Climate-Driven Insights: Predicting Black Pepper Yield and Quality with Long Short-term Memory Model

H. M. N. S. Subasinghe, K. A. S. H. Kulathilake

Department of Computing, Faculty of Applied Sciences, Rajarata University of Sri Lanka, Mihintale, Sri Lanka.

Abstract— **In the context of climate variability, predicting agricultural output remains a pressing challenge, particularly for high-value crops like black pepper in Sri Lanka, a leading spice exporter. This study introduces a novel machine-learning approach to predict black pepper yield and quality, utilizing thirty years of detailed weather data from the Matale district. Employing Long Short-Term Memory (LSTM) networks, the complex dependencies between weather conditions—including rainfall, temperature, and humidity—and crop productivity are modelled. The analysis demonstrates that LSTM models can effectively forecast black pepper yield and quality by learning from historical weather patterns and corresponding crop performance data. The models achieved a mean absolute error of 18-20% for quality predictions and a mean squared error reflecting consistent model performance across different evaluations. By providing reliable yield and quality estimates, these models serve as valuable tools for farmers and policymakers to better plan and manage black pepper cultivation in response to anticipated climate conditions. Furthermore, the research highlights the potential for enhancing model accuracy by incorporating diverse regional data, thereby contributing to more resilient agricultural practices in the face of global climate change.**

Keywords— **black paper, pepper quality, yield prediction, climate change, machine Learning, LSTM.**

I. Introduction

Black Pepper is the most widely used spice in the world. Known as the land of spices, Sri Lanka is renowned for its for its black pepper, which is in high demand globally due to its unique qualities, particularly its excellent aroma and high content of volatile oils, oleoresins, and Piperine, compared to peppers from other countries. The total area dedicated to spice cultivation in Sri Lanka is 122,000 ha, of which 42,989 ha are allocated for pepper.

Correspondence: H. M. N. S. Subasinghe (E-mail: sachinthas96@gmail.com)
Received: 16-06-2024 **Revised**: 12-08-2024 **Accepted**: 09-09-2024
H. M. N. S. Subasinghe and K. A. S. H. Kulathilake are from Rajarata University of Sri Lanka, Mihintale (sachinthas96@gmail.com, kule@as.rjt.ac.lk)

DOI: https://doi.org/10.4038/icter.v18i2.7304

The 2025 Special Issue contains the full papers of the abstracts published at the 24th ICTer International Conference.

For commercial cultivations, cuttings are selected from terminal stems or ground runners. If cuttings are taken from lateral branches bush-type pepper plants can be produced. Moreover, the annual export value of the crop amounts to Rs. 458 million, highlighting Sri Lanka's potential to become one of the main quality black pepper providers worldwide [1]. Overall, a nursery period of 4-6 months should be maintained for plants. The duration of the reproductive period is 8 - 9 months. It takes 8 to 9 months for harvesting after flower initiation. Eventually, Pepper is harvested after 7-8 months of maturity. Therefore, the black pepper yield and quality data are collected annually [2].

However, climate change poses a significant challenge to spice farmers in Sri Lanka, a tropical island nation with diverse weather patterns. Pepper flourishes in tropical climates with relatively high humidity and minimal variations in day length throughout the year. However, black pepper is sensitive to excessive heat and dryness, making adequate rainfall and its distribution a crucial factor in determining the success of pepper cultivation and its overall productivity [3]. However, while there is growing research on this issue, consensus on how these changes specifically affect pepper yield and quality is still lacking. For example, some studies suggest that decreased rainfall can lead to lower yields and quality, while others indicate that excessive rainfall can also be detrimental [4]. These uncertainties are particularly relevant in Sri Lanka, where pepper cultivation often relies on rainfed conditions.

To the best of the authors' knowledge, there have been no prior studies in Sri Lanka have investigated predictive models for black pepper crop yield and quality. Given the diverse climate parameters across different regions in Sri Lanka and the country's vulnerability to climate change, pepper production is significantly impacted, which in turn affects both yield and quality, potentially diminishing the nation's foreign exchange earnings from this crucial crop. Recognizing the critical importance of black pepper as a cash crop and the potential effects of climate variability on its production, there is a pressing need to understand the relationship between rainfall patterns and pepper yield and quality.

This study aims to address this significant knowledge gap by developing a Long Short-term Memory (LSTM)based Machine Learning (ML) prediction model for black pepper yield and quality under different rainfall scenarios, specifically focusing on the Matale district in the Central

Province of Sri Lanka, often regarded as the cradle of spices. Focusing initially on the Matale region, this research aims to generate insights that can be extended to enhance pepper cultivation practices in broader geographical areas, with the potential for future expansion beyond a single district.

The proposed model used a combination of agronomic and climatic data to predict black pepper yield and quality. The climatic data included annual rainfall, annual rainy days (the distribution of rainfall), annual rainfall soon after the drought period, and the annual drought period. The agronomic data included pepper yield and quality parameters. The LSTM model was trained on a dataset that combines these variables to predict pepper yield and quality. Google Colab was used to develop the model. The data preprocessing and simulation parts were also conducted using the same platform.

The proposed model makes several contributions to the field of agricultural prediction. It provides reliable predictions of pepper yield and quality by using a combination of climatic and agronomic variables. Furthermore, to the best of our knowledge, this is the first attempt to employ an ML approach to predict pepper yield and quality, which has not been widely used in agricultural prediction in Sri Lanka.

