Chapter 02 – Literature Review

Research Project – ICT 4608

Bachelor of Information and Communication Technology (Honors)

Department of Information and Communication Technology
Faculty of Technology
Rajarata University of Sri Lanka

Details of the Research Project

Research Title : A Comparative Study of Machine Learning Models for Rice Price Prediction in Sri Lanka

Group Number : 02

Group Name : Code Zen

Submission Date : 2025/03/11

Details of the Group Members

Name	Registration ID	Index No.	Signature
A. Sanjayan	ITT/2020/097	1363	
Y.M.G.D.L Yapabandara	ITT/2020/119	1381	
A.R.D. Nawarathna	ITT/2020/065	1336	
B.M.N Gayathri	ITT/2020/030	1306	
C.M.V.S Upananda	ITT/2020/112	1376	

Details of Supervisor(s)

Name : Mr. Dhanushka Jayasinghe

Designation : Lecturer (Temporary)

Department/ Unit/ Institute : Information Communication Technology

Contact Details : 0713380704

Name : Mr. Wiraj Wickramaarachchi

Designation : Senior Lecturer

Department/ Unit/ Institute : Information Communication Technology

Contact Details : 0776329957

Introduction

Rice plays a vital role in Sri Lanka's economy, serving as the staple food for a majority of the population and contributing significantly to national food security, rural livelihoods, and economic stability. However, rice price fluctuations remain a persistent issue in Sri Lanka, driven by various factors such as weather variability, inconsistent paddy yields, global market dynamics, and currency exchange rate fluctuations. These uncertainties in price not only affect farmers' incomes but also influence consumers' purchasing power and national food policies. Thus, accurate forecasting of rice prices is crucial for effective agricultural planning, market stabilization, and policy formulation.

In this context, machine learning (ML) technologies have emerged as powerful tools capable of modeling complex, nonlinear relationships that influence rice price dynamics. ML models such as Random Forest (RF), XGBoost, ARIMA, and Long Short-Term Memory (LSTM) networks have been widely adopted in global research for agricultural yield and price forecasting. Despite these advancements, there is a limited focus on applying ML models for rice price forecasting in Sri Lanka that comprehensively considers influencing factors such as weather conditions, paddy yields, and foreign exchange rates.

Scope of the Review

The scope of this literature review covers three key areas:

1. **Existing studies on agricultural and rice price forecasting using machine learning models**, both globally and within Sri Lanka.
2. **The role of multiple influencing factors** — such as climate variables (rainfall, temperature, solar radiation), paddy yield, and currency fluctuations — in shaping rice prices.
3. **Theoretical frameworks and methodologies** that underpin machine learning-based forecasting models for agricultural commodities.

Significance of Reviewing the Literature

Reviewing existing literature is essential for identifying current research gaps, understanding effective methodologies, and assessing the applicability of various machine learning models to rice price forecasting in the Sri Lankan context. It provides critical insights into how multi-dimensional datasets can be integrated to enhance prediction accuracy, thus enabling policymakers, farmers, and traders to make informed decisions. Moreover, understanding past research facilitates the development of a robust, data-driven approach to address the challenges in rice price forecasting, contributing to the broader goals of food security and economic stability in Sri Lanka.

Conceptual Review

In the early studies, the focus was primarily on traditional statistical models, especially the ARIMA (AutoRegressive Integrated Moving Average) model, which became widely used for forecasting rice prices. ARIMA is a time series model that uses historical price data to predict future price trends, assuming that past values in the time series have predictive power. These early studies were pioneering in demonstrating how rice price forecasting could be done through a statistical approach. For example, the study "Rice Price Forecasting Using the ARIMA Model" (2023) applied ARIMA to predict rice prices in the Philippines. The study utilized time series data to examine the influence of various factors such as agricultural inputs (fertilizers, fuel, and machinery), market competition, and climatic conditions on rice price volatility. The ARIMA model showed its ability to forecast price fluctuations based on the historical data, but it had limitations when it came to handling more complex relationships between the variables. This was particularly evident in dealing with nonlinear dependencies that traditional statistical models like ARIMA could not fully address.[1]

Another significant study in the early phase was "The Agricultural Decision Support System for Forecasting Rice Prices" (2013). This study used ARIMA to forecast rice prices in different regions and demonstrated how time series analysis could help stabilize rice markets by predicting future price trends based on past patterns. While ARIMA was useful for capturing linear trends in price data, it was not as effective in addressing more intricate patterns in the data. This limitation became more apparent as more variables, such as weather conditions and economic factors, began to play a larger role in rice price fluctuations, motivating further research into more advanced techniques.[2]

