

Enhancing Crop Recommendation Systems Using Deep Learning Techniques on Soil & Environmental Data

1st Sunil Bag
Department of CSE
GIET UNIVERSITY
Gunupur, Odisha, India
23mtcse006.sunilbag@giet.edu

1st Neelamadhab Padhy
Department of CSE
GIET UNIVERSITY
Gunupur, Odisha, India
dr.neelamadhab@giet.edu

1st Rasmita Panigrahi
Department of CSE
GIET UNIVERSITY
Gunupur, Odisha, India
rasmita@giet.edu

Abstract— A large portion of farmers still follow the traditional farming practices in India, which are time-consuming and affects productivity. Effective farming practices are analyzing various factors like weather conditions, soil type, and irrigation water quality and then customizing the process, such as selecting the appropriate crops, fertilizers, etc. This impacts production directly. This work aims to develop a reliable crop recommendation model based on a deep learning technique, which can address the challenges farmers have during the crop selection process by taking environmental data and soil data into account. A crop recommendation system takes various parameters from the environment, weather, and Soil, analyses them, and provides accurate recommendations. Using an efficient crop recommendation system a farmer can make precise decisions, which helps farmers to increase productivity and reduce cost. Here we have used various Deep Learning classifiers to develop crop recommendation systems and have compared their performance matrices, using Deep Learning models LSTM, RNN, DENSE N/W, and ANN. We have also proposed an ensemble model, which is a combination of all the DL models. The proposed ensemble model outperformed all the DL classifiers with an accuracy of 0.95, Precision of 0.94, and Recall of 0.93.

Keywords: *Crop Recommendation, Machine Learning, Deep Learning, Performance metrics*

I. INTRODUCTION

Nearly half of the workforce in India is directly or indirectly working in the agriculture sector for their livelihood support, it is about 40% of the population and has a share of 20% of the GDP, India produces huge amounts of food, but we also have a good amount of population to feed, even if India is a leading producer for some of the agricultural product, India is not very advanced in this sector and there is huge potential to grow. As half the population can contribute only 20% to the GDP financial condition is not good for farmers. The world is gradually heading towards a food shortage as the population continues to grow. Today agriculture faces multiple challenges due to various factors including poor soil health, unpredictable weather conditions, global warming, and the management approach. Traditional methods of cultivation are no longer beneficial. Farmers are required to make decisions smartly for crop selection, irrigation, weed control, etc. to survive and make the most out of it, this

informed decision can be taken based on factors like climate data, soil parameters, irrigation water quality, historical data, etc. Crop type selection is one of the major criteria in cultivation to produce more crops. Various crop recommendation systems are available today with breakthroughs in advanced technology like Machine Learning, Artificial Intelligence, and Deep Learning, to survive farmers must select climate-resilient crops depending on the weather and soil type. With the help of a recommendation (including fertilizer and Pesticide recommendation) system and yield prediction system, farmers can make precise decisions. These types of decisions benefit in the short term and the long term, for example, taking care of soil health in the long run.

It is extremely difficult to accurately make crop recommendations due to various factors. With accurate environmental data and weather forecasting, by leveraging advanced technology like AI, ML, and Deep Learning this can be achieved, an optimized deep learning model can deliver excellent results if we can provide the most accurate data. A crop recommendation system can assist farmers in making decisions about what to plant and the optimal timing for planting. In this research we have developed an efficient crop recommendations system based on Deep Learning techniques, we have used various Deep Learning classifiers and have compared their output matrixes, and used DL models are LSTM, RNN, DENSE N/W, and ANN. We have also proposed an ensemble model for the crop recommendation.

In these models of recommendation systems used parameters are rainfall, humidity, temperature, soil pH, and NPK, these are the soil and environmental features. The used dataset consists of 22 different crop types and all these parameters.

