SOC variation among elevation zones and agro-ecological regions; (ii) identify elevation and soil chemical variables (nitrogen, pH, and sodium) associated with SOC; and (iii) compare regression and Random Forest models.

2 LITERATURE REVIEW

SOC is recognized as one of the key indicators of soil quality (Das & Biswas, 2025). Elevation affects SOC mostly through its impact on temperature; lower temperatures slow down decomposition and contribute to increased carbon storage (Jobbágy & Jackson, 2000). In the case of tea plantations in Sri Lanka, SOC has been studied as regards management, productivity, and carbon storage (Wanniarachchi et al., 2012; Wijeratne, 2014; Dandeniya et al., 2022), and increased amounts of SOC were recently noted following conversion of marginal tea plantations into timber plantations (Rotawewa & Lokupitiya, 2023).

Nitrogen was considered due to the relationship between carbon and nitrogen mineralization (Khalil et al., 2005); however, the study was carried out in subtropical upland soils and cannot be compared with tea estates. Multiple linear regression is easily interpreted, and Random Forest (Breiman, 2001) captures non-linear patterns; it is recommended to use them together (Minasny et al., 2017). High uncertainties in soil carbon models stress the importance of model validation (Sulman et al., 2018).

However, Sri Lankan studies often focus on individual estates, selected districts, or specific land-use comparisons. A multi-region estate-level analysis of elevation, nitrogen, pH, and other soil variables therefore provides a useful complementary perspective.

3 METHODOLOGY

Data were collected through a systematic survey of the soil carried out at an estate level by the Tea Research Institute (TRI), Sri Lanka. The survey involved major tea-growing areas located in the Central, Southern, and Western provinces of the country and 24 agro-ecological regions belonging to three elevation zones. The sampling technique employed in this study involved a stratified multi-stage sampling design (rainfall zone, elevation zone, agro-ecological region, district, plantation company, and estate).The sample size was calculated using the Slovin formula with a 5% margin of error, given by

$$n = \frac{N}{1 + Ne^2} \quad (1)$$

where n is the sample size, N is the population size, and e is the margin of error (0.05). This yielded a working sample of 196 tea estates (23 Low Country, 65 Mid Country, and 108 Up Country). The soil variables measured included SOC (C%), pH in water (pHw) and in calcium chloride (pHCa), cation exchange capacity, nitrogen (N%), phosphate, potassium, magnesium, calcium, and sodium (Na, mg/kg), with composite soil samples analysed in TRI-certified laboratories.

All analyses were performed in R (version 4.3.0). Simple linear regression (SLR) of SOC on each predictor was followed by one-way Analysis of Variance (ANOVA) with Tukey's Honestly Significant Difference (HSD) post hoc test across elevation and agro-ecological categories, and Pearson correlation among variables. Backward stepwise multiple linear regression (MLR) based on the Akaike Information Criterion (AIC) gave a multivariable model, with diagnostics including the Shapiro–Wilk test and the Variance Inflation Factor (VIF). Random Forest regression with 500 trees was used for

variable-importance ranking and benchmarking. Performance was summarised by adjusted R^2 and Root Mean Square Error (RMSE); no independent test set or cross-validation was used. The significance level was $\alpha = 0.05$.

4 RESULTS AND DISCUSSION

The overall mean SOC content across the 196 estates was 2.69% (Standard deviation (SD) = 1.10%; range = 1.06–6.97%). Fig. 1 shows a right-skewed distribution, reflecting a small number of very high-carbon Up Country estates. As shown in Table 1, mean SOC increased from 1.59% in Low Country to 2.09% in Mid Country and 3.29% in Up Country estates, consistent with elevation-related conditions affecting SOC accumulation, although the cross-sectional design does not establish causality.

SLR of SOC on elevation explained 61.97% of the observed variance ($R^2 = 0.6197$; $p < 0.001$), as given by

$$SOC = 0.8048 + 0.001646 \times Elevation \quad (2)$$

Soil nitrogen percentage (N %) was also strongly associated with SOC and explained 60.81% of the observed variance ($R^2 = 0.6081$; $p < 0.001$), as given by

$$SOC = 0.5836 + 8.2993 \times N\% \quad (3)$$

Table 1: SOC by Elevation Category

Elevation Zone	n	Mean SOC (%)	Std. Dev.	Elevation Range (m)
Low Country	23	1.59	0.381	< 600
Mid Country	65	2.09	0.535	600–1200
Up Country	108	3.29	1.080	> 1200

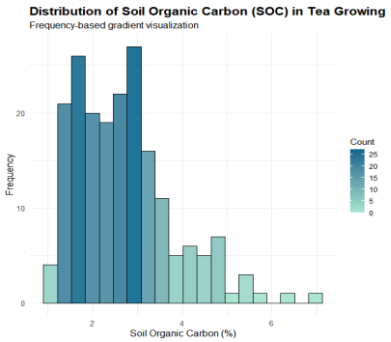


Figure 1: Distribution of SOC Across 196 Tea Estates

One-way ANOVA confirmed highly significant differences in SOC across elevation categories ($F = 58.52$, $p < 0.001$) and agro-ecological regions ($F = 8.25$, $p < 0.001$). Tukey's HSD comparisons showed that Up Country estates had significantly higher SOC than both Mid Country and Low Country estates ($p < 0.001$), while the Mid–Low Country difference was only marginally significant ($p = 0.051$), as illustrated in Fig. 2.

The elevation effect agrees with the temperature-controlled decomposition described by Jobbágy and Jackson (2000), and the nitrogen association is consistent with coupled carbon–nitrogen dynamics (Khalil et al., 2005). Sri Lankan studies likewise link SOC with productivity and land management (Dandeniya et al., 2022; Rotawewa & Lokupitiya, 2023) but are site-specific.

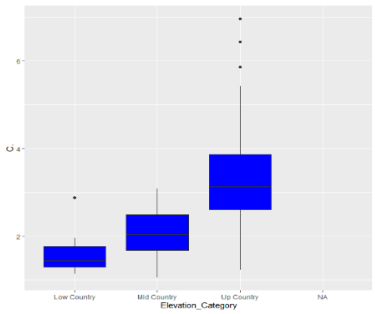


Figure 2: Boxplot of SOC Distribution by Elevation Category

As elevation covaries with climate, vegetation, and management, it should not be treated as the sole cause.

Pearson correlation analysis showed strong positive associations of SOC with elevation ($r = 0.787$, $p < 0.001$) and nitrogen ($r = 0.780$, $p < 0.001$), a moderate positive association with sodium ($r = 0.316$, $p < 0.001$), and weak negative associations with pHw ($r = -0.193$, $p = 0.006$) and pHCa ($r = -0.044$, not significant).

Backward stepwise regression using AIC retained elevation, nitrogen, pHw, and pHCa, while sodium was dropped for lack of significance. The final model is given by

$$SOC = 0.7301 + 0.0007928(Elevation) + 6.3463(N\%) - 0.4498(pHw) + 0.3775(pHCa) \quad (4)$$

Table 2 compares all fitted models. The stepwise model had an adjusted R^2 of 0.6895 and RMSE of 0.6138, compared with an adjusted R^2 of 0.6880 for the full five-predictor model. All VIF values were below 3, suggesting no substantial multicollinearity, although the opposite pHw and pHCa signs call for cautious interpretation. The Shapiro–Wilk result ($W = 0.931$, $p < 0.001$) indicated departure from residual normality, possibly due to the skewed SOC distribution.

The Random Forest model used elevation, N%, pHw, pHCa, and Na as predictors (500 trees). Its apparent R^2 was 0.7049 with RMSE = 0.5968, and variable importance ranked elevation first, followed by nitrogen, sodium, pHCa, and pHw.

Table 2: Comparison of Predictive Models

Model	Predictors	Adj. R^2	RMSE
SLR: Elevation only	Elevation	0.6167	0.8412
SLR: N% only	N%	0.6060	0.6913
Stepwise MLR (final)	Elevation, N%, pHw, pHCa	0.6895	0.6138
Random Forest	Elevation, N%, pHw, pHCa, Na	0.7049	0.5968

These values are slightly better than those of the linear models but indicate in-sample fit, as no held-out test set was used. The Random Forest model has no explicit equation, whereas the stepwise model uses variables already measured in routine soil tests.

5 CONCLUSION

This study examined estate-level variation in SOC across Sri Lankan tea-growing environments. Mean SOC was higher at higher elevation, and SOC was strongly associated with elevation and nitrogen. The stepwise regression model explained a substantial proportion of the variation (adjusted $R^2 = 0.6895$), and the Random Forest model gave a slightly higher apparent fit ($R^2 = 0.7049$). Elevation and soil nitrogen should therefore be prioritised in estate-level

SOC assessment, although the observational design does not establish causal effects.

The study has two main limitations. First, model performance was assessed on the data used for fitting, without an independent test set or cross-validation, so the models are not externally validated. Second, only elevation and measured soil chemical variables were considered. Future research should apply repeated k-fold cross-validation or an independent dataset and include management and climatic variables.

