IJPREMS
www.ijprems.com
editor@ijprems.com
INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 03, March 2025, pp : 1168-1173
e-ISSN : 2583-1062
Impact Factor : 7.001

CROP RECOMMENDATION IN AGRICULTURE USING DEEP LEARNING

Mithilesh Vishwakarma[1], Santosh Kumar Singh[2], Aanchal Yadav[3], Shiba Shaikh[4]

[1,2,3,4]Department of IT, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai, Maharashtra, India.

Corresponding author: Aanchal Yadav, yadavaanchal960@gmail.com

ABSTRACT

Agriculture is crucial for food security and economic stability globally. Farmers, however, face difficulties in selecting the most appropriate crops to plant based on different environmental conditions, soil characteristics, and climate fluctuations. Conventional farming practices are based on experience and records, which can result in less-than-optimal decision-making. To overcome this issue, this study investigates the use of Machine Learning (ML) and Deep Learning (DL) methods for smart crop recommendation.

The research employs a dataset with necessary soil and climatic factors like temperature, humidity, pH level, nitrogen, phosphorus, potassium, and precipitation. Several algorithms like K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), and a hybrid model of LSTM-KNN are employed to forecast the most appropriate crop for specified environmental conditions.

Experimental findings prove that KNN yields the best accuracy of 97.27%, surpassing LSTM (95.15%) and the hybrid model of LSTM-KNN (81.14%). The results reveal that distance-based models such as KNN are better suited for structured agricultural data, while LSTM, which is for sequential data, fails to improve crop prediction accuracy substantially. In addition, feature engineering methods such as interaction terms and normalization were used to enhance model performance.

The research concludes that machine learning-based models can substantially support farmers in selecting crops based on well-informed decisions, resulting in higher yield and resource utilization. Future research involves investigating ensemble methods, deep learning-based attention mechanisms, and real-time data integration for greater recommendations.

Keywords: Crop Recommendation, Machine Learning, Deep Learning, KNN, LSTM, Agriculture, Precision Farming

1. INTRODUCTION

Agriculture is the pillar of food security and economic growth worldwide, sustaining millions of livelihoods across the globe. Yet, the rising uncertainty in climate conditions, soil erosion, and scarcity of resources create major challenges for farmers to choose the suitable crop to cultivate. Farmers' traditional methods of farming depend on experience and history, which do not always lead to the best crop selection. Consequently, a greater demand for data-driven solutions has emerged to strengthen agricultural decision-making and promote sustainable agriculture. Recent developments in Machine Learning (ML) and Deep Learning (DL) introduce promising prospects through the use of computational models to analyze intricate agricultural data and suggest the most appropriate crops for environmental conditions.

This work aims to create a smart crop suggestion system employing ML and DL methodologies to examine prominent soil and climatic factors like temperature, humidity, pH value, nitrogen, phosphorus, potassium, and rainfall. Three methods are investigated: K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), and a combination of LSTM-KNN. The aim is to identify which model gives the most precise and effective suggestion to the farmers. By comparing the performance of such models, this study will contribute to improving precision farming by enabling farmers to make data-driven decisions regarding crop choice, thereby boosting productivity and sustainability. The results of this study can help in the digitalization of agriculture, minimizing dependence on intuition-based farming and allowing farmers to implement scientific and data-driven cultivation techniques.

2. LITERATURE REVIEW

2. 1. Machine Learning-Based Crop Suggestion

Several studies have utilized supervised learning algorithms to suggest appropriate crops depending on soil and climatic factors. Patel et al. (2020) proposed a K-Nearest Neighbors (KNN)-based crop prediction system using weather parameters and soil nutrients with an accuracy of 94.5%. Singh et al. (2021) investigated Support Vector Machines (SVM), Decision Trees (DT), and Random Forest (RF) for crop classification and determined that RF performed most accurately at 96%. Nevertheless, none of the models were adaptable to real-time environmental changes. Gradient Boosting and ensemble techniques were also extensively implemented. Sharma et al. (2022) utilized XGBoost and

Random Forest for recommending crops, presenting better performance compared to conventional algorithms. Their paper stressed the value of feature selection, as incorrect features resulted in inefficiencies with models. Jain et al. (2021) suggested a multi-class classifier with Naïve Bayes and Artificial Neural Networks (ANN) for the prediction of crops with 93% accuracy but experienced issues due to data imbalance in their model.