II. LITERATURE REVIEW

As we know how important is to select the right crop type for better yield, there are various studies have been conducted till now in this area. Below are some of the research works by various authors. Et al. Eliazer, M., [2] Model developed using TCN (Temporal Convolutional Networks), the model performed better than LR and DT. Et al Alkhudaydi, T., [3] Developed models based on deep learning that tackled water quality and crop recommendation, they used various algorithms including SVM and DL with an accuracy of 96.8 and 97.5 respectively. Et al. Okonkwo, Ngozi Ukamaka, [4] The Deep Learning model promised better crop

recommendation results, The model delivered a 99% accuracy, Highest compared to LR, DT, and KNN. At first, it was 45% with continuous trading the accuracy improved. Et al. Mythili, K., [5] Here they have used DNN, DT, KNN, R-Forest, Neu-net, and PSO MDNN, the crop recommendation model used Particle Swarm based hyperparameter tuning and delivered a result of 94.49% accuracy, better than Dec-Tree, KNN, R-Forest, New-Net. This model recommends crops based on the soil and climate parameters. The main goal is to be helpful for small-scale farms even if for the smallest crop plot. All these are achieved by using deep learning and classifiers with optimization. Et al. [6] With the help of image processing and other satellite data the proposed system archives the accurate crop recommendation and converts the collected data into meaningful parameters using MATLAB, this model utilized deep learning models CNN, RNN, and HNN, the dataset was collected from the department of agriculture Tamil Nadu. Among these three models, HNN delivered the best accuracy. Et al. Banavlikar, Tanmay, [7] In this paper authors used the proposed recommendation model using ANN, which is a prototype model with a small dataset and can extend into a larger dataset with modification. This also can integrate with irrigation systems for smart farming. Et al. Dharani, M. K., Thamiselvan, [8] This study focuses on AI technologies, especially ANN, DNN, and RNN. It shows how RNN and Hybrid networks are better than other networks in improving the accuracy of crop yield prediction. CNN was more accurate than ANN - around 87%, RNN with LSTM and Hybrid network is better than another network with the result of 89% accuracy. The hybrid network combines different networks with additional algorithms to achieve an impressive accuracy level of 90%. Et al. [9] Used a variety of machine learning techniques to assist small-scale farmers with an optimization-based Deep Learning crop recommendation system that analyses the input from meteorological and chemical factors such as PH, Nitrogen, phosphorus, Potassium, rainfall, temperature, and humidity. The model delivered an accuracy of 91.21% with KNN over CNN, RF, and DT Naïve Bays. Et al. Jyothika, Pasupuleti, [10] Proposed a crop recommendation model using a Random Forest classifier, Decision Tree Classification Algorithm, K Nearest Neighbor (KNN) Algorithm, and Deep Sequential model. This model provides the crop appropriateness data for the crop recommendation. Et al. Punitha, A., [11] In this article authors have presented the Gorilla Troops Optimization with Deep Learning-based Crop Recommendation and Yield Prediction model (GTODL-CRYPM), with an accuracy of 99.88% which is the most accurate compared to other models used in the study. This is a model focused on crop recommendation and yield prediction, using Gorilla troops optimization with Deep Learning. Et al. [1] In this paper three extra modules are proposed with the crop recommendation module using Deep Learning techniques. Used algorithms are K-nearest neighbor, Naïve Bayes, Decision Tree, Bagging Classifier, AdaBoost Classifier, Gradient Boosting Classifier, XGBoost Classifier, Random Forest, and Ibgm Classifier. Out of all the models most accurate result is delivered by the Random Forest algorithm.

III. PROPOSED METHODOLOGY

System implementation for a crop recommendation system based on deep learning involves multiple stages, we have divided it into four stages, namely 1) Data Gathering and Processing, 2) Model Training, 3) Performance Comparison, and 4) Loss. This crop recommendation system is developed leveraging data-driven deep learning models LSTM, RNN, DENSE NET, and ANN. Soil properties and Environmental data are considered parameters. The model is described below in detail. The used dataset is collected from online sources, The data set consists of seven parameters Temperature, Humidity, Soil pH, Rainfall, Nitrogen, Potassium, and Phosphorus. The data set has 3100 records with 22 unique crop types. During the data processing phase, we have cleaned the obtained raw data. This is a very important step to make sure we are gating unbiased data in a suitable format for training, testing, and validating the models. Here we remove the duplicate data, null data, chunk data, handled missing values, and handled outlier values. The cleaned data was then divided into three modules for training, testing, and validating. We have divided the data set into 80%, and 10%, 10% for training, testing, and validation of the models respectively.