REFERENCES

- Brady, N. C., & Weil, R. R. (2016). *The nature and properties of soils* (15th ed.). Pearson Education.
- Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.
- Dandeniya, W. S., Withana, P. A., Suriyagoda, L. B. D., Dissanayake, R. J., Ranasinghe, C. N., & Thalagoda, S. (2022). Importance of soil organic carbon levels for increasing land productivity of tea smallholdings in Kandy and Kegalle districts in Sri Lanka. *Journal of the Soil Science Society of Sri Lanka*, *26*.
- Das, S., & Biswas, A. (2025). Variability of organic carbon contents in tea soils of Chattogram Valley. *Bangladesh Journal of Soil Science*.
- Jobbágy, E. G., & Jackson, R. B. (2000). The vertical distribution of soil organic carbon and its relation to climate and vegetation. *Ecological Applications*, *10*(2), 423–436.
- Khalil, M. I., Hossain, M. B., & Schmidhalter, U. (2005). Carbon and nitrogen mineralization in different upland soils of the subtropics treated with organic materials. *Soil Biology and Biochemistry*, *37*(8), 1473–1481.
- Minasny, B., Malone, B. P., McBratney, A. B., Angers, D. A., Arrouays, D., Chambers, A., Chaplot, V., Chen, Z.-S., Cheng, K., Das, B. S., Field, D. J., Gimona, A., Hedley, C. B., Hong, S. Y., Mandal, B., Marchant, B. P., Martin, M., McConkey, B. G., Mulder, V. L., O'Rourke, S., Richer-de-Forges, A. C., Odeh, I., Padarian, J., Paustian, K., Pan, G., Poggio, L., Savin, I., Stolbovoy, V., Stockmann, U., Sulaeman, Y., Tsui, C.-C., Vågen, T.-G., van Wesemael, B., & Winowiecki, L. (2017). Soil carbon 4 per mille. *Geoderma*, *292*, 59–86.
- Rotawewa, B., & Lokupitiya, E. Y. K. (2023). Conversion of marginalised tea lands into timber plantations and its impact on soil organic carbon content: A case study in Central Highlands of Sri Lanka. *Proceedings of the 27th International Forestry and Environment Symposium*.
- Sulman, B. N., Moore, J. A. M., Abramoff, R., Averill, C., Kivlin, S., Georgiou, K., Sridhar, B., Hartman, M. D., Wang, G., Wieder, W. R., Bradford, M. A., Luo, Y., Mayes, M. A., Morrison, E., Riley, W. J., Salazar, A., Schimel, J. P., Tang, J., & Classen, A. T. (2018). Multiple models and experiments underscore large uncertainty in soil carbon dynamics. *Biogeochemistry*, *141*(2), 109–123.
- Wanniarachchi, S. O., Perera, P. D. P. S., Dharshana, M. A. C., & Zoysa, A. K. N. (2012). Soil organic carbon as an indicator of soil quality: A study in selected tea plantations in Galle District. *Proceedings of the International Forestry and Environment Symposium*.
- Wijeratne, T. (2014). Carbon sequestration potential of tea plantations in Sri Lanka. *228th Experiment and Extension Forum*, Tea Research Institute of Sri Lanka.

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

A Comparative Study of Glycemic Control in Gestational and Pre-Gestational Diabetes in Batticaloa District: A Case Study

Weerasinghe TK[1], Panahatipola PMOP[2]
[1,2]*Department of Mathematical Sciences, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka.*
weerasinghekaveesha8@gmail.com[1] , oshadi@wyb.ac.lk[2]

ABSTRACT

Gestational Diabetes Mellitus (GDM) is a significant pregnancy-related health concern that can adversely affect maternal and neonatal outcomes. This study examines maternal demographic and clinical factors associated with GDM and Fasting Blood Glucose (FBG) status among pregnant women attending Batticaloa General Hospital, Sri Lanka. However, direct comparative data on clinical features and demographic factors between GDM and non-GDM pregnancies in local tertiary care settings remain limited in Sri Lanka. Data from 130 pregnant women were analyzed, and two binary logistic regression models were developed, with GDM and elevated FBG considered as separate outcome variables. The GDM model identified Body Mass Index (BMI), FBG, Diastolic Blood Pressure (DBP), heart rate, and family history of diabetes as significant factors associated with GDM, while selected maternal characteristics were also associated with elevated FBG. By examining both GDM and elevated FBG within the same study population, this research provides locally relevant hospital-based evidence from the Batticaloa District. The study contributes to existing evidence by applying binary logistic regression to identify maternal factors associated with different indicators of abnormal glucose status. The findings demonstrate the potential value of routinely available maternal characteristics in supporting early assessment and identification of women at increased risk of GDM and elevated FBG.

KEYWORDS: Fasting blood glucose, Gestational diabetes mellitus, Logistic regression, Maternal risk factors

1 INTRODUCTION

Gestational Diabetes Mellitus (GDM) is a common metabolic condition characterized by elevated blood glucose during pregnancy. It is an important maternal health concern due to its adverse pregnancy outcomes and increased risk of Type 2 diabetes later in life. Its prevalence varies across populations and is influenced by maternal and demographic factors (Rajput et al., 2014; Nanda et al., 2011). Fasting blood glucose (FBG) is a commonly measured indicator of maternal glucose levels and can provide useful information about the risk of GDM. Previous studies have shown that higher fasting glucose levels during early pregnancy are associated with a greater likelihood of developing GDM and experiencing adverse pregnancy outcomes (Riskin-Mashiah et al., 2009; Cosson et al., 2017). Therefore, examining maternal characteristics together with glucose levels can help identify women who may be at greater risk of GDM. In this study, FBG is considered both as a continuous measure and according to its status as normal or elevated. Comparing FBG status between GDM and Non-Gestational Diabetes Mellitus (NGDM) (pre-existing diabetes) mothers may help determine whether elevated fasting glucose levels are associated with GDM status.

This study aims to identify factors associated with GDM status and FBG status. It compares GDM and NGDM mothers based on clinical characteristics

such as BMI, age, diastolic blood pressure (DBP), systolic blood pressure (SBP), heart rate, gestational age, and family history to identify potential risk factors for GDM. However, elevated fasting blood glucose levels may occur in women with GDM as well as in women with NGDM; therefore, this study also aims to identify factors associated with status and elevated FBG status.

2 LITERATURE REVIEW

Niu et al. (2024) developed and validated a logistic regression model for early GDM prediction using data from 6,000 pregnant women. Age, BMI, HbA1c, uric acid, triglycerides, and HDL cholesterol were identified as significant predictors. The model achieved Area Under the Curve (AUC) values of 0.803 in the development dataset and 0.782 in the validation dataset, demonstrating good predictive performance. These findings highlight the importance of combining demographic, anthropometric, and biochemical factors to improve early GDM risk assessment. Furthermore, the validation results indicate that the model maintained reasonable predictive performance when applied to separate data. This study supports the use of logistic regression for identifying important risk factors and developing interpretable predictive models that may assist in the early identification of women at increased risk of GDM.

Riskin-Mashiah et al. (2009) investigated the relationship between first-trimester fasting blood glucose (FBG) levels and the subsequent development of GDM. Their findings showed that higher early-pregnancy FBG levels were associated with an increased risk of developing GDM. This highlights the importance of early glucose assessment for identifying women at risk. However, FBG alone may not provide sufficient predictive accuracy. Combining FBG with maternal characteristics, such as age, BMI, and medical history, may improve GDM risk assessment. Furthermore, examining family history alongside FBG may help clarify the associations between maternal characteristics and GDM status. This supports further research on GDM prediction.

Nanda et al. (2011) compared logistic regression and neural networks for first-trimester GDM prediction. Logistic regression achieved an AUC of 0.78 and a sensitivity of 74.3%, demonstrating good predictive performance. Their findings highlighted the usefulness of combining maternal characteristics to identify women at increased risk of developing GDM. Furthermore, logistic regression provides interpretable results through coefficients and odds ratios, helping researchers understand the associations between maternal risk factors and GDM status. This interpretability supports better risk assessment and early pregnancy screening. Overall, the findings support logistic regression as a useful method for early GDM prediction and investigating multiple maternal risk factors.

3 METHODOLOGY

Clinical secondary data were collected from all eligible pregnant women (N = 130) presenting at 12–24 weeks of gestation at Batticaloa General Hospital between April and June 2025, based on available medical records during the study period. This study compares GDM and NGDM mothers based on clinical characteristics and uses binary

logistic regression to identify potential risk factors associated with GDM status. In addition, elevated fasting blood glucose levels(≥125 mg/dL) can be observed among women with GDM as well as in women with NGDM. Therefore, this study uses binary logistic regression to identify the potential risk factors associated with elevated FBG status. Statistical analyses were conducted using R software, while two separate binary logistic regression models were developed using Python in the Google Colaboratory software.

4 RESULTS AND DISCUSSION

The study compared the GDM and NGDM groups based on clinical indicators and assessed potential risk factors for both GDM status and elevated FBG status. Binary logistic regression was used to identify significant predictors and evaluate the performance of the developed models.

Table 1: The ability of Family history to predict the risk of GDM

Family	NGDM	GDM	X^2	p
Yes	40	20	14.7	0.00
No	23	47		

Table 2: The ability of each indicator to predict the risk of GDM

	NGDM	GDM	p
Age	29.1	30.4	0.33
FBG	130.1	125.0	0.00
BMI	30.4	31.5	0.00
SBP	110.9	114.5	0.02
DBP	78.1	77.0	0.06
Heartrate	112.0	115.1	0.01

Table 1 shows a significant association between family history of diabetes and GDM status based on the chi-square test, indicating that GDM status differed significantly according to whether a family history of diabetes was present. Table 2 shows significant differences in FBG, BMI, systolic blood pressure, and heart rate based on the independent samples t-test, while age and diastolic blood pressure show no significant differences. Overall, these findings highlight the importance of considering family history and relevant maternal clinical characteristics when investigating factors associated with GDM status.

