International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Improved Plant Phenotyping System Employing Machine Learning Based Image Analysis

[1]Chinnala Balakrishna, [2] Koushik Reddy Chaganti, [3]C. Sasikala, [4] G.Sirisha, [5]Shravani Amar, [6] Ravindra Changala, [7] Koppuravuri Gurnadha Gupta

Submitted: 27/01/2024 **Revised**: 05/03/2024 **Accepted**: 13/03/2024

Abstract: The objective of this paper was to propose an advanced platform for phenotyping plants using machine learning and image analysis. Phenotyping is the process of analyzing and measuring the physiological characteristics of a part or whole plant like the shape of a leaf, the color of the flower, or the structure of the root. It helps to understand the genetic and environmental factors that influence plant growth and productivity. It is used to provide new plant breeding programs. The images will be pre-processed to standardize their size and format, and relevant features will be extracted for use in training a machine learning model. The model will be trained to classify the images based on their phenotypic traits and will be validated for accuracy. The trained model will then be integrated into the phenotyping platform for automatic analysis and classification of new images. The result will be a tool that can aid in the study of plant phenotypes – crop yield prediction, type classification, crop growth, etc.

Keywords: *Image analysis, Machine learning, Phenotyping, Plant Phenotypes, physiological.*

1. Introduction

The foundation of any plant breeding selection method is phenotyping. Modern plant phenotyping, however, measures complex growth and yield-related variables with more accuracy and precision. The dynamic and local interactions between different phenotypes and the above- and below-ground environment are what give rise to the plant phenotype. Phenotype quantification comprises structural elements such as plant biomass, root shape, leaf properties, and fruit attributes. Functional attribute ideas like photosynthetic efficiency are far more complicated. Human-assisted disease diagnosis is ineffective and unable to keep up with the high demands due to the wide geographic distribution of agricultural areas, low education levels of farmers combined with inadequate knowledge, and lack of access to plant pathologists. The agriculture sector must be automated in today's technologically advanced globe for emerging nations like India. However, problems such as the spread of illnesses that may have been stopped by early detection hinder both production and food quality. To overcome the shortfall of human assistance in early disease detection, it is imperative to fabricate an advanced plant phenotyping platform with technology to monitor the overall development of plants. This well-furnished website is easily accessible to farmers aids them to examine whether a food crop is acquiring sufficient nutrients on time, and discloses useful information to grow any crop in an ecosystem. There have been advances made in using robotics and computer vision systems to address a variety of issues in the agriculture sector. Image processing has the ability to support precision farming techniques, weed and pesticide technology, monitoring plant growth, and management of plant nutrition. is getting enough nutrients on time and provides information that is useful for any crop to develop in an ecosystem. The guidelines for crops and fertilizers help produce large yields while being environmentally friendly. Crop pathologists and farmers who grow crops can communicate with each other thanks to an advanced wide-area network and cloud-based backend services that enable data calculations, centralized storage, and data analytics. The underlying technology constructs a neural network that categorizes labeled photos to achieve accuracy in the detection of plant diseases using unobserved data. The study of images has made substantial use of convolution neural networks (CNNs). To create an end-to-end platform with the many agricultural solutions needed by farmers, our suggested solution relies on a created cloud and API service. Deep Convolutional Neural Networks (CNNs) have consistently been the best architecture for computers for computer vision and image processing since 2012, when "Alex Net" won the ImageNet competition. Huge picture data sets, improvements in CNN algorithms, and computing power have all contributed to the revolutionary CNN capabilities. Because of open-source platforms like TensorFlow, AI has improved, become more widely used, and is now more affordable. Our idea makes use of Web and AI technologies to provide farmers with a crop diagnosis system that fully replicates the expertise (or "intelligence") of plant pathologists.

