INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

*** Corresponding author**.

Tel: +91-9900114849
naidila@jssateb.ac.in

An artificial intelligence based approach for increasing agricultural yield

Naidila Sadashiv[1]*, **Ankur Garg[2], Avinash Kumar[2], Ashwani Verma[2], Chandan Adhikari[2], S M Dilip Kumar[3]**

1 Associate Professor, Department of CSE, JSS Academy of Technical Education, Bangalore, 560060, India. Tel.: +91-9900114849
2 Department of CSE, JSS Academy of Technical Education, Bangalore, 560060, India
3 Professor, Department of CSE, University Visvesvaraya College of Engineering, Bangalore, 560001, India

Abstract

Background/Objectives: Identification of a suitable crop based on soil and climatic condition along with plant disease detection are very much essential as they boost the agricultural yield. Monitoring these parameters is not largely carried out as they are very laborious and require expertise in the field, thus curtails the overall productivity. To address these challenges an artificial intelligence based techniques were applied to predict a suitable crop and also to detect the plant leaf disease at an early stage are been presented. **Methods/Statistical analysis**: Crop predictions are carried out by leveraging two machine learning approaches, Logistic Regression (LR) and Support Vector Machine (SVM) on the agriculture dataset. Convolution Neural Network (CNN) with ResNet152 architecture is used for the detection of plant disease. An open dataset of 54,306 photos of healthy and diseased crops are considered during the performance evaluation. **Findings**: Crop predictions based on LR and SVM have achieved an accuracy of 93% and 97% respectively across a class of 13 crops. CNN based model predicts disease among 38 different categories from 14 specific crops with an accuracy of 96%. **Novelty/Applications**: The proposed approach will be beneficial to farmers to identify suitable crops and have plant leaf disease under control.

Keywords: CNN; crop prediction; leaf disease detection; logistic regression; machine learning; SVM

1 Introduction

India primarily depends on agriculture and plays a major role in India's growth. Agriculture is one of the Indian economy's most important fields. India produces the largest quantity of wheat, spices, spice, rice and bajra. According to the 2019-20 Survey on Indian Economy, more than 60% of India's total workers are employed in the agricultural sector and totally backs roughly 19-20% to the GDP. This sector is the main source of income for rural households. Hence, agriculture is considered as a major pillar of the Indian Economy. India has more than 1.2 billion people and to feed everyone

its agriculture sector occupies nearly 55% of the total land area. The productivity of agriculture depends on identifying the suitable crop to be grown, parameters of the soil, such as soil moisture, nitrogen, surface temperature, potassium, crop rotation, phosphorus, and on the natural climate conditions like heat, humidity and rainfall[1–3]. Crop disease is an important factor that affects productivity and is a major concern for Indian framers. Efforts in the application of technological advances to improve crop growth have been made in the field of agriculture however, there still exists a need to address these issues more efficiently in order to grow more healthy crops and improve the overall yield[4–6]. Automated and technology-enabled farming will draw a large number of people to farming leading to a resurgence in food productivity.

Thus, a framework for the identification of a suitable crop for cultivation and also addressing leaf disease issues are very essential to improve crop yield. Given these issues, the proposed work provides an Artificial Intelligence (AI) - based solution to the farmers irrespective of their experience to increase their crop productivity. Artificial Intelligence is renovating in various industries. A subset of Artificial Intelligence is machine learning which tends to give systems the ability to learn and enhance automatically from knowledge and experience without explicit programming. The proposed system relies on machine learning algorithms, Logistics Regression and Support Vector Machine to predict a suitable crop for the soil based on soil parameters such as humidity, temperature, moisture, rainfall and pH. A plant shows some visible symptoms of the disease on its leaves. The noticeable features like form, size, dryness, wilting, are very helpful in identifying the state of the plant. The proposed system uses a Convolutional neural network to identify plant leaf disease. The underlying algorithm uses neural networks that have layers of neurons with a connectivity pattern. CNN comes with a combination of improved computing capabilities and availability of large data sets of images which results in high accuracy of the result[7]. Along with CNN, the ResNet architecture has also been used and it provides both better performance and improved train time. Various machine learning algorithms-based solutions are explored in the agricultural sector for crop prediction, as well as leaf disease detection and some of these works, are been discussed here.

The artificial neural network is used for the prediction of crop yields by sensing the various environmental and soil parameters[8]. Parameters include water depth, the form of soil, temperature, pressure, rainfall, phosphorus, nitrogen, potassium, and organic materials. The effect of these parameters was analyzed and tested empirically on the crop yield rate. The work also proposed an effective crop based on its estimation to advance crop production volume.

Comparison of different techniques of data mining was carried out[9] and based on precision results applied the technology to collect and analyze the data to make a prediction of suitable crops. Techniques included data collection, decision tree and data filtering. Crop predictions were performed and also proposed approaches for utilization of space between the plants to grow other plants without affecting the crop yield.

Authors have hypothesized an explorative data study and explored the essence of the specific predictive analytics for yield prediction[10]. Based on a test dataset collected, different regression techniques were used to define each property and evaluate it. Specific regression methods discussed include the regression of linear, multiple linear, nonlinear, logistic, polynomial, and slope. A comparative study of data analytics of the various algorithms was carried to determine the efficient approximation method.

Authors[11] have worked on rice production prediction and have used Bayesian networks considering precipitation, range of temperatures, and reference crop evapotranspiration for a particular season. BayesNet and NaiveBayes classifiers were used and were evident from the result that the performance of BayesNet was better than NaiveBayes.

The use of texture statistics was proposed[12] to detect the disease of the leaves in plants by first translating the color transformation structure of RGB into Hue Saturation Value (HSV) space since HSV is a powerful color descriptor. Covering and removal of green pixels were performed at the pre-computed threshold stage followed by segmentation using 32X32 image size. Such segments were used to examine texture via the occurrence of color matrices. Finally, the characteristics of texture were contrasted with the characteristics of regular-leaf texture to detect disease in the leaf.