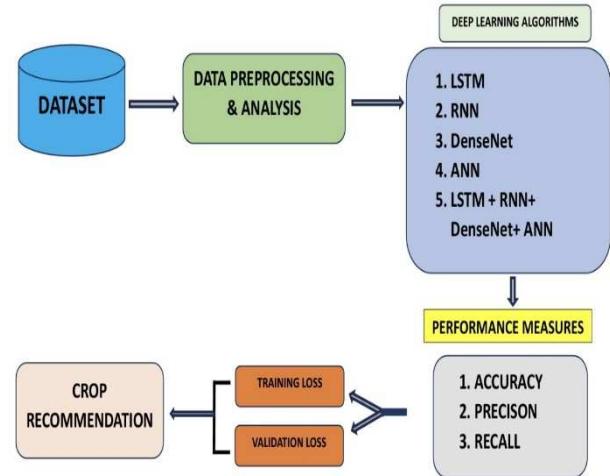


Fig. 1. Proposed model for Crop Recommendation Systems Using Deep Learning Techniques

Fig. 1 represents the Proposed Model for the Crop Recommendation System.

IV. RESULT AND DISCUSSIONS

TABLE I. PERFORMANCE METRICS OF DIFFERENT DL MODELS

Model	Accuracy	Precision	Recall
LSTM	0.91	0.92	0.91
RNN	0.93	0.94	0.93
DENSE N/W	0.94	0.95	0.94
ANN	0.93	0.95	0.93
ENSEMBLE	0.95	0.94	0.93

Table 1 shows the performance metrics of different Deep Learning Models where the performance metrics were Accuracy, Precision, and Recall.

A. Long Short-Term Memory (LSTM)

LSTM is a recurrent neural network (RNN) architecture that is ideal for sequence prediction problems due to its capacity to hold long-term dependencies. It is made up of memory cells and gates that regulate the flow of information through the cell over time, allowing for the recognition of patterns in sequential data.

B. Recurrent Neural Network (RNN):

RNNs are neural networks built to cope with sequential data by keeping a hidden state that stores information about prior inputs. They process data sequentially while keeping a recollection of previous inputs throughout time.

C. Neural Network:

Dense neural networks, sometimes referred to as feedforward neural networks, are made up of layers of neurons, with each layer's neurons coupled to every subsequent layer's neuron. These networks are employed to make predictions and discover intricate patterns in data.

D. Artificial Neural Network (ANN):

ANNs are a type of machine learning model inspired by the biological neural networks found in human brains. They are made up of interconnected nodes (neurons) organized into layers that process input data and provide output predictions using a system of weighted connections and activation functions.

E. Ensemble Classifiers:

Ensemble classifiers use numerous separate models to increase predictive accuracy. Voting, bagging (bootstrap aggregating), and boosting are common ensemble approaches that combine the outputs of numerous models to create a final prediction.

F. Training and Validation Loss:

TABLE II. TRAINING AND VALIDATION LOSS FOR DIFFERENT DL MODELS

Model	Training loss	Validation loss
LSTM	0.382006	0.387762
RNN	0.363584	0.386185
DENSE N/W	0.138803	0.150516
ANN	0.131376	0.136864
ENSEMBLE	0.11242	0.101211

In the above Table.II we used the ensemble learning approach which is the combination of all deep learning classifiers.i.e. Ensemble=LSTM+RNN+DENSE+ANN. We estimated the training as well as validation loss of the proposed model which is the merging of all the individual models. Missing Training and Validation Loss for Ensemble. In this proposed model, we estimated the performance metrics based on the individual model's performance.

RQ 1: How do different deep learning architectures perform in predicting suitable crops based on soil and

environmental conditions, and can ensemble methods enhance predictive accuracy and reliability?

We used the 4 deep-learning approaches to combine one model for crop recommendation. Our objective was to enhance the performance of the model. The performance metrics help to determine the model's robustness. The parameters like accuracy, precision, and recall are used. During the implementation we also consider the training and testing loss across the different epochs to determine whether the model is overfitted or underfitted. To prove the robustness, we used several data visualization techniques like- loss curves, accuracy curves, confusion matrices, ROC curves, and precision-recall curves. The findings will provide farmers and agronomists with a trustworthy suggestion tool to enhance crop productivity through data-driven decisions. By comparing individual models with an ensemble model performance, we determined many models as well as decision-making processes. By combining these analyses and visualizations, you can create a robust research study that is not.

