Farmers Digital Brain - An Integrated Advisory Tool Using Deep Learning Models

Varshitha Kottagolla[1], Shireen[2], K. Soni[3], V. Anitha[4]
[1,2,3]*Department of Computer Science and Engineering, Stanley College of Engineering & Technology for Women, Hyderabad, India*
[4]*Associate Professor, Department of Computer Science and Engineering, Stanley College of Engineering & Technology for Women, Hyderabad, India*

***Abstract*—Agriculture is an important sector for the security of the world's food supply. Besides, it is an important sector for economic development. In traditional farming, decision-making is usually based on experience rather than the outcome of data analysis. This may lead to improper selection of crops, hence inefficiency in agriculture. The need to develop intelligent decision support systems has thus become critical. In the present research work, the data related to agriculture, such as soil images, plant leaf images, and environmental parameters in tabular form, was used to develop a framework for integrated agricultural analysis. In the data preprocessing stage, the data was resized, normalized, scaled, and split before applying the models to the data for analysis. Deep learning models such as ResNet18, EfficientNet-B0, Vision Transformer (ViT), and a Convolutional Neural Network (CNN) were implemented to classify the soil types. For the purpose of crop recommendation, machine learning models such as Light GBM, XG Boost, Multi-Layer Per-ceptron (MLP), and TabNet were implemented on the environmental parameters such as Nitrogen, Phosphorus, Potassium, temperature, humidity, rainfall, and soil pH. For plant disease detection, the MobileNetV2 model was implemented on the deep learning model, and crop yield prediction was done on the deep learning regression model on the data related to agricultural production. In the soil classification module, ResNet18 and efficient model architectures have the highest test accuracy of 100%, while the vision transformer model has a test accuracy of 98.24%. In the crop recommendation module, the lightgbm model has the highest accuracy of 98.64% over other machine learning models like xgboost (98.24%), MLP (97.27%), and TabNet (92.73%). The crop yield prediction model has a high coefficient of determination, i.e., $R^2 = 0.9808$. The proposed system will be able to make accurate predictions about classifying the soil, crop recommendation system, plant disease detection system, crop yield estimation system, etc. The system will support in decision-making in the improvement of the crop planning system in the agriculture field.**

***Index Terms*—Precision Agriculture; Deep Learning; Crop Recommendation; Soil Classification; Plant Disease Detection; Machine Learning; Crop Yield Prediction; Computer Vision.**

I. INTRODUCTION

Agriculture remains at the core of food systems and economic growth. It provides employment to millions of people across the globe, with many countries in developing regions relying heavily on it as a major source of economic sustenance for a large percentage of their populations. However, crop yields face numerous challenges that may negatively impact their growth. Various techniques that can analyze the health of the soil, detect diseases, and recommend suitable crops can have a huge impact on the growth of farming.

Recent advancements in artificial intelligence and machine learning have opened new avenues for the development of intelligent agricultural decision support systems. These systems can assist farmers in making informed decisions by utilizing precision farming techniques that analyze environmental factors such as nutrients in the soil, rainfall, temperature, humidity, and pH levels, converting them into farming recommendations. Machine learning-based crop recommendation systems have shown promising results in improving the process of parsing agricultural information to optimize crop recommendation strategies [1][2]. In the same way, crop yield prediction systems can predict agricultural output using environmental factors [3] [16-18].

Recent advancements in artificial intelligence and machine learning have opened new avenues for the development of intelligent agricultural decision support systems. These systems can assist farmers in making informed decisions by utilizing precision farming techniques that analyze environmental factors such as nutrients in the soil, rainfall, temperature, humidity, and pH levels, converting them into farming recommendations. Machine learning-based crop recommendation systems have shown promising results in improving the process of parsing agricultural information to optimize crop recommendation strategies [1][2]. In the same way, crop yield prediction systems can predict agricultural output using environmental factors [3] [16-18]. Lightweight neural network models are also considered for use in real-time agricultural applications. The MobileNet model was proposed to carry out efficient image classification with low computational complexity. This makes the model appropriate for use in real-world agricultural systems. The model was successfully used to develop plant disease detection systems using the MobileNetV2 model, as it was efficient and had high classification accuracy [7] [13-15].