4.1 Binary Logistic Regression Model for Predicting GDM Status

A binary logistic regression model was developed to predict GDM status, with GDM coded as 1 and NGDM coded as 0 for women with a previous history of diabetes. The model examined the relationship between GDM status and maternal demographic and clinical factors.

Table 3: Results of the Binary Logistic Regression Analysis for GDM

	β	p	OR	95%CI
Age	0.03	0.21	1.03	0.9-1.0
BMI	0.20	0.04	1.22	1.1-1.3
FBG	0.13	0.00	1.14	1.0-1.2
SBP	0.04	0.28	1.04	0.9-1.1
Family	0.99	0.02	5.45	3.1-9.5
DBP	0.17	0.03	1.05	1.0-1.1
Heartrate	0.09	0.00	1.09	1.0-1.1
Constant	28.14	0.00		

Table 3 shows that BMI, FBG, family history, DBP, and heart rate were significant predictors of GDM ($p < 0.05$). Family history showed the strongest effect (OR = 5.45), while BMI, FBG, DBP, and heart rate were also associated with higher odds of GDM. Age and SBP were not significant predictors. These findings suggest that maternal clinical characteristics and family history may contribute to understanding GDM risk. However, the results should be interpreted in the context of the study sample and the limitations of the available data.

4.2 Binary Logistic Regression Model for Predicting Fasting Blood Glucose

A binary logistic regression model was developed to predict elevated FBG (≥125 mg/dL) using first-trimester clinical indicators. The model examined the relationship between elevated FBG status and relevant maternal clinical factors, including age, BMI, diabetes type, SBP, DBP, family history of diabetes, and heart rate. Table 4 shows that BMI, diabetes type, family history of diabetes, and heart rate were significant predictors of elevated FBG status ($p < 0.05$).

Table 4: Results of the Binary Logistic Regression Analysis for FBG

	β	p	OR	95%CI
Age	0.02	0.58	1.0	0.9-1.0
BMI	0.19	0.04	1.2	1.1-1.4
Type	0.54	0.01	1.7	1.1-2.5
SBP	0.01	0.31	1.0	0.9-1.0
DBP	0.28	0.06	1.3	0.9-1.7
Family	0.10	0.00	1.1	1.0-1.1
Heartrate	0.87	0.01	2.3	1.2-4.7
Constant	18.1	0.00		

Heart rate showed the strongest association (OR = 2.3), followed by diabetes type (OR = 1.7), BMI (OR = 1.2), and family history (OR = 1.1). These results suggest that maternal clinical characteristics and diabetes-related history may be associated with the likelihood of elevated FBG. Age, SBP, and DBP were not statistically significant predictors in this model.

Overall, these findings highlight the importance of considering multiple maternal factors when examining elevated FBG status and provide a basis for further research on glucose regulation during pregnancy.

Overall, BMI, FBG, heart rate, and family history of diabetes were important predictors, highlighting the value of first-trimester clinical indicators in identifying women at risk of GDM and elevated FBG.

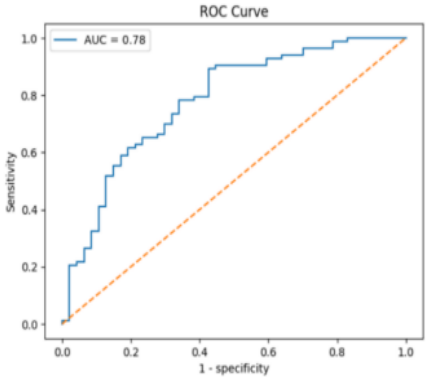


Figure 1: The Receiver Operating Curve (ROC) Curve of GDM Type

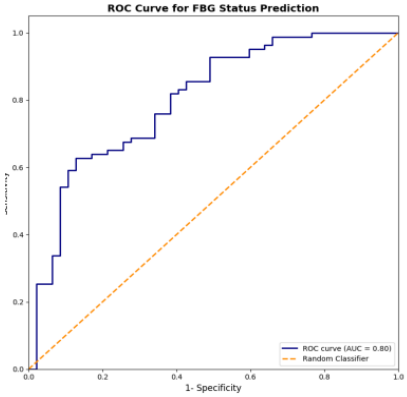


Figure 2: The ROC Curve of FBG Status

5 CONCLUSION

This research project identifies key first-trimester clinical indicators associated with GDM and elevated FBG among pregnant women. The findings highlight the importance of BMI, heart rate, FBG, blood pressure, diabetes history, and family history in early risk assessment. While this study evaluated individual predictor contributions, potential interaction effects were not explicitly modeled and warranted further investigation in larger prospective cohorts. These findings suggest that considering both individual risk factors and their combined effects may improve the identification of women at higher risk of GDM and abnormal glucose levels. The study demonstrates the potential of routinely collected clinical indicators to support early identification and risk assessment, providing a foundation for improved screening and timely intervention.

REFERENCES

- Cosson, E., Carbillon, L., Valensi, P., and High Council for Public Health Working Group. (2017). High fasting plasma glucose during early pregnancy: A review about early gestational diabetes mellitus. Diabetes & Metabolism, 43(6), 489–496.
- Nanda, S., Savvidou, M., Syngelaki, A., Akolekar, R., and Nicolaides, K. H. (2011). Prediction of gestational diabetes mellitus by maternal factors and biomarkers at 11 to 13 weeks. Prenatal Diagnosis, 31(2), 135–141
- Niu, X., Zhang, Y., Li, Y., et al. (2024). Establishment of gestational diabetes risk prediction model and clinical verification. Diabetology & Metabolic Syndrome, 16, Article 4.
- Rajput, R., Yadav, Y., Nanda, S., and Rajput, M. (2014). Prevalence of gestational diabetes mellitus and associated risk factors at a tertiary care hospital in Haryana. Indian Journal of Medical Research, 140(6), 728–733.
- Riskin-Mashiah, S., Younes, G., Damti, A., and Auslender, R. (2009). First-trimester fasting hyperglycemia and adverse pregnancy outcomes. Diabetes Care, 32(9), 1639–1643.

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Modelling and Forecasting Daily USD/LKR Exchange Rate Dynamics Using ARIMA and EGARCH Models

Nawarathna NWP[1] and Aponsu GMLM[2]
[1,2]*Department of Mathematical Sciences, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*
pavithranawarathna323@gmail.com[1], lakshana@wyb.ac.lk[2]

ABSRACT

This study investigates the dynamics and forecasting of daily USD/LKR exchange rate returns over a 26-year period from January 2000 to August 2026 using econometric models. Exchange rate returns exhibit several stylized characteristics, including volatility clustering, asymmetric responses to shocks, and non-normal, heavy-tailed distributions, which present challenges for accurate modelling and forecasting. To account for these features, an Autoregressive Integrated Moving Average (ARIMA) model was employed to capture temporal dependence in exchange rate returns, while an Exponential Generalized Autoregressive Conditional Heteroskedasticity (EGARCH) model was used to model time-varying volatility and asymmetric effects. Prior to model estimation, extreme observations were treated to reduce the influence of outliers on the distributional properties of the return series. Descriptive and diagnostic analyses indicated significant departures from normality, including positive skewness and excess kurtosis. Among the volatility specifications considered, the EGARCH(1,1) model provided the most appropriate representation of asymmetric volatility dynamics, while the ARIMA(3,0,3) model demonstrated satisfactory forecasting performance for exchange rate returns. The findings highlight the importance of accounting for both temporal dependence and asymmetric volatility when modelling the USD/LKR exchange rate. The results provide useful insights for exchange rate forecasting, financial risk management, and policy-related analysis in the Sri Lankan foreign exchange market.

KEYWORDS: ARIMA, EGARCH, Exchange Rate, Volatility

1 INTRODUCTION

Exchange rates exhibit complex behavior characterized by volatility clustering, asymmetric shocks, and heavy-tailed return distributions. These features pose significant challenges for accurate modeling and forecasting, which are critical for market participants and policymakers to manage financial risks effectively. Traditional symmetric volatility models often fail to capture the nonlinearities and asymmetries commonly observed in exchange rate return series. To address this, advanced econometric models such as the EGARCH have been developed, capturing both leverage effects and persistent volatility in the data. Additionally, linear time series models like ARIMA are useful to model temporal dependencies and provide baseline forecasts. This study applies both EGARCH and ARIMA models to a long daily exchange rate sample, examining their ability to characterize the conditional variance and forecast returns accurately. The comparative analysis highlights the strengths and limitations of each approach in managing the inherent volatility and forecasting challenges of exchange rate movements.