Moreover, the study "Forecasting Rice Production of Bangladesh Using ARIMA and Artificial Neural Network Models" (2020) compared ARIMA with Artificial Neural Networks (ANN). While ARIMA was still the standard for forecasting, the integration of ANN highlighted the need for methods that could better capture nonlinear relationships between multiple influencing factors, such as weather, production levels, and economic conditions. The ARIMA model, despite its usefulness, was showing clear signs of being outpaced by more complex, adaptive methods.[3]

Mid-Period Studies:

As computational resources improved and data became more abundant, machine learning models began to emerge as alternatives to traditional statistical methods like ARIMA. During the mid-period studies, models such as Artificial Neural Networks (ANN) and Random Forest were employed to handle the increasing complexity of rice price forecasting. These models were capable of handling nonlinearities and interactions between a large number of variables, something that ARIMA could not efficiently process.[4]

The study "Machine Learning for Price Prediction for Agricultural Products" (2021) applied ANN to predict agricultural prices, marking a significant shift towards machine learning. The study found that ANN models could better capture the nonlinear relationships between variables like weather patterns, rice yield, and economic indicators. By using historical data to train the model, ANN could detect hidden patterns that were not immediately apparent using traditional methods like ARIMA. The ANN models, unlike ARIMA, were able to incorporate complex interactions between variables, which significantly improved prediction accuracy. This was an important milestone in the development of forecasting models, as it demonstrated that machine learning models could outperform traditional models in predictive tasks.[5]

Another important study during this period was "Comparison of Linear Regression and Random Forest Algorithms for Premium Rice Price Prediction" (2024). This research compared the effectiveness of Random Forest, a machine learning model based on ensemble learning, with Linear Regression, a traditional statistical model. The study showed that Random Forest outperformed Linear Regression in rice price prediction due to its ability to account for the complex relationships between various influencing factors, including rice production levels, market conditions, and government policies. The strength of Random Forest lies in its ability to combine multiple decision trees to generate predictions, making it far more capable of capturing intricate patterns in data compared to simpler models like Linear Regression.[6]

In the study "Forecasting Rice Price Volatility Using Machine Learning-Based Models" (2019), the authors employed Random Forest and XGBoost to forecast price volatility. These models were more robust than ARIMA because they could handle volatility and erratic price fluctuations, often caused by external shocks like natural disasters, policy changes, or global economic crises. The study highlighted how machine learning models could capture the sudden, unpredictable changes in rice prices that ARIMA struggled to forecast accurately.[7]

Recent Studies:

In the recent studies, the evolution of rice price forecasting reached a new level of sophistication with the advent of Long Short-Term Memory (LSTM) networks, XGBoost, and Fuzzy Time Series Models (FTS). These modern models are highly efficient at handling large datasets and complex, high-dimensional interactions between variables, which is critical for making accurate predictions in uncertain agricultural markets.[8]

The study "Enhanced Forecasting of Rice Price and Production in Malaysia Using Multivariate Fuzzy Time Series Models" (2024) introduced Fuzzy Time Series (FTS) models for rice price forecasting. FTS models, particularly Fuzzy Vector Autoregressive Models (FVAR), handle uncertainty and imprecision in the data, which is a common challenge in agricultural markets. The study showed that integrating fuzzy numbers with traditional time series methods improved the accuracy of price forecasting by accounting for the uncertainty in agricultural data, such as unpredictable weather events or market fluctuations. This approach represents a step forward in dealing with complexity and uncertainty in forecasting, as it allows models to work effectively even with imprecise or incomplete data.[9]

Another key development in the recent phase is the use of LSTM networks, which are a type of Recurrent Neural Network (RNN) designed for sequential data, such as time series. The study "Rice Price Prediction

Using Machine Learning Models for Sustainable Agriculture" (2024) applied LSTM, along with Random Forest and XGBoost, to predict rice prices. The LSTM model was particularly well-suited for handling long-term dependencies in time series data, such as seasonal patterns and historical trends that are crucial in agricultural forecasting. This model's ability to remember information over extended periods allowed it to capture long-term trends in rice price fluctuations more accurately than earlier models.[10]