The article is organized as follows. Section II reviews relevant literature, highlighting the advancements and challenges in using ML for agricultural yield prediction, thereby setting the stage for the innovative approach proposed in this study. Section III details the methodology, describing the phases of data collection, preprocessing, and model development, essential for understanding the predictive modelling process. Section IV presents the results and discusses the effectiveness of the LSTM models, demonstrating their capability to accurately forecast agricultural outcomes under varied climatic conditions. Finally, Section V concludes by summarizing the key findings and outlining potential future work to enhance model accuracy and applicability in different agricultural settings.

II. RELATED WORK

ML techniques have transformed the field of crop yield prediction, enhanced accuracy, and enabled more reliable forecasting. These methods are indispensable for modern agricultural practices, aiding in crucial decisions related to crop management and sustainability.

Among the popular algorithms, Support Vector Machines (SVM), Random Forests (RF), and deep learning approaches like Convolutional Neural Networks (CNN) have been widely adopted [5]- [8]. Studies, such as those by You et al. (2017), demonstrate the effectiveness of CNN for predicting soybean yields in the United States, utilizing sequences of remotely sensed images. This showcases the capability of deep learning approaches to process and analyze complex spatial and temporal agricultural data.

Jeong et al. (2016) further validated the application of Random Forest models in predicting yields for crops such as wheat, maize, and potatoes, achieving a Root Mean Square Error (RMSE) significantly lower than that obtained through traditional multiple linear regression models. This illustrates the superiority of ML techniques over conventional statistical methods in handling multidimensional and heterogeneous datasets typically encountered in agricultural research. Moreover, the comparison among machine learning algorithms reveals that techniques such as RF are not only more precise but also offer robust versatility across different scales and environmental conditions.

However, the deployment of these advanced ML models in crop yield prediction is not without challenges. These models require significant computational resources for training and deploying, which may not be feasible in resource-limited settings. Addressing these limitations, Mohan et al. (2018) utilized self-organizing maps to enhance model efficiency by reducing data dimensionality, demonstrating a practical approach to applying ML under constrained conditions.

Moreover, the comparative studies conducted by Oguntunde et al. (2018) which pitted SVM against MLR in predicting rice yield influenced by climate variables in Southwest Nigeria, underscore the capability of SVM to capture complex nonlinear relationships better than MLR. This comparison not only highlights the advanced capabilities of machine learning models but also their adaptability to diverse agricultural environments, making them invaluable tools for local and regional agricultural planning and management.

The integration of ML in crop yield prediction also faces issues related to model selection and algorithmic complexity. The choice of the appropriate model and its parameters significantly impacts the accuracy and applicability of the predictions. Khaki & Wang (2019) addressed this by designing a sophisticated deep learning model that outperformed conventional regression and shallow network approaches, indicating the critical role of model architecture and depth in handling agricultural data effectively.

Despite these technological advancements, significant gaps remain in the application of ML to crop yield prediction. Current models often fail to integrate multidisciplinary data comprehensively, such as soil health indicators, pest and disease infestations, and micro-climate variations, which are crucial for accurate yield forecasting [9].

Furthermore, the black-box nature of many deep learning models poses challenges for interpretability and trust among users, particularly in sectors where understanding model decision-making is critical.

During this research, it became evident that existing literature on the application of Long Short-term Memory (LSTM) models to predict agricultural yield and quality, specifically for black pepper, is limited. The scarcity of comprehensive studies in this domain presented a challenge in sourcing relevant information and highlights a significant gap in current research. This study aims to address this gap by providing climate-driven insights into black pepper yield and quality prediction, thereby contributing to the emerging body of knowledge in this area.

III. METHODOLOGY

A. Study Phase

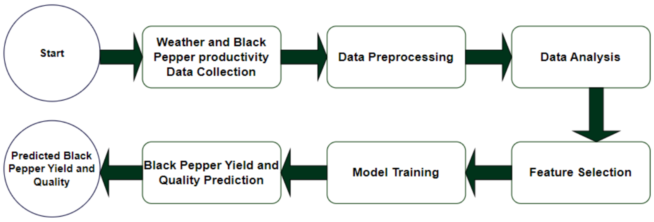

Fig.1 Study Phase

The study outlined in the flow chart depicted in Figure 1 involves a structured approach to predict pepper yield and quality through five distinct phases. Initially, weather (climate) and productivity data (agronomy) are collected, serving as the foundation for analysis. In the data preprocessing phase, this data undergoes normalization to ensure a uniform representation and is reshaped into the appropriate format for subsequent processing. Next, data analysis is undertaken to identify specific weather parameters that significantly impact pepper yield and quality, optimizing the predictive model's focus and accuracy. The feature selection step further refines the model by selecting the most relevant features that should be input into the deep learning framework. With the relevant features and parameters at hand, the model training phase involves developing and training an LSTM model tailored for this purpose. Finally, the prediction phase utilizes the trained model to make forecasts on pepper yield and quality, thereby assessing the model's effectiveness and accuracy in real-world scenarios.

B. Data Collection

Data collection for this research was conducted at the Central Research Station of the Export Agriculture Department in Matale and the Natural Resource Management Center (NRMC) of the Agriculture Department. The research involved the compilation of two distinct datasets: weather and black pepper productivity data related to black pepper cultivation. The weather dataset encompassed a range of environmental parameters, including annual average temperature (℃, [24.64 – 26.75]), annual average relative humidity (%, [74.23 – 80.89]), annual rainfall (mm, [1115.8 – 2733.5]), the number of rainy days (number of days, [91 – 152]), the duration of drought period in February and March (number of days, [12 – 30]), and rainfall immediately following drought periods (mm, [751.6 – 1643.3]). These measurements were crucial for analysing the environmental factors affecting agricultural productivity.