An automated classification[13] method based on pixels, was designed to classify unhealthy areas in images of leaves using linear SVM. The work also discussed other recognition methods, which also rely mainly on the data related to the colors of the leaf, and how the presented algorithm could be extended to other functions.

The work on plant disease has provided a method to detect disease in plants using random forest classifiers[14]. First preprocessing was done to reduce the image size to a standard unit and then used HoG, a feature descriptor to detect objects and extracting features. The RGB value of the images was transformed into the grayscale to compute the Hue value. The work has demonstrated that random forest provides more accuracy than SVM, LR with even fewer images.

Soya bean leaf disease detection method based on a transfer learning was carried out by using pre-trained AlexNet and GoogleNet CNN[15]. The work has identified three soybean diseases and has been validated by a fivefold cross-validation method.

Convolutional neural networks based model for identification of Apple leaf diseases that mainly consists of Mosaic, Brown spot, Rust and Grey spot are been carried out using GoogLeNet Inception structure and Rainbow concatenation. Techniques such as data augmentation and image annotation were performed on the real-time images to construct the apple dataset[16].

In these related works, crop prediction is been carried out considering a few of the soil and environmental conditions and some of the works are on leaf disease detection. The proposed work proposes a complete solution from a farmer's perspective right from identifying a suitable crop to identifying the disease and suggest appropriate fertilizers as remedies.

1.1 Objectives

The key contributions of this paper are as follows:

- Machine learning-based approaches to predict suitable crop for cultivation.
- CNN based plant leaf disease detection to increase agricultural yield.
- Exploration of the pattern in the dataset to check its applicability for evaluation.
- Performance evaluation of the models for crop prediction and leaf disease detection.

2 Proposed System

This section presents the methodology used for crop prediction and plant leaf disease detection.

2.1 Crop Prediction

Following steps are carried for crop prediction:

- Input
- Data Acquisition
- Data Pre-processing
- Machine Learning Model
- Output

1. Input: The forecast of the harvest relies on temperature, humidity, moisture, rainfall, and pH to envisage the yield precisely. Entirely these aspects are site dependent and thus tell the user which crop is most suitable.

2. Data Acquisition: Involves the collection of raw agricultural data for analysis from a government repository managed by the Department of Agriculture Cooperation & Farmers Welfare under the Government of India.

3. Data Pre-Processing: The data is transformed to a state so that the analysis can be made to get better results than on the raw data.

4. Machine Learning Model: SVM and Logistic Regression are used in the proposed work to forecast the harvest. The prophecy is grounded on previous crop production statistics by identification of observable climate and soil constraints and comparison with existing situations to foresee crop accuracy and practicality.

5. Output: The utmost gainful crop is projected by the model using Logistic Regression and SVM algorithm. The end-user is also provided with recommendations of crop along with suitable fertilizers for the crop.

2.1.1 Support Vector Machine (SVM)

SVM is supervised learning in machine learning and the prevalent classification algorithms correspondingly used for regression analysis. SVM supports the kernel method also known as the SVM kernel that helps us to confront nonlinearity.

Alike a human being classifies a set object considering all their bodily and pictorial features and then identifies them based on erstwhile familiarity, a machine using SVM classifies a new object grounded on the data given. The more structures in the data the easier it is to recognize and differentiate both. SVM's goal is to draw a stripe that best splits the two statistics groups taken into consideration. The sideline is the area that splits the two scattered dashed lines as revealed in Figure 1. Further the width of the margin M the improved the classes are separated. The support vectors are the dots and triangles scattered in Figure 1 which are the statistical arguments that move through each of the dashed lines. These arguments are called vectors of support since they fit into the margins and thus to the classifier aforementioned. Such support vectors are fundamentally the data points that lie adjacent to the borderline of either class that is expected to be in either class. Then the SVM generates a hyperplane H with the equivalent width separating the two classes at the ideal remoteness between the two classes. The line is substituted by a hyperplane which splits multidimensional spaces in the event of more than two features and numerous dimensions.
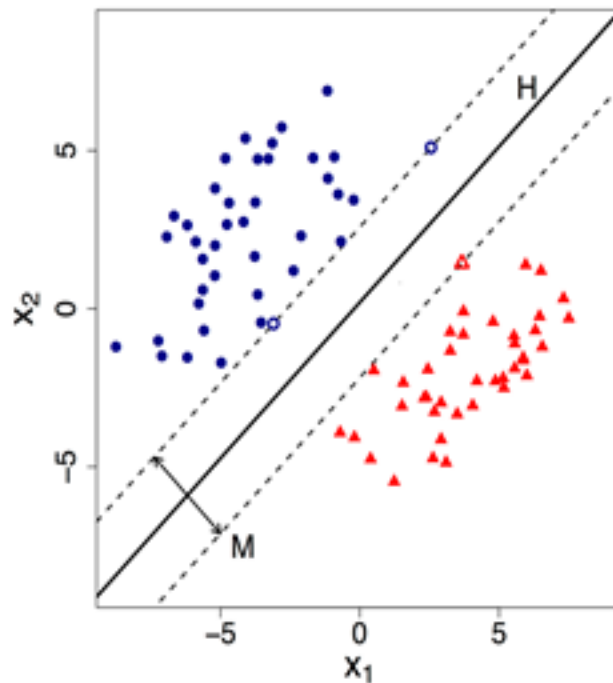
Fig 1. Linear SVM classification of response triangles and circles based on predictors (X1 and X2)

The first phase in the working of the SVM would be to load the set of data consisting of temperature, moisture, humidity, rainfall and pH. Next, describe the goal variable followed by splitting the dataset into a test set and training set. Later pre-processing is performed by standardizing the data that scales values between 0-1 in the dataset. For many machine learning estimators, adjustment of a dataset is an elementary requirement. Then we transform and fit the training data and fit the testing data into the model. After fitting the training and testing data we initialize the support vector machine with the parameter with decision function shape as one vs one which determines the probability of each class. Finally, fit the training data into the model and predict the output.

2.1.2 Logistic Regression (LR)

Logistic Regression is a supervised model for classification, it calculates the logits (or Score) using the linear regression model and uses it to train a classification model in the later stages. This trained classification model performs the multi-classification task. Multinomial logistic regression is a type of logistic regression for the prediction of an objective variable with a class of more than two. The logistic regression is used with the softmax function and the cross-entropy as the loss function.