Further, the study "Economic Indicators and Their Impact on Rice Production in Sri Lanka Using Machine Learning" (2024) employed machine learning models to understand the relationship between economic indicators (like GDP, inflation, and exchange rates) and rice production. The study confirmed that economic factors significantly influence rice prices, and that integrating this data into forecasting models, alongside environmental data, leads to more accurate predictions.[11]

Finally, "Machine Learning Models for Rice Price Forecasting Using Economic and Environmental Data" (2024) used advanced machine learning techniques, including Random Forest, XGBoost, and LSTM, to predict rice prices in Sri Lanka. By integrating economic, climatic, and production data, these models significantly improved forecasting accuracy. This research underscores the growing importance of integrating multiple data sources and using advanced machine learning techniques to address the complexities of rice price prediction in modern agricultural markets.[12]

Theoretical Framework

1. **ARIMA (Autoregressive Integrated Moving Average) Model**

The ARIMA model is based on the premise that the future value of a time series depends linearly on its past values and past errors. It captures the trend and seasonality of data, making it effective for short-term forecasting in time series data. ARIMA assumes that the data series is stationary, and if it is not, it can be made stationary by differencing.

Early studies, such as 'Rice Price Forecasting Using the ARIMA Model (2023)' and 'The Agricultural Decision Support System for Forecasting Rice Prices (2013)', applied ARIMA to predict rice prices. These studies demonstrated ARIMA's effectiveness in forecasting linear trends and seasonal fluctuations. However, ARIMA struggles with complex, nonlinear relationships between variables such as weather conditions and economic indicators, which later motivated the transition to more advanced methods.

ARIMA provides a solid theoretical foundation for time series analysis, but its limitations in handling nonlinear dependencies highlight the need for more complex models, especially as data complexity increases in agricultural forecasting.

2. Artificial Neural Networks (ANN)

ANN models are designed to simulate the human brain's information processing capabilities. They are composed of layers of neurons (input, hidden, and output layers), which allow them to model complex, nonlinear relationships between inputs and outputs. ANN models can be trained to learn patterns in data, making them adaptive and capable of improving over time.

The 'Forecasting Rice Production of Bangladesh Using ARIMA and Artificial Neural Network Models (2020)' highlighted that ANN could better capture nonlinear dependencies between multiple influencing factors, such as weather, economic conditions, and production levels. ANN has outperformed traditional ARIMA models in rice price prediction due to its ability to handle complex interactions between variables.

ANN provides a powerful approach for modeling complex systems where nonlinear relationships exist, making it an essential tool for rice price forecasting in dynamic agricultural markets.

3. Random Forest and XGBoost (Ensemble Learning Models)

Random Forest is an ensemble method that combines multiple decision trees to make predictions, averaging the outputs to improve accuracy and reduce overfitting. XGBoost, another ensemble technique, optimizes the gradient boosting method by building strong predictive models from weak learners. Both methods are effective in handling high-dimensional data and improving prediction robustness.

Studies like 'Comparison of Linear Regression and Random Forest Algorithms for Premium Rice Price Prediction (2024)' and 'Forecasting Rice Price Volatility Using Machine Learning-Based Models (2019)' have shown that Random Forest and XGBoost outperform ARIMA in handling the volatility and price fluctuations caused by external shocks, such as natural disasters, policy changes, or economic crises. These methods excel at capturing complex relationships between rice production, weather, and market conditions.

These ensemble learning models are highly effective in addressing the challenges of data complexity and market volatility, which makes them crucial for improving rice price forecasting in unpredictable agricultural environments.

4. Fuzzy Time Series Models (FTS)

Fuzzy Time Series Models integrate fuzzy logic with time series forecasting to handle uncertainty and imprecision in data. These models use fuzzy numbers to represent uncertain or vague data, which is common in agricultural data, where precise measurements are often difficult to obtain.

In 'Enhanced Forecasting of Rice Price and Production in Malaysia Using Multivariate Fuzzy Time Series Models (2024)', FTS models were applied to improve rice price forecasting. The integration of fuzzy logic with time series methods accounted for the uncertainty in agricultural data, such as unpredictable weather pattern or market fluctuations, leading to more accurate predictions.

FTS models are well-suited for forecasting in situations where data uncertainty is high, making them an essential tool in agriculture forecasting, where conditions are often uncertain or incomplete.