Additionally, black pepper productivity data was collected, focusing on black pepper yield (kg/ha, [620 – 1450]), and key quality parameters such as bulk density (g/l, [533 – 584]), Piperine content (%, [4.9 – 6.4]), and oleoresin content (%, [12.9 – 17.2]). Data gathering methodologies included the use of automated weather stations for recording climatic data and manual sampling coupled with laboratory analysis for agronomic data. The data collection process for this study spanned a period of 30 years, from January 1992 to December 2022. The original weather dataset consisted of a total of 7671 daily data records, covering the period from 2002 to 2022. The black pepper productivity dataset comprised 30 annual records, since 1992 to 2022 being the period of observation.

To create a dataset that would allow for the analysis of black pepper productivity and weather parameters, the 20-year daily weather dataset was aggregated into 20-year annual data records. It was done by taking the average of the daily values for each year, resulting in a dataset with 20 annual records.

C. Data Preprocessing

Normalization of the dataset was achieved through the application of the MinMax Scaler in Equation (1), which adjusts the data values to a scale ranging from 0 to 1.

$$x_{normalization} = \frac{x - \min(x)}{msx(x) - \min(x)} \qquad (1)$$

Where x represents an original value, min (x) is the minimum value in the feature, and max(x) is the maximum value. This transformation ensures uniformity in the scale of all input features, which is crucial for maintaining consistency across the dataset. Following the normalization, the input data was reshaped into a three-dimensional array conforming to the input requirements of the proposed deep learning model, specifically in the format (batch_size, timesteps, input_dimension).

D. Data Analysis

Data analysis was conducted in two steps. Firstly, an analysis of the weather dataset was carried out to understand the monthly weather distribution and to identify any anomalies or deviations from the average monthly rainfall distribution in the Matale district. Subsequently, black pepper productivity data and weather data were analysed to identify weather patterns and trends that occur during the growth season of black pepper and are associated with higher yields. This analysis also aimed to understand which weather variables significantly impact pepper productivity and enhance the interpretability of the model.

1) Weather dataset analysis

The objective of this data analysis is to assess the compatibility of the weather data collected over 30 years with the standardized average rainfall distribution published by the Department of Export Agriculture, which is based on 40 years of research cited in [10]. Additionally, the distributions of the average minimum and maximum temperatures recorded in the weather dataset were compared with the standard average minimum and maximum temperature distributions published in [11]. The results of the analysis are presented in Figure 2, which indicates that the patterns in the collected data align with the standard distributions.

2) Black Pepper productivity dataset analysis

Correlation tests were conducted with black pepper productivity data and weather data to identify weather patterns and trends during the growth season of Black pepper that correlate with higher yields. The pepper productivity dataset included 30 records of both pepper yield data and pepper quality data. Statistical analysis was performed using SPSS software.

Three quality parameters of Black pepper were evaluated: Bulk Density (g/l), Oleoresin Content (%), and Piperine

Content (%) of the seeds. The measurement of Bulk Density was carried out using a 1-liter volume cup, weighing the seeds within. High quality pepper was defined as having a Bulk Density exceeding 550g/l, a Piperine content over 5%, and an Oleoresin content above 14%, with Bulk Density deemed the most influential on overall quality.

The study examined the relationship between black pepper yield and weather parameters, including annual rainfall, annual average temperature, and annual average humidity. Additionally, the study analyses the correlation between black pepper quality parameters, specifically bulk density, Piperine content, and Oleoresin content, and the same set of weather parameters. The analysis was conducted separately for each parameter to identify any potential patterns or trends.

The study revealed a weak positive linear correlation between annual average temperature and black pepper yield, suggesting a minor positive impact of temperature on pepper productivity. Furthermore, the analysis of pepper quality parameters, including bulk density, piperine content, and oleoresin content, showed a weak positive linear relationship with annual average temperature, indicating a slight improvement in pepper quality with increasing temperature. Therefore, the overall effect of annual average temperature on black pepper productivity was relatively low, suggesting that other weather factors may play a more significant role in determining pepper yield and quality.

Conversely, the analysis of the relationship between rainfall and black pepper productivity parameters revealed a significant positive linear correlation, suggesting that increased rainfall is associated with improved pepper productivity.

Figure 3 depicts these correlations to further support this finding. All the graphs illustrate a clear and consistent pattern of positive correlation between rainfall and pepper productivity. These results suggest that rainfall is a key factor in determining black pepper productivity, and that optimal rainfall conditions can significantly enhance pepper production.

E. Feature Selection

The correlation matrix depicted in Figure 4 identified rainfall parameters highly correlated with each black pepper productivity variable (yield and quality). These identified features are significant for predicting pepper yield and quality. According to Figure 4, a higher correlation was observed between annual rainfall, annual rainy days, and rainfall soon after the drought period with a weak influence on black pepper yield, bulk density, piperine content, and oleoresin content. Therefore, the results indicate that RH has a minimal impact on black pepper yield and quality.

Black Pepper yield is marked with the yellow square. In addition to these rainfall parameters, the drought period showed a considerable correlation with black pepper quality.

