The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM)
ISSN: 2602-3199

The Eurasia Proceedings of Science, Technology, Engineering and Mathematics (EPSTEM), 2025

Volume 38, Pages 861-870

IConTES 2025: International Conference on Technology, Engineering and Science

Deep Learning-Based Crop Recommendation Using Soil and Environmental Parameters

Fatma Ozge Ozkok
Erciyes University
FOMOTECH R&D

Abstract: Crop identification is one of the priorities for increasing the productivity of agriculture, as well as the economic stability of the crops. Most farmers have traditional knowledge along with their experience, which leads to ineffective crop identification. In this context, this paper introduces a proper crop identification technique using deep learning techniques. The proposed system uses a Long Short-Term Memory network to determine the best crops according to the soil and environmental properties. The system is intended to be developed in two significant steps. At the beginning, the ANOVA test is used to determine the effect of the input variables on the system. Irrelevant variables reduce the efficiency of the system. Based on the test, the dominating properties of the crops, according to the environment, have been recognised as potassium, phosphorus, nitrogen, water, and rainfall. Then, in the second stage, the system is optimised based on several experiments involving variations in the number of layers, number of neurons, as well as the training epochs. The proposed approach provides promising insights for the development of an efficient decision-support system that assists the farmer in the selection of suitable crops.

Keywords: Deep learning, Crop recommendation system, Soil and environmental analysis, Precision agriculture, Decision support system

Introduction

Agriculture is a fundamental source of income around the world. However, many farmers face economic difficulties and falling crop yields due to poor crop selection, changing weather patterns and soil characteristics. Especially small and medium-sized farmers suffer financial loss due to a lack of knowledge and improper crop selection. This situation not only lowers farmers' income but also threatens the long-term sustainability of agriculture. Poor planning and unsuitable crop selection endanger future food security by reducing the availability of fruits and vegetables. In recent years, agriculture has increasingly relied on IoT devices, sensors, UAVs, and satellites. Data obtained from the devices include soil parameters, weather conditions, temperature, humidity information, and plant and soil images. The data provides important information on the monitoring of agricultural lands. However, the increasing amount of data makes analysis more difficult.

Deep learning methods using multi-layered filters can analyse datasets of various sizes and attributes. The methods using historical and daily or novel data can recommend new insights. This allows the system to extract patterns and relationships in the data. Although deep learning has a wide range of applications, it provides particularly valuable applications in the agricultural sector. For example, monitoring crop health, yield prediction, disease detection, irrigation optimisation and crop recommendation based on agricultural land conditions.

The crop recommendation systems assist farmers in choosing the appropriate crop based on soil and climate conditions. These conditions can include soil nutrient content, pH, moisture, and environmental parameters.

Deep learning systems using the data can determine the appropriate crop based on historical and real-time data. This can increase income and also help the sustainability of farming.

The study aims to present a plant recommendation system using environmental data, such as temperature, humidity, relative humidity, rainfall, and soil data, such as Nitrogen content rate in the soil, Phosphorus content rate in the soil, Potassium content rate in the soil, and pH value of the soil. The system uses deep learning architectures to suggest plants that are suitable for the land. The results were evaluated by comparing different deep learning architectures.

This study consists of five sections. Section 2 is a literature review. Section 3 describes the deep learning methods, such as LSTM, and the datasets used in this study. Section 4 presents the experimental results and a discussion of the model's performance. Section 5 summarises the conclusions and recommendations of the study.

Related Works

Crop selection recommendations form one of the most important processes involved in yield enhancement in agriculture. Choosing the right crop leads to superior yield and better crop quality. While traditional field practices have been used for crop selection, current technological advancements have introduced automatic processes in which field details, along with climate information, are used to determine the appropriate crop. These automatic techniques involve data mining algorithms such as SVM, Random Forest, single-layer or few-layer artificial neural networks, or deep learning models consisting of multiple layers.

