Forecasting Mango Crop Yield in Miyazaki using Deep Learning

Arini Robbil Izzati Ulinnuha[a], Makoto SAKAMOTO[b], Ken MASUYA[c], Ryuusuke KAWAMURA[d]

Abstract

Mangoes from Miyazaki Prefecture are known for their high-quality, which are famous for fully ripeness, rich flavor, and high price. The premium mango is an important agricultural product of the region that must be able to survive in different seasons. This adaptation ensures consistency in quality and yield but introduces complexities in forecasting due to the sensitive interplay of climatic factors. This study explores the use of Long Short-Term Memory, LSTM deep learning models to improve forecasting mango crop yield in Miyazaki Prefecture. By analyzing extensive historical data on climate and past yields from 2002 to 2019, the LSTM model has been designed to analyze the complex factors affecting crop production. The results demonstrate that the LSTM models can effectively incorporate multiple variables that influence agricultural production, offering a robust tool for farmers and agricultural stakeholders to optimize their crop yield predictions and management practices. The strongest factor influencing crop yield is the cultivation area, with an R-squared value of 0.9433. For further analysis, forecasting Miyazaki mango crop yield has been performed using LSTM deep learning with a univariate approach, focusing on the correlation between cultivation area and mango crop yield. The best result in univariate forecasting is obtained with the number of batch size of 2 and root mean square error, RMSE in mango crop yield is 75.803. To improve these results, it is recommended to add more data and relevant features and optimize the settings which minimize the RMSE. The use of the LSTM model demonstrates the potential of deep learning techniques in forecasting agricultural production.

Keywords: Deep learning network, Long Short-Term Memory, Forecasting, Miyazaki mangoes

1. INTRODUCTION

The demand for mangoes from Miyazaki Prefecture, hereinafter referred to as Miyazaki mangoes, reflects not only their quality but also their status as a symbol of luxury and careful farming practices. Each mango is harvested according to strict quality standards such as specific weight and sugar content. This ensures that only the best fruit can sell in the market. The level of quality control has made Miyazaki mangoes a premium fruit with a very high selling value. The quality of this fruit is certainly obtained with great effort. The harvesting method using nets has made it possible to harvest high quality ripe mangoes. When mangoes are fully ripe, their sugar content increases and they become tastier, while ripe fruits are more likely to fall off. When the fruit is close to full ripeness, each fruit is surrounded by a net, and the falling ripe fruit is caught by the net [1)]. Irrigation and fertilization are carefully managed according to the needs of the mango trees.

The climate in Miyazaki is ideally suited for mango cultivation due to its warm temperatures, sufficient sunlight duration, appropriate rainfall, and high humidity, all of which synergize to create optimal conditions for growing high-quality mangoes. Despite these advantageous conditions, mango farming in Miyazaki faces several challenges. The main challenge in cultivating this crop yield from its pronounced sensitivity to climatic fluctuations, requiring the implementation of meticulous adaptation strategies to ensure the maintenance of consistent quality and crop yield. Consequently, there is a growing need to develop sophisticated methods to accurately predict and improve agricultural outcomes.

In the contemporary era, technology is fundamentally transforming agricultural practices. Advanced tools and methods are progressively being incorporated into agriculture to enhance efficiency and predictability. A key domain where technology has significantly penetrated is crop yield forecasting. Precise predictions, grounded in weather and other environmental variables, are essential for strategic planning and resource allocation. These forecasts allow farmers to proactively manage risks associated with climatic variability, thereby optimizing crop yields, and minimizing potential losses.

This research explores the deployment of Long Short-Term Memory, LSTM deep learning models as a viable solution to the prevalent challenges in agricultural forecasting. Known for their ability to efficiently process time series data, LSTM models are well suited to capture complex dependencies within historical climate and crop yield datasets. Using these datasets, the study develops a predictive model for Miyazaki mango production. The evaluation with annual datasets aims to determine the impact of

a) Master Student, Graduate School of Engineering, Mechanical Systems and Informatic Course
b) Professor, Center for Science Engineering Education
c) Associate Professor, Mechanical Engineering Program
d) Professor, Mechanical Engineering Program

climate change on Miyazaki mango production. This advanced model benefits farmers and provides critical insights to government and agricultural stakeholders. It enhances decision-making processes in Miyazaki's agricultural sector. This helps farmers and stakeholders prepare for mango production by understanding natural climate patterns.

