**Abstract**

When raising crops, farmers encounter a number of difficulties, including erratic irrigation, subpar soil, etc. A significant portion of farmers, particularly in India, lack the knowledge necessary to choose the right crops and fertilizers. Additionally, both farmers and consumers suffer large losses when crops fail due to illness. Testing the soil is important since it enables the measurement of soil fertility and, consequently, the forecast of crops. The pH of a soil is a gauge of its acidity and alkalinity. pH scales run from 0 to 14, with 7 considered neutral, acidic below that, and alkaline beyond that. We have suggested a system that will have a pH-reading device, and we will estimate the amount of nitrogen (N), phosphorus (P), and potassium (K) in the soil based on the pH of that soil. The use of Deep Learning has not been adequately investigated, despite recent advancements in the automated identification of certain diseases utilizing Machine Learning approaches. Such models are also difficult to utilize due to the high-quality data used in their training, a shortage of processing capacity, and the models' limited generalizability. In order to solve some of these problems and perhaps enhance crop productivity, we develop an open-source, user-friendly online application. We support recommendations for crops, fertilizers, and plant disease prevention in particular. We think that by doing this, farmers will be able to produce more crops with a given amount of land, and crop damage will be less likely to occur.

**Keywords**: Crop Recommendation, Fertilizer Recommendation, Plant Disease Classification, Machine Learning, Deep Learning, Image pre-processing.

**Introduction**

Around 58 percent of the population of our nation is employed in agriculture, making it one of the most significant sources of revenue. A farmer's average monthly income in 17 states, according to the 2016–17 Economic Survey, is Rs. 1700, which leads to farmer suicides and the use of agricultural land for non–agricultural use. Farmers typically choose crops that won't produce well in a certain soil or that are inappropriate for the season in which they will be planted, which is the cause of this. Lower yields are always the outcome of poor crop selection. If the family's only source of income is this, it is tough to make ends meet. Numerous variables, including climatic, geographic, organic, and economic considerations, have an impact on crop output.

On the other hand, conventional farming techniques are useless. It does not effectively utilize all of the resources at hand. Traditional approaches typically lead to soil nutrient depletion and tiredness because production is the main priority. The earth is exhausted since only some crops are produced there. Depending on the crop, there are different pH ranges that are optimal for plant growth. Due to the greater accessibility of nutrients in these pH levels, most plants flourish there. The pH of the soil affects the amount of nutrients that are readily available for plant growth. In more alkaline soil, phosphorus and other micronutrients are present in trace amounts.

The pH of the soil may reveal whether it is suitable for cultivation or whether it needs to be changed for the best possible growth of the plants. When used in conjunction with other factors, pH can help with fertilizer recommendations and the development of the best soil type for the area. The optimum ratio of nitrogen, potassium, and phosphorus is essential for the right kind of soil growth.

Farmers are confused about which crop to grow in a given season because of the volatility of the climate. Due to seasonal climatic variations as well as the importance of essential elements like soil, water, and air, the use of different fertilizers is similarly unpredictable. Crop yields are steadily declining in this setting. Offering farmers a user-friendly recommender system is the problem's answer. In this study, we present a model that takes these issues into account. The suggested methodology enables crop selection based on environmental and economic criteria in an effort to increase crop yields and meet the nation's rising food demand. In this study, we suggest a method that aids farmers in identifying plant ailments, suggesting the optimum crop for their soil, and advising fertilizers to assist them obtain the most yield possible.

**Literature review**

**Crop Recommendation**

In [1], the authors use - Random forests, Artificial Neural Nets, Support Vector Machines, etc. and conclude that Random forests work best for their dataset in crop recommendation. They also create a mobile application system which takes in

location data using GPS and predicts the crop yield for a given crop, in addition to recommending crops based on area and soil quality as input.

Similarly, [2] uses a majority voting on an ensemble of CHAID, Naive Bayes, K-NN and Random Trees for crop recommendation.