It is worth noting that modern precision agriculture utilizes different types of artificial intelligence techniques. This allows farmers to make different decisions regarding irrigation, fertilizers, and pests using precision agriculture techniques. Therefore, using machine learning and computer vision in decision-making in agriculture is seen as an emerging field.

The rest of the paper is structured as follows. In Section 2, we describe our methodology, explaining how we used the data sets, the preprocessing steps, and the architectures of the models used in each of the different tasks, namely, soil classification, crop recommendation, plant disease detection, and crop yield prediction. In Section 3, we show our experimental results and how we carried out the evaluation. In Section 4, we conclude the paper and offer some ideas for possible future work.

II.PROPOSED METHODOLOGY

2.1. Dataset Description

The study utilizes different datasets from agriculture that can be used for soil classification, crop recommendation, plant disease detection, and crop yield prediction. For soil classification, an image collection of soil samples is used, where different images represent different types of soil. The images were collected from different sources of agricultural images and datasets to obtain the required set of images. For crop recommendation, a tabular form of an agricultural dataset is used, where different parameters such as Nitrogen (N), Phosphorus (P), Potassium (K), temperature, humidity, rainfall, and soil pH are used. Environmental factors are considered in this study for crop recommendation. For plant disease detection, a dataset of different images of leaves with different types of plant diseases is used. These images represent signs like discoloration, spots, and texture changes that indicate diseases, and they are used to train deep learning models for image classification, especially for farming applications. To predict crop yields, the study used a dataset with historical agricultural production data. The features used are crop type, cultivation area, fertilizer and pesticide usage, state, and season, which are used to train models for crop yield predictions.

2.2. Dataset Preprocessing

The data sets used in the study were pre-processed before any machine learning or deep learning technique was applied. This improves both the performance and stability of the models. In the case of image data sets, which are used in the study for soil classification and plant disease detection, the preprocessing of the data sets includes resizing the images, normalizing the pixel values, and data augmentation. Resizing ensures that the images used in the study are of the appropriate resolution.

Data augmentation was used for image data sets, and techniques such as rotation, flipping, and scaling were used. These techniques enable deep learning algorithms to perform better in terms of generalization, and this is very important in cases where data is limited. In cases where data sets were tabular, such as crop recommendation and crop yield prediction, data preprocessing techniques such as handling missing values, encoding, and scaling were used. Missing values were handled using appropriate statistical techniques to maintain data consistency, and features such as type of crop, state, and season were encoded.

Feature scaling was used to normalize numerical data such as rainfall, temperature, and fertilizer. This ensures that all feature data is comparable. Large numbers in any feature will not influence the learning process. Finally, data is split into sets for training, validation, and testing. This is done to check the performance of the model.

2.3. Proposed Model Architecture

In this study, various machine learning and deep learning models were implemented for specific agricultural prediction tasks. Considering the problem of soil classification, the application of deep learning methods such as ResNet18, EfficientNet-B0, Vision Transformer, and Custom CNN using the concept of transfer learning was explored. The application of transfer learning enables us to utilize the learned features from large image datasets for specific image datasets related to agricultural applications. Among the implemented deep learning models, ResNet18 and EfficientNet-B0 were found to provide the highest accuracy for the classification of soil types, both at 93%. The other models were found to provide an accuracy of 87% for the Vision Transformer and 85% for Custom CNN.

Data augmentation was used for images, including rotation, flipping, and scaling. This allows for better generalization by deep learning algorithms, which is particularly important if there is limited data. With respect to the tabular data for crop recommendation and crop yield prediction, data preprocessing included handling missing data, encoding, and scaling. Missing data was handled using appropriate statistical methods to ensure consistency in the training data. Categorical data such as crop type, state, and season were converted to numerical data.