2. STUDY AREA

The study area in this research is in Miyazaki Prefecture, such as Miyazaki City, Saito City, and Nango-cho. The species of mango cultivated in Miyazaki is *Mangifera indica 'Irwin'.* The historical data of climate and crop yield was taken from Japan Meteorological Agency [2] and the Ministry of Agriculture, Forestry, and Fisheries Japan [3] as for the statistical data from 2002 to 2019. The variables observed annually include rainfall, daily minimum, mean, and maximum temperature, mean relative humidity, mean and maximum wind speed, sunshine duration, global solar radiation, cultivation area, and crop yield.

The warm temperatures and long sunshine duration in Miyazaki are highly conducive to mango growth. In Miyazaki, the average daily mean temperature is 17.7°C [2], and the annual sunshine duration is 2121.7 hours [2]. These conditions enhance the pigmentation of Miyazaki mango skin. The red color of Miyazaki mango skin is an indicator of higher quality, with deeper red tones indicating superior fruit. The favorable climatic conditions in Miyazaki contribute to an accelerated ripening process for mangoes. Compared to other regions, Miyazaki mangoes are harvested earlier, from April to July.

3. EVALUATION METHOD

3.1 Regression analysis

Regression analysis is used to examine and determine the relationship between a dependent variable and explanatory variables. In these regression models, mango crop yield is treated as the dependent variable, while climatic factors and cultivation area serve as the explanatory variables. The linear regression model calculates the impact of each explanatory variable on crop yield and evaluates the R^2 values over a 15 years period. The type of regression used in this study is univariate. This is due to the reason for exploratory analysis in the research, as an initial exploratory step to identify variables that have a significant relationship with the dependent variable. It is also intended to focus on specific relationships, which is the study of the effect of a specific explanatory variable on the dependent variable. This evaluation is visualized by combining histograms and line graphs with an additional secondary axis. Through regression analysis, several variables showed both positive and negative coefficients, indicating correlations between crop yield and the respective climatic factors and cultivation area.

3.2 Deep learning LSTM Method

Google Colaboratory is used as a deep learning development environment. Keras, an open-source library in the Python language for deep learning is used. Using a Long Short-Term Memory, LSTM network model with the ability to efficiently process time series data, the dependence between past climatic factors and crop yield can be captured. In addition, Optuna framework for hyperparameter tuning to ascertain the optimal parameters for training the LSTM network model. By employing hyperparameter tuning, it is possible to identify the optimal set of parameter values that minimize loss or enhance model performance, resulting in an optimal network model configuration [4]. Subsequently, data preparation, modeling, and forecasting are performed using the LSTM networks. The data preparation phase involves critical processes such as data scaling and partitioning the dataset into training and testing subsets. The training data is divided into batches and mini-batch learning is used, which learns batch by batch. The batch size affects the training time and the performance of the network. Therefore, an effect of batch size on network performance is examined. During the modeling phase, the LSTM model architecture is constructed, appropriate loss functions and optimization algorithms are selected, and the model is trained using the prepared data. The Adam algorithm for updating weights and biases between neurons in the network is selected. Early stopping which terminates training data before overfitting, in which the generalization performance of the network model decreases, and dropout which randomly erases neurons except the output layer with a certain probability. Finally, the forecasting phase involves generating predictions and evaluating the model's performance using metrics such as mean squared error, MSE and root mean squared error, RMSE.

4. ANALYSIS AND RESULTS

4.1 Identification of factors with regression analysis

The regression analysis identifies three key factors that significantly impact mango crop yield such as cultivation area, mean relative humidity, and annual rainfall. Cultivation area has the strongest effect ($R^2 = 0.9433$) with a positive relationship. Mean relative humidity ($R^2 = 0.5858$) also has a positive impact, as higher humidity helps maintain necessary moisture for growth. Annual rainfall ($R^2 = 0.2448$) shows a moderate positive effect, indicating that adequate rainfall is essential for proper hydration and fruit production.

4.2 Effect of batch size on training

The analysis is performed for the largest influencing factor on crop yield, which is cultivation area. At this stage, Optuna hyperparameter tuning is used to select the best LSTM parameters. Early stopping helps to prevent from overfitting during model training. It monitors the model performance on validation data and stops the training data if there is no improvement in loss after a specified number of epochs, known as “patience”. In this study, “patience” and the maximum number of epochs are set to 10 and 100, respectively. This means that training will stop if the loss does not improve for 10 consecutive epochs before reaching 100 epochs. The batch sizes used in this training are 2, 4, and 6. This is done to find out which results are the best for the predicted value of Miyazaki mango crop yield in the time span 2017 to 2019. From the given batch size variation, the difference value between the actual data and the predicted data is obtained. The RMSE is then calculated to determine the accuracy of the prediction of the actual data. The RMSE results are shown in Table 1.