[1] [1] Associate Professor, Department of CSE(AIML & Cyber Security), Guru Nanak Institute of Technology(A), Hyderabad.
[2] Research Scholar, Dept. of Computer and Information Systems, University of the Cumberlands, KY USA.
[3]Associate Professor, Department of CSE, Srinivasa Ramanujan Institute of Technology (A), Anantapuramu.
[4] Assistant Professor, Dept of ET, CVR College of Engineering (A), Hyderabad.
[5] Assistant Professor, Department of CSE, Mallareddy University, Hyderabad.
[6] Associate Professor, Department of Computer Science and Engineering, Guru Nanak Institutions Technical Campus, Hyderabad.
[7] Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur Dist., Andhra Pradesh.
** Corresponding Author Email: rtit2009@gmail.com*

2. Existing Work

In the field of artificial intelligence, remarkable works have introduced the domain of image analysis, both in general and as applied to the problem of plant phenotyping. Images are considered a measurement tool, and the automated processing of images allows for greater throughput, and reliability at all scales of measurement (from microscopic to field level). Although a picture may be considered a measurement in certain ways, the more fascinating problem is extracting quantitative measures from images that have inherent significance for biologists. The throughput of plant phenotyping devices will skyrocket when such a process is made totally automatic. This will allow for widespread plant growth monitoring and break the existing phenotyping bottleneck caused by human measuring of individual plants. Building a plant phenotyping tool is often an iterative process where the approach and the environment are both modified until the performance is satisfactory and predictable. When deciding on the light source spectrum, this should consider the sensor's spectral sensitivity, numerical dynamic (number of bits), and spatial resolution.

3. Limitations in Existing System

The main limitation of the existing system is it can only capture the images that are, and it can only perform image analysis. During the preprocessing and storage of leaf images, more information is lost. The major problem in the use of deep learning models for plant disease detection is the insufficiency of datasets. In addition, environmental conditions are not considered; hence, the accuracy rate obtained will be higher than that obtained in a practical application. In real-world settings, the distribution across classes is skewed and unbalanced. The last choice in terms of capture is the sensor, which takes the plant's spectrum sensitivity, numerical dynamic, and spatial resolution into account. Therefore, there is no all-purpose device for phenotyping plants. Designing a plant phenotyping system entails balancing the demands of method, goal, and environment.

4. 4.Proposed Work

Disease Classifier-

The Classifier employs a trained deep Convolutional Neural Network (CNN) model to identify the kind of disease from photographs that are supplied through a desktop and operate on a dynamically connected platform that is a standalone application operating in the Cloud. The Classifier automatically classifies the submitted photos into the appropriate illness category using the CNN model that was generated by the Deep CNN Trainer. Low accuracy frequently happens when a user uploads irrelevant photographs or images with low image quality or an underlying ailment that the trained CNN model is not yet aware of. The inclusion of additional illness categories that may be kept for the next training sessions is made possible by expert assistance in cases with poor categorization scores. The Classifier can begin automatically identifying the new illness after the Training Database contains a significant number of photos representing the new disease category and a high classification accuracy is attained. It enables us to increase the precision of automated responses to covered illnesses over time as more farmers interact and provide photos; while utilizing the scarce expert skills to expand coverage for new diseases. The deep CNN model employed by the Classifier to classify diseases more accurately is progressively improved by consecutive runs of this training program, which utilizes a bigger training dataset.

Crop yield predictor-

Nutritional parameters like Nitrogen, Phosphorus, Potassium, and pH of the soil are supplied to the web application, where the queries of plant growth status will be resolved. So that they get suggestions of the plant status and whether they had to supply the necessary amount of nutrients. Python-based web application framework, flask designs the website and also deploys our Machine Learning model. In exchange for the nutrient values present in any ecosystem, the crop yield is evaluated and it is decided whether the given crop has been grown in that environment or not.

Advantages over Existing System

At a very early stage itself, it detects the symptoms of diseases . Fast and accurate results. The detection of objects in images is realized accurately by the specification of a set of features precisely characterizing the object. Easy to use. Utilizes data and algorithms to forecast agricultural productivity. By knowing the expected yield in advance, the right amount of farmers can allocate the labor, equipment, fertilizers, pesticides, and water resources to optimize productivity and minimize waste.