2 LITERATURE REVIEW

Forecasting exchange rate dynamics is central to financial econometrics, with ARIMA and GARCH models frequently utilized for their respective strengths in

mean and volatility modeling. Magaji and Garba (2022) applied an integrated ARIMA-GARCH model to analyze the Nigerian naira to US dollar exchange rate, illustrating how this hybrid approach captures linear dependencies and conditional heteroscedasticity, enhancing forecasting accuracy. Muhamad Shah and Khairol Azmi (2026) evaluated the Malaysian ringgit and found ARIMA effective for autocorrelation while GARCH models excelled in capturing volatility clustering and leverage effects, advocating for GARCH-type models like EGARCH in cases of pronounced asymmetry. Dou (2025) focused on the JPY/USD exchange rate using ARIMA, demonstrating that careful selection of lag structures yields reliable point forecasts, with ARIMA(3,0,3) achieving superior mean forecast accuracy. These findings underscore the importance of addressing linear and nonlinear dynamics in exchange rate forecasting. ARIMA is adept at modeling mean structures and providing unbiased forecasts, while GARCH-family models are crucial for capturing volatility features. The literature supports a combined or comparative modeling strategy for achieving thorough and accurate exchange rate forecasts, as shown in our analysis of EGARCH, and ARIMA models.

3 METHODOLOGY

The analysis applies two main models to daily exchange rate returns from January 2000 to August 2026. (The daily exchange rate data, obtained from Stlouisfed.Org (2025), was subjected to cleaning and analysis for modeling with ARIMA and EGARCH frameworks.) First, EGARCH(p,q) model was used to handle return asymmetry and heavy tails after trimming extreme outliers, which greatly reduced kurtosis. To capture the optimal conditional variance dynamics, a grid search was performed over the candidate space $p \in [1,2], q \in [1]$. EGARCH(1,1) was selected as the best-fitting model, with all EGARCH parameters being highly significant, indicating that the model captures the asymmetric effects of shocks on volatility and exhibits stability and persistence. It effectively tracks volatility clusters and adapts standard errors during volatile periods. Second, ARIMA(p, d, q) model was used to model linear dependencies in the cleaned return series where optimal parameters were identified after conducting a grid search over the candidate space $p \in [1,3], d \in [0,1]$ $q \in [1,3]$. ARIMA(3,0,3) was selected as the best model based on autocorrelation analysis. Its parameters are statistically significant and stable. In comparison, the ARIMA model achieves better predictive accuracy, with a perfect zero bias proportion and lower Theil inequality coefficient, indicating superior overall forecast performance, while the EGARCH model better captures volatility dynamics. Together, these models provide a robust framework combining volatility and mean structure for forecasting exchange rate returns.

4 RESULTS AND DISCUSSION

4.1 Descriptive Statistics and Diagnostic Baseline

The empirical analysis begins with an evaluation of the summary statistics for the cleaned return series (R_CLEAN) across the 6,954 daily observations. R_CLEAN refers to the cleaned daily return series of the USD/LKR exchange

rate after preprocessing steps, including trimming extreme outliers. This cleaning procedure was done to reduce the kurtosis and achieve convergence in the modeling process. The cleaned return series exhibits characteristics such as significant positive skewness, high kurtosis (fat tails), and strong non-normality. As documented in Table 1, the sample exhibits a mean return of 0.000202 and a standard deviation of 0.003901.

The data reflects a distinct positive skewness of 1.693823, indicating that daily exchange rate shocks are asymmetric with a longer right tail.

The kurtosis is recorded at 58.74762, confirming that even after trimming extreme outliers to achieve convergence, the distribution remains highly leptokurtic with significant fat-tailed behavior compared to a normal distribution (kurtosis of 3).

Table 1: Descriptive Summary Statistics

	R_CLEAN
Mean	0.000202
Median	0.000000
Maximum	0.050000
Minimum	-0.050000
Std.Dev.	0.003901
Skewness	1.693823
Kurtosis	58.74762
Jaque-Bera	903809.5
Probability	0.000100
Sum	1.404329
Sum Sq.Dev.	0.105793
Observations	6954

R_Clean: The cleaned daily return series of the USD/LKR exchange*

Finally, the Jarque-Bera statistic reaches 903,809.5 with a p-value of 0.000000, overwhelmingly rejecting the null hypothesis of normality. This non-normal, asymmetric profile provides strong empirical justification for employing an advanced asymmetric volatility model like EGARCH(1,1).

4.2 Estimation Results of the ARIMA(3,0,3) and EGARCH(1,1) Models

Both competitive models were estimated using Maximum Likelihood Estimation in Eviews, successfully achieving mathematical convergence. The parameter estimates for both models are presented in Table 2.

For the ARIMA(3,0,3) framework, all linear lag structures are statistically significant at the 1% level ($p < 0.001$), demonstrating strong linear dependencies in daily returns. The model is structurally stable, as all inverted AR and MA roots fall strictly within the unit circle.

For the EGARCH(1,1) framework, all conditional variance parameters are highly significant ($p < 0.001$). The asymmetric leverage parameter, gamma(C(4)), is estimated at 0.035137, statistically confirming that positive and negative shocks have asymmetric impacts on impending exchange rate risk.

Crucially, the volatility persistence parameter, beta(C(5)), is established at 0.971444. Because beta < 1, the EGARCH model satisfies strict covariance stationarity, proving that while daily volatility shocks are highly persistent, the system remains stable over time.

Table 2: Model Parameter Estimation Results

Model	Parameter	Coefficient	z / t-Statistic	p-value
ARIMA(3,0,3)	Constant (C)	0.000201	2.169009	0.0301
	AR(1)	-0.158963	-3.733219	0.0002
	AR(3)	0.352455	7.536900	0.0000
	MA(3)	-0.259388	-5.525902	0.0000
EGARCH(1,1)	Constant (C)	-6.45E-06	-0.899608	0.3683
	Omega	2.51E-08	34.47801	0.0000
	Shock Alpha	0.397264	108.1730	0.0000
	Persistence Beta	0.971444	907.1040	0.0000
	Asymmetry Gamma	0.035137	10.12351	0.0000

4.3 Comparative Forecast Evaluation

Static out-of-sample forecasts were generated simultaneously for both models across the full sample timeline to evaluate predictive accuracy.

The performance matrix is compiled side-by-side in Table 3.

Table 3: Out-of-Sample Static Forecast Performance Matrix

Forecast Evaluation Metric	EGARCH(1,1) Model	ARIMA(3,0,3) Model
RMSE	0.003904	0.003861
MAE	0.001620	0.001663
Theil Inequality Coefficient	0.993409	0.858653
Bias Proportion	0.002061	0.000000
Variance Proportion	0.997902	0.750581

4.4 Discussion of Forecast Metrics

The comparative evaluation reveals a clear trade-off between modeling linear conditional means versus non-linear conditional variance.

Theil Inequality and Variance Optimization: The most prominent finding is the behavior of the Theil Inequality Coefficient, which drops noticeably from 0.993409 in EGARCH down to 0.858653 in ARIMA. This improvement is explained by the Variance Proportion, which declines from 0.997 to 0.750581 under the ARIMA framework. While EGARCH effectively maps dynamic risk bands and volatility clusters, it over-allocates error variance because it lacks structural mean lags. ARIMA(3,0,3) captures the underlying autoregressive patterns of daily currency trades much more efficiently.