Linear Model: The linear equation of the model is like the linear equation in the linear regression model.

$$y = B0 + B1 * x \tag{1}$$

Where the x is the input value, y is the output for the input values, B0 and B1 are the intercept and coefficient to each input respectively in (1).

The coefficient will update in the training phase and this is done with the help of parameter optimization. The Logits (B1*x) are the linear model outputs and are dependent on coefficients or weights.

Softmax Function: The Softmax function computes the probabilities for the specified score. It calculates the probabilities of each target class over all possible target classes. With the Softmax function, the greatest probability value for high scores and less probability for the remainder is returned.

Logistic Regression tries to find a relationship between the different features and the probabilities of a target. It is usually used for Binomial classification containing only two classes. The proposed work deals with Multinomial classification with 13 different crop classes as targets. The first phase of the LR working would be to load the set of data consisting of temperature, moisture, humidity, rainfall and pH. The dataset is split into two parts: the training set and the testing set with the ratio of 80:20.

Initialize the logistic regression function parameters and the class is set as multinomial for multi-class classification. Once the model is trained and set with different parameters, the model will predict the crop according to different input values.

2.2 Plant disease detection

The approach used in designing plant disease detection is Deep Convolution Neural Network. A CNN model is made up of an input layer, an output layer and several hidden layers. The hidden layers consist mainly of the convolution layer, dropout Layer, Rectified Linear Unit, pooling layer and normalization layers. In this work, the input to the model is a leaf image, and the output is the name of a class. Input leaf images are taken from the open repository that has leaf images of cereal crops - Soybean, cash crops, vegetables –pepper bell and fruits which includes Apple, Corn, Orange, Peach and Raspberry. The model detects the different diseases among the leaf dataset, such as apple scab, powdery mildew, cedar apple rust and black rot in the case of apple leaves and accordingly a remedy has been recommended displayed.

Transfer learning is applied by following a completely trained model applied to a defined class of images and re-train for new ones. Earlier research has shown that for a variety of applications, transfer learning is successful[17,18] and has much lower computational requirements and is beneficial in the proposed model.

Overview of CNN layers:

- Input Layer: The plant leaf image of size 224x224x3 is provided as an input to the model. Data augmentation is done to increase the dataset size and is transferred to the CNN model. Since large volumes of data are essential for deep-learning models, artificial data generation by extending the original data set are performed by multiple transformations, including rotations, skewing, transposing, cropping, and zooming while the picture label is retained.
- Convolutional Layer: The key process of the convolution layer is convolution which is a mathematical operation. The input image is passed through a convolutional layer to abstract the image to a feature map. To generate the output function maps, the input is transformed with the filters, called kernels. The convolution output can be denoted as,

$$x_j^1 = f\left(\sum_{i \varepsilon M_j} x_i^{l-1} * k_{ij}^l + b_j^l\right) \text{ (2)}$$

Here, x_j represents the set of output feature maps, M_j represents the set of input maps, k_{ij} represents the kernel for convolution, b_j represents the bias term, l represents the layers.

- ReLU Layer: The Rectified Linear Unit layer is a rectifier used to introduce non-linearity to the system. It performs linear operations during every convolution process. Hence, it is introduced after every convolution layer and sets all negative values for the activation to 0. It carries out a thresholding operation for every element defined by the max function, and it speeds up the training process.

$$f(x) = max(0, x) \tag{3}$$

- Max-Pooling Layer: This layer divides the input into several overlying modules. This layer provides the maximum of the elements in each module to create a reduced volume output while maintaining the essential input information.
- Dropout Layer: The fundamental principle of the dropout layer is to deactivate or drop the input elements with a certain probability to allow the individual neurons to learn features less dependent on their environment. This process only occurs during the time of training.
- Batch Normalization Layer: Batch Normalization adds the framework of normalization within each layer to reduce the issue of internal covariate change. Regardless of this standardization of layers between each fully connected layer, the range of input distribution for each layer remains the same regardless of the previous layer's changes. All inputs centered around 0 are standardized. That way, there is not much improvement in every input layer. And, at the same time, layers in the network will benefit from back-propagation, without having to learn from the previous layer. That fastens the network.
- Fully Connected Layer: Every neuron in this layer is connected to all the neurons in the previous layer in the completely connected layer, thus integrating all of the features acquired from the previous layer to promote classification. At the output, this layer generates an N-dimensional vector, where N is the size of groups. In our case, the number N is 38 and represents the different categories of disease.
- Output Layer: The output layer comprises a softmax layer that is followed by the classification layer. The softmax layer outputs a distribution of probability based on which the model of the network classifies disease as a class with the maximum value of probability.
- ResNet based Deep CNN model: ResNet[19] implements skip connections to match the input from the preceding layer to the next layer without any input changes. This architecture is used for image classification, detection, and localization. In

addition to using the architecture of a set of stacked convolutional layers, finally, it is followed by a fully connected layer, it uses skip connections or shortcuts ResNet uses 3x3 convolutions and employs heavy use of batch normalization. In the proposed work, ResNet-152 which has 152 layers is used.

- Training ResNet: The dataset is broken down into training and test sets (80:20). The current weights of the ResNet framework were retrained in the proposed model to classify the image datasets of the plant disease using the vast amount of visual information already obtained by ResNet. We froze the model to the last layer and remove the pre-trained model's last predicting layer and replace it with our predicting layers. These layers include information concerning the grouping of the extracted characteristics into probability and class labels. The last fully connected is of the same size as class numbers.

3 Experimental Evaluation

This section discusses the experimental setup and platform used to carry out the testing of the proposed model. Finally, the experimental results are analyzed and presented.

3.1 Experimental setup

This experiment is performed on Windows 10 platform powered by Intel(R) Core (TM) i5-9300H CPU @2.40Ghz that was accelerated by an NVIDIA GTX1050 GPU. Our System comprises 8GB RAM and high-speed inbuilt Intel GPU. The front end is developed in Python using the libraries Tkinter and PIL. Crop Prediction has been carried out based on the SVM and LR machine learning techniques while leaf disease detection is a CNN-based model.