5. Literature Survey

This Paper defines a huge problem for plant science and agricultural development is ensuring that crop production is adequate to meet the needs of a human population that is anticipated to increase to more than 9 billion by 2050. This objective is difficult mainly because crop output is only increasing at an average rate of 1.3% annually and cannot keep up with population expansion. High-yielding, stress-tolerant plants can be selected much faster and more effectively if the genotype and phenotype are linked. Breeders can have access to new technology like next-generation DNA sequencing to potentially accelerate the rate of genetic improvement through molecular breeding. However, Our capacity to analyse the genetics of quantitative variables relevant to growth, yield, and stress adaptability is constrained by the lack of availability of phenotyping tools. Long before DNA and molecular markers were discovered, plant breeders and farmers made decisions based on phenotypes. The likelihood of finding the best genetic variation increases with the number of selection contexts and crossings performed[1].

The interaction between a plant and its environment leads to the development of dynamic phenotypes in plants. For the improvement of fundamental plant science and its translation into the application, such as breeding and crop management, it is crucial to comprehend these activities that take place over the course of a plant's lifespan in a constantly changing environment. It became necessary for the plant research community to precisely assess the various features of an ever-increasing variety of plants to aid in their adaptation to areas with scarce resources and low-input agriculture. We describe the emergence of plant phenotyping as a multidisciplinary field in this overview [2]. This paper defines Genetics, botany, and agronomy all depend heavily on plant phenotype, but present methods for measuring phenotypic traits have some drawbacks in terms of cost, performance, and space-time coverage.

Computer vision has undergone phenotypic examination due to the quick advancement of image technologies, computing power, and algorithms. For the reasons listed above, scientists are committed to creating image-based plant phenotyping techniques as a supplement to, or even a replacement for, manual measurement. However, several variables can influence the application of computer vision technology to examine plant phenotypic features, including the research setting, imaging system, research object, feature extraction, model selection, and more. There isn't a review study available right now to compare and extensively analyze these techniques. this review introduces the average plant as a result. Phenotyping techniques based on computer vision in detail, including their theory, scope, outcomes, and comparisons. In light of the technical advancement in plant phenotyping during the past 20 years (from 2000 to 2020), this paper evaluates more than 200 papers in great detail. This paper discusses a variety of subjects, including modern phenotyping techniques, plant databases, and imaging technology. In this review, we divide plant phenotyping into two major categories: whole-plant phenotyping and phenotyping of plant organs. Additionally, we review each group's research, analyze it, and talk about the shortcomings of the present paradigms and potential future research area[3]. The measuring of observable plant features is known as plant phenotyping. Current field phenotyping techniques require a lot of manual labour and are prone to mistakes. For plant breeding to proceed more quickly, automated, non-invasive high throughput plant phenotyping is essential. High throughput plant phenotyping is significantly hampered by occlusions and less-than-ideal sensing conditions, and the majority of cutting-edge 3D phenotyping algorithms mainly rely on heuristics or manually adjusted parameters. We provide a novel model-based optimization strategy for determining plant physical attributes from plant units known as phytomers to overcome this issue. The suggested method involves drawing data from an underlying probability distribution to sample parameterized 3D plant models. Then it optimizes, bringing this probability distribution's mass closer to the model's actual values[4]. The annual losses sustained by the horticulture and agriculture sectors worldwide can be reduced with the accurate and timely detection of plant diseases. Many researchers have utilized thermal imaging to investigate how disease impacts a plant's thermal profile since it is a quick and non-destructive approach to check for diseased plant sections. Environmental factors, such as leaf angles and the depth of the canopy sections exposed to the thermal imaging camera, have been reported to have an impact on the thermal image of a diseased plant. In this study, researchers combined depth information with thermal and visible light imaging data to develop a method for remotely identifying plants infected with the tomato powdery mildew fungus, Oidium neolycopersici[5].

6. Proposed Methodology

The model that is proposed by us to detect and classify the infected plant leaves consists of 4 phases.