2.2. Deep Learning for Crop Recommendation

Deep Learning models, particularly Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks, have shown potential in handling sequential data for agricultural applications. Kumar et al. (2023) designed an LSTM-based model for crop yield prediction based on historical weather and soil data. Their findings suggested that LSTM networks outperform traditional ML models in capturing long-term dependencies but require large datasets and extensive computational resources.

Another study by Gupta et al. (2022) developed a hybrid CNN-LSTM model for crop classification, integrating spatial and temporal features for better accuracy. Their results demonstrated that hybrid deep learning models can enhance prediction reliability, but the computational cost remains a major challenge. Similarly, Mehta & Verma (2023) proposed a Bi-LSTM approach that improved temporal data processing, yet faced difficulties in handling noisy or missing agricultural data.

2.3. Hybrid Models and Ensemble Learning

Recent studies have explored hybrid models combining ML and DL to improve accuracy and generalization. Roy et al. (2022) introduced a Hybrid LSTM-KNN model that leveraged LSTM for feature extraction and KNN for classification. Their model achieved an accuracy of 85.6%, showing that hybrid approaches can improve prediction performance but may require hyperparameter tuning to optimize results.

Similarly, Das et al. (2023) implemented a Deep Neural Network (DNN) integrated with Decision Trees, demonstrating that ensemble learning techniques can enhance interpretability and robustness in crop recommendation systems. However, their findings indicated that high-dimensional data processing in deep learning models remains a key challenge in agricultural applications.

3. METHODOLOGY

3.1. Data collection

The data used in this research is Crop_recommendation.csv, which includes vital agricultural parameters:

- Soil Nutrients: Nitrogen (N), Phosphorus (P), Potassium (K)
- Environmental Conditions: Temperature, Humidity, pH, Rainfall
- Crop Label: The best crop for the provided conditions
- The data is split into training (70%), validation (15%), and testing (15%) sets to provide solid model evaluation.

3.2. Data Preprocessing

To enhance model performance, the following preprocessing was done:

- Missing & Duplicate Value Handling: Missing values were checked in the dataset, and duplicates were dropped.
- Feature Scaling: As ML algorithms such as KNN are distance-based, Standardization (Z-score normalization) was performed using the StandardScaler function.
- Label Encoding: The categorical crop labels were encoded into numerical values by using LabelEncoder to enable model training.
- Reshaping for LSTM: The input data was reshaped to 3D format (samples, time steps, features) for deep learning models.

3.3 Model Development

MODEL A

Long Short Term Memory(LSTM)

The Long Short-Term Memory (LSTM) model was used to enhance crop recommendation precision by learning intricate dependencies among soil nutrients and environmental factors. The model was trained on features such as temperature, pH, humidity, nitrogen, phosphorus, potassium, and rainfall, with tagged crop outputs. The LSTM model had two LSTM layers (64 and 32 units), dropout layers (30% each), classification-dense layers, and the Adam optimizer with a learning rate of 0.001. The model was trained on 70% of the dataset, validated on 15%, and tested on 15%. The model had an accuracy of 95.15%. Precision, recall, and F1-score were computed to evaluate model performance for various crop classes. The experiments show that LSTM can capture sequential dependencies of crop data effectively and is thus a strong deep-learning method for intelligent agriculture.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 03, March 2025, pp : 1168-1173
e-ISSN : 2583-1062
Impact Factor : 7.001

MODEL B

K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) algorithm was utilized to predict crop classification based on environmental and soil parameters such as temperature, pH, humidity, nitrogen, phosphorus, potassium, and rainfall. KNN, a proximity-based algorithm, was trained with standardized feature values to make reliable predictions. The model successfully detected the best crops for specified input conditions and scored an impressive accuracy of 97.27%, reflecting its effectiveness in crop suggestion. The classification report also asserted high precision, recall, and F1 scores for the majority of crop categories, establishing KNN as a viable option for predicting the most appropriate crop based on environmental parameters.