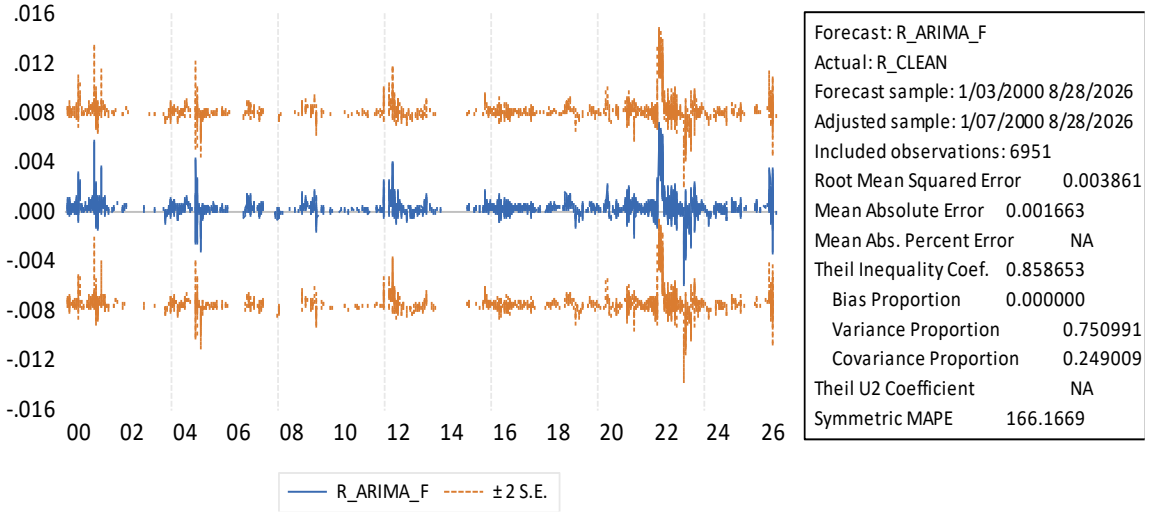


Figure 1: Forecast of $ARIMA(3,0,3)$ Model

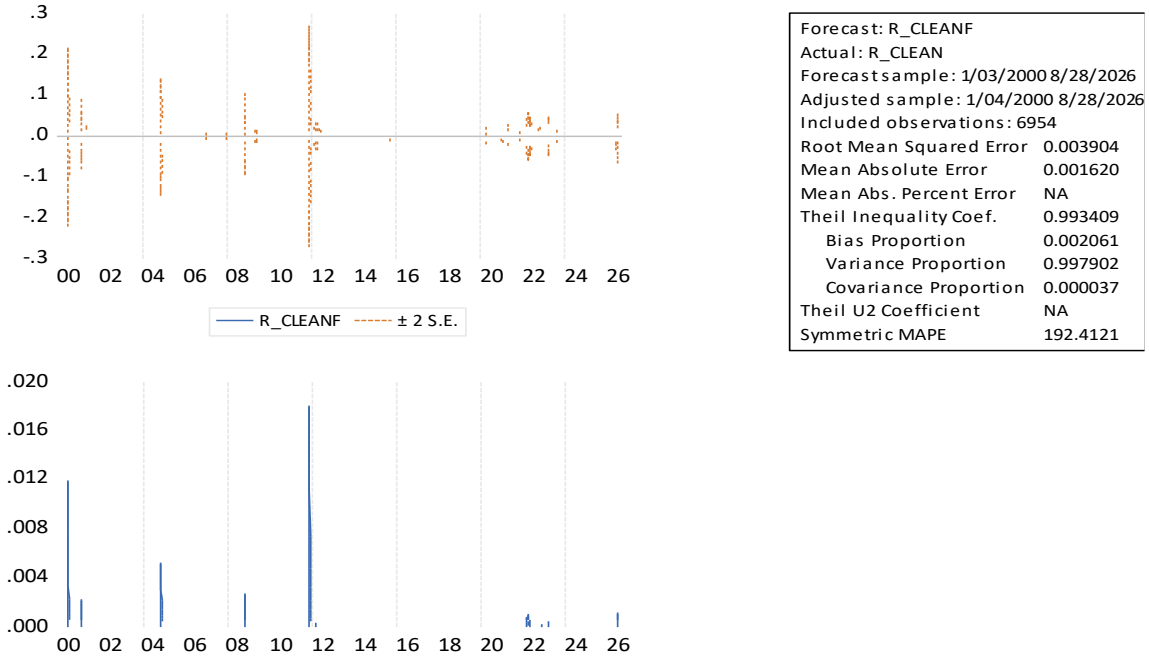


Figure 2: Forecast of $EGARCH(1,1)$ Model

Systemic Bias: The ARIMA model displays a Bias Proportion of 0.000000, implying absolute systemic balance where the forecast mean sits perfectly flat and centered across the true data trajectory without overshooting or undershooting trends.

The Mean Absolute Error (MAE) vs. Root Mean Squared Error (RMSE) Trade-Off: EGARCH minimizes average absolute proximity during normal trading days, yielding a slightly superior MAE (0.001620) compared to ARIMA (0.001663). However, because the RMSE heavily penalizes large isolated tracking errors, ARIMA wins on RMSE (0.003861 vs. 0.003904). This confirms that linear lag structures absorb sudden market shocks better than pure variance adaptations.

Ultimately, while the EGARCH framework provides crucial asymmetric risk data, the ARIMA(3,0,3) model serves as the statistically superior tool for directional daily exchange rate forecasting within this empirical timeline (Fig. 1 and Fig. 2).

5 CONCLUSION

The comparative analysis demonstrates that while the EGARCH(1,1) model excels at capturing volatility clustering, asymmetry, and persistence in the exchange rate return series, the ARIMA(3,0,3) model provides superior overall predictive performance for the mean return dynamics. The ARIMA model achieves a perfect zero bias and significantly lowers the Theil Inequality Coefficient, reflecting more accurate and reliable forecasts of daily returns. Therefore, for practical forecasting purposes, the ARIMA(3,0,3) model is the preferred choice. However, incorporating the volatility insights from the EGARCH model can complement this approach by modeling the risk and uncertainty inherent in exchange rate movements. Together, these approaches offer a comprehensive understanding of both the conditional mean and variance of exchange rate returns.

REFERENCES

- Dou, H. (2025). Forecasting the JPY/USD exchange rate using the ARIMA model. *Advances in Economics, Management and Political Sciences*, *193*(1), 247–255.
- Magaji, B., & Garba, J. (2022). Forecasting the exchange rate of Nigerian naira to United State' dollar using ARIMA-GARCH model. *Dutse Journal of Pure and Applied Sciences*, *8*(3b), 87–96.
- Muhamad Shah, R. A., & Khairol Azmi, N. N. (2026). Forecasting the alaysian ringgit (MYR) exchange rate: Arima vs Garch. *Voice of Academia*, *22*(1), 1–16.
- *Sri Lankan rupees to U.S. dollar spot exchange rate*. (2025). Stlouisfed.Org. https://fred.stlouisfed.org/series/DEXSL

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Hybrid DNN-FNN Neural Networks Model for Potato Yield Prediction: A Case Study in India

Warnakula WPM[1], Napagoda NADN[2]

[1,2]*Department of Mathematical Sciences, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*

madhubhashani2001@gmail.com[1] , namaleen@wyb.ac.lk[2]

ABSTRACT

Agriculture is essential for food security and economic development in developing countries. Potato (Solanum tuberosum) is an important crop in India, contributing to nutrition, agricultural production, and food security. However, accurate potato yield prediction is challenging because soil, weather, and agro-climatic factors interact in complex and nonlinear ways. Existing studies have mainly focused on individual machine learning or deep learning models, with limited attention to combining Deep Neural Networks (DNN) and Feedforward Neural Networks (FNN) with systematic feature selection. Our study proposed a Hybrid Deep Neural Network and Feedforward Neural Network (DNN-FNN) model for potato yield prediction in India. Random Forest (RF), XGBoost, and Recursive Feature Elimination (RFE) were evaluated as feature selection methods within the proposed framework. The study used 2,198 weekly observations with 47 soil, weather, and agro-climatic features. Data preprocessing included categorical encoding, Winsorization for outlier treatment, and Z-score standardization. DNN, FNN, and Hybrid DNN-FNN models were trained using each feature selection method, resulting in nine predictive models. Among these, the Hybrid DNN-FNN model combined with RF feature selection achieved the best performance, with an R-squared (R^2) of 81.64% and a Mean Absolute Error (MAE) of 0.17. The findings demonstrate that combining hierarchical deep feature learning with the ability to capture complex relationships among input features, supported by effective feature selection, improves 10-week-ahead potato yield forecasting accuracy. The proposed framework can support farmers, agronomists, and policymakers in data-driven crop management and provide a foundation for applying hybrid deep learning approaches to other crops.

KEYWORDS: Deep learning, Hybrid DNN-FNN, Random Forest, Recursive feature elimination, XGBoost

1 INTRODUCTION

Agriculture is essential to food security, economic growth, and sustainable development, particularly in developing countries. It plays a vital role in the economies of many developing countries, including India, and supports the livelihoods of millions of people. With the growing global population and increasing climate variability, accurate crop yield forecasting has become important for effective agricultural planning and resource management. Potato is an important food crop that provides essential nutrients and contributes to agricultural production and food security. Therefore, accurate prediction of potato yield is important for improving crop management, supporting resource allocation, and food security. However, the major challenge in potato yield prediction is the number of factors that influence the growth and development of crops. These factors can be related to the soil (pH levels, nutrient levels, moisture), weather conditions (temperature, rainfall, humidity), and agroclimatic indices (growing degree days, frost days). Agroclimatic indices further increase the complexity of these relationships.

However, traditional statistical methods, such as Multiple Linear Regression, may be less suitable for modelling complex non-linear relationships among agricultural variables. Previous studies have demonstrated the application of Deep Neural Networks (DNN) and Feedforward Neural Networks (FNN) for potato yield prediction (Singha & Swain, 2021). Existing studies have mainly investigated individual machine learning or deep learning architectures, while comparatively limited attention has been given to combining DNN and FNN architectures specifically for potato yield prediction. In addition, a systematic comparison of feature selection approaches such as Random Forest (RF), XGBoost, and Recursive Feature Elimination (RFE) within individual and hybrid neural network models remains limited. This creates a research gap in developing a systematic hybrid framework that combines DNN and FNN with feature selection for potato yield prediction.

The primary objective of this study is to develop and evaluate neural network models for predicting Potato yield in India. By analyzing a dataset of over 2,000 observations with 47 soil, weather, and agro-climatic variables, this study evaluates individual and hybrid models together with feature selection techniques. The resulting framework provides a basis for accurate yield forecasting and supports data-driven agricultural management.

2 LITERATURE REVIEW

In this section, a detailed overview of the key findings of previous research studies is presented. Recent studies founded on such models are summarized, and their intent, approach, results, and limitations for potato yield forecasting are examined.

Karbasi et al. (2026) developed a hybrid DNN model combined with Satin Bowerbird Optimization (SBO) for potato yield prediction using soil characteristics from Canada. The study used soil data collected during the 2017 and 2018 growing seasons and applied feature selection techniques to identify important soil properties. The results showed that the DNN-SBO model achieved strong prediction performance. However, the study focused on soil characteristics and did not compare DNN, FNN, and hybrid DNN–FNN architectures for potato yield prediction.

Sivanandhini and Prakash (2020) used FNN and Recurrent Neural Network (RNN) models to predict crop yield using data for Maharashtra State, India, for the year 2013. The results showed that neural networks effectively captured nonlinear relationships with high prediction accuracy. However, the study did not investigate hybrid architectures or feature selection techniques.