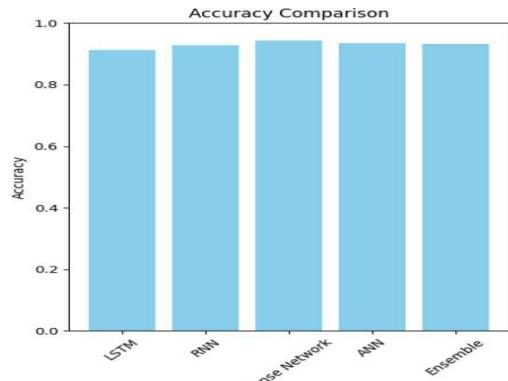


Fig.2. : Accuracy comparison of DL models

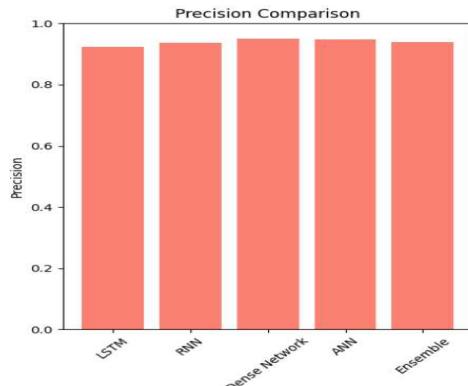


Fig.3. Precision comparison of DL models

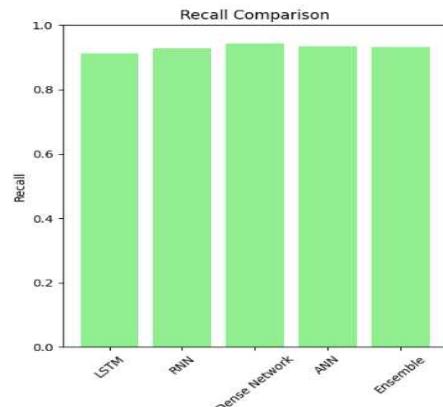


Fig.4. Recall the comparison of all 4 deep learning classifiers

The above-mentioned fig.2,3, and 4 are meant for the performance metrics of the deep learning classifiers and these are accuracy, precision, and recall

G. Ensemble Combination

In this paper, we used a novel approach to combine the different deep learning models into a single model which are: -LSTM, RNN, Dense Networks, and ANN. We used the majority voting method where the individual model was able to predict a label and the Ensemble classifier received the most votes. This model is used to enhance the prediction capabilities.

Step 1: Training with History Tracking:

We used history = model.fit(...) when training each model to record accuracy and loss at each epoch.

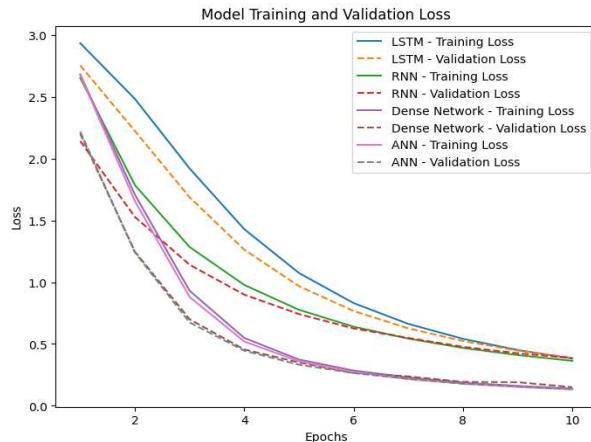
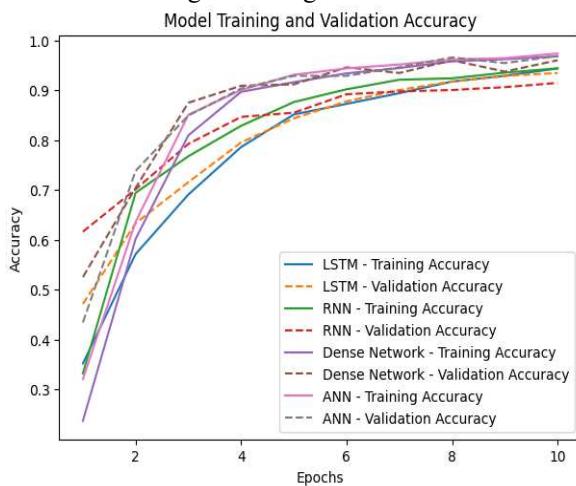