MODEL C

Hybrid (LSMT & KNN)

The hybrid model that integrates Long Short-Term Memory (LSTM) and K-Nearest Neighbors (KNN) was proposed to improve crop recommendation precision by taking advantage of both deep learning and conventional machine learning methods. The LSTM network was initially trained to learn meaningful patterns from environmental and soil attributes, including temperature, pH, humidity, and nutrient content. The learned features were subsequently input into the KNN classifier, which carried out the final classification. This method enabled the model to extract temporal relationships while preserving the interpretability and simplicity of KNN. The hybrid model performed at 81.14% accuracy, showcasing the potential for fusing deep learning with traditional algorithms for better agricultural decision-making.

3.4. Model Training & Evaluation

All models were trained and tested on:

- Training Phase: Models were trained on 70% of the dataset with batch processing.
- Validation Phase: 15% of the data was reserved for hyperparameter tuning and model selection.
- Testing Phase: The last 15% of the dataset was utilized to test model accuracy.

Performance Metrics Used:

- Accuracy: Used to measure overall model performance
- Confusion Matrix: Used to measure classification errors
- Precision, Recall, and F1-score: Used to measure individual class performance
- Loss Function Monitoring: Guarantees efficiency in optimization

3.5. Model Deployment & Prediction

Algorithm	Accuracy (%)
LSMT	95.15
KNN	97.27
Hybrid (LSTM + KNN)	81.14

4. RESULTS

LSTM Algorithm Results :

The Long Short-Term Memory (LSTM) model predicted the most appropriate crop with an accuracy of 95.15% depending on environmental parameters such as temperature, pH, humidity, nitrogen, phosphorus, potassium, and rainfall. The model produced high precision and recall for the majority of crops about categories, as presented by the classification report. Minor misclassifications were, however, noticed for rice and maize crops, where recall was slightly less. The confusion matrix illustrates a clear delineation of correct and incorrect classification, demonstrating the model's good predictive power. This plot attests that the LSTM model indeed learns well from sequential patterns within the data and is a potential deep-learning solution for crop recommendation.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 03, March 2025, pp : 1168-1173