James et al. (2025) developed an integrated FNN model for greenhouse tomato yield prediction in Nigeria using climatic, soil, and agronomic factors. The model demonstrated effective prediction performance after preprocessing, regularization, and cross-validation. However, the study focused on an FNN model and did not compare it with other deep learning or hybrid architectures.

Oikonomidis et al. (2022) developed hybrid deep learning models for soybean yield prediction across nine US states using weather, soil, management, and historical yield data from 1980 to 2018.

The study compared CNN–DNN, CNN–XGBoost, CNN–RNN, and CNN–LSTM models with XGBoost, with the CNN–DNN model achieving the best predictive performance. However, the study focused on CNN-based hybrid architectures.

Ebrahimy et al. (2023) investigated potato yield prediction using remote sensing data and machine learning algorithms. They used county-level data from the United States from 2000 to 2012 and field-level data from Madison, Wisconsin, from 2016 to 2020, along with models such as RF, XGBoost, DNN, and other machine learning algorithms. The results showed that the performance of several models improved, particularly DNN and RF. However, the study was limited by the small amount of field-level yield data and the possibility of noise being introduced through synthetic data generation.

A review of the literature indicates that there are no studies related specifically to a hybrid DNN–FNN architecture for potato yield prediction in India. It concludes that there is a research gap in directly comparing these two architectures on the dataset. Traditional statistical methods can be effective under specific model assumptions and functional relationships; however, their performance may be limited in highly volatile and high-dimensional agricultural environments characterized by complex and nonlinear patterns. However, neural networks are able to learn these patterns automatically from massive data. Comparative analysis of DNN, FNN, and hybrid models provides decision-making information to farmers, agronomists, and policymakers to select appropriate tools for potato yield forecasting.

3 METHODOLOGY

In this study, a potato yield dataset obtained from Kaggle was utilized. The dataset contains potato yield data collected from potato-growing regions in India, covering the period from November 1983 to December 2025. It consists of 47 soil, weather, and agro-climatic variables, including temperature, rainfall, humidity, wind speed, solar radiation, atmospheric pressure, soil pH, soil texture, organic matter, nutrient levels, moisture content, bulk density, irrigation, and other agricultural factors. Potato yield was considered the target variable. The dataset is publicly available at https://www.kaggle.com/datasets. Standardization was applied to bring the input features onto a common scale.

Initially, a systematic preprocessing step is applied to ensure the dataset's reliability and quality. It is initiated by checking missing values, duplicate records, and statistical outliers using the Interquartile Range (IQR) method. Categorical variables are label-encoded, and extreme values are treated using Winsorization.

Then, to identify the most informative predictors, three feature selection techniques, RF, XGBoost, and RFE, are applied, retaining features above a ranking threshold corresponding to the top 50% of ranked variables for model development.

Then, three neural network architectures, namely DNN, FNN, and Hybrid DNN–FNN, are proposed. The DNN consists of multiple hidden layers for learning complex nonlinear relationships, while the FNN processes information sequentially through fully connected layers. The basic neuron

operation in both architectures is expressed as in (1).

$$z_j = \sum_{i=1}^{n} w_{ij} x_i + b_j \qquad (1)$$

where x_i represents the input feature, w_{ij} is the corresponding weight, and b_j is the bias term. The architecture of the Hybrid model consists of a DNN branch and an FNN branch combined through concatenation. As shown in Table 1, the three models are proposed.

For yield forecasting, the selected features are used as inputs, and the models predict potato yield for each of the next 10 weeks. The dataset is split into training, validation, and test sets of 80%, 10%, and 10% in order to preserve generalization and avoid data leakage.

Model training is conducted with early stopping that is based on validation loss optimization. Further, model performance is evaluated with Mean Absolute Error (MAE) and R-Squared (R^2). Finally, all nine models forecast the next 10 weeks, with results and loss curves plotted to assess convergence and stability.

Table 1: Architectures of DNN, FNN And Hybrid DNN-FNN

Characteristic	**DNN**	**FNN**	**Hybrid Model**
Layers	Dense (2), Dropout	Dense (2), Dropout	DNN+FNN branches
Activation functions	ReLU (hidden), Linear (output)	ReLU (hidden), Linear (output)	Swish (hidden), Linear (output)
Regularization	Dropout (0.3,0.4)	Dropout (0.1)	Batch Normalization

4 RESULTS AND DISCUSSION

The preprocessing step was completed and ensured there were no missing values, zeros, or duplicate records. The feature importance results obtained using RF, XGBoost, and RFE are presented in Table 2.

According to Table 3, the Hybrid DNN–FNN achieved its best performance using RF feature selection (R^2 = 81.64%, MAE = 0.1704), indicating outstanding prediction accuracy and generalizability.

Since the Hybrid DNN–FNN model with RF feature selection achieved the highest performance among all nine configurations, its training and forecasting results are presented in detail. In Fig. 1, the steady alignment between the training and validation loss curves indicates stable optimization and strong generalization.

In Fig. 2, the forecasted values for the next 10 weeks closely follow the trend of recent historical yield observations. The differences between the actual and forecasted values indicate the forecasting errors, with several weeks showing some deviations between the actual and forecasted values, as presented in Table 4.

However, our study is limited by the use of a single dataset from potato-growing regions in India, which may limit the generalizability of the results to other regions and growing conditions. In addition, the forecasting performance may be affected by the availability and variability of weather, soil, and agricultural factors in the dataset.

Table 2: Feature Importance for Yield Prediction

RF	XGBoost	RFE
feelslikemax_C	feelslikemax_C	feelslikemax_C
Calcium_mg_kg	Calcium_mg_kg	Calcium_mg_kg
Magnesium_mg_kg	Magnesium_mg_kg	Magnesium_mg_kg
sealevelpressure_mb	sealevelpressure_mb	sealevelpressure_mb
windspeed_km/h	windspeed_km/h	windspeed_km/h
Clay_%	Clay_%	Clay_%
Sulfur_mg_kg	Sulfur_mg_kg	Sulfur_mg_kg
Temp_Shock (TS)	Temp_Shock (TS)	Soil_Temperature_C
precipitation_mm	Soil_Temperature_C	Rainfall_mm
solarenergy_MJ/m2	precipcover%	precipcover%
conditions	Irrigation_Applied_mm	Irrigation_Applied_mm
precipcover%	solarenergy_MJ/m2	solarenergy_MJ/m2
	Rainfall_mm	Potassium_mg_kg
	cloudcover%	cloudcover%
	Sand_%	Sand_%

Table 3: Comparison of the nine models

Architecture	Feature Selection	MAE	R^2
DNN	XGBoost	0.2893	0.7312
FNN	XGBoost	0.2849	0.7419
Hybrid	XGBoost	0.1971	0.8031
DNN	RFE	0.3113	0.7074
FNN	RFE	0.2632	0.7664
Hybrid	RFE	0.1835	0.8112
DNN	RF	0.3027	0.7149
FNN	RF	0.2692	0.7620
Hybrid	RF	0.1704	0.8164

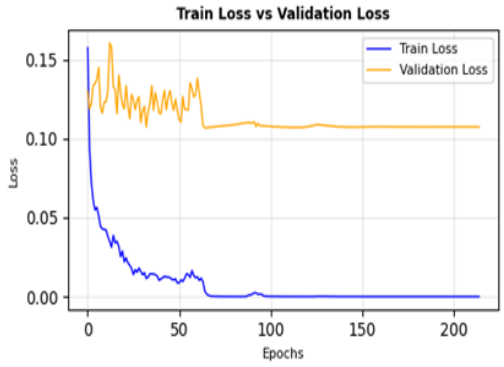


Figure 1: Training History for Hybrid DNN-FNN (RF-Based) model

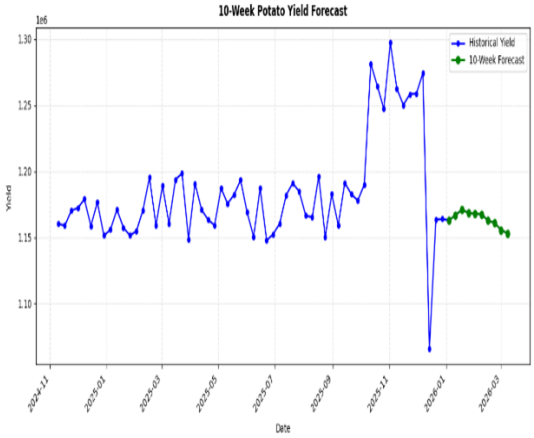


Figure 2: 10-Week Forecast for Hybrid DNN-FNN (RF-Based) model

Table 4: Predicted values of the Hybrid DNN-FNN (RF-Based) model

Date (2025)	Actual	Hybrid DNN-FNN (RF-Based)
26-Oct	1247234	1187753.50
02-Nov	1297161	1204980.75
09-Nov	1262453	1155750.50
16-Nov	1250080	1198468.00
23-Nov	1258264	1178944.50
30-Nov	1258561	1164141.38
07-Dec	1274235	1159349.50
14-Dec	1065808	1184673.12
21-Dec	1163413	1224273.88
28-Dec	1163859	1167666.50

5 CONCLUSION

Our study developed and evaluated deep learning models for potato yield prediction using environmental, soil, and agricultural data. The dataset contained 2,198 entries with 47 variables, which were preprocessed through encoding, outlier handling, standardization, and feature selection. Accurate crop yield prediction can support sustainable agricultural management and informed decision-making.