Fig.5. Model training and Validation loss

The above-mentioned Fig 5. shows the training and validation loss per epoch, which aids in detecting overfitting or underfitting in the models. These graphs provide a clear image of model performance trends, enabling fine-tuning of training parameters like epoch number, learning rate, and architecture alterations based on how the training and validation lines converge or diverge.



The above-mentioned Fig 6. demonstrates the training and validation accuracy per epoch for each model. This allows us to examine how well each model learns and generalizes over time.

H. Confusion Matrix

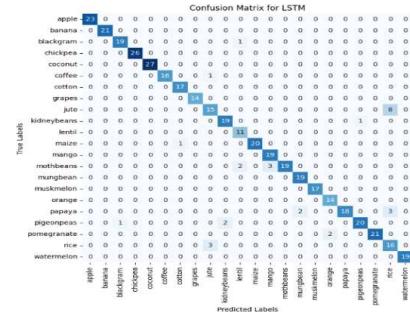


Fig.7. Confusion matrix for LSTM

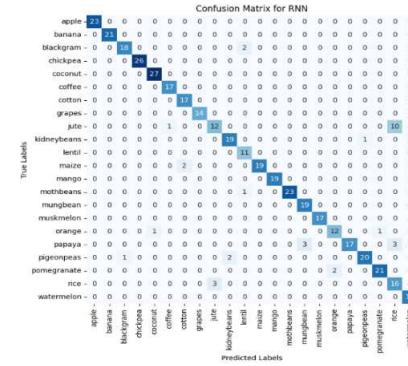


Fig.9. Confusion matrix for RNN

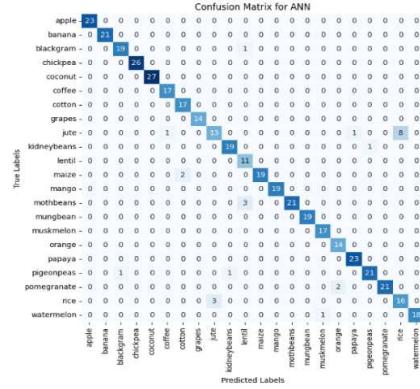


Fig.10. Confusion matrix for ANN

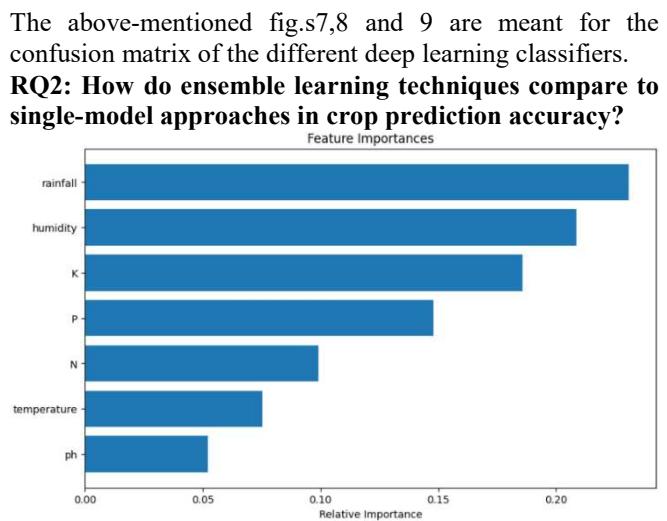


Fig.11. Feature importance of various learning classifiers

Fig.11 represents the feature importance metrics for different classifier

DNN Test Accuracy: 0.94

Random Forest Test Accuracy: 0.99

Voting Classifier Test Accuracy: 0.99

RQ 3. What role do environmental and soil features play in predicting crop suitability, and how can feature importance be visualized?