In one of the data mining studies, the soil of the region around Kasur in Pakistan was studied. This is because agriculture is the primary source of livelihood for 65% of the population in that region. Factors like pH, soil type, electrical conductivity (EC), and potassium content were analyzed in that research. Among the various evaluated algorithms: OneRule (OneR), C4.5 decision tree, Naive Bayes, and Best First Tree (BF), the highest success percentage of 97.63% was demonstrated by the Naive Bayes algorithm (Arooj et al., 2018). In another study, a dataset of 2,200 samples combining rainfall, climate, and fertilizer data from India was used. Rule-based algorithms such as PART, Decision Table, and JRip were tested to provide optimal crop recommendations, with PART achieving 99.4% accuracy and 98.33% precision (Parameswari et al., 2021). A group of machine learning models was executed, with the results of their classification accuracies ranging between 90.13% and 99.93%. Among them, the Random Forest algorithm turned out to be much better compared to the other methods and gave the overall highest accuracy (Ashwitha & Latha, 2022).

A study using ANN has developed AI-based product recommendation systems that utilize sensor data such as soil moisture, temperature, and humidity. Artificial neural networks are used to recommend the most suitable products, increasing efficiency in both product selection and automation applications such as smart irrigation (Banavlikar et al., 2018). In another study, an ANN-based crop recommendation system was developed using soil and climate data from the Hadonahalli and Durgenahalli districts of Karnataka, India. The system evaluated crop suitability and achieved 97% accuracy, outperforming a decision tree classifier (92%) (Madhuri & Indiramma, 2021). In another research work, the IoT and AI-based smart agriculture was proposed. Real-time data on temperature, humidity, soil moisture, pH, and water level sensors are utilized by the MLP algorithm to predict the right type of crop. The architecture can automatically control irrigation, and users can also monitor the conditions of a web interface (Jagadeeswari & Manikandababu, 2022).

A crop recommendation framework is presented by combining the metaheuristic algorithm, Improved Distribution-based Chicken Swarm Optimization, and the deep learning algorithm Weight-based Long Short-Term Memory (LSTM) and the IDCSO-WLSTM algorithm. The proposed algorithm is found to be better compared to ANN in terms of recall, precision, accuracy, and execution time (Kiruthika and Karthika, 2025). Majumdar et al. also developed a novel hybrid algorithm called DSSA-AWLSTM. In this algorithm, the Dynamic Salp Swarm Algorithm (DSSA) is employed to extract the most relevant features, while an Adaptive Weighting Long Short-Term Memory Network Ensemble (AWLSTM) is used to determine the most suitable crop from the public crop recommendation dataset (Majumdar et al., 2025). Vani and Guruprakash proposed a crop recognition system named REALNET–LSTM. This method first applied statistical techniques for feature extraction, then used ResNet and its output LSTM in the encoder, followed by a fully connected layer in the decoder. The method demonstrated superior performance compared to traditional data mining methods and ANN (Vani and Guruprakash, 2025).

It has been observed from these studies that data mining, ANN, and deep learning techniques are becoming increasingly effective for crop recommendation system applications. Although conventional techniques are useful, computational techniques always offer better accuracy, precision, and efficiency. Recent studies confirm that deep learning techniques are more effective for crop recommendation tasks than conventional data mining and ANN techniques.

Method

Dataset and Proposed Method

In this section, the dataset and the LSTM model used are explained.

Dataset

The dataset was provided by Indian Chamber of Food and Agriculture for precision agriculture (ICFA,2020). It contains data on the ratio of Nitrogen, phosphorus, and Potassium content in soil, the pH value of soil, temperature, humidity, and rainfall, along with the label for each data entry. It contains 2200 rows and 8 columns with data for 22 classes, which are apple, banana, black gram, chickpea, coconut, coffee, cotton, grapes, jute, kidney beans, lentil, maize, mango, muskmelon, moth beans, papaya, pomegranate, rice, watermelon, and orange. The above distribution is presented in Fig. 1.