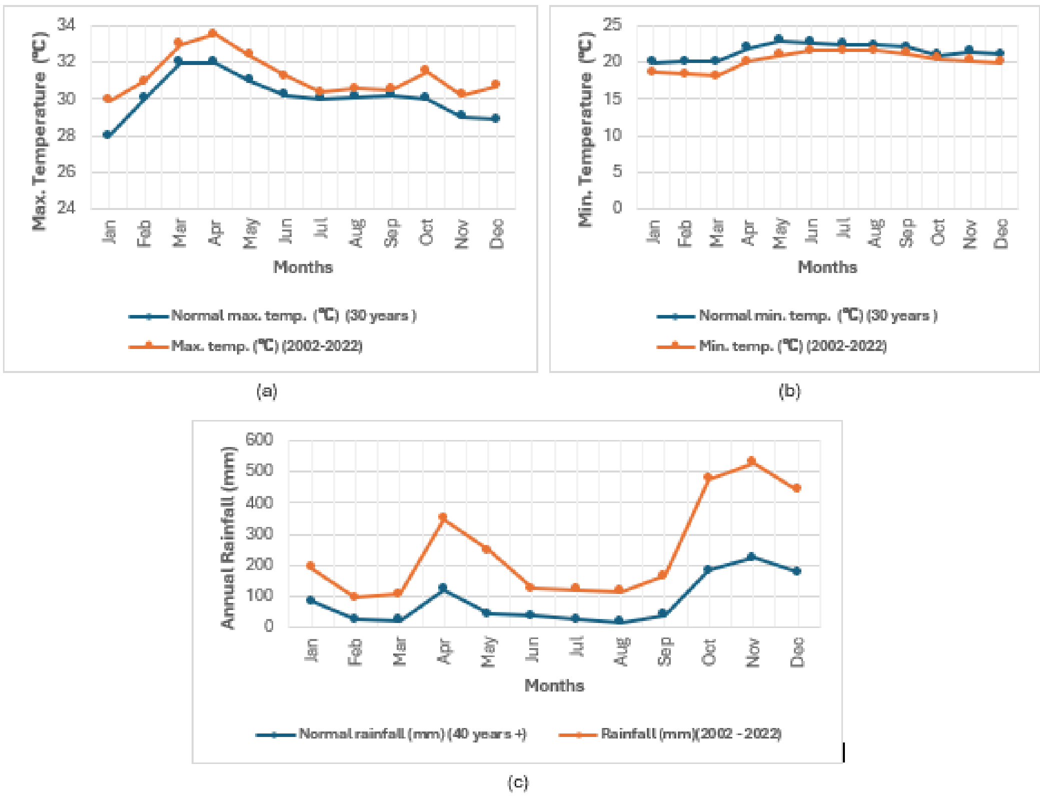

Fig. 2 Comparison of average data distributions of selected parameters in the weather dataset with the standard weather distributions for Matale District, published by the Department of Export Agriculture: (a) Comparison of the maximum temperature distribution with the standard maximum temperature distribution, (b) Comparison of the minimum temperature distribution with the standard minimum temperature distribution, (c) Comparison of the rainfall distribution with the standard rainfall distribution.

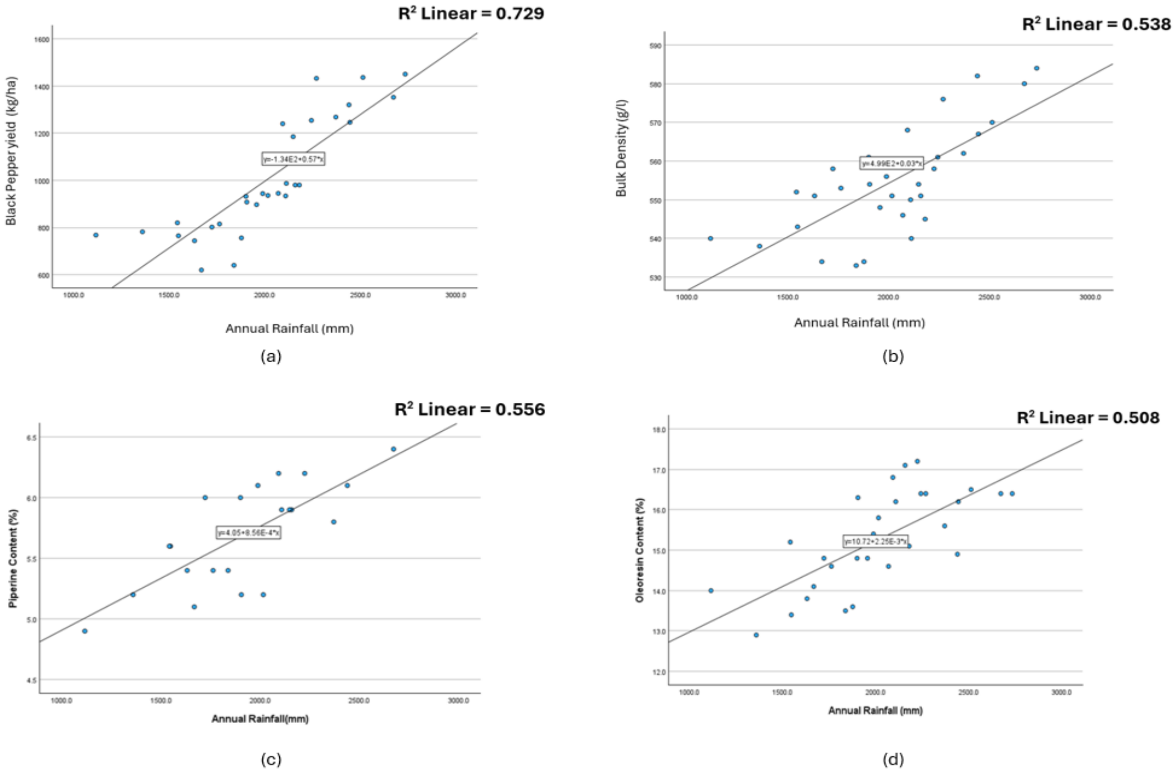

Fig. 3 The analysis of annual rainfall with black pepper yield and quality parameters; (a) black pepper yield and annual rainfall correlation graph, (b) Bulk density and annual rainfall correlation graph, (c) Piperine content and annual rainfall correlation graph, (c) Oleoresin content