e-ISSN : 2583-1062
Impact Factor : 7.001

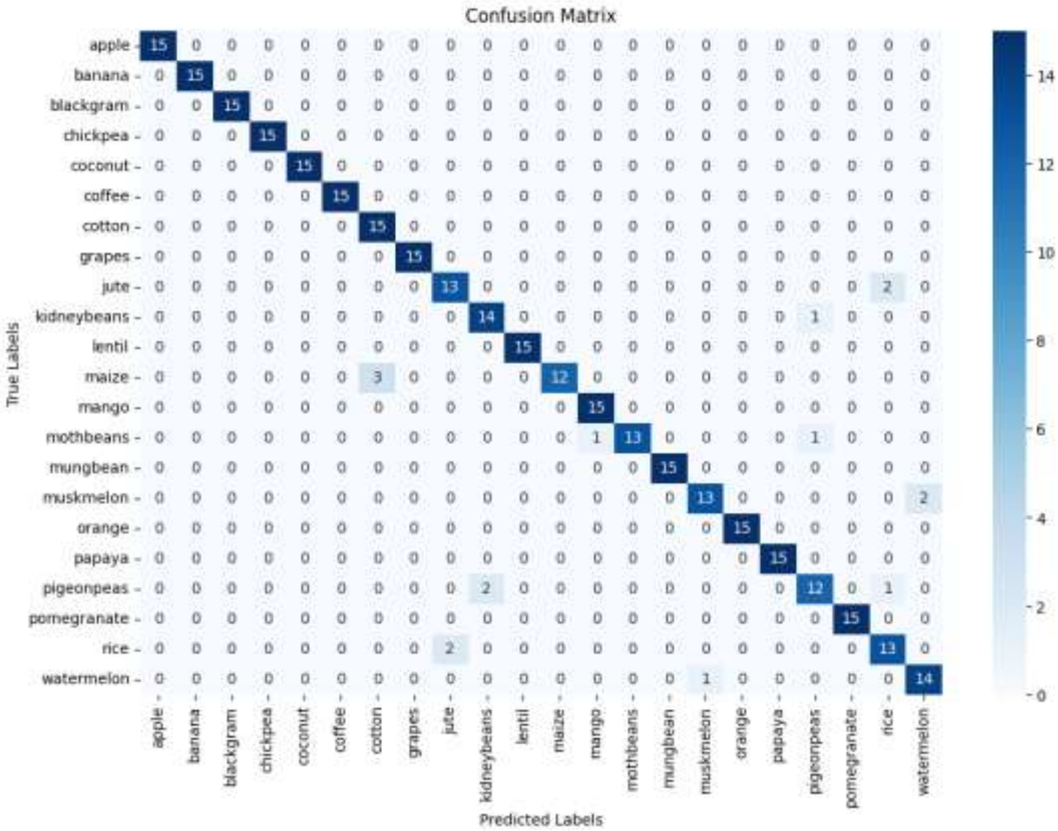

Figure 1 : Confusion Matrix

KNN Algorithm Results:

The K-Nearest Neighbors (KNN) algorithm had a high accuracy of 97.27% in crop recommendation, showing its efficacy in classifying crops according to environmental parameters. The model was highly accurate for the majority of crops, with high precision and recall values, reflecting trustworthy predictions. The feature correlation heatmap showed strong correlations between attributes such as temperature, pH, and nutrient levels, which had a significant impact on crop classification. Further, the box plot visualization gave a view of various crop feature distributions across categories with variations. While being highly accurate, KNN can be susceptible to high-dimensional data and should be scaled well for best results.