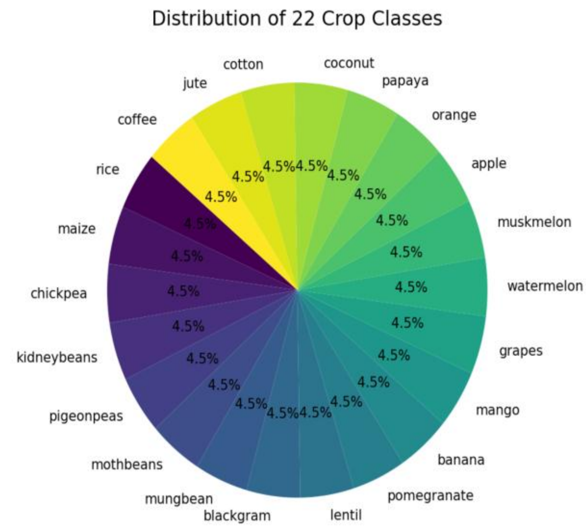

Figure 1. Distribution of crop classes in the dataset

LSTM

The Long Short-Term Memory, or LSTM, refers to a sophisticated form of RNN that has been developed with the purpose of finding and tracing patterns in data that span a long period. However, due to its unique structure, which has a memory cell, the LSTM has the capability to retain data for a relatively long period. This contradicts a normal RNN, where the gradient either vanishes or diverges with each step of the training process, resulting in either losing memory or losing the gradient. The main gates within each unit necessary during the processing of decision-making of which data to store, update, and delete are referred to as forget gates, input gates, and output gates. This form of technology can be applied where data necessary through time or a sequence is involved (Chauhan et al., 2024; Yadav & Thakkar, 2024; Orosoo, 2025).

The overall structure of an LSTM cell is represented in Figure 2. Every LSTM cell contains the previous state of the cell, denoted by C_{t-1}, and the current state, denoted by C_t. The current input to the cell is given by X_t, along with the previous hidden state h_{t-1} are combined and passed to three gates: the forget gate, the input gate, and the output gate. The forget gate decides on which parts of Ct-1 are forgotten, whereas the input gate decides on what is added on top of C_t. These two gates function simultaneously in an operation that results in the combination of previous information with new information, thus an update on C_t. Finally, the output gate decides which parts of

C_t will affect h_t in that given time step. It then passes this information on to the next step. The capability described above enables the LSTM network to learn dependencies that are long and temporal patterns existing in the data (Zhou et al., 2024).

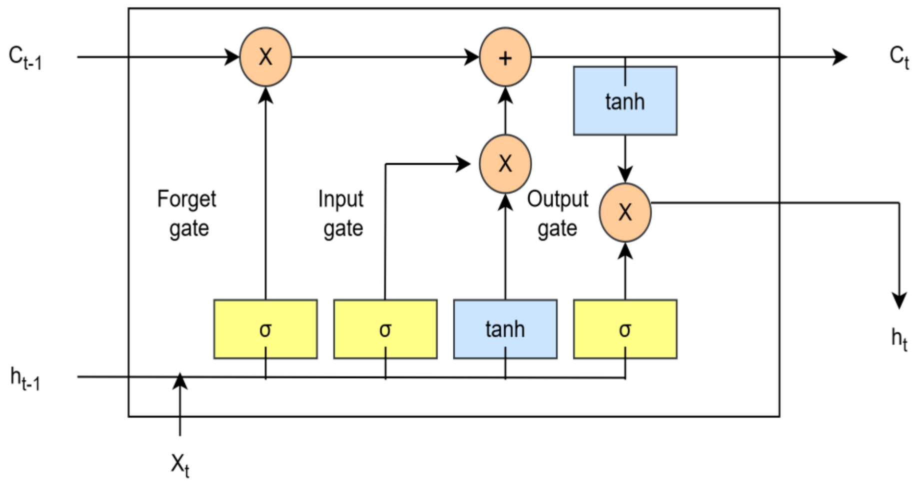

Figure 2. The structure of an LSTM unit (Zhou et al., 2024).

Even though it cannot be regarded as a common time series from the given dataset, it has the capacity to handle complex nonlinear associations between the variables.

Results and Discussion

This section presents and discusses the results obtained from the experiments. Firstly, an F-test analysis was carried out to determine the most effective components of the data that contributed to the enhancement of the model. Secondly, the section witnessed the performance of some experiments with the goal of trying to enhance the Long Short-Term Memory network structure by either layer or neuron modification. These experiments and their comparisons are discussed in detail.

ANOVA F-Test Results

To evaluate the statistical significance and usefulness of variables in a given dataset, an F-test analysis of ANOVA was performed before constructing the proposed LSTM model.