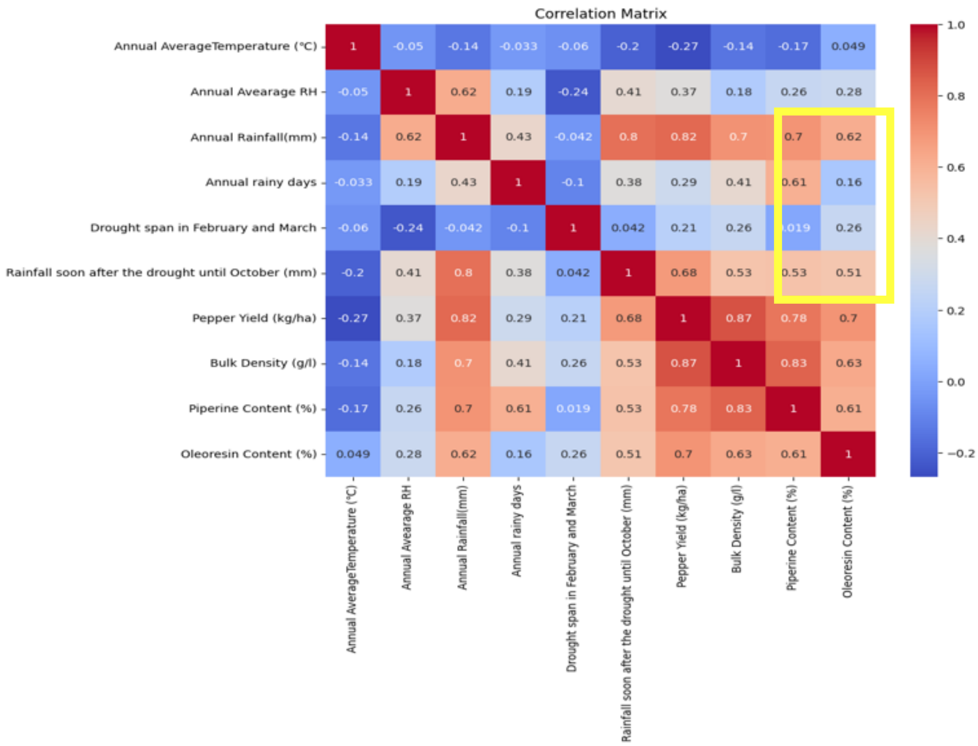

Fig. 4 Correlation matrix (rainfall parameters highly correlated with pepper yield and quality are highlighted)

Hence, the parameters of annual rainfall, annual rainy days, and rainfall soon after the drought period were used as the input features to predict black pepper yield. Meanwhile, for predicting black pepper quality, the input features considered were annual rainfall, annual rainy days, rainfall soon after the drought period, and the drought period.

F. Model Design

Based on the outcomes of the feature selection process and insights from the literature review, an LSTM model was selected for predicting both the yield and quality of black pepper. An LSTM model is a type of Recurrent Neural Network (RNN) architecture designed to process and predict sequence data. It is particularly suited for this task due to its ability to capture complex temporal relationships within the data, that necessitate capturing context and prolonged connections within sequential data, such as forecasting time series, analysing natural language, and identifying speech patterns benefit from architectures like LSTMs. Unlike traditional time-series models that rely on a fixed time window to consider past values, the LSTM can learn from data with variable time intervals, allowing it to effectively discern long-term patterns and trends. This capability is crucial for predicting black pepper yield and quality, where variable weather conditions and environmental factors significantly influence outcomes.

The structure of the proposed LSTM model is depicted in Figure 5. The model maintains a consistent structure and hyperparameters, but is trained on two distinct datasets to predict pepper yield and quality. It includes a layer with 64 units and utilizes the ReLU activation function to facilitate faster training without compromising accuracy. Adam is chosen as the optimizer for its robustness in managing sparse gradients in noisy datasets, and the Mean Squared Error (MSE) serves as the objective function, aiming to minimize the average squared differences between the predicted and actual values.

The model is trained separately on identified features to predict the black pepper yield and quality. The yield prediction model was trained using data on annual rainfall, annual rainy days, and rainfall immediately following the drought period, targeting the prediction of black pepper yield in kg/ha. Conversely, the pepper quality prediction model was trained using data on annual rainfall, annual rainy days, rainfall immediately after the drought period, and the annual drought period, predicting the quality (bulk density) in g/l.

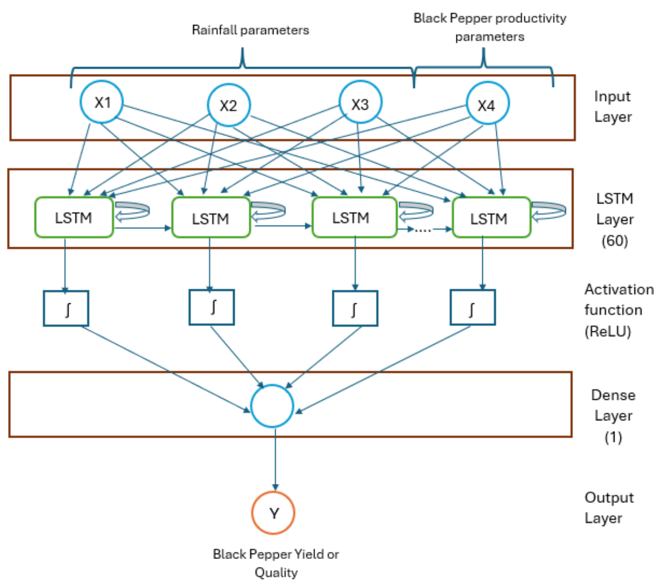

Fig. 5 Black Pepper Yield Prediction Model

Model training was conducted on the Google Colab platform. To train the model, relevant fields from the weather dataset and black pepper productivity dataset were utilized. Initially, the LSTM model was trained to predict black pepper yield and quality using datasets that contained only 25 records, with 5 records reserved for testing. This led to overfitting the model. To address the issue, an LSTM-based data simulation mechanism was employed. This approach generated 30 simulated records for the quality prediction model and 60 for the yield prediction model. Consequently, the total number of training data increased to 55 for the quality prediction model and 85 for the yield prediction model.

