FORECASTING MALAYSIAN RICE PRODUCTION USING HISTORICAL CLIMATE DATA AND MACHINE LEARNING ALGORITHMS

NURHAFIZAH BINTI MOHD YUNOS

UNIVERSITI TEKNOLOGI MALAYSIA

TABLE OF CONTENTS

TITLE **PAGE**

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Rice is a key food crop in Malaysia and plays an important role in national food security and rural livelihoods. However, rice production is affected by climate-related challenges such as unpredictable rainfall, rising temperatures, and extreme weather events. These issues make it increasingly important to develop accurate forecasting methods to support agricultural planning and decision-making.

Recent studies have shown that historical climate data, when combined with machine learning (ML) techniques, can improve the accuracy of crop yield predictions. Unlike traditional statistical models, ML methods can capture complex patterns in large datasets and are well-suited for modelling the relationship between climate variables and rice yields.

This literature review explores previous research on rice yield forecasting, with a focus on the Malaysian context. It begins by discussing trends in rice production and the effects of climate on yield. It then reviews different types of data used in forecasting, compares classical statistical models with modern ML approaches, and highlights common evaluation methods. The chapter also identifies research gaps and provides a foundation for the methodology used in this study.

2.2 Overview of Rice Production in Malaysia

Rice (Oryza sativa) is a staple food for the majority of Malaysians and an essential component of national food security. Despite its importance, Malaysia remains only partially self-sufficient in rice production, with a self-sufficiency level (SSL) that has fluctuated around 70%–75% in recent years (Ministry of Agriculture and Food Security, 2023). The government continues to implement policies and programs to boost local production and reduce reliance on imports.

Malaysia's rice production is geographically concentrated in key granary areas such as the Muda Agricultural Development Authority (MADA) region in Kedah, the Kemubu Agricultural Development Authority (KADA) in Kelantan, and the Integrated Agricultural Development Areas (IADAs) in Sabah and Sarawak. These areas benefit from irrigation infrastructure, making double-cropping systems possible and contributing significantly to national output.

National production levels are influenced by a range of agroecological and socioeconomic factors, including land availability, fertiliser use, irrigation access, and especially climatic conditions such as rainfall patterns and temperature variation. Paddy yield per hectare has generally improved due to modernisation and government subsidies, but it remains vulnerable to climate variability and extreme weather events (Shamshiri et al., 2018).

To address production challenges, the Malaysian government has introduced several strategic plans, including the National Agro-Food Policy 2.0 (2021-2030), which aims to modernise agriculture, promote sustainable practices, and increase the rice SSL to 80% by 2030 (Ministry of Agriculture and Food Industries, 2024). Investments in smart farming technologies, early warning systems, and precision agriculture are among the key strategies being promoted.

2.3 Climate Variables Affecting Rice Production

Rice production is highly sensitive to climatic conditions, particularly in tropical regions like Malaysia, where paddy cultivation often depends on seasonal rainfall and temperature regimes. Key climatic variables such as rainfall, temperature, humidity, solar radiation, and wind speed significantly influence rice growth stage, including germination, tillering, flowering, and grain filling (Sparks, 2009). Deviations from optimal climate conditions, such as droughts, floods, or extreme heat, can lead to substantial yield losses.

In Malaysia, studies have shown that rainfall variability plays a critical role in determining rice yield, especially in rain-fed areas without reliable irrigation infrastructure (Gumel et al., 2017). The onset and distribution of the monsoon season, both the Southwest and Northeast monsoons, directly affect planting schedules and water availability during critical growth stages. Delayed rainfall or prolonged dry spells have been associated with reduced yields, particularly in states like Kelantan, Sabah, and Sarawak (Tan et al., 2021).

Temperature is another crucial factor, as rice is sensitive to both low and high-temperature extremes. High temperatures above 35°C during flowering stages can cause spikelet sterility and grain abortion, significantly reducing yields (Yoshida, 1981). Conversely, minimum temperatures below 20°C can also suppress germination and tillering. Recent trends in Malaysia indicate rising average temperatures and more frequent heatwaves, posing a long-term risk to stable paddy production (Shamshiri et al., 2018).