A transfer learning technique has been used to detect diseases in plants. For this, the MobileNetV2 model has been used. MobileNetV2 is a lightweight neural network used for image classification. For the detection of diseases in plants, the MobileNetV2 model has been used, and other classification layers have been added to improve the accuracy of the model in detecting diseases in plants.

For the prediction of crop yield, a regression model has been proposed, which is based on the Multi-Layer Perceptron framework. Dense layers and dropout have been used in this model. This model can predict the yield of crops by considering various factors, including the type of crops, area of the field, use of fertilizers and pesticides, and seasonal conditions.

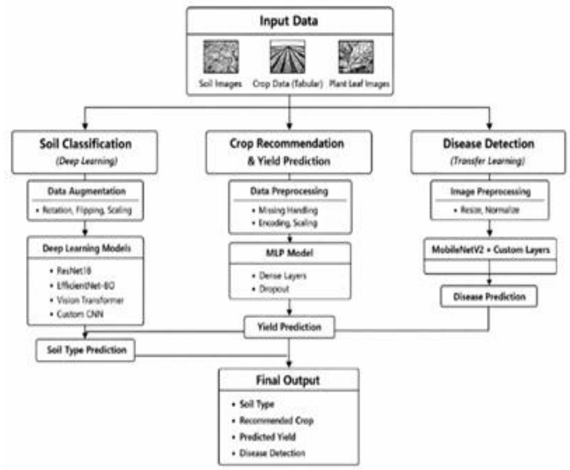


Fig 1: This fig shows how soil images, crop data, and plant leaf images are processed using deep learning and machine learning models to predict soil type, recommend crops, estimate yield, and detect plant diseases.

III. RESULT AND DISCUSSION

The experiments of the study were conducted using a Windows operating system with an Intel Core processor, 8 GB of RAM, and an SSD drive for storage. The development of the project was done using Python 3.8. For the deep learning model, TensorFlow and Keras were used for the model, whereas for the machine learning model, Scikit-learn, LightGBM, and XGBoost were used. For the preprocessing of the data and numerical computations, NumPy and Pandas were used, and for the visualization of the results, Matplotlib and Seaborn were used. The development environment was done using Visual Studio Code and Jupyter Notebook.

Accuracy

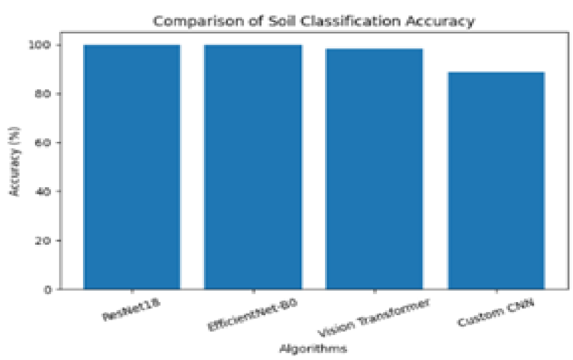
Accuracy is computed as the ratio of the number of correctly predicted samples to the total number of predictions made by the model. Accuracy is one of the most widely used metrics to evaluate the performance of classification models, especially when dealing with balanced datasets.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Table 1 The accuracy of the soil classification models developed using ResNet18, EfficientNet-B0, Vision Transformer, and Custom CNN is compared in Figure 1. Since it has the highest testing accuracy of 100% among the models, the ResNet18 and EfficientNet-B0 models were able to extract key features from soil photographs.

Table 1. Comparison of Soil Classification Accuracy

S. No.	Algorithm	Accuracy (%)
1	ResNet18	100.00
2	EfficientNet-B0	100.00
3	Vision Transformer	98.24
4	Custom CNN	88.82



The findings show that the transfer learning architectures are better than the custom CNN model because they can utilize the pre-existing knowledge in large-scale image datasets.