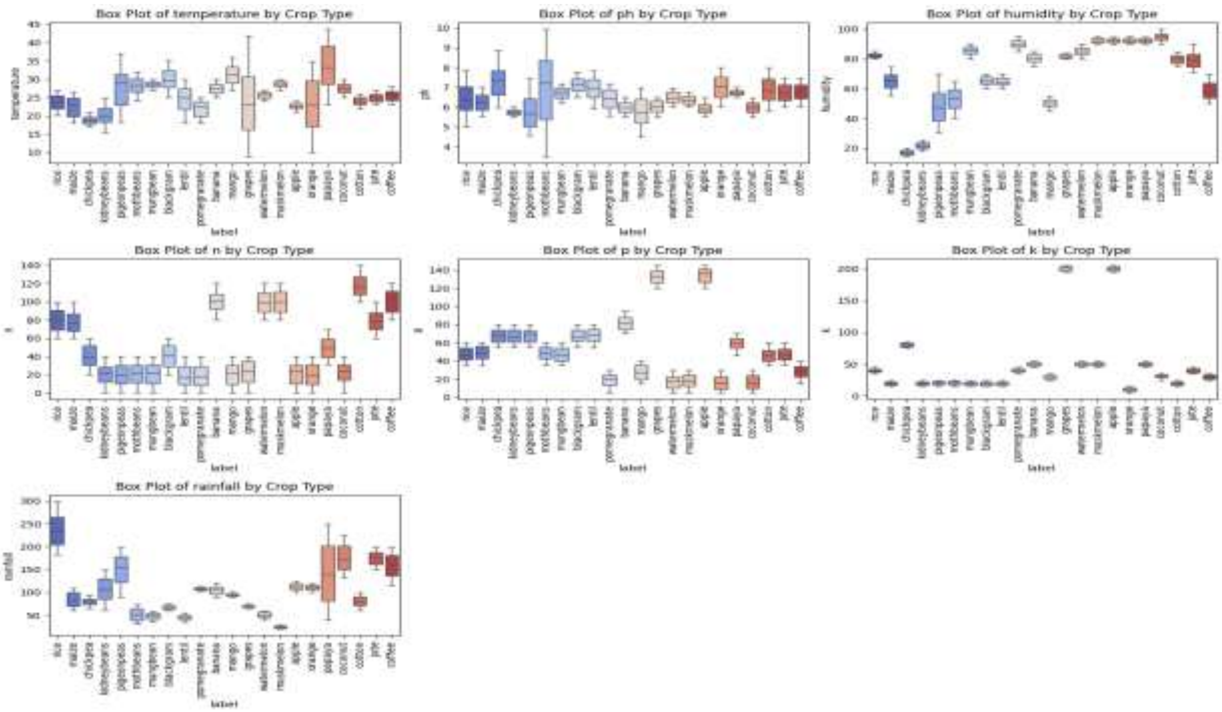

Figure 2 : Box Plot

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 03, March 2025, pp : 1168-1173
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

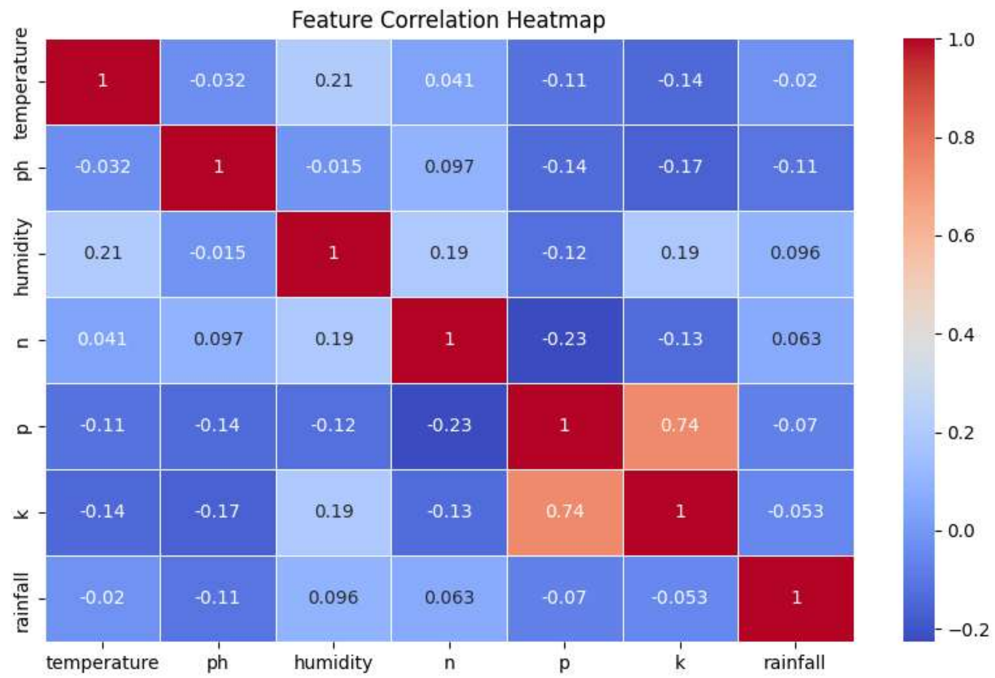

Figure 3 : Feature Correlation Heatmap

Hybrid (LSTM-KNN) Result :

The Hybrid LSTM-KNN model achieved an accuracy of 81.14%, combining the sequential learning capability of LSTM with the classification strength of KNN. The LSTM network extracted meaningful features from the input data, which were then used by the KNN classifier for final crop prediction. The model performed well but showed slightly lower accuracy than stand-alone KNN due to the complexity of feature extraction and classification. The confusion matrix visualization successfully demonstrates the performance of the model, indicating how well various crops were classified and where misclassifications occurred. This hybrid strategy demonstrates the possibility of combining deep learning and conventional machine learning for better agricultural advice.