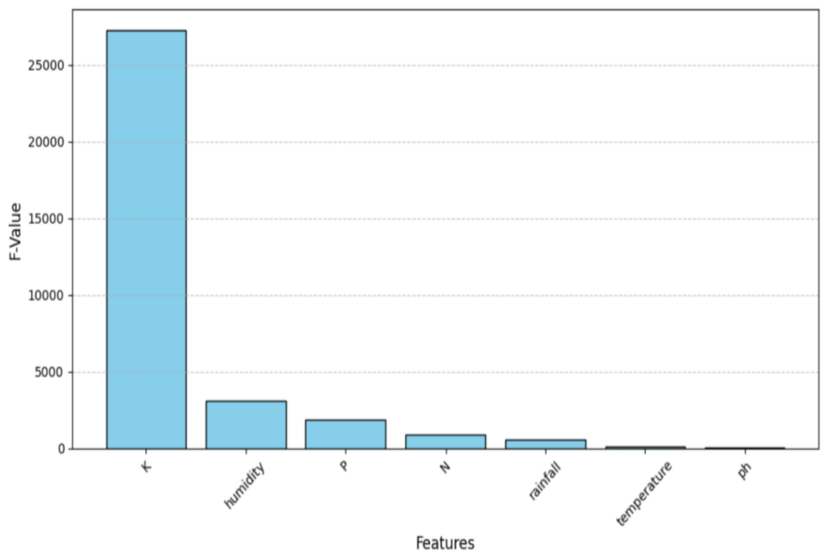

Figure 3. Feature importance based on ANOVA F-test

The results of the ANOVA F-test are presented in the figure based on the calculated F-values. The ANOVA F-test results showed F-values of 60.344 for pH, 102.187 for temperature, 605.528 for rainfall, 897.568 for nitrogen (N), 1885.658 for phosphorus (P), 3103.709 for humidity, and 27238.362 for potassium (K). The F-value in ANOVA indicates the ratio of variance between groups to the variance within groups. A higher F-value represents a stronger influence of that variable on the dependent variable, implying that differences in that feature contribute more significantly to variations in the model output. Therefore, potassium (K), phosphorus (P), nitrogen (N), humidity, and rainfall, which had the highest F-values, were selected as input features.

LSTM Architecture Optimisation

In this section, three experiments were conducted to investigate the effects of the number of LSTM layers, the number of neurons per layer, and the number of epochs on model performance. In addition, the classification performance for different crop types was evaluated to show how well the model predicted each class. In the first experiment, the number of LSTM layers was varied to analyse its impact while keeping the other hyperparameters constant. The fixed hyperparameters were as follows: the learning rate was set to 0.001, the batch size was 16, the dropout rate was 0.2, the number of epochs was 20, and each layer contained 32 units. The number of LSTM layers was changed from one to five, and the obtained training results are presented in Figure 4, while the test accuracies for each model are shown in Figure 5.