[3] data mining algorithms are used on agriculture data. The main criterion for this categorization is that if the pH value is greater than 8.5, the soil is unsuitable for crop cultivation; otherwise, it is. To overcome this problem the proposed system will give necessary suggestion to increase or decrease the pH value of soil

[4] the proposed system is related to increasing the net yield rate of the crop, based on the parameter related to the soil and atmosphere. The model gives the Crop prediction which can be carried out by using the "Bayesian algorithm". Data mining is used to extract a large amount of data from the data set and analyze that data to predict the crop yield and suggest the crop. The limitation of this includes atmospheric prediction that is not accurate.

**Plant Disease Detection**

The paper [5] used AlexNet and GoogLeNet with and without transfer learning on the PlantVillage dataset to achieve 99.35% accuracy. They also visualize activations and test on scraped data from Bing and Google Search. the authors use VGG, ResNet, Inception-V3 on an augmented version of PlantVillage dataset with 87K images, and conclude that VGG is the best for their settings.

Reference [6] provides an excellent review of over hundred papers which use Deep Learning for plant disease detection and classification. They point out that a majority of the papers use PlantVillage dataset for their task, and deploy Image-net based pre-trained models (VGG, ResNet, Inception, DenseNet, etc.) as their model backbones. They also mention several visualization techniques used for this task, like - heatmaps, saliency maps, feature maps, activation visualization, segmentation maps, etc.

[7] Plant Leaf Disease Detection and Classification Based on CNN with LVQ Algorithm - Melike Sardogan, Adem Tuncer, Yunus Ozen. Early disease detection is critical in agriculture for efficient crop yield. The diseases bacterial spot, late blight, septoria leaf spot, and yellow curved leaf have an impact on tomato crop quality. Automatic plant disease classification methods also aid in taking action after detecting symptoms of leaf diseases. This paper describes a Convolutional Neural Network (CNN) model and Learning Vector Quantization (LVQ) algorithm-based method for detecting and classifying tomato leaf disease. The dataset contains 500 images of tomato leaves with four disease symptoms. We created a CNN model for automated feature extraction and classification. Colour information is being extensively employed in plant leaf disease research .The filters in our model are applied to three channels depending on RGB components. For training the network, the LVQ was fed the output feature vector of the convolution component. The experimental findings show that the proposed approach accurately detects four forms of tomato leaf diseases.

**Fertilizer Recommendation**

A lot of research has been done in fertilizer recommendation and a majority of them [19]–[22] use the N, P, K, pH values of soil sometimes in addition with depth, temperature, weather, location, precipitation. The usual approach is to use rule based classification, but some approaches [8] also use clustering on fertilizer data using K-Means and Random Forests for recommendation.

[9] shows a case study related to wireless sensor networks for crop monitoring, growth and measurement of meteorological factors. The paper suggests farmers for application of specific pesticides and insecticides in stressful conditions. There was no focus on soil nutrients, the level of soil fertility and monitoring the crop growth or suggesting the crop for the next season in the above proposed paper. The solution to the issues in agriculture trends is proposed in this paper. The study suggests that farmers need to increase the fertility of soil and measure all parameters which are required to grow a crop in healthy condition.

**Methodology**

The following subsections describe our application and the machine learning involved in our experiments, with the implementation details as well as dataset and training information. First, we describe our application with the help of a flowchart, as well as block diagrams which explain how the application user-interface is designed. Then, we move on to our Machine Learning experiments, where we describe various models that we use, and other experimental details. Both the subsections are further split into nested subsections, namely - crop recommendation, fertilizer recommendation, and plant diseases detection. The machine learning subsection also includes an explanation of how we use LIME for interpretation, while the application subsection deals with the news feed implementation description.

**Crop Recommendation**

Dataset Description: The dataset was acquired from the Kaggle competition. It consists of 7 features namely -

N: Ratio of Nitrogen content in the soil,

P: Ratio of Phosphorus content in the soil,

K: ratio of Potassium content in the soil,

temperature: Temperature in degree Celsius,

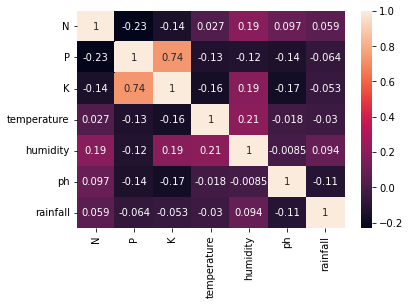
Humidity: relative humidity in %,

ph: ph value of the soil,

rainfall: rainfall in mm.