The Hybrid DNN–FNN model with RF feature selection achieved the best performance (R^2 = 81.64%, MAE = 0.17), demonstrating its effectiveness in potato yield prediction. The findings can support data-driven agricultural management, timely planning, and food security decision-making.

This study was limited by the availability and characteristics of the dataset, including its study period, geographical coverage, and selected variables. Future studies could incorporate additional data sources, such as satellite imagery and weather forecasts, and evaluate other hybrid and ensemble models to improve potato yield prediction.

REFERENCES

- Ebrahimy, H., Wang, Y., and Zhang, Z. (2023). Utilization of synthetic minority oversampling technique for improving potato yield prediction using remote sensing data and machine learning algorithms with small sample size of yield data. *ISPRS Journal of Photogrammetry and Remote Sensing, 201*, 12–25.
- James, I., James, I., Ibanga, U., Udoeka, I., and Asuquo, D. (2025). An integrated feedforward neural network for categorical prediction of greenhouse tomato yield under Nigeria's climatic, soil, and agronomic parameters. *Indonesian Journal of Computer Science, 14(6),* 10180–10193.
- Karbasi, M., Randhawa, G. S., Farooque, A. A., Ali, M., Jamei, M., Khosravi, K., Afzal, H., Malik, A., and Zaman, Q. U. (2026). Utilizing soil characteristics and hybrid machine learning for interpretable potato yield prediction: A study with satin-bowerbird optimization and deep neural network. *Field Crops Research, 338*, 110311.
- Oikonomidis, A., Catal, C. and Kassahun, A. (2022). Hybrid deep learning-based models for crop yield prediction. *Applied Artificial Intelligence, 36*(1), 2031822.
- Singha, C., & Swain, K. C. (2021). Rice and potato yield prediction using artificial intelligence techniques. In *Internet of Things and Analytics for Agriculture, Volume 3*, 185-199.
- Sivanandhini, P. and Prakash, J. (2020). Crop yield prediction analysis using feed forward and recurrent neural network. *International Journal of Innovative Science and Research Technology, 5*(5), 1092–1096.

Proceedings of the 18th Symposium on Applied Science, Business & Industrial Research – 2026
ISSN 2989-0683, ISBN 978-624-6674-56-4

Exchange Rate Movements and Tourism Demand in Sri Lanka: Evidence from a Vector Error Correction Model

Sewwandi JAB[1], Aponsu GMLM[2]
[1.2]*Department of Mathematical Sciences, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka.*
bhagyasewwandi302@gmail.com[1] , lakshana@wyb.ac.lk[2]

ABSTRACT

This study examines the relationship between exchange rate movements and tourism demand in Sri Lanka using monthly data from January 2000 to March 2020. The sample ends at the onset of the COVID-19 pandemic to focus on pre-pandemic tourism dynamics and avoid the subsequent disruption. A Vector Error Correction Model (VECM), supported by unit-root and Johansen cointegration analysis, is employed to examine long-run equilibrium relationships and short-run adjustment dynamics among tourist arrivals, the Sri Lankan rupee/US dollar (LKR/USD) exchange rate, interest rate, inflation, gold price, and oil price. The Johansen procedure identifies one cointegrating relationship. The long-run estimates indicate that the exchange rate and interest rate are statistically significant, while inflation, gold price, and oil price are not significant. Since an increase in the LKR/USD exchange rate represents a depreciation of the Sri Lankan rupee, a 1% depreciation is associated with a 2.45% increase in tourist arrivals, while a 1% increase in the interest rate is associated with a 0.70% decrease in arrivals. The negative and significant error-correction coefficient (−0.264, $p < 0.01$) indicates a speed of adjustment of approximately 26.4% per month toward long-run equilibrium. In the short run, lagged tourist arrivals are significant at the 1% level, while the exchange-rate term is significant at the 10% level ($p = 0.056$). Pairwise Granger-causality tests indicate that exchange rate and gold price have predictive precedence for tourist arrivals. The findings highlight the importance of exchange-rate movements and other macroeconomic determinants in understanding Sri Lankan tourism demand.

KEYWORDS: Cointegration, Exchange rate movements, Macroeconomic determinants, Tourism demand, VECM

1 INTRODUCTION

Tourism is a major contributor to Sri Lanka's economic growth, employment, and foreign exchange earnings, and tourism demand is shaped by several economic factors, of which exchange rate fluctuations are particularly important because they affect travel costs and destination competitiveness. Other macroeconomic factors, such as interest rates, inflation, gold prices, and oil prices, may also influence tourism demand. Previous studies on Sri Lanka report mixed findings on the exchange rate–tourism demand relationship, and many consider exchange rates in isolation from other macroeconomic factors.

This study therefore examines the impact of exchange rate fluctuations on Sri Lankan tourism demand while jointly considering interest rate, inflation, gold price, and oil price. Using monthly data from January 2000 to March 2020, it employs a Vector Error Correction Model (VECM) and related econometric tests to investigate long-run and short-run relationships, causality, and the relative importance of exchange rate fluctuations in determining tourism demand, providing insights for policymakers and tourism authorities.

2 LITERATURE REVIEW

Rathnayake (2018) found that exchange rate volatility negatively affects

tourist arrivals to Sri Lanka, while Nisthar and Nufile (2019) reported a significant exchange rate–tourism demand relationship; Sanjaya (2025), by contrast, found no significant effect using annual data. Wijayawardhana (2024) identified exchange rates, inflation, and past arrivals as important determinants using machine learning. Using unit root and bivariate cointegration tests on data from 1970 to 2015, Mohamed Aslam (2016) found a positive and significant long-run and short-run relationship between the exchange rate and tourism demand in Sri Lanka. Internationally, Dogru et al. (2019) applied linear and nonlinear autoregressive distributed lag (ARDL) cointegration techniques to U.S. bilateral tourism trade with Canada, Mexico, and the U.K., and reported evidence supporting the Marshall–Lerner condition rather than the J-curve theory, indicating that currency movements materially affect tourism flows. Given these mixed findings and the limited use of monthly, cointegration-based frameworks considering a broader set of macroeconomic variables, this study applies a VECM to examine the long- and short-run effects of exchange rate, interest rate, inflation, gold price, and oil price on Sri Lankan tourism demand.

2.1 Theoretical Background

Cointegration Theory (Engle & Granger, 1987) underlies the examination of long-run equilibrium among non-stationary variables, and the VECM captures both long-run relationships and short-run adjustment. Tourism demand (LTOURIST) is accordingly modelled using exchange rate, interest rate, inflation, gold price, and oil price, all in natural logarithms so coefficients can be read as elasticities.

3 METHODOLOGY

The study adopts a quantitative research design based on monthly secondary time-series data covering January 2000 to March 2020, providing a series suitable for cointegration-based analysis. Data were obtained from the Sri Lanka Tourism Development Authority the Central Bank of Sri Lanka, the Department of Census and Statistics, and the World Bank, as summarized in Table 1.

Tourism demand (LTOURIST) is the dependent variable; the independent variables are exchange rate (LEXCHANGE), interest rate (LINTEREST), inflation (LINFLATION), gold price (LGP), and oil price (LOP), all in natural logarithms. The functional and econometric models are:

$$TOURIST_t = f(EXCHANGE_t, INTEREST_t, INFLATION_t, GP_t, OP_{t)} \quad (1)$$

$$LTOURIST_t = \beta_0 + \beta_1 LEXCHANGE_t + \beta_2 LINTEREST_t + \beta_3 LINFLATION_t + \beta_4 LGP_t + \beta_5 LOP_t + \varepsilon_t \quad (2)$$

Stationarity was assessed with the Augmented Dickey-Fuller (Dickey & Fuller, 1979) and Phillips-Perron tests; lag length was chosen using the Akaike (AIC), Schwarz (SC), and Hannan–Quinn (HQ) information criteria; and the Johansen (1988) trace and maximum eigenvalue statistics tested for cointegration. As the series are I(1) and cointegrated, a VECM was estimated:

$$\Delta Y_t = \alpha + \sum_{i=1}^{p} \beta_i \Delta X_{t-i} + \lambda ECT_{t-1} + \varepsilon_t \quad (3)$$

where Δ is the first-difference operator, ECT is the error correction term, and λ is the speed-of-adjustment coefficient. Short-run causality was assessed with Wald and Granger causality tests; impulse response functions (IRFs) trace the response of tourism demand to one-standard-deviation shocks in each variable, and forecast error variance decomposition (FEVD) apportions its forecast error variance among these shocks. Model adequacy was checked with the Breusch-Godfrey Lagrange multiplier (LM) test, an autoregressive conditional heteroskedasticity (ARCH) test, the Jarque-Bera (JB) test (Jarque & Bera, 1980), and the cumulative sum (CUSUM) stability test (Brown et al., 1975).

4 RESULTS AND DISCUSSION

4.1 Descriptive Statistics and Trends

As shown in Table 2, gold price shows the highest mean and largest standard deviation, indicating the greatest fluctuation, while the exchange rate is the most stable series. Exchange rate and inflation are normally distributed (Jarque-Bera $p > 0.05$); the others are not, which does not affect the cointegration-based analysis that follows.