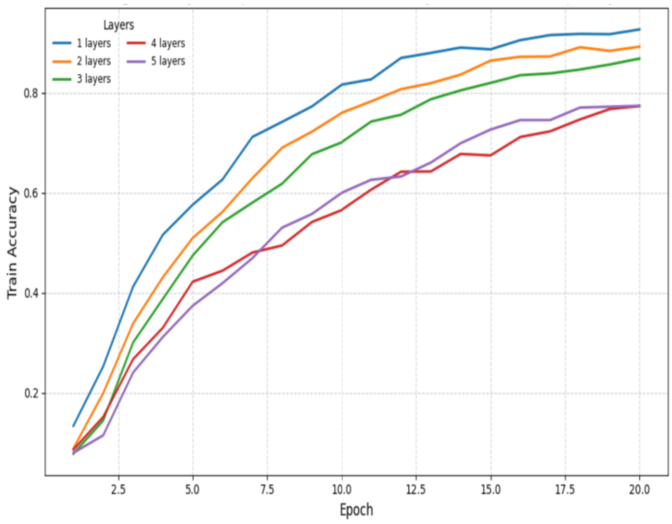

Figure 4. Train accuracy

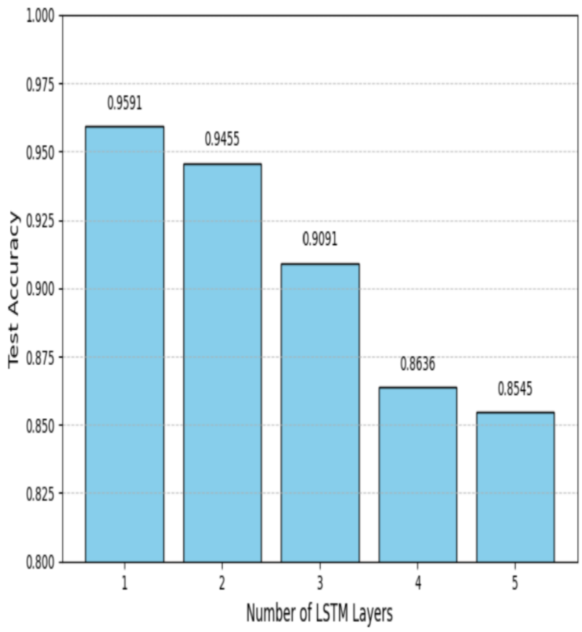

Figure 5. Test accuracy

As shown in Figure 4, the per-epoch training curves indicate that the one-layer LSTM reaches the highest accuracy among the tested configurations. Figure 5 presents the test accuracies, which are 0.9591 for one layer, 0.9455 for two layers, 0.9091 for three layers, 0.8636 for four layers, and 0.8545 for five layers. Taken together, these results show that the one-layer LSTM provides the best overall performance; therefore, the subsequent experiments were conducted using a one-layer LSTM architecture.

In the second experiment, the number of neurons per layer was varied to analyse its impact while keeping the other hyperparameters constant. The fixed hyperparameters were as follows: the learning rate was set to 0.001, the batch size was 16, the dropout rate was 0.2, the number of epochs was 20, and the number of LSTM layers was fixed at one. The number of neurons per layer was changed to 16, 32, 64, 96, and 128, and the obtained training results are presented in Figure 6, while the test accuracies for each model are shown in Figure 7.