Analyzing and visualizing the importance of various features in crop prediction. Fig.12 discusses the visualization importance of crop prediction. We plotted the visualization of the different features' impact and their score

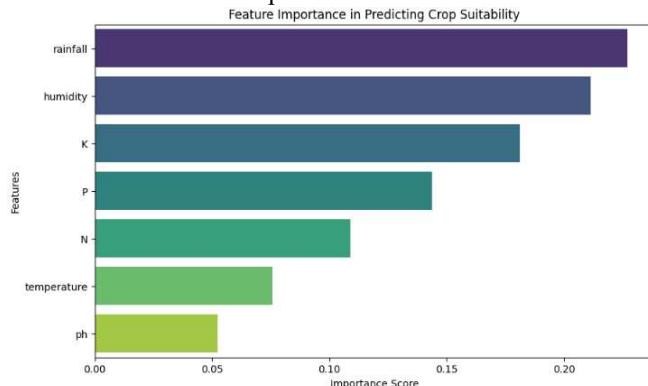


Fig.12. Visualizing the importance of features in crop prediction

I. Transfer Learning

RQ4: How can transfer learning be utilized to improve crop prediction models on limited datasets?

To answer this research question, we start with a pre-trained model (e.g., from similar or generic datasets) and fine-tune it with your crop suggestion dataset. Use image classification models such as ResNet or VGG, which may be converted for tabular data with feature extraction. We use the DenseNet model for the transfer learning approach. We plot training and validation loss curves to assess the effectiveness of transfer learning. Compare the performance parameters (accuracy, precision, recall) of models trained from scratch vs those employing transfer learning.

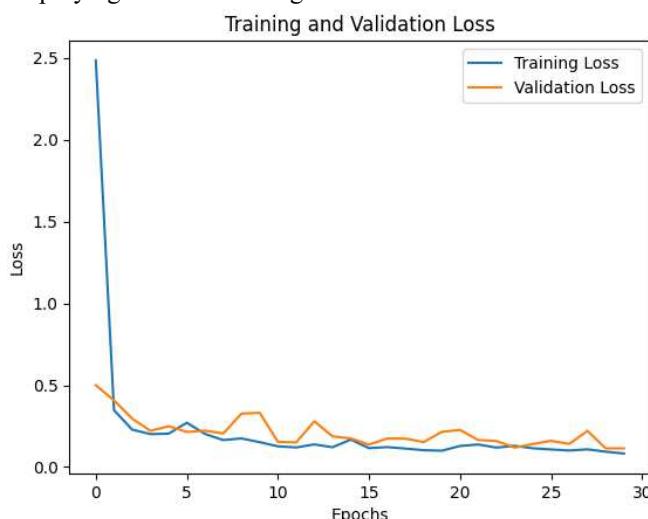


Fig.13 depicts the losses that occurred during the training and validation losses.

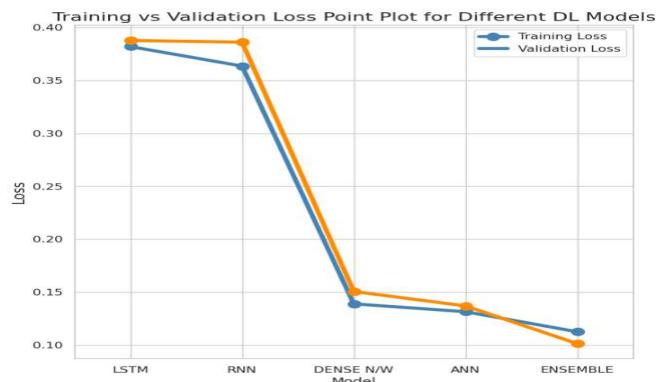
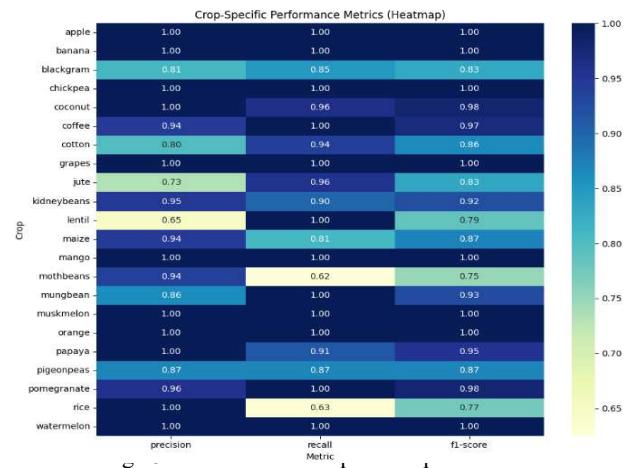