Using these 7 features, the aim is to forecast the crop type. There are 2200 samples overall, and there are a total of 22 class labels, some of which include rice, maize, coffee, muskmelon, etc. The dataset is properly balanced and does not require any extra imbalance handling method because there are 100 samples in each class.

Data Analysis :



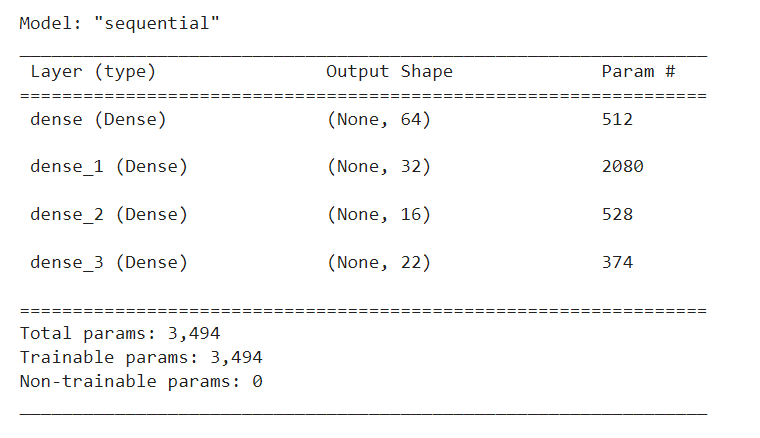
demonstrates the Pearson correlation between the variables, which is necessary for feature selection in order to eliminate duplicate and pointless properties. A redundant variable is one that can be obtained from another if two variables have a high correlation (value close to 1); otherwise, they do not. Only potassium and phosphorus exhibit a strong association in the presented figure which is 0.74 .

Approach: The dataset is divided in half in an 80:20 ratio, and after the split, it was normalized to have a mean value of 0 and a standard deviation of 1.

Method 1 : Using Artificial Neural Network

1. I built an ANN model that classifies the crop type using 64 features as input, and the final layer comprises 22 neurons that are softmax activated.
2. Compiled the model with following parameters:
   1. Optimizer : Adam
   2. Metrics : Accuracy
   3. Loss : Categorical cross entropy
3. Train the model with following parameters:
4. Batch size : 64
5. Epochs : 100
6. Callbacks : ModelCheckpoint, EarlyStopping, ReduceLROnPlateau

4. Tes the model performance on the unseen set.



Method 2 :Voting Classifier

A voting or ensemble classifier is developed using 7 different classifiers they are as follows:

1. MLPClassifier
2. LogisticRegression
3. CatBoostClassifier
4. RandomForestClassifier
5. DecisionTreeClassifier
6. GaussianNB
7. AdaBoostClassifier

With the exception of CatBoost, which is taken from the catboost library, we use the sklearn library for all the models. For the sake of our training, all other settings are set to their default values. Used the same classifiers for both inference and training.

**Crop Disease Detection**

Dataset Description: We take into account the PlantVillage dataset while detecting plant leaf disease. The collection includes 87,000 RGB images of both healthy and ill crops, each with a range of 38 class labels. There are 14 crops mentioned, and there are a total of 26 illnesses. The dataset has been expanded to add more diversity, and the photos have been transformed into 256\*256 pixels after being normalized by dividing them by 255. The training to validation ratio for the dataset is 75:25.

Approach:

1. We used the residual attention network to perform classification tasks. It has 11.3 million parameters. This model has two major advantages they are :
2. Residual connection : It is a gateless or open-gated variation of the HighwayNet, which was the first functionally complete, extremely deep feedforward neural network with hundreds of layers—much deeper than conventional CNN networks.
3. Attention network : Attention is a method that simulates cognitive attention in neural networks. The purpose of the impact is to encourage the network to give greater attention to the little but significant portions of the input data by enhancing some and reducing others.

2. Compiled the model with following parameters:

* 1. Optimizer : Adam (learning rate = 0.0001)
  2. Metrics : Accuracy
  3. Loss : Categorical cross entropy

3. Train the model with following parameters:

1. Batch size : 32
2. Epochs : 25
3. Callbacks : ModelCheckpoint, EarlyStopping, ReduceLROnPlateau

**Fertilizer Recommendation**

Dataset Description: We employ a proprietary dataset with 4 features—Crop, N, P, K, and pH—for fertilizer recommendation. There are 22 different crops with their perfect N, P, and K values, including rice, maize, coffee beans, mothbeans, soybeans, and coffee.