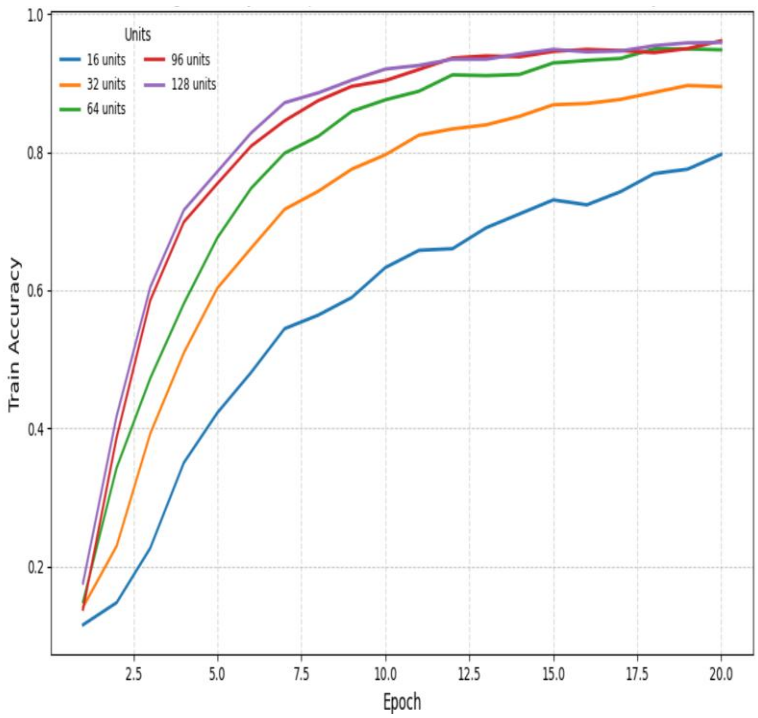

Figure 6. Train accuracy

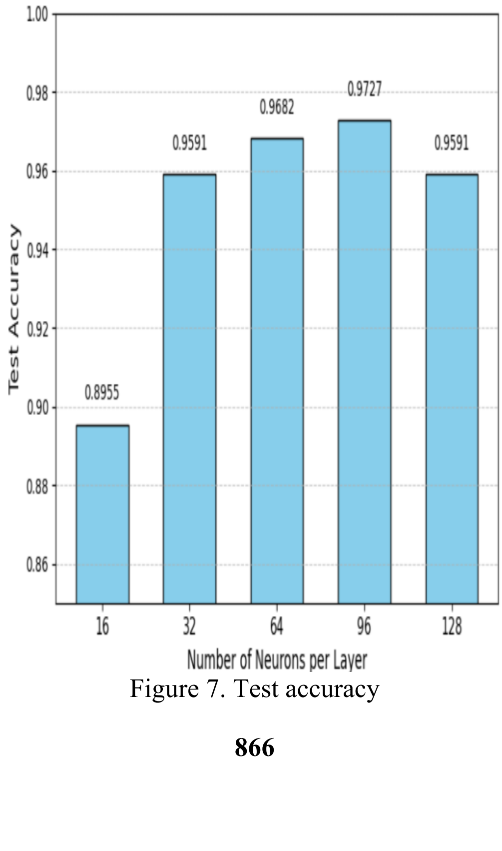

Figure 7. Test accuracy

As shown in Figure 6, the per-epoch training curves demonstrate that the model with 96 neurons per layer achieves the highest training accuracy among the tested configurations. Figure 7 presents the test accuracies, which are 0.8955 for 16 neurons, 0.9591 for 32 neurons, 0.9682 for 64 neurons, 0.9727 for 96 neurons, and 0.9591 for 128 neurons. These results clearly indicate that the LSTM model with 96 neurons provides the best overall performance; therefore, this configuration was selected for the following experiment.

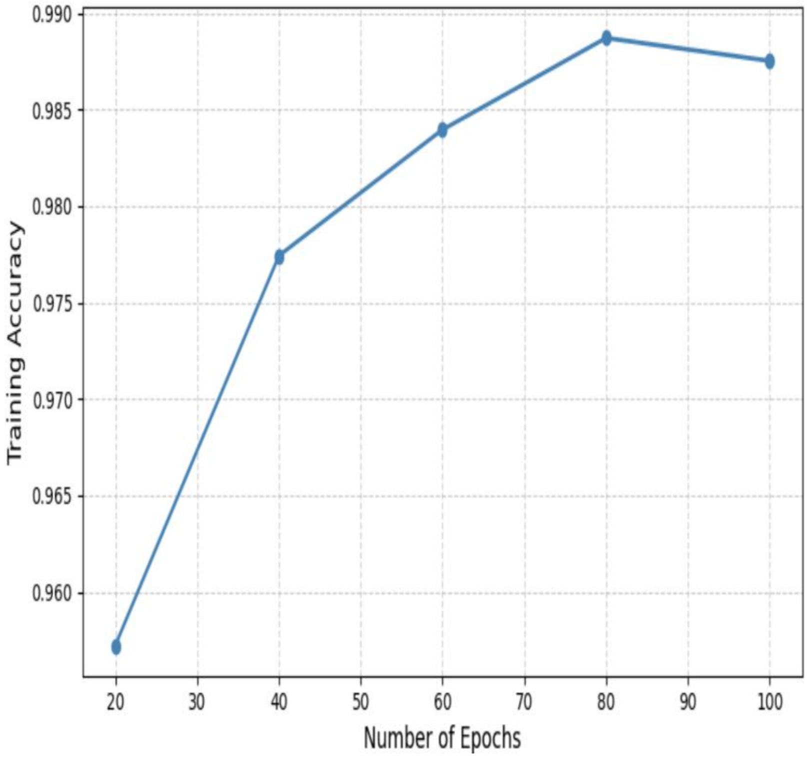

Figure 8. Train accuracy

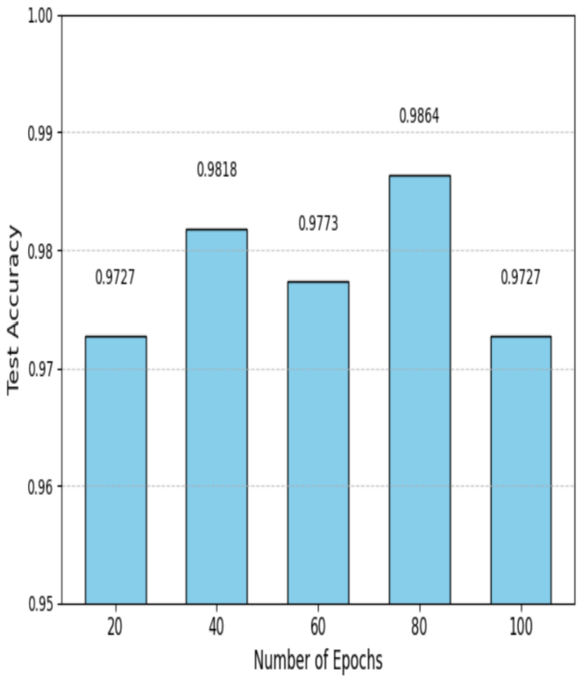

Figure 9. Test accuracy

In the third experiment, the number of epochs was varied to analyse its impact while keeping the other hyperparameters constant. The fixed hyperparameters were as follows: the learning rate was set to 0.001, the batch size was 16, the dropout rate was 0.2, the number of neurons per layer was 96, and the number of LSTM layers was fixed at one. The number of epochs was adjusted to 20, 40, 60, 80, and 100. The corresponding training results are illustrated in Figure 8, and the test accuracies for each configuration are presented in Figure 9.