- Dataset Collection
- Image Preprocessing
- Segmentation
- Selection of Classifier

The following datasets can be used to detect plant diseases. The majority of these datasets are collections of images of various plant species that have been annotated with details about the presence and type of disease. The Plant Village Dataset was created by researchers at EPFL and Penn State [7]. 54,000 controlled-environment photographs of both healthy and damaged plant leaves were obtained for this collection. The photos display the 14 unique crop species in addition to different mold, rust, and leaf blight species. Plant phenotyping data sets: [8] The Plant Phenotyping team at Leibniz University Hannover created this data collection. They offer a variety of datasets, including the "Sugar Beet" dataset and the "Arabidopsis" dataset, for the goal of discovering plant diseases .Agricultural Plant Disease Dataset (APD2): Based on the crop and disease, the 57643 images of both healthy and ill crops have been divided into 31 groups. The 57643 images of healthy and diseased rice crops in this collection are categorized into 31 groups based on the crop and disease.

The dataset has been collected from the Kaggle. The below figure shows the various diseases of plant from the dataset:-

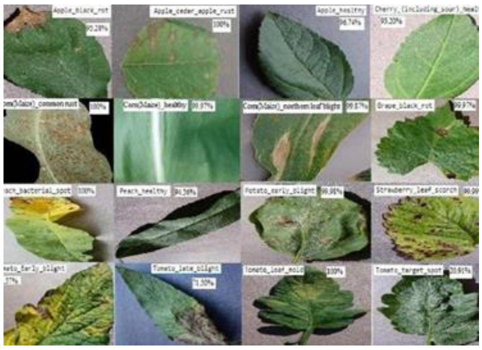


Fig.1 Sample images from dataset

The dataset consists of both healthy and infected leaves which covers diseases like black rot, rust, bacterial spot, early blight, late blight, leaf scorch, target spot, mosaic virus of different crops like apple, potato, tomato, grape, strawberry, corn.

Plant disease classifier using CNN:

The steps to be followed before training our model

Data Preprocessing: Image resizing: To provide a constant input size for the CNN, the input images are typically downsized to a fixed resolution. The input characteristics are pre-processed to an acceptable size and normalized such that they lie within the same range. Occasionally, image augmentation techniques (such flipping, rotating, zooming, etc.) are applied to increase the diversity of the data and reduce overfitting.

Data augmentation: Methods like random rotation, flipping, and translation are frequently used on the images to broaden the diversity of the training data and enhance the generalization of the model.

The Classifier employs a trained deep Convolutional Neural Network (CNN) model to identify the kind of disease from photographs that are supplied through a desktop and operate on a dynamically connected platform that is a standalone application operating in the Cloud. The Classifier automatically classifies the submitted photos into the appropriate illness category using the CNN model that was generated by the Deep CNN Trainer [9]. Low accuracy frequently happens when a user uploads irrelevant photographs or images with low image quality or an underlying ailment that the trained CNN model is not yet aware of. The inclusion of additional illness categories that may be kept for the next training sessions is made possible by expert assistance in cases with poor categorization scores. The Classifier can begin automatically identifying the new illness after the Training Database contains a significant number of photos representing the new disease category and a high classification accuracy is attained. It enables us to increase the precision of automated responses to covered illnesses over time as more farmers interact and provide photos, while utilizing the scarce expert skills to expand coverage for new diseases. The deep CNN model employed by the Classifier to classify diseases more accurately is progressively improved by consecutive runs of this training program, which utilizes a bigger training dataset. The capacity of convolutional neural networks (CNNs) to learn hierarchical information from images makes them a popular choice for classifying plant diseases [10]. There are various processes involved in training a CNN, including as data preprocessing, designing the model architecture, and choosing the loss function. Here, I'll give an overview of the standard procedures and some important elements that CNNs for classifying plant diseases use.

Building the CNN model: A CNN model usually contains the following layers:

Convolutional Layers: These layers convolution the input before sending the results to the following layer. [11] They are used to find local conjunctions of the previous layer's features.

Activation functions: Rectified Linear Units, or ReLUs, are frequently employed to induce non-linearity.

Pooling layers: Following convolution, pooling layers reduce computation by down sampling the spatial dimensions of the feature maps and limiting overfitting. A popular option is max-pooling.