The farmer is advised to use fertilizer based on the pH range, N, P, or K value deficiency, and other factors.

Approach: We have used rule-based classification– To provide a plant the optimal fertilizer, a classification technique that uses fuzzy logic rules for class prediction. A fertilizer is advised based on how far a plant is from its ideal N, P, or K value. Based on whether the N/P/K values are high or low, we now recommend 6 different types of fertilizer.

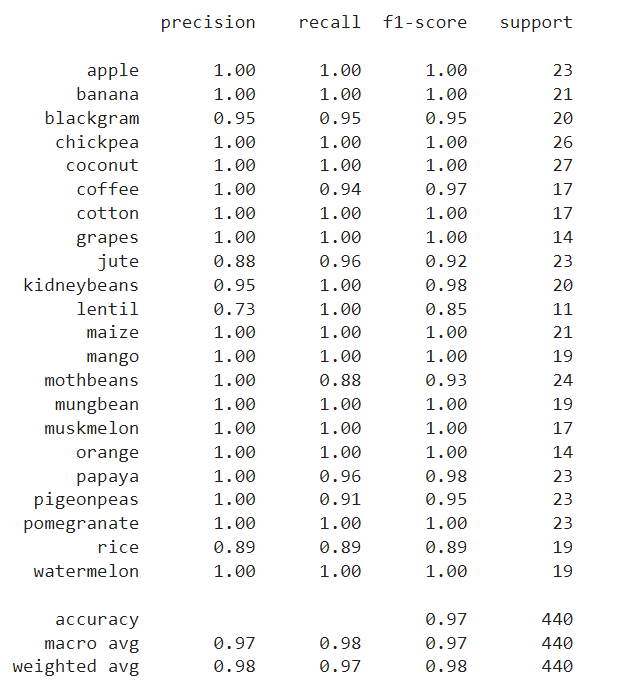
**Results and Discussions**

**Crop recommendation**

The results for our crop recommendation experiments are shown in the table.

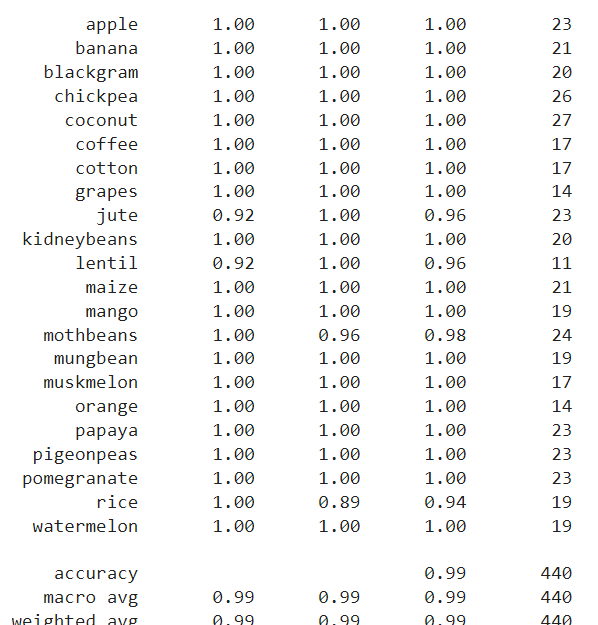
| **Model Type** | **Train-set accuracy** | **Test-set accuracy** |
| --- | --- | --- |
| Artificial Neural Network | 99.2 | 97.5 |
| Voting Classifier | 99.5 | 99 |

Voting Classifier models appear to perform best, however ANN performance can be enhanced by boosting the sample size.