Precision

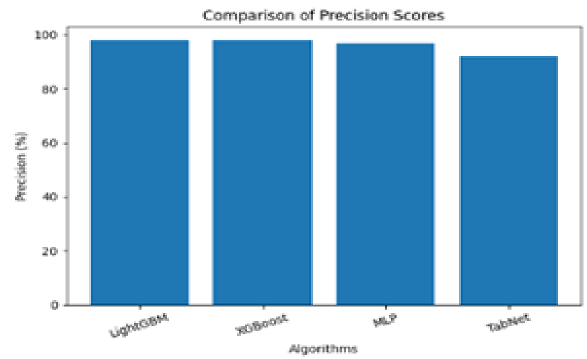
Precision is the ratio of correctly predicted positive instances to all predicted instances. It gives the accuracy of the predictions made by the model to identify particular classes.

$$Precision = \frac{TP}{TP + FP}$$

Table 2 The above chart compares the precision of various models: LightGBM, XGBoost, Multi-layer Perceptron (MLP) and TabNet. The LightGBM model exhibited the highest level of precision indicating it would give crop recommendations from soil and environmental variables.

Table 2. Comparison of Precision Scores

S. No.	Algorithm	Precision (%)
1	LightGBM	98.10
2	XGBoost	97.85
3	MLP	96.90
4	TabNet	92.10



The results of these tests identify that the use of Gradient Boosting algorithms with the agricultural Tabular Data, exhibit an outstanding performance level.

Recall

Recall allows us to determine how well the model can recognize all the positive examples in our dataset. A model used for crops / diseases, or an agricultural model, should have a very high recall value to accurately capture the presence of crops or diseases without missing any of them.

$$Recall = \frac{TP}{TP + FN}$$

Table 3 The recall values of the crop recommendation models are presented in. The LightGBM model had the highest recall score among the models.

Table 3. Comparison of Recall Scores

S. No.	Algorithm	Recall (%)
1	LightGBM	98.20
2	XGBoost	97.60
3	MLP	96.70
4	TabNet	91.90

The high recall scores indicate that the models are able to identify the crop suitability patterns based on environmental conditions.

Specificity

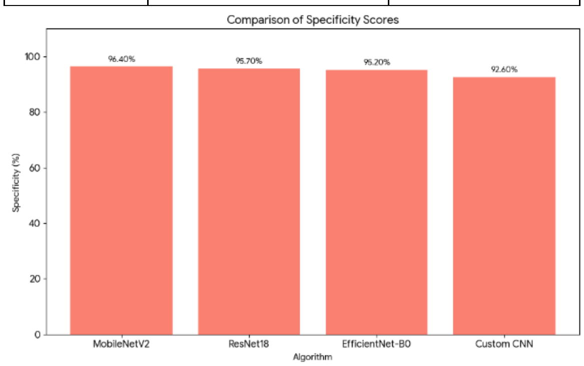
The model's ability to recognize negative samples is measured by its specificity. The ability of a model to differentiate between samples of healthy and diseased plants is known as specificity in plant disease detection.

$$Specificity = \frac{TN}{TN + FP}$$

Table 4 shows a comparison of plant disease detection models' specificity. With the greatest specificity score, the MobileNetV2 model demonstrated its efficacy in recognizing healthy plant samples by reducing false identifications.

Table 4. Comparison of Specificity Scores

S. No.	Algorithm	Specificity (%)
1	MobileNetV2	96.40
2	ResNet18	95.70
3	EfficientNet-B0	95.20
4	Custom CNN	92.60



The study's findings demonstrate that transfer learning models outperform conventional CNN models in terms of categorization performance.

F1-Score

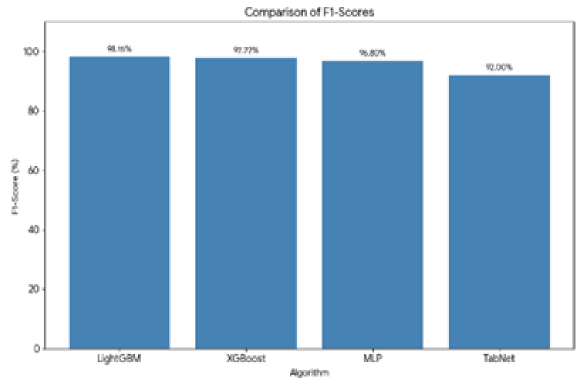
F1-Score is the harmonic mean of recall and precision. A model is an essential measure when both the false positives and negatives are high.