3.2 Dataset

The methodology used in Crop Prediction is SVM and LR and the dataset contains 5 variables with 5600 lines of data values under each variable. The dataset is divided into training and testing datasets as 80% and 20% respectively. The training set has 4480 lines and the remaining 1120 lines are used for testing. Testing based on random values by the user was also carried out. Over 50K images of healthy and infected leaf data set from Kaggle are considered for training and testing purpose[20–22].

3.3 Experimental results and analyses

This section discusses the results achieved for crop prediction and plant leaf disease detection.

3.3.1 Results of crop prediction based on SVM and LR

During pre-processing of the crop prediction data set correlation matrix and data histogram are calculated to understand the pattern in the data set as presented in Figures 2 and 3. Later, different values for temperature, moisture, humidity, rainfall and pH are taken from the user and a suitable crop is predicted based on the given condition.

- Correlation Matrix
  The matrix of correlations is used to approximate the historical linear association between collections of variables in the dataset of crop prediction models. To swiftly check correlations between different columns, the data set is visualized as a heatmap in Figure 2. The columns in the dataset are Temp, Humidity, Soil Moisture, Rainfall and pH. Heatmap is interpreted as the brighter the color, the larger the degree of the association and vice versa. Value of the correlation should be less for a better dataset. As we can see from Figure 2, the correlation between temp and rainfall is the least and the correlation between any two different pairs of variables is less which means the dataset is perfectly chosen and contains a variety of values. In the main diagonal, the color is lightest which means they have the highest correlation and should have to as the variables are the same in the diagonal that means the dataset is perfectly classified.

- Data Histogram
  A histogram is a figurative representation of the distribution of numerical data. The probability distribution of the columns within the data set is represented as data histograms in Figure 3. Here, five different parameters were taken to predict the crop and are represented in the form of a frequency distribution graph. It shows the range of values in the dataset for a particular type of variable with the frequency of each value for that variable.

The x-axis denotes the magnitude of the particular column and y-axis reflects the value's frequency. The range of the x-axis denotes the range of the particular variable while the y-axis denotes the number of values present in the dataset for x-axis value.
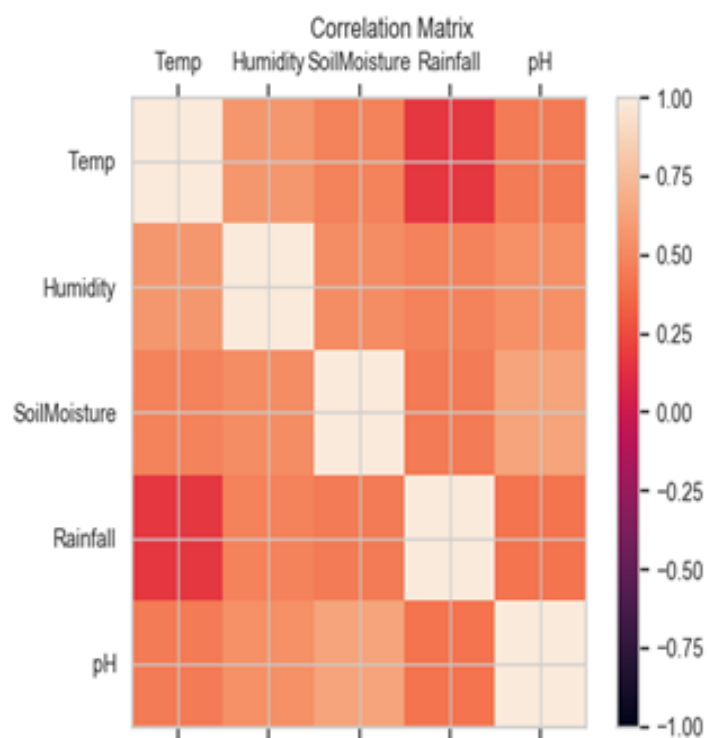
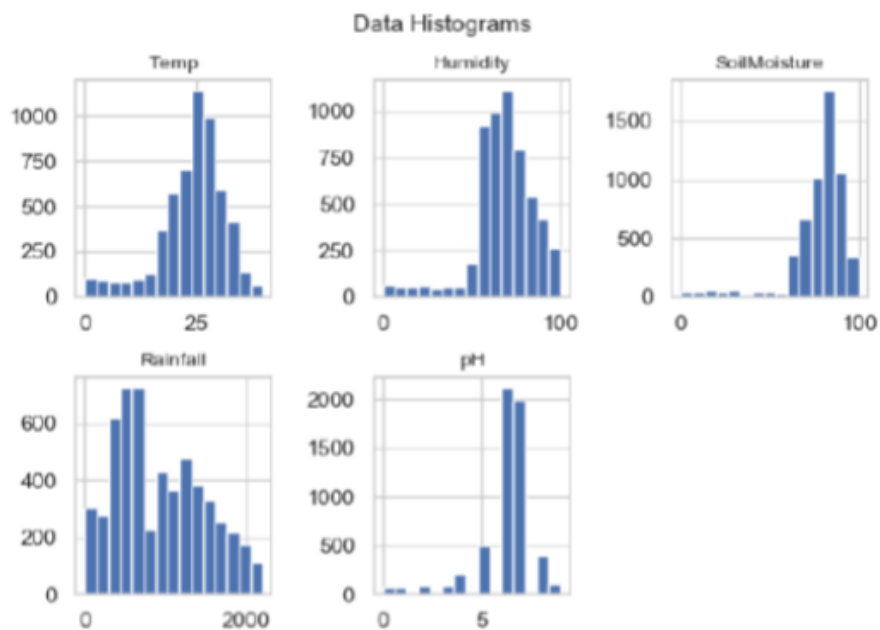**Fig 2.** Correlation matrix for crop prediction dataset

Fig 3. Data histogram – temperature, humidity, soil moisture, rainfall and pH for crop prediction dataset

For example, the value ranges from 0 to 100 for Humidity, and the frequency is between 0-1500, means for some value between 0 and 100 on the x-axis, the frequency maybe 1500 or less.