ANN classification report

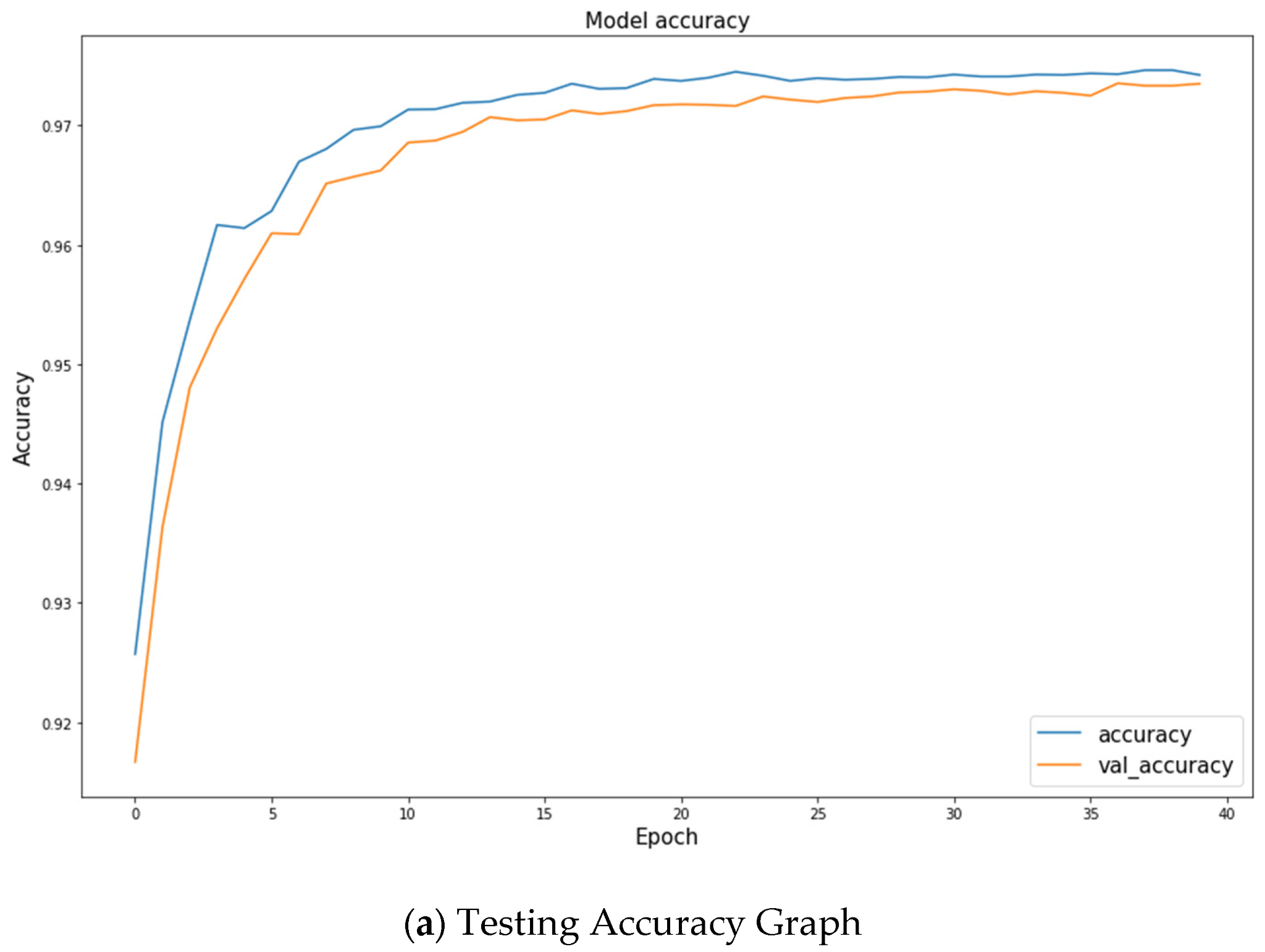
We can deduce from the provided figure that the overall performance of the ANN was confined to the rice, lentil, and jute classes.



Voting Classifier classification report

It demonstrates the generalisation of voting classifiers across all classes. save for jute and rice.

**Crop Disease Detection**



With a small gap between the train and test sets, the residual attention model did not overfit the dataset. On the train-set and test-set, it achieved 98.6% and 98.2% accuracy, respectively. However, by fine-tuning the model's hyperparameter with an optimization technique, the model's performance can be improved up to 99.5%. The main benefit of this model is that it can also handle data that is not distributed, however it is computationally expensive and needs a GPU setup to be trained.