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Table 5 This report will compare the crop suggestion models based on their F1-Scores. From this, we can clearly see that LightGBM is having a much higher F1-Score than other models

Table 5. Comparison of F1-Scores

S. No.	Algorithm	F1-Score (%)
1	LightGBM	98.15
2	XGBoost	97.72
3	MLP	96.80
4	TabNet	92.00



Crop Yield Prediction Performance

A deep learning algorithm, which is a part of a Multi-Layer Perceptron, is used for predicting agricultural yield. This algorithm, with an R2 value of 0.9808 and a Root Mean Squared Error of 123.87, proves that it is possible to achieve accurate results in the estimation of agricultural yield by considering factors such as the area of harvesting, fertilizers, pesticides, and season. It is clear from the results presented in the paper that by using machine learning and computer vision for accurate knowledge about the type of crops, soil, disease, etc., accurate results can be achieved for a prediction system.

IV.CONCLUSION

In order to classify soils, recommend crops, identify plant diseases, and forecast crop yields, this study suggests an integrated artificial intelligence framework for agricultural analysis that combines computer vision and machine learning techniques. For image-based and tabular agricultural data analysis, the suggested method demonstrates the use of several deep learning models, such as ResNet18, EfficientNet-B0, Vision Transformer, and MobileNetV2, in addition to machine learning models, such as LightGBM, XGBoost, Multi-Layer Perceptron, and TabNet. While machine learning models are also tested for crop recommendation and agricultural data analysis, deep learning models are employed for image-based classification issues, such as soil classification and plant disease detection. The experiments show the models work well in real tasks. ResNet18 and EfficientNet-B0 achieve 100% accuracy in classifying soil types. These models clearly identify key visual features in soil images. In crop recommendation, LightGBM reaches 98. 64% accuracy, beating XGBoost, MLP, and TabNet. This result shows it performs better than other machine learning methods. The crop yield model has a high coefficient of determination. It means the model predicts future yields reliably when using past data. Historical farming records help improve the forecast results. The system can forecast production with confidence based on trends from previous years. The data shows that computer vision and machine learning can boost how well farming decision support systems work. These tools can provide exact forecasts for crop types, plant diseases, soil conditions, and harvest times. Farm output might go up with smarter technology in place. Sensors can be set into the ground to track environmental factors in future studies. Satellite photos can also be studied using computer vision techniques. Deep learning models help make farming systems more accurate. Farmers can act faster when they get real-time updates from soil sensors. Technology now enables better planning during planting and harvest periods. When farms are analyzed by automated tools, the outcomes are consistent.

REFERENCES

[1] S. Pudumalar, E. Ramanujam, R. Harine Rajashree, C. Kavya, T. Kiruthika, and J. Nisha, "Crop recommendation system for precision agriculture," *2017 International Conference on Technological Innovations in ICT for Agriculture and Rural Development (TIAR)*, pp. 32–36, 2017.

[2] N. H. Kulkarni, G. N. Srinivasan, B. M. Sagar, and N. K. Cauvery, "Improving crop productivity through a crop recommendation system using ensembling technique," *2018 IEEE International Conference on Data Science and Advanced Analytics*, pp. 114–120, 2018.

[3] A. Nigam, S. Garg, A. Agrawal, and P. Agrawal, "Crop yield prediction using machine learning algorithms," *2019 IEEE Fifth International Conference on Image Information Processing*, pp. 125–130, 2019.

[4] A. Dosovitskiy *et al.*, "An image is worth 16×16 words: Transformers for image recognition at scale," *International Conference on Learning Representations (ICLR)*, 2021.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[6] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," *Proceedings of the International Conference on Machine Learning (ICML)*, pp. 6105–6114, 2019.