Although the core architecture remains the same, the training configurations differ between the two applications; for yield prediction, the model is trained for 23 epochs, while for quality prediction, it is trained for 27 epochs. Figures 6 and 7 illustrate the learning curves of the LSTM model, demonstrating the model's generalizability on unseen

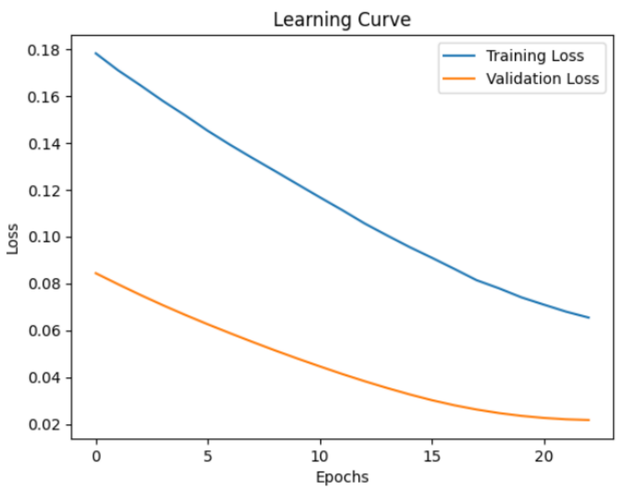

Fig. 6 Pepper yield prediction LSTM model

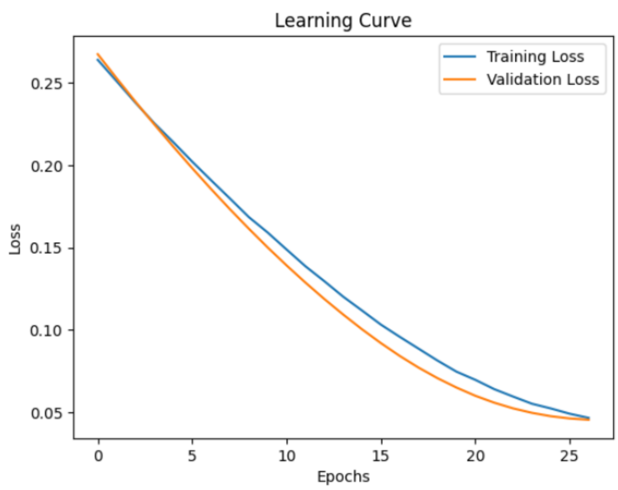

Fig. 7 Pepper quality prediction LSTM model.

data. Notably, the pepper quality prediction model exhibits better generalizability compared to the yield prediction model.

IV. RESULTS AND DISCUSSION

A) Evaluation Matrices

The performance of the LSTM model was assessed using two commonly employed evaluation metrics in the field of machine learning: R-squared (R^2) and Mean Absolute Error (MAE). These metrics are widely used in regression tasks due to their ability to provide a comprehensive and robust assessment of a model's performance. MAE is robust to outliers, a common issue in regression tasks, and is sensitive to the magnitude of errors. R^2 provides a measure of the model's ability to capture the underlying patterns in the data. A higher R-squared value indicates that the model is a better fit for the data, meaning that it captures more of the underlying patterns in the data. Therefore, the use of MAE and R squared as evaluation metrics in machine learning, particularly in regression tasks, is a widely accepted and effective practice.

Equation (2) describes the R^2 equation:

$$R\,^2 = 1 - (SSE - SST) \quad (2)$$

where SSE is the sum of the squared errors (i.e., the difference between the predicted values and the actual values), and SST is the total sum of squares (i.e., the sum of the squared differences between the actual values and the mean of the actual values).

MAE in Equation (3) measures the average absolute difference between predicted and actual values.

$$MAE \ = (1/n) * \ \Sigma|y\_true \ - \ y\_pred| \quad (3)$$

y_actual is the actual value, and *y_predict*ed is the predicted value, and *n* is the total number of data points.

B) Model Evaluation Results

In this experiment, the proposed LSTM model was trained twice: (i) with the original training dataset, and (ii) with the original dataset supplemented by simulated records. After each training phase, the LSTM model was evaluated using a test dataset. The test dataset for the yield prediction model consisted of 8 records, while the test dataset for the quality prediction model consisted of 6 records. The results are reported in Table 2.

The results of the experiment indicate that, for the yield prediction model, the R-squared value is 0.197 when trained with the original training data, and 0.262 when trained with the original and simulated data. For the quality prediction model, the R-squared value is 0.317 when trained with the original training data, and 0.358 when trained with the original and simulated data. This indicates that the model can explain about 31.7% of the variability in the quality data using the original training data, and 35.8% of the variability when the simulated data is included.

TABLE 1
MODEL EVALUATION RESULTS

Prediction Model	Training with Original Training Data		Training with Original and Simulated Data	
	R^2	MAE	R^2	MAE
Yield Prediction	0.197	0.187	0.262	0.147
Quality Prediction	0.317	0.192	0.358	0.189

For the yield prediction model, the MAE is 0.187 when trained with the original training data, and 0.147 when trained with the original and simulated data. This suggests that the model is making predictions that are, on average, 18.7% away from the actual yield values using the original training data, and 14.7% away from the actual values when the simulated data is included. For the quality prediction model, the MAE is 0.192 when trained with the original training data, and 0.189 when trained with the original and simulated data.