**Conclusion and future work**

In this study, we present the "Farmer's Assistant," a user-friendly web application system built on machine learning and deep learning. Our system successfully enables the provision of a number of features, including the recommendation of crops using Artificial Neural Networks and Voting Classifiers, the recommendation of fertilizers using a rule-based system, and the detection of crop diseases using Residual Attention model on leaf images. The user can input data using forms on our user interface and receive responses immediately.

Only 22 crops are supported by our crop recommendation system at the moment, but we want to be able to apply sophisticated machine learning systems in the future to make recommendations that are even more precise.

After examining the plant disease model, we discovered that model performance declines when we use data outside of the domain. As a result, the best course of action is to generalize the model by gathering additional data from various distributions, training the model, and then expanding the data set.

Last but not least, we also want to offer fine-grained segmentations of the dataset's diseased section. Currently, a lack of such data prevents this from being possible. However, we may incorporate a segmentation annotation tool inside our application so that users can point out any shortcomings. Additionally, we can identify the unhealthy regions in the image by using some unsupervised methods. In our forthcoming work, we plan to include these features and fill in these gaps.

**References**

[1] S. M. Pande, P. K. Ramesh, A. Anmol, B. R. Aishwarya, K. Rohilla and K. Shaurya, “Crop Recommender System Using Machine Learning Approach”, 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), 2021, pp. 1066-1071, doi: 10.1109/ICCMC51019.2021.9418351.

[2] S. Pudumalar, E. Ramanujam, R. H. Rajashree, C. Kavya, T. Kiruthika and J. Nisha, “Crop recommendation system for precision agriculture”, 2016 Eighth International Conference on Advanced Computing (ICoAC), 2017, pp. 32-36, doi: 10.1109/ICoAC.2017.7951740

[3] Journal Article Mrs. N. Hemageetha, Dr. G.M. Nasira, “Analysis of soil condition based on pH value using Classification Technique”, IOSRJCE, Volume 18, Issue 6, Nov-Dec 2016. https://www.iosrjournals.org/iosrjce/papers/Vol18-issue6/Version3/I1806035054.pdf

[4] International Journal of Computer Science and Informatics. Jay Gholap, Anurag Ingole, Jayesh Gohil, Shailesh Gargade and Vahida Attar, "Soil Data Analysis Using Classification Techniques and Soil Attribute Prediction", IJCSI, Vol. 9, Issue 3, No 3, ISSN: 1694-0814, May 2012. https://citeseerx.ist.psu.edu/viewdoc/download? doi=10.1.1.402.2833&rep=rep1&type=pdf

[5] A. V. Panchal, S. C. Patel, K. Bagyalakshmi, P. Kumar, I. Raza Khan, M. Soni, “Image-based Plant Diseases Detection using Deep Learning”, Materials Today: Proceedings, 2021, ISSN 2214-7853, https://doi.org/10.1016/j.matpr.2021.07.281.

[6] M. H. Saleem, J. Potgieter, K. M. Arif, “Plant Disease Detection and Classification by Deep Learning”, Plants, 2019, 8, no. 11: 468. <https://doi.org/10.3390/plants8110468>

[7] Melike Sardogan, Adem Tuncer, Yunus OzenPlant Leaf Disease Detection and Classification Based on CNN with LVQ Algorithm, 10th December-2018 https://ieeexplore.ieee.org/abstract/document/85 66635

[8] K.S. Subramanian, “Design and Implementation of Fertilizer Recommendation System for Farmers”, 2020

.

[9] Luca Bencini, Davide Di Palma, Giovanni Collodi, G. Manes and Antonio Manes, "Agricultural monitoring based on wireless sensor network technology: Real long life deployments for physiology and pathogens control.". Third International Conference on Sensor Technologies and Applications. IEEE, 2009.

[10] S Pavani and AS Beulet. Heuristic prediction of crop yield using machine learning technique. International Journal of Engineering and Advanced Technology (IJEAT), pages 135–138, 2019.

[11] K Sriram Rakshith, G Deepak, M Rajesh, KS Sudharshan, S Vasanth, and N Harish Kumar. A survey on crop prediction using machine learning approach. Int. J. Res. App. Sci. Eng. Techno.(IJRASET)., 7(4):3231–3234, 2019.