5. Long Short-Term Memory (LSTM) Networks

LSTM networks are a type of Recurrent Neural Network (RNN) that can capture long-term dependencies in sequential data. Unlike traditional RNNs, LSTM uses memory cells to retain information over extended periods, making them particularly effective at learning from time series data that involves long-term seasonal patterns or historical trends.

The study 'Rice Price Prediction Using Machine Learning Models for Sustainable Agriculture (2024)' demonstrated that LSTM networks were highly effective for predicting rice prices over long periods, especially in capturing seasonal cycles that are crucial for agricultural forecasting. LSTM outperforms ARIMA in handling long-term dependencies.

LSTM networks are powerful for forecasting long-term trends, which is particularly important in agricultural markets where seasonality and historical trends play a major role in determining future prices.

6. Economic Theory in Forecasting

Economic theory suggests that rice prices are influenced not only by historical data but also by macroeconomic factors such as GDP, inflation, and exchange rates. Integrating these economic indicators into forecasting models allows for a more comprehensive understanding of market dynamics.

Studies like 'Economic Indicators and Their Impact on Rice Production in Sri Lanka Using Machine Learning (2024)' integrated economic data with machine learning models to improve forecasting accuracy. This integration confirmed that economic factors significantly affect rice prices and that including them in forecasting models leads to better predictions.

The integration of economic factors is crucial for improving forecasting models, as macroeconomic variables significantly affect rice production and market conditions. Machine learning models like Random Forest and XGBoost are particularly well-suited for incorporating these factors into rice price forecasting.

Conclusion / Summary

Publication Year	Research Title	Used Methodologies	Results	Limitations
2024	**Rice Price Forecasting Using the ARIMA Model**	ARIMA Time Series Forecasting	Rice prices projected to rise from 2024 to 2028; Regular well-milled rice by 5.68% annually; Special rice 1.71-1.80% annually.	ARIMA limited in handling nonlinear data; lacks comparison with alternative models.
2022	**Impact of COVID-19 Lockdown on Rice Prices in India**	ARIMA, ANN, ELM models for time series forecasting	ELM outperformed others; ARIMA intervention model predicted INR 0.92/kg increase during lockdown.	Focused on pandemic context; may not generalize to normal conditions.
2024	**ANN-based Paddy Yield Forecasting in Sri Lanka**	ANN with BGD, SGD, MBGD optimization techniques	BGD achieved highest accuracy; ANN effective for yield forecasting.	Traditional models like ARIMA inadequate for nonlinear relationships.
2024	**Feature-engineered ML Models for Rice Yield Prediction**	LR, SVM, KNN, RF, Feature Selection	RF achieved high accuracy with fewer features; supports SDGs for food security.	Limited environmental and economic variable integration.
2023	**ML-based Decision Support System for Crop Profitability**	LSTM, Time Series Analysis, Deep Learning	Identified profitable crops based on market trends, costs, and climate; improved farmer income potential.	Focused on profitability, not direct price forecasting.
2024	**Impact of Economic Indicators on Rice Production in Sri Lanka**	SVM, LR, GPR	Economic factors (GDP, inflation) significantly correlated with rice production; improved prediction accuracy.	Gap in applying hybrid models like ANFIS for better handling of nonlinearities.

2013	**Agricultural Decision Support Systems for Rice Market Forecasting**	ARIMA, SVM, RF, Neural Networks, Data Warehouse	Enhanced rice price and production forecasting; ML models outperformed traditional models.	Lack of hybrid model integration (ANFIS suggested).
2020	**Rice Price Forecasting Using Econometric and ML Models in Sri Lanka**	Regression, ARIMA, RF, XGBoost, LSTM	Identified key economic and environmental determinants; ML models improved accuracy over traditional ones.	Limited Sri Lanka-specific advanced AI studies.
2021	**Machine Learning Modelling of Weather-Paddy Yield in Sri Lanka**	RF, MLR, PR	RF best for capturing nonlinear weather-yield relationships; high accuracy in yield forecasting.	Limited integration of non-climatic factors like economics.
2013	**Statistical Models for Climate-Crop Yield Relations**	Time-series, Cross-section, Panel models	Effective in capturing climate impacts on yields; highlighted limitations of collinearity and data issues.	Needs better data integration and adaptation strategy incorporation.
2021	**ML-based Rice Price Prediction Research Gap Review**	Comparative analysis of ML models (RF, XGBoost, LSTM)	Identified gap in using advanced ML for rice price forecasting; need for multi-factor integrated models.	Lack of real-time, multi-dimensional forecasting models.
2021	**ML in Agricultural Price Prediction: Review**	ANN, SVM, Decision Trees, Regression (Ridge, Lasso), ARIMA, LSTM	Neural networks and SVM most effective; Decision Trees and XGBoost highly accurate; RMSE, MAPE used for evaluation.	Issues with data availability, seasonality, external market influences.
2023	**ARIMA-based Rice Price Study in Philippines**	ARIMA, Econometric Analysis	Predicted price rise 2024–2028; identified key determinants like fuel, fertilizer, climate.	No integration of ML models; limited nonlinear analysis.