Logged tourist arrivals trend upward overall, with a dip around 2001, growth from 2010, a 2017–2018 peak, and a sharp 2019–2020 decline coinciding with the Easter Sunday attacks and COVID-19. The exchange rate shows persistent Rupee depreciation, interest rate tracks the 2001 and 2008 recessions, and gold price shows a long-run uptrend. These patterns suggest all series are non-stationary in levels, motivating formal unit root testing.

4.2 Unit Root Tests and Cointegration

Augmented Dickey-Fuller and Phillips-Perron tests show all six variables are non- stationary in levels ($p > 0.05$, e.g. tourist arrivals 0.657/0.216) but stationary after first differencing ($p < 0.05$), so all are treated as I(1). With lag 2 selected by the majority of the likelihood ratio (LR), final prediction error (FPE), and HQ criteria, the Johansen test — Trace statistic 99.80

Table 1: Data Sources

Variable	Measurement	Source
Tourist Arrivals	Number of international tourist arrivals	Sri Lanka Tourism Development Authority
Exchange Rate	Average monthly exchange rate (LKR/USD)	Central Bank of Sri Lanka
Interest Rate	Monthly interest rate (%)	Central Bank of Sri Lanka
Inflation Rate	Colombo Consumer Price Index inflation rate (%)	Department of Census and Statistics
Gold Price	International gold price	World Bank Commodity Price Database
Oil Price	International crude oil price	World Bank Commodity Price Database

Table 2: Descriptive Statistics of Variables (n = 243)

Variable	Mean	Median	Max.	Min.	Std. Dev.	JB (p)
LTOURIST	11.108	10.931	12.442	9.372	0.699	0.003
LEXCHANGE	4.764	4.723	5.221	4.282	0.214	0.213
LINTEREST	2.329	2.319	3.075	−0.693	0.370	0.000
LINFLATION	5.204	5.201	5.766	4.754	0.245	0.147
LGP	11.449	11.768	12.593	9.919	0.801	0.000
LOP	4.053	4.133	4.895	2.928	0.502	0.002

(> 95.75 critical, p = 0.026) and Maximum Eigenvalue 53.46 (> 40.08, p = 0.0009) — rejects no cointegration but not at most one cointegrating equation, confirming one cointegrating relationship (r = 1) and supporting VECM estimation.

4.3 Long-Run Relationship: VECM Results

Table 3 reports the long-run estimates. The exchange rate coefficient is significant (t = −3.45) and, after normalization, positive: a 1% Rupee depreciation is associated with an approximately 2.45% rise in tourist arrivals in the long run, consistent with improved price competitiveness. Interest rate is also significant (t = 4.77): a 1% rise in interest rates is associated with a 0.70% fall in arrivals. Inflation, gold price, and oil price are not significant in the long run. The normalized cointegrating equation is:

$$LTOURIST = 2.454 \cdot LEXCHANGE - 0.695 \cdot LINTEREST + 0.103 \cdot LINFLATION - 0.013 \cdot LGP + 0.127 \cdot LOP + 0.143 \quad (4)$$

The error correction term is negative and highly significant (−0.264, p < 0.01): about 26.4% of any short-run deviation is corrected each month, implying full adjustment in roughly 3.8 months (R^2 = 0.187, n = 241). Only lagged D(LTOURIST) is significant at 1% (p = 0.0006) and D(LEXCHANGE) is weakly significant at 10% (p = 0.056); interest

Table 3: Long-Run Cointegrating Equation and Adjustment Coefficient

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LTOURIST(-1)	1.000	–	–	–
LEXCHANGE(-1)	−2.454	0.712	−3.448	<0.01
LINTEREST(-1)	0.695	0.146	4.774	<0.01
LINFLATION(-1)	−0.103	0.297	−0.347	0.729
LGP(-1)	0.013	0.234	0.058	0.954
LOP(-1)	−0.127	0.197	−0.643	0.520

Table 4: Pairwise Granger Causality Test Results (Lags: 2)

Null Hypothesis	F-Stat.	Prob.
LOP does not Granger-cause LTOURIST	1.701	0.185
LINTEREST does not Granger-cause LTOURIST	2.763	0.065
LINFLATION does not Granger-cause LTOURIST	3.217	0.042
LTOURIST does not Granger-cause LINFLATION	4.806	0.009
LGP does not Granger-cause LTOURIST	7.284	0.001
LEXCHANGE does not Granger-cause LTOURIST	10.671	0.000

rate, inflation, gold price, and oil price are not significant in the short run, and individual Wald χ^2 tests confirm the same pattern (all $p > 0.10$ except exchange rate, $p = 0.056$, and lagged arrivals, $p = 0.100$).

4.4 Granger Causality, Impulse Response and Variance Decomposition

Exchange rate ($p < 0.001$) and gold price ($p = 0.001$) Granger-cause tourist arrivals unidirectionally (Table 4); inflation and arrivals show bidirectional causality ($p = 0.042$ and $p = 0.009$); interest rate shows only weak evidence ($p = 0.065$); and oil price shows no causal relationship. Consistent with this, a one-standard-deviation exchange rate shock raises arrivals from month 2, peaking at month 3 (0.039) before declining, while an interest rate shock produces an increasingly negative, persistent response (−0.093 by month 10); other shocks produce only small responses. Arrivals explain 100% of their own forecast error variance in month 1, falling to 66.8% by month 10 as interest rate (27.5%) and exchange rate (3.9%) gain explanatory power.

4.5 Diagnostic and Stability Tests

Residuals show no ARCH effects ($F = 0.059$, $p = 0.809$) and insignificant Q-statistics at all 36 lags, indicating white-noise behaviour. The Jarque-Bera test rejects normality ($JB = 20.514$, $p < 0.001$), reflecting moderate skew and kurtosis, though with 241 observations the estimators remain asymptotically consistent. The Breusch-Godfrey test detects serial correlation at 2 lags ($F = 5.934$, $p = 0.003$); combined with the satisfactory ARCH and CUSUM results the CUSUM statistic stays within the 5% bounds throughout the model is judged adequate for inference, though this is noted as a limitation.

5 CONCLUSION

This paper examined the impact of exchange rate fluctuations on Sri Lankan tourism demand using monthly data from January 2000 to March 2020 within a VECM. The Johansen test confirmed one cointegrating relationship, and the error correction term was negative and highly significant (−0.264, $p < 0.01$), with about 26.4% of any disequilibrium corrected each month. Exchange rate and interest rate were the key long-run determinants. Rupee depreciation raises arrivals,

consistent with Purchasing Power Parity (Cassel, 1918), while higher interest rates lower them while inflation, gold price, and oil price were not significant. In the short run only exchange rate and lagged arrivals showed weak (10%) significance, and Granger causality confirmed exchange rate and gold price predict arrivals, with a bidirectional link between inflation and arrivals; diagnostics confirmed the model is stable.

These findings carry policy implications. Because exchange rate movements have a significant, economically meaningful long-run effect on arrivals, a stable and predictable exchange rate through sound macroeconomic policy can support Sri Lanka's price competitiveness, while monetary authorities should weigh the demand-dampening effect of higher interest rates alongside their primary objectives. Continued investment in tourism infrastructure and market diversification would sustain growth independent of exchange rate cycles. As the sample ends just before COVID-19, future research could apply ARDL, nonlinear ARDL (NARDL), or time-varying parameter models to capture structural breaks and post-pandemic dynamics.

REFERENCES

- Brown, R. L., Durbin, J., & Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society: Series B (Methodological)*, *37*(2), 149–192.
- Cassel, G. (1918). Abnormal deviations in international exchanges. *The Economic Journal*, *28*(112), 413–415.
- Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, *74*(366), 427–431.
- Dogru, T., Işık, C., & Sirakaya-Turk, E. (2019). The balance of trade and exchange rates: Theory and contemporary evidence from tourism. *Tourism Management*, *74*, 12–23.
- Engle, R. F., & Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation and testing. *Econometrica*, *55*(2), 251–276.
- Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, *6*(3), 255–259.
- Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, *12*(2–3), 231–254.
- Mohamed Aslam, A. L. (2016). Dynamic impact of exchange rate on tourism demand in Sri Lanka. *KALAM – International Research Journal*, *10*(1), 90–95.
- Nisthar, S., & Nufile, A. A. M. (2019). An analysis of the relationship between tourist arrivals and exchange rates: An empirical study in the context of Sri Lanka. *KALAM – International Research Journal*, *12*(2), 81–90.
- Rathnayake, G. I. (2018). Effect of exchange rate volatility on Sri Lanka's inbound tourist flow. *Sri Lanka Journals Online*, *48*(2), 41–81.
- Sanjaya, K. J. M. (2025). The impact of exchange rate fluctuation on tourist arrivals of Sri Lanka. *International Journal of Multidisciplinary Research and Analysis*, *8*(5), 2454–2467.
- Wijayawardhana, A. (2024). Estimating the potential impact of shocks to tourism demand in recent history of Sri Lanka using a machine learning modeling approach. Sri Lanka Journals Online, 54(1), 17–32.

Faculty of Applied Sciences
Wayamba University of Sri Lanka
Kuliyapitiya

ISBN 978-624-6674-55-7 (Paper)
ISBN 978-624-6674-56-4 (Online)
ISSN 2279-1558 (Paper)
ISSN 2989-0683 (Online)