Evaluation of the accuracy and errors are carried out with the confusion matrix and evaluation graph for the proposed crop prediction model based on SVM and LR algorithms.

- Confusion Matrix
  A confusion matrix is sometimes used to define the performance of a classifier on a test data set on which the true values are identified. Throughout the specified estimates, X-axis represents the all-possible expected label or projected value while the Y-axis represents the real label or actual value. Those findings are reflected by the heat map. As the value in the heatmap moves from 0 (light blue) to 100 (dark blue), the outcomes also move from false-negative to false positive to true negative to true positive at the darkest part in Figures 4 and 5.
- a) Prediction Results based on SVM
  Figure 4 represents that most of the crops predicted by the SVM algorithm are exactly the same as required or trained. The values of the true label and predicted label are the same and this denotes that the accuracy of the algorithm is very high and has achieved 97% accuracy.
  Example: From Figure 4, taking the crop pea from the true label and moving horizontally we get a value 1 that corresponds to the crop oats in the predicted label. This shows that there is 1 instance when the true value should be pea but the classifier predicted the value as oats.
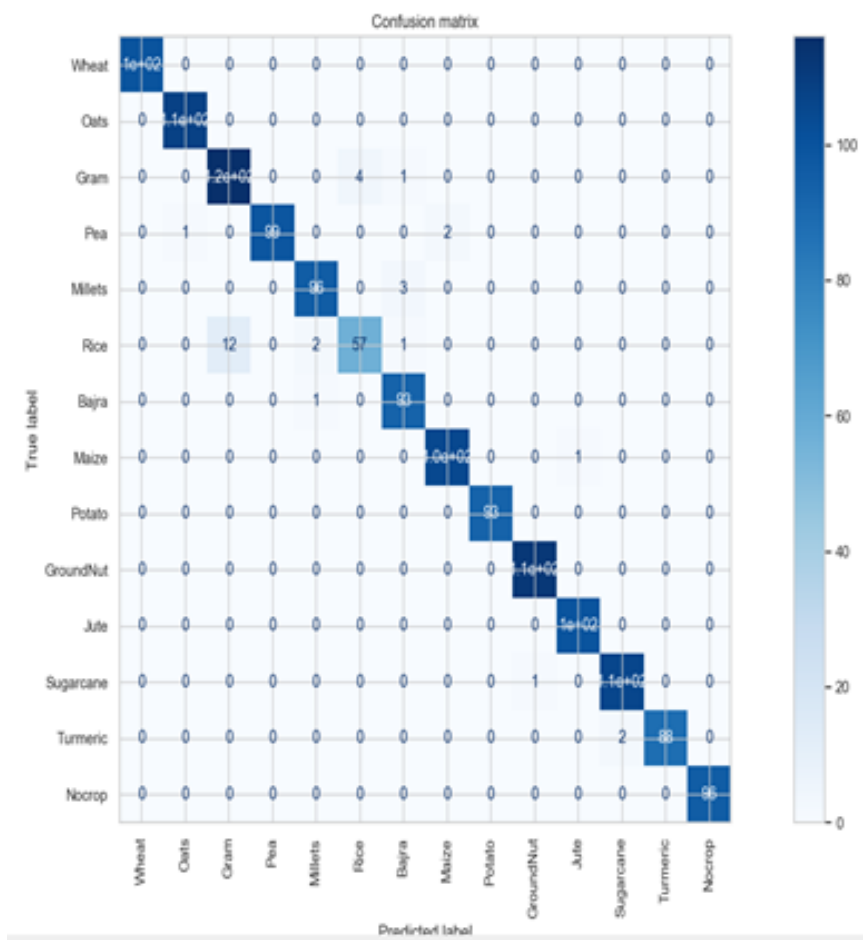

Fig 4. Confusion matrix for SVM

This example shows that for just 1 instance out of thousands of such instances the value is predicted wrong for pea while the other values are predicted correctly.

- b) Prediction Results based on LR
  Figure 5 represents the confusion matrix for the LR algorithm. In the main diagonal, the color is bright and the values of the fields in the diagonal are very high.
  This represents that most of the crops predicted by the LR algorithm are correct and are as per the trained model. This means that the values of a true label and predicted label are the same and this denotes that the accuracy of the algorithm is quite high and has achieved an accuracy of 93%.
  Also, the error value is quite less and there are very few cases when the predicted and true label values are different which again justifies the accuracy of 93%.
  Example: From Figure 5, taking the crop maize from the true label and moving horizontally we get a value 1 that corresponds to the crop gram in the predicted label. This shows that there is 1 instance when the true value should be maize but the classifier predicted the value as gram.
  This example shows that for just 1 instance out of thousands of such instances the value is predicted wrong for maize while the other values are predicted correctly. The main diagonal represents all the correct values obtained.
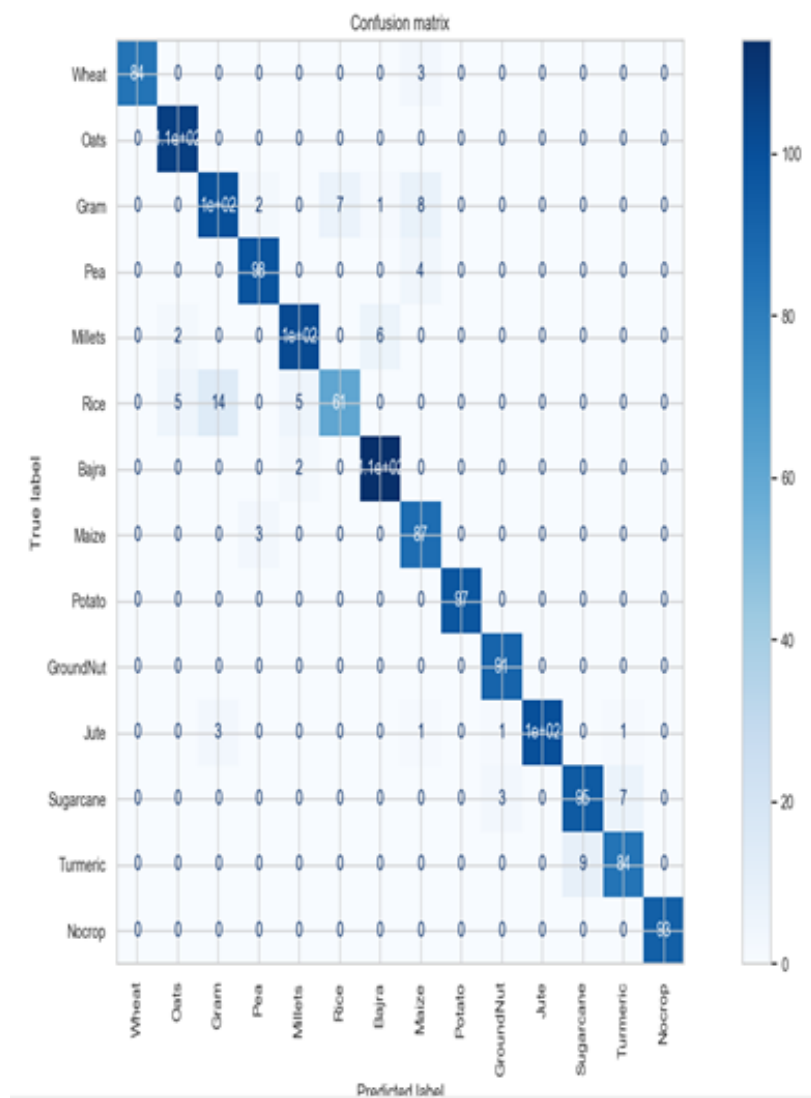