Humidity and solar radiation also impact photosynthetic efficiency and transpiration rates. While high relative humidity is typical in Malaysia and generally supports growth, excessive humidity may increase pest and disease risks, such as rice blast or sheath blight. Solar radiation influences biomass accumulation and grain filling; reduced sunshine during prolonged rainy seasons has been correlated with yield declines in the Muda and KADA granaries (Herath et al., 2020).

Given these sensitivities, forecasting rice yield in Malaysia requires a strong understanding of how these climatic determinants interact with local agricultural practices and rice varieties. The use of historical climate data, particularly rainfall and temperature trends, is essential for developing accurate and site-specific prediction models.

2.4 Role of Data Analytics in Agriculture

The rise of digital agriculture has opened up new possibilities for data-driven decision-making. With increasing access to real-time data from satellites, weather stations, and agricultural surveys, analytics is becoming a vital tool in farming. Predictive analytics, in particular, helps anticipate future outcomes based on patterns found in historical and real-time data.

In recent years, machine learning (ML) has gained momentum in agriculture due to its ability to handle complex, non-linear relationships—something traditional statistical models often struggle with. ML algorithms can analyse large, multi-variable datasets to identify patterns and make accurate forecasts that inform better decisions on the ground.

By combining historical climate data with crop yield records, ML enables the development of robust forecasting tools that support smarter planning at both the policy and farm level. This shift aligns with global trends toward precision agriculture and smart farming, where technology helps maximise productivity and sustainability.

2.5 Classical Statistical Forecasting Approaches

Before the rise of machine learning, classical statistical models were widely used for crop yield forecasting due to their interpretability and relatively low computational requirements. These models typically assume linear relationships and rely heavily on time series data or explanatory variables such as temperature, rainfall, or cultivated area.

One of the common approaches is Multiple Linear Regression (MLR), which models the relationship between a dependent variable (such as rice yield) and multiple independent variables (e.g., rainfall, temperature, and fertiliser use). While easy to implement and interpret, MLR is often limited by its assumption of linearity and sensitivity to multicollinearity and outliers (Oguntunde et al., 2018).

Other classical approaches include Generalised Linear Models (GLM) and panel regression models, which allow for the inclusion of fixed and random effects when analysing data from multiple locations or years. These methods have been used in multi-location studies to estimate the effects of climate variables on rice production across different agroecological zones (Joshi et al., 2011).

Although these statistical methods are still relevant for baseline analysis and comparisons, they often struggle to capture non-linear and complex interactions between variables. This has led to a growing interest in machine learning models that are more flexible and better suited for large and noisy datasets.

2.6 Applications of Machine Learning in Malaysian Agriculture

Although machine learning has been widely applied to agriculture around the world, its use in Malaysia remains limited. Most local studies have relied on basic statistical models or general trends, often using coarse satellite data instead of detailed, state-level datasets.

Only a handful of studies have attempted to forecast rice production in Malaysia using ML, and those that do often use simple linear regressions without incorporating detailed climate data. There is also minimal exploration of advanced ML algorithms like Random Forest, SVR, or LSTM in a localised text. Furthermore, few Malaysian studies utiutilisegh-quality data sources like NASA's POWER dataset, which provides detailed and validated climate data tailored for agriculture.

2.7 Gaps in Existing Research

Despite global progress in ML applications for agriculture, several key research gaps remain in the Malaysian context:

a) Limited Use of ML in Local Forecasting: Many existing studies still rely on traditional methods, overlooking the potential of advanced machine learning techniques.

b) Weak Integration of Climate Data: Most models use a narrow range of climate variables or outdated sources, which limits forecasting accuracy.

c) Lack of Model Comparisons: Few studies compare the performance of multiple ML models to determine which works best under Malaysian conditions.

d) Underuse of Time-Series Methods: Techniques like LSTM, which are capable of modmodellingasonal trends and long-term dependencies, are rarely explored in local rice forecasting efforts.