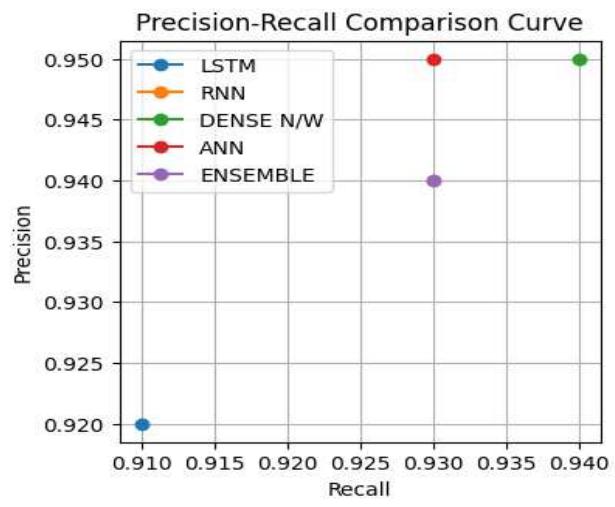
Fig.14. Training and validation loss of different classifiers

Fig.14 depicts the losses that occurred during the training and validation losses of different deep learning classifiers.

RQ5: What insights can we get from comparing model performance indicators across different crops, and how can we adapt our recommendation model to account for crop-specific accuracy?



We have created precision-recall curves for crops with different performances, especially those with moderate scores, to examine the trade-offs between them.



In the above Fig. 16, we used the PR curve for different crops. An average precision (AP) score, which provides a single statistic summarizing the precision-recall relationship, is included in each PR curve. Stronger forecast confidence for that crop is suggested by higher AP scores, which show that the model can maintain high precision while obtaining high recall. Fig. 15. represents the Heatmap of different classifiers to represent the highly correlated features that help to predict the crops

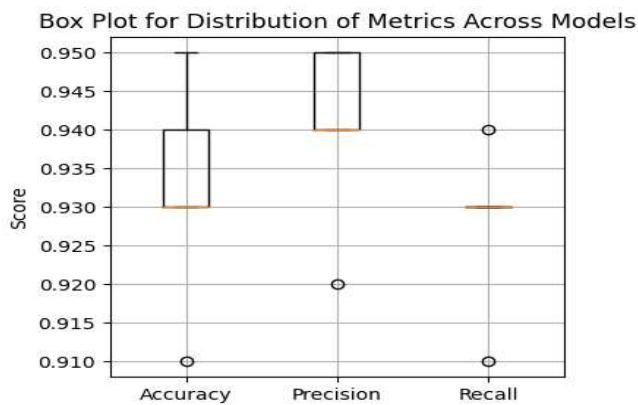


Fig.17. Precision and recall measure

In the above Fig.17, it demonstrates that each metric (accuracy, precision, recall) is shown as a box plot across models. Our objective is to present the shows the spread or variability within each statistic, which helps figure out how consistent different models are.

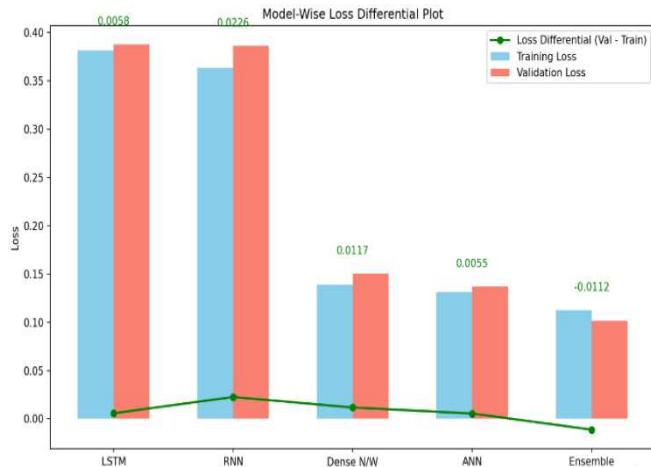


Fig.18. Model loss behavior

Fig. 18 demonstrates the model-wise loss differential between the training as well as the validation loss of deep learning classifiers. It provides the whether the model tends to overfit or not. Each bar we used to represent the training and validation loss. The line shows the difference (validation loss - training loss).