Fig 5. Confusion matrix for LR

- Performance Evaluation Metrics for Crop Prediction Models
  The models are evaluated using the five-primary metrics: Mean absolute error (MAE), mean square error (RMSE), Mean square error (MSE), R-Squared and accuracy. They all require two lists as parameters, one is predicted values and the other, the true values.
  MAE: For a set of predictions, MAE calculates the average magnitude of errors/mistakes, without considering their direction. This is the easiest to comprehend.
  RMSE: It is quite similar to MSE, but the result is squarely rooted to make it more interpretable. MSE: MSE is the square mean discrepancy between the expected values and real values. It is identical to MAE but noise is amplified.
  R-Squared: It is a statistical fit measure that indicates how much variation of a dependent variable in a regression model is from the independent variables.
  Accuracy: Accuracy is the ratio of the number of accurate observations to a total number of sample data.
- Performance Evaluation based on SVM
  Accuracy in the case of SVM is 97%. MSE and MAE in the case of SVM are low which indicates that the model evaluated using SVM is highly accurate and R-Squared value is also high indicating the same. RMSE value is also small, 0.5 in the case of SVM. Lower the value of RMSE indicates a smaller number of errors. All these parameters as presented in Figure 6 clearly state that the SVM has higher accuracy.
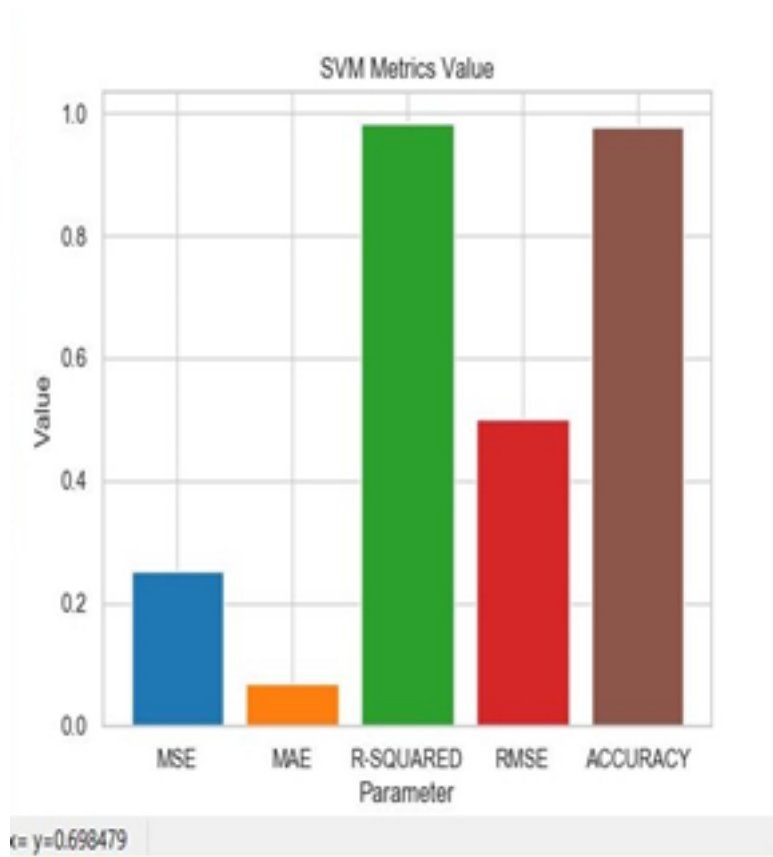