Overall, the results suggest that the models were able to explain a significant proportion of the variability in the dependent variables and made accurate predictions, with the inclusion of simulated data improving the performance of both models. It is indicated by the higher R-squared values and lower MAE values. This suggests that the simulated data can capture additional patterns in the data that are not captured by the original training data, leading to more accurate predictions.

C) Prediction Results

To determine the prediction accuracy, a dataset consisting of five records representing reported black pepper yield and quality from 2018 to 2022 was selected. The parameters chosen as input features for the LSTM model are depicted in Tables 2 and 3.

TABLE II
PREDICTION DATASET FOR DETERMINING THE BACK PEPPER YIELD

Year	Input features			Black Pepper Yield (kg/ha)
	Annual Rainfall (mm)	Annual Rainy Days	Rainfall soon after the Drought (mm)	
2018	1900.4	139	1391.2	932
2019	1359	121	910.8	782
2020	1541.8	115	995.9	820
2021	2439.3	138	1556.8	1320
2022	2147.6	121	1517.5	1185

The results, as shown in Table 4, indicate that the model performs reasonably well in predicting both black pepper yield and quality. For black pepper yield, the model

ABLE III
PREDICTION DATASET for DETERMINING THE BACK PEPPER QUALITY

Year	Input features				Black Pepper Quality (g/l)
	Annual Rainfall (mm)	Annual Rainy Days	Drought Span	Rainfall soon after the Drought (mm)	
2018	1900.4	139	23	1391.2	561
2019	1359	121	21	910.8	538
2020	1541.8	115	23	995.9	552
2021	2439.3	138	27	1556.8	582
2022	2147.6	121	16	1517.5	554

Table 4 lists the prediction results for the test records provided in Tables 2 and 3.

TABLE IV
LSTM MODELS' PERFORMANCE EVALUATION

Year	Black Pepper yield (kg/ha)		Black Pepper quality (g/l)	
	Actual	Predicted	Actual	Predicted
2018	932	1000	561	550.66
2019	782	823.43	538	547.43
2020	820	1049.10	552	549.56
2021	1320	1280.57	582	558.43
2022	1185	1180.5	554	554.30

predictions closely matched the actual values, particularly in the years 2021 and 2022, where the differences were minimal (e.g., 1320 kg/ha actual vs. 1280.57 kg/ha predicted in 2021 and 1185 kg/ha actual vs. 1180.5 kg/ha predicted in 2022). However, the model showed a larger discrepancy in 2020, with the predicted yield being significantly higher than the actual value (1049.10 kg/ha predicted vs. 820 kg/ha actual). In terms of black pepper quality, the model demonstrated high accuracy, with the predictions for 2020 and 2022 almost identical to the actual values (552 g/l actual vs. 549.56 g/l predicted in 2020 and 554 g/l actual vs. 554.30 g/l predicted in 2022). These results highlight the model's robust performance and its potential utility in agricultural forecasting, despite some variability in prediction accuracy across different years.

The predictive capabilities of the LSTM model in assessing the quality of black pepper were rigorously evaluated, demonstrating noteworthy accuracy across multiple years. The model consistently produced reliable estimates closely aligning with actual measurements, particularly in the later years of the study. For instance, in 2022, the predicted quality was 554.30 g/l, almost indistinguishable from the actual quality of 554 g/l, showcasing the model's precision. Furthermore, the model effectively captured slight fluctuations in quality metrics, as evidenced by its predictions in 2020 and 2021, which closely approximated the true values. This level of accuracy underscores the model's potential as a valuable tool for stakeholders in the agricultural sector, providing them with critical insights that can help optimize cultivation practices and enhance product quality.

Although the model provides acceptable prediction results, it also has some limitations. Currently, it does not consider additional agricultural data sources such as soil characteristics, farming practices, and crop management techniques. Incorporating these sources could yield a more comprehensive understanding of the factors affecting black pepper yield and quality. The study also faces challenges due to the difficulty of collecting 30-year weather data from various districts. However, incorporating extensive weather data and pepper productivity records from other districts could enhance the model's accuracy in predicting outcomes under different weather conditions and across various pepper-growing regions. These limitations highlight areas for further research and development in the field of black pepper yield and quality prediction.

V. CONCLUSION AND FUTURE WORK

This study successfully leveraged LSTM models to predict black pepper yield and quality in the Matale district of Sri Lanka, demonstrating the potent application of machine learning in agricultural forecasting. The models achieved notable prediction accuracy, with the yield and quality predictions closely aligning with actual yields and quality, particularly in recent years. Utilizing a detailed 30-year weather dataset, the models provided robust predictions with a mean absolute error of 18-20% for quality and consistently reliable performance across various assessments. These findings underscore the LSTM's effectiveness in capturing complex temporal relationships between weather conditions and crop productivity, offering valuable tools for farmers and policymakers to enhance agricultural planning and response strategies. The study's significance lies in its potential to transform agricultural practices by enabling more precise and informed decision-making, ultimately contributing to sustainable farming practices and economic stability in regions dependent on agriculture. Despite the promising outcomes, the research identifies avenues for future enhancement, particularly in broadening the dataset to include additional agricultural factors such as soil conditions and crop management practices. Future work could also expand the geographical scope of the study to validate the model's effectiveness across different climatic zones, thereby refining its applicability and accuracy in national agricultural contexts.