2.8 Addressing the Gaps

This research aims to fill these gaps by:

a) Applying machine learning models—Random Forest, SVR, and LSTM—to forecast rice production using local Malaysian data.

b) Incorporating a wide range of climate variables from high-resolution datasets like NASA POWER.

c) Including lagged features and time-based indicators to account for seasonality and temporal trends.

d) Comparing the performance of different models to identify the most accurate and reliable approach for Malaysian rice forecasting.

By addressing these issues, the study hopes to contribute meaningful insights to the growing field of smart agriculture in Malaysia and support more informed, climate-resilient decision-making.

REFERENCES

Gumel, D. Y., Abdullah, A. M., & Sood, A. M. (2017). Assessing Paddy Rice Yield Sensitivity to Temperature and Rainfall Variability in Peninsular Malaysia Using DSSAT Model. *International Journal of Applied Environmental Sciences*, *Volume 12*(Number 8), 1521-1545. https://d1wqtxts1xzle7.cloudfront.net/53999765/05_57752-IJAES_ok_1521-1545_new1-libre.pdf?1501252351=&response-content-disposition=inline%3B+filename%3DAssessing_Paddy_Rice_Yield_Sensitivity_t.pdf&Expires=1751260365&Signature=U8qJxCrpny92s3AeD6sFE6AX2E9Saq

Herath, G., Hasanov, A. S., & Park, J. (2020). Impact of Climate Change on Paddy Production in Malaysia: Empirical Analysis at the National and State Level Experience. In *Proceedings of the Thirteenth International Conference on Management Science and Engineering Management* (pp. 656-664). Advances in Intelligent Systems and Computing. https://www.researchgate.net/publication/333905137_Impact_of_Climate_Change_on_Paddy_Production_in_Malaysia_Empirical_Analysis_at_the_National_and_State_Level_Experience

Joshi, Prakash, N., & Lall, K. (2011). Effect of climate variables on yield of major food-crops in Nepal -A time-series analysis-.

Ministry of Agriculture and Food Industries. (2024, August 13). *National Agrofood Policy - Ministry of Agriculture and Food Security*. KPKM. Retrieved June 30, 2025, from https://www.kpkm.gov.my/en/agro-food-policy/national-agrofood-policy

Ministry of Agriculture and Food Security. (2023). *Perangkaan AgroMakanan Malayisa 2022*. Bahagian Dasar dan Perancangan Strategik Kementerian Pertanian dan Keterjaminan Makanan.

https://www.kpkm.gov.my/images/08-petak-informasi/penerbitan/perangkaan-agromakanan/perangkaan-agromakanan-2022.pdf

Oguntunde, P. G., Lischeid, G., & Dietrich, O. (2018). *Relationship between rice yield and climate variables in southwestNigeria using multiple linear regression and support vector machine analysis*. International Journal ofBiometeorology. https://www.researchgate.net/publication/320403178_Relationship_between_rice_yield_and_climate_variables_in_southwest_Nigeria_using_multiple_linear_regression_and_support_vector_machine_analysis

Shamshiri, R. R., Weltzien, C., Hameed, I. A., Yule, I. J., Grift, T. E., Balasundram, S. K., Pitonakova, L., Ahmad, D., & Chowdhary, G. (2018). IJABE. *Research and development in agricultural robotics: A perspective of digital farming*, *Vol 11*(No 4 (2018)), 1-14. 10.25165/j.ijabe.20181104.4278

Sparks, D. L. (Ed.). (2009). *Advances in Agronomy*. Elsevier Science.

Tan, B. T., Fam, P. S., & Firdaus, R. B. R. (2021). Impact of Climate Change on Rice Yield in Malaysia: A PanelData Analysis. *Agriculture 2021*, *2021*(11), 569. ResearchGate. 10.3390/agriculture11060569

Yoshida, S. (1981). *Fundamentals of Rice Crop Science*. The International Rice Research Institute. http://books.irri.org/9711040522_content.pdf