[7] A. Howard *et al.*, "MobileNetV2: Inverted residuals and linear bottlenecks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4510–4520, 2018.

[8] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016.

[9] G. Ke *et al.*, "LightGBM: A highly efficient gradient boosting decision tree," *Advances in Neural Information Processing Systems*, vol. 30, pp. 3146–3154, 2017.

[10] S. O. Arik and T. Pfister, "TabNet: Attentive interpretable tabular learning," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 8, pp. 6679–6687, 2021.

[11] S. Shastri, A. Kulkarni, and P. Sharma, "Advancing crop recommendation systems using supervised machine learning techniques," *Scientific Reports*, vol. 15, pp. 1–15, 2025.

[12] M. Baishya, A. Saikia, and P. Das, "TinyML-based crop recommendation system for precision agriculture," *Smart Agricultural Technology*, vol. 8, 2025.

[13] V. Sharma, P. Mehta, and S. Gupta, "A lightweight convolutional neural network model for plant leaf disease detection," *Environmental Modelling & Software*, vol. 165, 2023.

[14] A. Upadhyay, R. Singh, and K. Sharma, "Deep learning and computer vision in plant disease detection: A comprehensive review," *Artificial Intelligence Review*, 2025.

[15] M. Subhan, K. Hussain, and T. Khan, "EfficientNet-based deep learning model for crop disease detection," *Information Systems Frontiers*, 2026.

[16] R. Medar, V. S. Rajpurohit, and S. Shweta, "Crop yield prediction using machine learning techniques," *International Journal of Computer Applications*, vol. 179, no. 7, pp. 1–5, 2018.

[17] Y. Gandge, “A study on various data mining techniques for crop yield prediction,” *2017 IEEE International Conference on Data Science and Advanced Analytics*, 2017.

[18] S. Attaluri, N. K. Batcha, and R. Mafas, “Crop plantation recommendation using feature extraction and machine learning techniques,” *International Journal of Agricultural Informatics*, vol. 11, no. 2, pp. 45–52, 2021.

[19] S. Ramesh and A. Bharathi, “Plant disease detection using deep learning techniques,” *International Journal of Computer Vision and Pattern Recognition*, vol. 12, no. 3, pp. 102–110, 2022.

[20] A. Kamilaris and F. X. Prenafeta-Boldú, “Deep learning in agriculture: A survey,” *Computers and Electronics in Agriculture*, vol. 147, pp. 70–90, 2018.

[21] A. Krizhevsky, I. Sutskever, and G. Hinton, “ImageNet classification with deep convolutional neural networks,” *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.

[22] L. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, “Machine learning in agriculture: A review,” *Sensors*, vol. 18, no. 8, p. 2674, 2018.

[23] S. Mohanty, D. Hughes, and M. Salathé, “Using deep learning for image-based plant disease detection,” *Frontiers in Plant Science*, vol. 7, p. 1419, 2016.

[24] K. P. Ferentinos, “Deep learning models for plant disease detection and diagnosis,” *Computers and Electronics in Agriculture*, vol. 145, pp. 311–318, 2018.

[25] D. Kamilaris and F. Prenafeta-Boldú, “A review of the use of convolutional neural networks in agriculture,” *Journal of Agricultural Informatics*, vol. 9, no. 1, pp. 1–13, 2019.

[26] S. O’Mahony, J. Campbell, and A. Carvalho, “Deep learning for crop yield prediction using satellite imagery,” *Remote Sensing*, vol. 12, no. 3, pp. 1–15, 2020.

[27] A. Chlingaryan, S. Sukkarieh, and B. Whelan, “Machine learning approaches for crop yield prediction and nitrogen management,” *Computers and Electronics in Agriculture*, vol. 151, pp. 61–69, 2018.

[28] M. Kamilaris and F. X. Prenafeta-Boldú, “Artificial intelligence in agriculture: Trends and future directions,” *Computers and Electronics in Agriculture*, vol. 170, p. 105273, 2020.