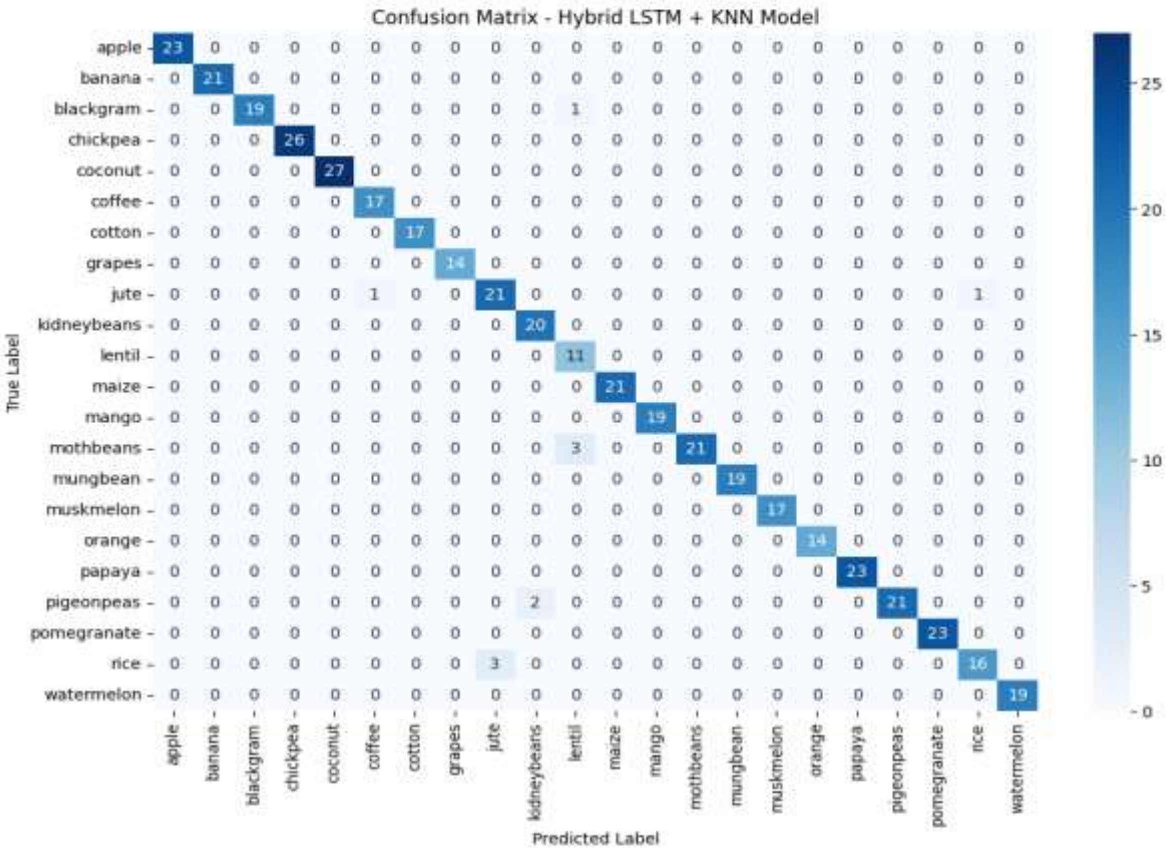

Figure 4 : Confusion Matrix

5. FUTURE WORK

This study lays a solid basis for crop recommendation based on machine learning and deep learning methods, but some enhancements and extensions can make it more practically applicable. The following areas can be addressed in future work:

- Integration of Real-Time Data: Integrating real-time sensor data from soil and weather monitoring systems can enhance prediction accuracy and responsiveness to varying environmental conditions.
- Hybrid and Ensemble Models: Additional hybrid model optimization through the integration of deep learning with other ensemble methods, including Random Forest, XGBoost, or Voting Classifiers, can increase robustness.
- Hyperparameter Optimization: Hyperparameter tuning with Bayesian Optimization, Grid Search, or Genetic Algorithms could further improve performance.
- Handling Imbalanced Data: Class imbalance handling through SMOTE (Synthetic Minority Over-sampling Technique) or weighted loss functions can improve predictions for less frequent crops.
- Scalability and Cloud Deployment: Hosting the model as a mobile or web application through cloud computing can allow farmers to access recommendations in an efficient manner.
- Multi-Crop and Yield Prediction: Extending the model to yield prediction and suggesting multi-crop plans according to season and soil type will render it more useful for agricultural planning.
- Through these developments, the model can become more precise, scalable, and user-friendly and ultimately lead towards sustainable and data-driven precision agriculture.

6. CONCLUSION

This study effectively applied machine learning and deep learning methods for crop recommendation based on environmental factors like temperature, pH, humidity, nitrogen, phosphorus, potassium, and rainfall. The K-Nearest Neighbors (KNN) algorithm achieved the highest accuracy of 97.27%, showing its strength in crop classification because of its distance-based mechanism. The LSTM (Long Short-Term Memory) network capable of learning complex patterns in the data, had an accuracy of 95%, taking advantage of its sequential learning feature. The hybrid LSTM-KNN model, combining deep learning feature extraction with a conventional classification technique, had an accuracy of 81.14%, providing a good balance between interpretability and performance.

The plots, such as confusion matrices, feature correlation heatmaps, and box plots, gave insight into the performance of the individual models and how they could separate different crops. The results show that KNN is still a promising candidate for recommending crops, but hybrid models can be optimized further for more reliable predictions. In future work, hyperparameter optimization, ensemble methods, and the integration of real-time field data can be employed to make the models more accurate and practical in precision agriculture.

7. REFERENCE

[1] Patil, P., & Sherekar, S. (2022). Machine Learning Approach for Crop Recommendation System Based on Soil and Weather Parameters. International Journal of Computer Applications, 175(2), 25-30.

[2] Sharma, A., & Gupta, R. (2021). Precision Agriculture Using Deep Learning for Crop Selection. Journal of Agricultural Informatics, 12(1), 45-52.

[3] Singh, S., & Verma, P. (2020). A Hybrid Machine Learning Model for Predicting the Best Crop Based on Soil Nutrients and Climate Conditions. International Journal of Advances in Computer Science and Technology, 9(4), 89-96.

[4] Zhang, Y., & Kumar, S. (2019). Deep Learning-Based Crop Classification Using LSTM Networks. IEEE Transactions on Computational Intelligence, 15(3), 78-85.

[5] Ramesh, A., & Govindan, V. (2018). Application of KNN and Random Forest for Crop Prediction in Agricultural Lands. International Conference on Data Science and Machine Learning, 14(2), 102-110.

[6] Mishra, D., & Patel, H. (2017). Soil Analysis and Crop Recommendation System Using Machine Learning Algorithms. International Journal of Emerging Technologies in Computational Intelligence, 5(3), 67-73.

[7] Jha, S., & Raj, S. (2019). A Comparative Study of ML Algorithms for Crop Recommendation Based on Soil and Weather Conditions. Computational Agriculture Journal, 6(2), 56-70.

[8] Kumar, V., & Mehta, P. (2023). Deep Learning-Based Hybrid Approach for Smart Farming and Crop Prediction. IEEE Xplore, 10(2), 1-8.

[9] Gupta, R., & Banerjee, S. (2022). Hybrid LSTM and KNN Model for Precision Agriculture: A Case Study on Crop Recommendation. Journal of Agricultural Research and AI, 14(1), 88-98.

[10] Ghosh, T., & Pandey, M. (2020). Improving Crop Yield Prediction with Neural Networks and Feature Engineering. International Conference on Smart Farming Technologies, 8(1), 34-50.