2024	**Rice Price Prediction in Indonesia: LR vs. RF**	Linear Regression, Random Forest	RF achieved 98.69% accuracy; identified factors like production, policies, competition.	Limited to specific region; lacks broader application testing.
2024	**Multivariate Fuzzy Time Series for Rice Prediction in Malaysia**	Fuzzy Vector Autoregressive (FVAR), Trapezoidal Fuzzy Numbers (TrFNs)	Improved accuracy over conventional VAR models; effective for price and production forecasting.	Still emerging; requires validation in other contexts.

References

[1] B. Mark et al., “Rice Price Forecasting Using the ARIMA Model,” IRE Journals |, vol. 7, 2024.

[2] W. D. R. Somawardhana, G. A. Deegala, Fernando, G. C. Dassanayake, and A. Perera, “Vee-Bissa: The agricultural decision support system,” 2013 IEEE 8th International Conference on Industrial and Information Systems (ICIIS), pp. 449–453, Dec. 2013.

[3] A. Sultana and M. Khanam, “Forecasting Rice Production of Bangladesh Using ARIMA and Artificial Neural Network Models,” Dhaka University Journal of Science, vol. 68, no. 2, pp. 143–147, Oct. 2020.

[4] “(PDF) Machine Learning for Price Prediction for Agricultural Products,” ResearchGate.

[5] Irfan Rasyid Muchtar and Afiyati Afiyati, “Comparison of Linear Regression and Random Forest Algorithms for Premium Rice Price Prediction (Case Study: West Java),” Jurnal Indonesia Sosial Teknologi, vol. 5, no. 7, pp. 3122–3132, Jul. 2024.

[6] C. Hathurusingha, N. Abdelhamid, and D. Airehrour, “Forecasting Models Based on Data Analytics for Predicting Rice Price Volatility: A Case Study of the Sri Lankan Rice Market,” Journal of Information & Knowledge Management, vol. 18, no. 01, p. 1950006, Mar. 2019.

[7] M. Bilal, M. A. Alrasheedi, Muhammad Aamir, S. Abdullah, Siti Mariam Norrulashikin, and Reza Rezaiy, “Enhanced forecasting of rice price and production in Malaysia using novel multivariate fuzzy time series models,” Scientific Reports, vol. 14, no. 1, Dec. 2024.

[8] Aminda Amarasinghe et al., “Advancing food sustainability: a case study on improving rice yield prediction in Sri Lanka using weather-based, feature-engineered machine learning models,” Deleted Journal, vol. 6, no. 11, Nov. 2024.

[9] E. J. K. P. Nandani and T. T. S. Vidanapathirana, “Forecasting Paddy Yield in Sri Lanka Using Back-propagation Learning in Artificial Neural Network Model,” Journal of the University of Ruhuna, vol. 12, no. 2, pp. 110–120, Dec. 2024.

[10] A. Alagalla and L. Weerasinghe, “Best Profitable Crops Prediction with Profit, Cost, and Farmland Optimization using Machine Learning,” International Journal of Computer Information Systems and Industrial Management Applications, vol. 15, pp. 550–565, 2023, Accessed: Jan. 31, 2025.

[11] W. Shi, F. Tao, and Z. Zhang, “A review on statistical models for identifying climate contributions to crop yields,” Journal of Geographical Sciences, vol. 23, no. 3, pp. 567–576, Apr. 2013.

[12] I. Mahmud, P. R. Das, M. H. Rahman, A. R. Hasan, K. I. Shahin, and D. M. Farid, “Predicting Crop Prices using Machine Learning Algorithms for Sustainable Agriculture,” 2017 IEEE Region 10 Symposium (TENSYMP), pp. 1–6, Sep. 2024.