As shown in Figure 8, the per-epoch training curves demonstrate that the model trained for 80 epochs achieves the highest training accuracy among the tested configurations. Figure 9 presents the test accuracies, which are 0.9727 for 20 epochs, 0.9818 for 40 epochs, 0.9773 for 60 epochs, 0.9864 for 80 epochs, and 0.9727 for 100 epochs. These results clearly indicate that training the LSTM model for 80 epochs provides the best overall performance; therefore, this configuration was used in the final evaluation and confusion matrix analysis.

In the final experiment, the model was trained using the optimal parameters obtained from the previous experiments. The results, illustrated in Figure 10, show that most crop classes were correctly identified by the model. Specifically, three rice samples and one coffee sample were misclassified as jute, while one jute sample was incorrectly predicted as rice. Apart from these minor errors, the model accurately recognized 19 out of 22 crop types, demonstrating strong classification performance and generalization capability.

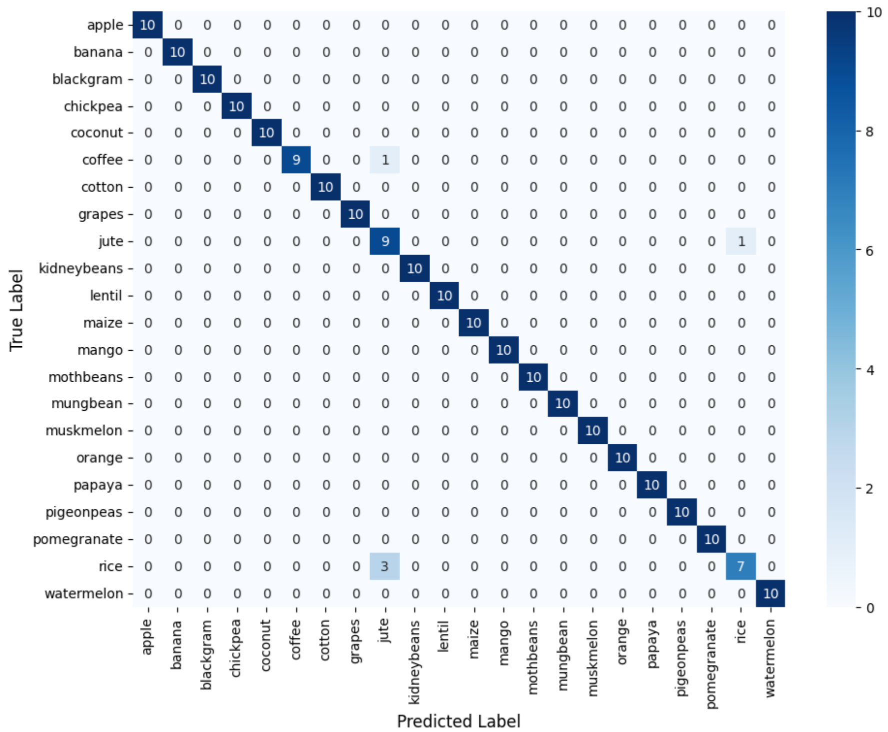

Figure 10. The confusion matrix of the proposed model

Conclusion

The suitable selection of crops is one of the most important processes in agriculture. Therefore, the proposed system is a crop-recommending model which is based on LSTM and soil and environmental data. Initially, the significant factors influencing crop growth were determined using the ANOVA F-test. The result demonstrates that potassium, phosphorus, nitrogen, humidity, and rainfall are dominant in the data. Later, a variety of experiments were carried out to choose the best LSTM model parameters. The one-layer LSTM model, with 96 neurons and 80 epochs, achieved the highest accuracy. These results show that the proposed LSTM-based system is effective for data-driven crop recommendation.

Recommendations

In this study, an LSTM-based deep learning model was used to predict suitable crops based on environmental and soil parameters. Although the LSTM architecture demonstrated strong performance, future research could explore alternative recurrent neural network models, such as GRU (Gated Recurrent Unit) or BiLSTM (Bidirectional LSTM) to further improve prediction accuracy. Additionally, attention mechanisms or hybrid models that combine LSTM with convolutional layers could enhance the model's ability to capture spatial and temporal dependencies in the data. Future studies could also utilise larger and more diverse datasets, as well as real-time data collection methods, to improve the generalisation and robustness of crop recommendation systems.