Table 1. Values of RMSE when using different batch size in LSTM method

Batch size	2	4	6
RMSE	75.803	85.140	116.002

The batch size parameter is customized to identify which configuration would give the most accurate predictions. As shown in Table 1, smaller batch size will give the lowest RMSE, indicating better predictive performance and the smaller batch size should be used [5]. The results of the experiments show that the proposed model with 2 batch size for training is more accurate, the proposed model with 4 batch size for training is the second best and 6 batch size is the third best. Therefore, for optimal prediction accuracy in forecasting Miyazaki mango cultivation, a batch size of 2 is used. While the Miyazaki mango forecasting results is shown in Figure 1.

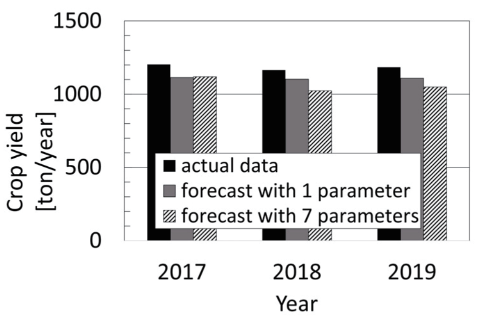

Fig.1 Miyazaki mango crop yield forecasting using LSTM on 2017-2019 based on annual historical data

The graph shows a comparative analysis of crop yield in tons per year over three years, featuring actual data alongside forecasts generated using different modeling parameters. The actual value of Miyazaki mango crop yield in 2017, 2018 and 2019 are 1203, 1165, and 1184 ton/year. Meanwhile, the forecasting result obtained using cultivation area are 1114, 1103, and 1109 ton/year and the forecast values with 7 parameters are 1120, 1023, and 1049 ton/year.

Initially, the prediction is conducted using a single parameter that exhibits the strongest interaction with crop yield, that is the cultivation area. Since the cultivation area can be managed and adjusted by farmers, this model allows for strategic planning and decision-making to optimize crop yield. Efforts to increase or decrease cultivation area have a direct impact on crop yield outcomes [6]. Subsequently, the analysis was extended to include seven parameters based on the obtained R^2 values. With a minimum threshold of 10%, further forecasting of Miyazaki mango crop yield is performed. This approach is chosen because considering multiple parameters can provide a more comprehensive and accurate prediction model that captures the complex interactions and factors influencing crop yield.

In 2017, the actual crop yield slightly exceeds the single-parameter forecast and significantly exceeds the seven-parameter forecast. The 2018 data continues this trend, with the actual crop yield again exceeding both forecasts. However, the single-parameter forecast remains closer to the actual data than the seven-parameter forecast, which shows a more significant deviation. The 2019 data continues this consistent pattern, where the actual crop yield exceeds both forecasts. In this case, the single-parameter model consistently outperforms the seven-parameter model over all three years due to the deviation values.

Based on Figure 1, the prediction trend is close to the actual value. The use of the LSTM method shows the potential of deep learning techniques in forecasting agricultural production. The ability of the model to handle sequential data and learn from it is a positive indication of its applicability in this domain. The model effectively utilizes historical data to generate forecasts. This demonstrates the model’s ability to leverage past information to make future prediction, which is crucial in agricultural forecasting.

For overall batch sizes examined in the previous step, the best prediction with the lower RMSE was selected based on experiments performed with 2 batch size. The single-parameter used to predict the crop yield is cultivation area. While the seven-parameters used to predict the crop yield are annual rainfall, daily maximum temperature, mean and minimum relative humidity, mean wind speed, solar global radiation, and cultivation area. The trend analysis of the crop yield shows a decline from 2017 to 2018, followed by an increase in 2019 with the RMSE obtained is

122.606. The RMSE suggests that the model can be improvement because the fluctuations in RMSE indicate that the model's predictions are inconsistent. There are some potential improvements such as adding more relevant data and features that affect crop yield and adjusting the hyperparameter tuning to find the optimal settings that minimize the RMSE [6].