V. CONCLUSION AND FUTURE WORK

In modern agriculture, making informed decisions is crucial at every stage to achieve higher yields and improved product quality. By analyzing and correlating various factors from the soil and environment, it becomes possible to make accurate and informed decisions. Choosing the right crop, based on soil and environmental properties, is a fundamental factor in achieving better yields. In this paper, we analyzed parameters such as rainfall, humidity, temperature, soil pH, and N.P.K. using advanced machine learning models, including LSTM, RNN, Dense Network, and ANN. We compared their results and developed a proposed ensemble model that shows promise and outperformed all the above individual models. The proposed model combines LSTM, RNN, Dense Net, and ANN. The model achieved performance metrics of 0.95 for Accuracy, 0.94 for Precision, and 0.93 for Recall.

REFERENCES

- [1] Durai, Senthil Kumar Swami, and Mary Divya Shamili. "Smart farming using machine learning and deep learning techniques." *Decision Analytics Journal* 3 (2022): 100041.
- [2] Eliazer, M., Jerrell Franklin C, and Shridhar Shrivastava. "Using Temporal Convolutional Networks (TCN) Deep Learning Model for Crop Recommendation." *Shridhar, Using Temporal Convolutional Networks (TCN) Deep Learning Model for Crop Recommendation* (May 11, 2024) (2024).
- [3] Alkhudaydi, T., Albalawi, M. Q., Alanazi, J. S., Al-Anazi, W., & Alfarshouti, R. M. (2023). Deep Learning for Combined Water Quality Testing and Crop Recommendation. *International Journal of Advanced Computer Science and Applications*, 14(4).
- [4] Okonkwo, NgoziUkamaka, F. C. Uwaechia, FadipeBayo Michael, and EzeekwudoCelestinaChinenye. "International Journal of Food Science Research and Development (IJFSRD)." *Journal Id* 1746: 6001.
- [5] Mythili, K., & Rangaraj, R. (2021). Deep learning with particle swarm-based hyperparameter tuning-based crop recommendation for better crop yield for precision agriculture. *Indian Journal of Science and Technology*, 14(17), 1325-1337.
- [6] SSL, Durai Arumugam, R. Praveenkumar, and V. Balaji. "An Intelligent Crop Recommendation System using Deep Learning." *International Journal of Intelligent Systems and Applications in Engineering* 11: 423-428.
- [7] Banavlikar, Tanmay, Aqsa Mahir, Mayuresh Budukh, and Soham Dhodapkar. "Crop recommendation system using Neural Networks." *International Research Journal of Engineering and Technology (IRJET)* 5, no. 5 (2018): 1475-1480.
- [8] Dharani, M. K., ThamilSelvan, R., Natesan, P., Kalaivaani, P. C. D., & Santhoshkumar, S. (2021, February). Review on crop prediction using deep learning techniques. In *Journal of Physics: conference series* (Vol. 1767, No. 1, p. 012026). IOP Publishing.
- [9] Kumar, D. Manoj, and S. Sheeba. "Towards an Efficient Deep Learning Approach for Crop Recommendation," *SPAST Reports* 1.3 (2024).
- [10] Jyothika, Pasupuleti, K. Venkata Ramana, and Ch Lakshmi Narayana. "Crop Recommendation System to Maximize Crop Yield Using Deep Neural Network." *Journal of Engineering Sciences* 12.11 (2021).
- [11] Punitha, A., and V. Geetha. "Gorilla Troops Optimization with Deep Learning based Crop Recommendation and Yield Prediction." *International Journal of Cognitive Computing in Engineering* (2024).