Scientific Ethics Declaration

* The author declares that the scientific ethical and legal responsibility of this article published in EPSTEM journal belongs to the author.

Conflict of Interest

* The authors declare that they have no conflicts of interest

Funding

*This publication is supported by the EU COST (European Cooperation in Science and Technology) Action CA22136 "Pan-European Network of Green Deal Agriculture and Forestry Earth Observation Science" (PANGEOS)

Acknowledgements or Notes

* This article was presented as an oral presentation at the International Conference on Technology, Engineering and Science (www.icontes.net) held in Antalya/Türkiye on November 12-15, 2025.

References

Arooj, A., Riaz, M., & Akram, M. N. (2018, February). Evaluation of predictive data mining algorithms in soil data classification for optimized crop recommendation. In *2018 International Conference on Advancements in Computational Sciences (ICACS)* (pp. 1-6). IEEE.

Ashwitha, A., & Latha, C. A. (2022). Crop recommendation and yield estimation using machine learning. *Journal of Mobile Multimedia*, *18*(3), 861-883

Banavlikar, T., Mahir, A., Budukh, M., & Dhodapkar, S. (2018). Crop recommendation system using Neural Networks. *International Research Journal of Engineering and Technology (IRJET)*, *5*(5), 1475-1480.

Chauhan, A., Kukkar, Y., Nagrath, P., Gupta, K., & Hemanth, J. D. (2024). Integrating LSTM and NLP techniques for essay generation. *Intelligent Decision Technologies*, *18*(1), 571-584.

ICFA (Indian Chamber of Food and Agriculture), 2020, https://www.kaggle.com/datasets/siddharthss/crop-recommendation-dataset/data, Accessed: 08.06.2025

Jagadeeswari, M., & Manikandababu, C. S. (2022, September). Artificial Intelligence based Crop Recommendation System. In *2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA)* (pp. 1127-1133). IEEE.

Kiruthika, S., & Karthika, D. (2023). IOT-BASED professional crop recommendation system using a weight-based long-term memory approach. *Measurement: Sensors*, *27*, 100722.

Madhuri, J., & Indiramma, M. (2021). Artificial neural networks based integrated crop recommendation system using soil and climatic parameters. *Indian Journal of Science and Technology, 14*(19), 1587-1597.

Majumdar, P., Mitra, S., & De, D. (2025). An effective crop recommendation system using a dynamic salp swarm algorithm with adaptive weighting based LSTM network. *SN Computer Science*, *6*(5), 549.

Orosoo, M., Raash, N., Treve, M., Lahza, H. F. M., Alshammry, N., Ramesh, J. V. N., & Rengarajan, M. (2025). Transforming English language learning: Advanced speech recognition with MLP-LSTM for personalized education. *Alexandria Engineering Journal*, *111*, 21-32.

Parameswari, P., Rajathi, N., & Harshanaa, K. J. (2021, October). Machine learning approaches for crop recommendation. In *2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)* (pp. 1-5). IEEE.

Vani, B. V., & Guruprakash, C. D. (2025). Representational alignment net-based long short-term memory model for crop recommendation system. *Proceedings of the Indian National Science Academy*, 1-20.

Yadav, H., & Thakkar, A. (2024). NOA-LSTM: An efficient LSTM cell architecture for time series forecasting. *Expert Systems with Applications*, *238*, 122333.

Zhou, M., Wang, L., Hu, F., Zhu, Z., Zhang, Q., Kong, W., ... & Cui, E. (2024). ISSA-LSTM: A new data-driven method of heat load forecasting for building air conditioning. *Energy and Buildings*, *321*, 114698.

Author(s) Information

Fatma Ozge Ozkok
Erciyes University
Department of Computer Engineering, Turkiye
FOMOTECH R&D, Erciyes Technopark, Turkiye
Contact e-mail: *fozgeozkok@erciyes.edu.tr*

To cite this article:

Ozkok, F.O (2025). Deep learning-based crop recommendation using soil and environmental parameters. *The Eurasia Proceedings of Science, Technology, Engineering and Mathematics (EPSTEM), 38,* 861-870.