ACKNOWLEDGEMENT

The authors would like to extend their sincere gratitude to the Central Research Station of the Export Agriculture Department in Matale and the Natural Resource Management Center (NRMC) of the Agriculture Department, for providing the dataset for this study.

REFERENCES

[1] The International Pepper Community, *Production and Processing of Pepper (P.nigrum L.)*. Jakartha, Indonesia: The International Pepper Community, 2016.

[2] P. N. Ravindran, *Black Pepper Piper nigrum*, 1st edition. London, 2000.

[3] "(PDF) CLIMATE CHANGE AND BLACK PEPPER PRODUCTION," ResearchGate. https://www.researchgate.net/publication/282286146_CLIMATE_CHANGE_AND_BLACK_PEPPE R_PRODUCTION

[4] K. Krishnamurthy, K. Kandiannan, C. Sibin, C. Bhageerathy, and B. Chempakam, "Trends in climate and productivity and relationship between climatic variables and productivity in black pepper (Piper nigrum)," *The Indian Journal of Agricultural Sciences 81(8)*, pp. 729–733, Aug. 2011, Available: https://www.researchgate.net/publication/267963994_Trends_in_climate_and_productivity_and_relationship_between_climatic_varia bles_and_productivity_in_black_pepper_Piper_nigrum

[5] A.Crane-Droesch, "Machine learning methods for crop yield prediction and climate change impact assessment in agriculture," *Environmental Research Letters*, vol. 13, no. 11, p. 114003, Oct. 2018, doi: https://doi.org/10.1088/1748-9326/aae159.

[6] C.S. Karthik, D.K. Ghosh (Lkn), Saon Banerjee, Apurba Bandyopadhyay, And P. Dinesh Kumar, "Black pepper (Piper nigrum L.) yield estimates using crop weather model," Journal of Agrometeorology, vol. 22, no. 3, pp. 372–376, Nov. 2021, doi: https://doi.org/10.54386/jam.v22i3.301

[7] R. Meenal, P. A. Michael, D. Pamela, and E. Rajasekaran, "Weather prediction using random forest machine learning model," Indonesian Journal of Electrical Engineering and Computer Science, vol. 22, no. 2, p. 1208, May 2021, doi: https://doi.org/10.11591/ijeecs.v22.i2.pp1208-1215.

[8] R. Cunha, B. M. Silva, and Marco, "A Scalable Machine Learning System for Pre-Season Agriculture Yield Forecast," Sep. 2018, doi: https://doi.org/10.1109/escience.2018.00131.

[9] J. Majumdar, S. Naraseeyappa, and S. Ankalaki, "Analysis of agriculture data using data mining techniques: application of big data," *Journal of Big Data*, vol. 4, no. 1, pp. 1–15, Jul. 2017, doi: https://doi.org/10.1186/s40537-017-0077-4.

[10] B.V.R.Punyawardena, *Rainfall and Agro-Ecological Zones of Sri Lanka*. Department of Export Agriculture , 2008, pp. 82–83.

[11] "Matale Climate by Month," *weather-and-climate*. https://weatherand-climate.com/average-monthly-RainfallTemperatureSunshine,matale-matale-district-lk,Sri-Lanka (accessed May 2023).

[12] Shakib Badarpura, Abhishek Jain, Aniket Gupta, and Deepali Patil, "Rainfall Prediction using Linear approach Neural Networks and Crop Recommendation based on Decision Tree," International Journal of Engineering Research and, vol. V9, no. 04, Apr. 2020, doi: https://doi.org/10.17577/ijertv9is040314.

[13] Jasmin Praful Bharadiya, Nikolaos Tzenios Tzenios, and M. Reddy, "Forecasting of Crop Yield using Remote Sensing Data, Agrarian Factors and Machine Learning Approaches," vol. 24, no. 12, pp. 29– 44, Apr. 2023, doi: https://doi.org/10.9734/jerr/2023/v24i12858.

[14] J. H. Jeong *et al.*, "Random Forests for Global and Regional Crop Yield Predictions," *PLOS ONE*, vol. 11, no. 6, p. e0156571, Jun. 2016, doi: https://doi.org/10.1371/journal.pone.0156571.

[15] K. Palanivel and C. Surianarayanan, "an approach for prediction of crop yield using machine learning and big data techniques," international journal of computer engineering and TECHNOLOGY, vol. 10, no. 3, Jun. 2019, doi: https://doi.org/10.34218/ijcet.10.3.2019.013.

[16] P. G. Oguntunde, G. Lischeid, and O. Dietrich, "Relationship between rice yield and climate variables in southwest Nigeria using multiple linear regression and support vector machine analysis," International Journal of Biometeorology, vol. 62, no. 3, pp. 459–469, Oct. 2017, doi: https://doi.org/10.1007/s004840171454-6.

[17] P. Mohan and K. Patil, "Deep Learning Based Weighted SOM to Forecast Weather and Crop Prediction for Agriculture Application," International Journal of Intelligent Engineering and Systems, vol. 11, no. 4, pp. 167–176, Aug. 2018, doi: https://doi.org/10.22266/ijies2018.0831.17.

[18] S. Khaki and L. Wang, "Crop Yield Prediction Using Deep Neural Networks," *Frontiers in Plant Science*, vol. 10, May 2019, doi: https://doi.org/10.3389/fpls.2019.00621.

[19] S. Roy, K. Pramod, and Maurya, "Temperature Prediction Using Various Machine Learning Algorithms," vol. 10, pp. 2320–2882, 2022,[Online]. Available: https://ijcrt.org/papers/IJCRT2204008.pdf

[20] You, X. Li, M. Low, D. Lobell, and S. Ermon, "Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 31, no. 1, Feb. 2017, doi: https://doi.org/10.1609/aaai.v31i1.11172.