Fig 6. Evaluation metrics for SVM

- Performance Evaluation based on LR
  Accuracy in the case of LR is 93%. MSE and MAE in the case of LR are higher than SVM but still not very high which indicates that the model evaluated using LR is also very accurate and the R-Squared value for LR is also high which indicates the same. RMSE value in the case of LR is a bit high as it is calculated based on the SME value of the algorithm. The more the value of SME, the more will be the value for RMSE. All these parameters clearly state that the LR also has higher accuracy but less than SVM as presented in Figure 7.
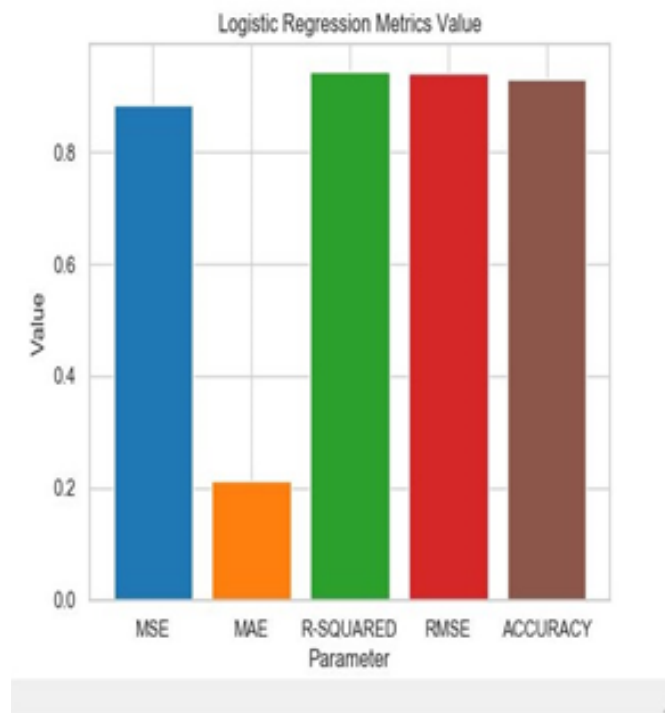
Fig 7. Evaluation metrics for LR

3.3.2 Results of leaf disease detection based on CNN

Fig 8. Sample diseased leaf images

Leaf Disease Detection is carried out using CNN for classifying the leaf image and predicting the disease in it. The image is uploaded and is fed into the CNN model. The model performs preprocessing and scaling on the image to remove noise from the image. The images in Figure 8 are sample diseased leaves and are uploaded in the GUI for preprocessing. The first image is an apple leaf with the cedar rust disease in it.

After uploading this image, the proposed model will check for the disease in it and will return the name of the plant to which this leaf belongs along with the name of the disease in it as shown in Figure 9.
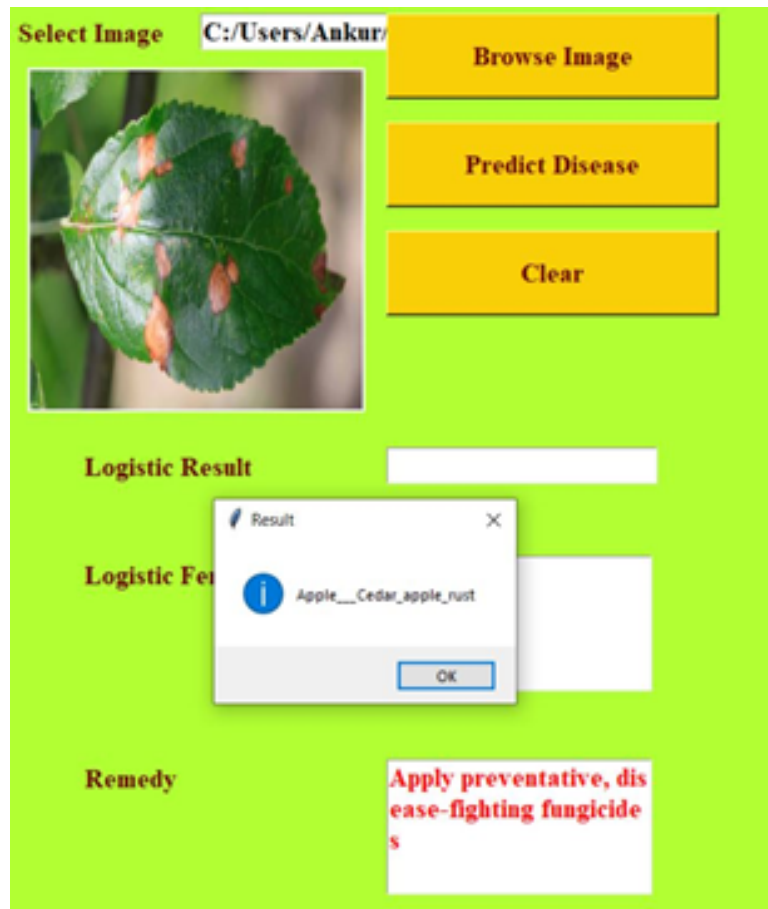

Fig 9. Detection of diseased leaf

A remedy is also recommended with the disease name to cure it before it will do any more harm to the crops and plants.

For a healthy leaf as in Figure 10, the proposed model will detect the same and responds with an appropriate message displaying the plant name with the keyword healthy and also no remedy will be provided as shown in Figure 11.
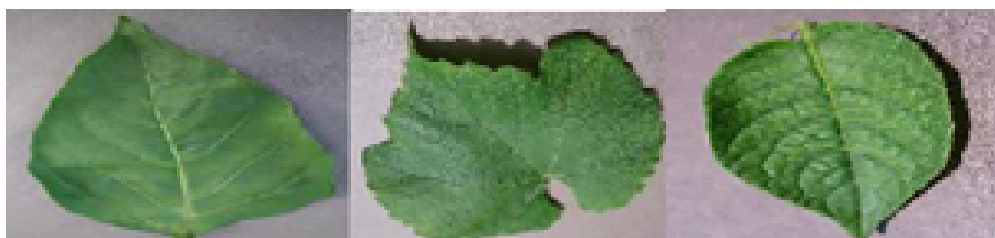

Fig 10. Sample healthy leaf images

Fig 11. Detection of a healthy leaf

For a randomly different image that is not a leaf image, it will respond with an appropriate message "Not a Leaf! Please upload a leaf image only".

4 Conclusion

An artificial intelligence-based crop prediction model using machine learning algorithms, which will help a farmer to identify a suitable crop based on the environmental parameters and soil nutrients. CNN based leaf disease detection across 14 crops using sample leaf images from the dataset was carried out. Results demonstrated an accuracy of 97% for the crop prediction based on the Support Vector Machine and with the Logistics Regression approach accuracy of 93% was achieved for the agricoop dataset. The CNN based leaf disease detection model achieved an accuracy of 96% for the large open dataset. AI-based crop prediction and leaf disease detection models will help for crop identification and thereby contribute towards the increase in agricultural yield and the country's economy.