4.3 Performance based on root mean square error

The impact of climatic variables on the crop yield of Miyazaki mangoes is of critical importance, relies on specific environmental conditions for optimal growth and fruit production. The R^2 values and linear regression analysis reveal significant correlations among these seven parameters, indicating that optimal crop yield is highly dependent on these combined factors. The cultivation area has a very strong positive impact on mango crop yield. Larger cultivation areas lead to higher yields [7]. While mean relative humidity has a significant positive impact on mango crop yield. Higher humidity levels help in maintaining optimal conditions for the growth of mangoes. It reduces water loss from the trees and supports better physiological functions [8]. Annual rainfall has a moderately positive impact on mango crop yield. This indicates that as the amount of rainfall increases, the yield of mangoes also tends to increase. Rainfall is crucial to provide the necessary water supply for mango trees, especially during critical growth periods [9]. Minimum relative humidity also has a positive impact, although less pronounced than mean relative humidity. Consistently higher minimum humidity levels can prevent dehydration of mango trees, especially during the night [10]. Mean wind speed has a slight negative impact on mango yield. Higher wind speeds can cause physical damage to the trees and flowers of mango when grown in the field and increase evaporation rates leading to water stress [11]. Mean global solar radiation has a slight negative impact on mango yield. While sunlight is necessary for photosynthesis, excessive radiation can lead to increased temperatures and water loss, stressing the trees [12]. The annual daily maximum temperature has a slight negative impact on mango crop yield. Higher maximum temperatures can stress the mango trees, leading to reduced yield. Extreme heat can negatively affect the photosynthetic process and increase evaporation rates, causing water stress [6].

Table 2. R^2 values of seven parameters which impact Miyazaki mangoes crop yield

Factor	R^2 in linear regression	Sign of slope in line
Cultivation area	0.9433	Positive
Mean relative humidity	0.5858	Positive
Annual rainfall	0.2448	Positive
Minimum relative humidity	0.1113	Positive
Mean wind speed	0.1028	Negative
Solar global radiation	0.0995	Negative
Annual daily maximum temperature	0.0978	Negative

5. CONCLUSION

The LSTM model demonstrates the potential of deep learning in agricultural forecasting by using historical data to predict future crop yields. Using Optuna hyperparameter tuning, the optimal LSTM parameters has been obtained for forecasting Miyazaki mango crop yields from 2017-2019. Different batch sizes have been examined, which showed that a batch size of 2 gives the lowest RMSE (75.803). The strongest positive influence on crop yield is the cultivation area with R^2 0.9433. Improvements of the model could be obtained by adding more data, more relevant parameters, and refining the hyperparameters to minimize the RMSE.

REFERENCES

1) Y. Ishiki, "Why Miyazaki mangoes could become the top brand in the industry: Comparison with Okinawa Mangoes", The Journal of General Industrial Research, Vol. 27, pp. 29-47, 2019 (in Japanese).
2) Japan Meteorological Agency, https://www.jma.go.jp/
3) Ministry of Agriculture, Forestry, and Fisheries Japan, https://www.maff.go.jp/
4) V. K. Kumar, K. V. Ramesh, V. Rakesh, "Optimizing LSTM and Bi-LSTM models for crop yield prediction and comparison of their performance with traditional machine learning technique", Applied Intelligence, Vol. 53, pp. 28291-28309, 2023.
5) D. Masters, C. Kuschi, "Revisiting small batch size training for deep neural networks, Graphcore Research, Bristol, UK, 2018
6) T. v. Klompenburg, A. Kassahun, C. Cata, Crop yield prediction using machine learning: A systematic literature review, Computer and Electronics in Agriculture, Vol. 177, 2020.
7) R. Vasuki, M. Soundariyan, Cost of Cultivation and Profitability of Mango Cultivation in Dharmapuri District, Tamil Nadu, EPRA International Journal of Agriculture and Rural Economic Researc (ARER), Vol. 10, 2022
8) D. Zhang, C.Wang, X. Li, Yield gap and production constraints of mango *(Mangifrera indica)* cropping systems in Tianyang Country, China, Journal of Integrative Agriculture, Vol. 18, pp. 1726-1736, 2019
9) S. Rajan, Phenological Responses to Temperature and Rainfall: A Case Study of Mango, Division of Crop Improvement and Biotechnology, Central Institute for Subtropical horticulture, Lucknow, India, 2012
10) J. Wei, G. Liu, D. Liu, Y. Chen, Influence of irrigation during the growth stage on yield and quality in mango *(Mangifera indica),* Plos One, 2017
11) A. B. Bhandari, T. P. Kharel, K. N. Reddy, Assessing wind damage and potential yield loss in mid-season corn using a geospatial approach. Frontiers in Agronomy, 2023
12) A. E. A. Shaban, M. I. M. El-Banna, A. A. Rashedy, Mitigation of Excessive Solar Radiation and Water Stress on 'Keitt' Mango *Mangifera indica* Trees Through Shading,Acta Sci. Pol. Hortorum Cultus, Vol. 20, No. 4, pp. 77-88, 2021