Future Scope

As future work, various other parameters, including the uncertainty in the prediction that impacts the crops yield along with the generated datasets using augmentation and image annotation techniques that cover a variety of crops can be incorporated.

References

1) Sellam V, Poovammal E. Prediction of Crop Yield using Regression Analysis. *Indian Journal of Science and Technology*. 2016;9(38):1–5. Available from: https://dx.doi.org/10.17485/ijst/2016/v9i38/91714.
2) Kumar KK, Kumar KR, Ashrit RG, Deshpande NR, Hansen JW. Climate impacts on Indian agriculture. *International Journal of Climatology*. 2004;24(11):1375–1393. Available from: https://dx.doi.org/10.1002/joc.1081.
3) Kumar R, Singh MP, Kumar P, Singh JP. Crop Selection Method to maximize crop yield rate using machine learning technique. In: 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM). 2015;p. 138–145. Available from: https://doi.org/10.1109/ICSTM.2015.7225403.
4) Rehman TU, Mahmud MS, Chang YK, Jin J, Shin J. Current and future applications of statistical machine learning algorithms for agricultural machine vision systems. *Computers and Electronics in Agriculture*. 2019;156:585–605. Available from: https://dx.doi.org/10.1016/j.compag.2018.12.006.
5) Shreyas J, Kumar SMD. A Survey on Computational Intelligence Techniques for Internet of Things. *International Conference on Communication and Intelligent Systems*. 2020;p. 271–282. Available from: https://DOI.org/10.1007/978-981-15-3325-9_21.
6) Srinidhi NN, Sagar CS, Chethan SD, Shreyas J, Kumar D, S. Machine Learning Based Efficient Multi-copy Routing for OppIoT Networks. *Advances in Computational Intelligence, Security and Internet of Things*. 2020;p. 288–302. Available from: https://DOI.org/10.1007/978-981-15-3666-3_24.
7) Khan A, Sohail A, Zahoora U, Qureshi AS. A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review*. 2020;53(8):5455–5516. Available from: https://dx.doi.org/10.1007/s10462-020-09825-6.
8) Snehal, Dahikar S, Rode VS. Agricultural Crop Yield Prediction Using Artificial Neural Network Approach". *International Journal of Innovative Research in Electrical, Electronic, Instrumentation and Control Engineering*. 2014;2(1). Available from: https://www.semanticscholar.org/paper/Agricultural-Crop-Yield-Prediction-Using-Artificial-Dahikar-Rode/aa36006afdd0ee0c9d08a13684d4e6255d6c63bc.
9) Pandey A, Purohit S, Jadhav S, Shah K. Optimum Crop Prediction using Data Mining and Machine Learning techniques". *International Journal for Research in Applied Science & Engineering Technology*. 2014. Available from: https://doi.org/10.22214/ijraset.2019.3436.
10) Nagini R, Rajnikanth TV, Kiranmayee BV. Agriculture Yield Prediction Using Predictive Analytic Techniques. In: and others, editor. 2nd International Conference on Contemporary Computing and Informatics. 2016. Available from: https://doi.org/10.1109/IC3I.2016.7918789.
11) Gandhi N, Armstrong LJ. Owaiz petkar Amiya Kumar Tripathy, "Rice Crop Yield Prediction in India using Support Vector Machines. *International Joint Conference on Computer Science and Software Engineering*. 2016;p. 1–5. Available from: https://doi.org/10.1109/JCSSE.2016.7748856.
12) Dhaygude SB, Kumbhar NP. Agricultural plant Leaf Disease Detection Using Image Processing. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*. 2013;2(1). Available from: https://www.ijareeie.com/upload/january/5_Agricultural%20plant.pdf.
13) Madhgoria S, Marekschikora W, Koch. Pixel-Based Classification Method for Detecting Unhealthy Regions in Leaf Images. 2009. Available from: https://www.semanticscholar.org/paper/Pixel-based-classification-method-for-detecting-in-adhogariaSchikora/9ae48c2979651f5163321837d8ecd8bc27d8ca51.
14) Ramesh S, Niveditha M, Pooja R, Bhat P, Shashank N, N, et al. Plant Disease Detection Using Machine Learning. In: and others, editor. International Conference on Design Innovations for 3Cs Compute Communicate Control. 2018. Available from: https://doi.org/10.1109/ICDI3C.2018.00017.
15) Jadhav SB, Udupi VR, Patil SB. Identification of plant diseases using convolutional neural networks. *International Journal of Information Technology*. 2020. Available from: https://dx.doi.org/10.1007/s41870-020-00437-5.
16) Jiang P, Chen Y, Liu B, He D, Liang C. Real-Time Detection of Apple Leaf Diseases Using Deep Learning Approach Based on Improved Convolutional Neural Networks. *IEEE Access*. 2019;7:59069–59080. Available from: https://dx.doi.org/10.1109/access.2019.2914929.
17) Kirchner A, Signorino CS. Using Support Vector Machines for Survey Research. *Survey Practice*. 2018;11(1):169–173. Available from: https://doi.org/10.29115/SP-2018-0001.
18) Revathi P, Hema M, Latha. Classification Of Cotton Leaf Spot Diseases Using Image Processing Edge Detection Techniques". *International Journal of Applied Research*. 2012;p. 169–173. Available from: https://doi.org/10.1109/INCOSET.2012.6513900.
19) Khairnar K, Dagade R. Disease Detection and Diagnosis on Plant using Image Processing A Review. *International Journal of Computer Applications*. 2014;108(13):36–38. Available from: https://dx.doi.org/10.5120/18973-0445.
20) Department of Agriculture, Cooperation and Welfare. 2020. Available from: http://agricoop.nic.in.
21) Kaggle Datasets. 2015. Available from: https://www.kaggle.com/abdallahalidev/plantvillage-dataset.
22) Thapa R, Zhang K, Snavely N, Belongie S, Khan A. The Plant Pathology Challenge 2020 data set to classify foliar disease of apples. *Applications in Plant Sciences*. 2020;8(9). Available from: https://dx.doi.org/10.1002/aps3.11390.