Fully connected layers: The final classification is performed using fully connected layers after numerous convolutional and pooling layers.

For multiclass or binary classification tasks at the output layer, respectively, Softmax or Sigmoid Layer are utilized.

Training the CNN model: The model is trained using the pre-processed images. The CNN is fed the images during training, the model's predictions are compared to the actual labels, and backpropagation is performed to update the weights of the model.

Validation and Testing: To correct the hyperparameters and avoid overfitting, a new validation set is used. [15] After it has been finalized, the model is tested against a test set to see how well it works.

After being trained and assessed, the model can be used to detect diseases in images of recently harvested plants. Your photograph is examined by CNN, which then predicts the health of the plant and, if a disease does exist, its type.

Convolutional Layer:
This is the major activity of CNN.[13,14] By designating the input to this layer as X, the weights (or filter) of the convolutional layer as W, and the output as Y, the operation in this layer can be mathematically described as follows.

$$Y[i,j,k]=\Sigma\Sigma\Sigma(X[a,b,c]*W[i-a,j-b,k-c]) \qquad (1)$$

where the sum of the input X's spatial extent and all of the input feature maps is determined (c). This operation is performed for each type k output feature map.

Non-linear Layer (ReLU):
Each convolutional layer's output is then subjected to a non-linear function. Typically, a Rectified Linear Unit (ReLU) is employed for this, which just applies the equation $f(x) = \max(0,x)$. If Z is the output of the convolutional layer, then the ReLU layer's output would be Y.

$$Y=\max(0,Z) \qquad (2)$$

Pooling Layer:
Pooling layers are used to reduce the spatial dimensions of the input. [14] The most well-known pooling technique is known as max pooling, and it entails dividing the input into a number of non-overlapping rectangles and calculating the maximum value for each of these sub-regions. If X is the input and Y is the output of the pooling layer, think about 2x2 max pooling.

$$Y[i,j,k]=\max(X[2i:2i+2,2j:2j+2,k]) \qquad (3)$$

Fully Connected Layer:
In the fully linked layer, neurons are entirely connected to every activation in the layer above. The output of this layer, which is a volume of size [1x1x10], has 10 values, each of which represents a class score. The class scores are calculated by this layer. If we designate the input to this layer as X, the weights as W, the bias as b, and the output as Y, the operation can be represented as follows.

$$Y=W*X+b \qquad (4)$$

The mathematical procedures presented above are simplified versions; the actual computation takes stride and padding into consideration. Additionally, more sophisticated CNN systems might employ layers of a different type [16].

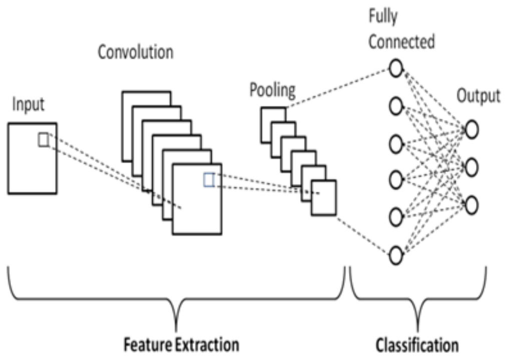


Fig.2 General CNN Architecture

Crop Yield Predictor and Crop Recommendation System:

Nutritional factors including nitrogen, phosphorus, potassium, and soil pH are sent to the online application, which will answer questions about the condition of the plant's growth. [17, 18] so that they may determine whether they need to feed the required amount of nutrients and the status of the plant. Flask is a web application framework built on Python that both develops the website and deploys our machine learning model. The crop yield is assessed in relation to the nutrient values present in any ecosystem, and it is decided whether or not the specified crop has been cultivated there.

Maximize

$$F(x) = [f_1(x), f_2(x), \ldots\ldots, f_n(x)] \qquad (5)$$

We recommend the crop based on these values in addition to evaluating the NPK parameters in this crop yield classifier.

Therefore, this predictor can be utilized as a crop recommender as well as a predictor of plant nutrition levels. The values will be retrieved from the database, which contains a variety of temporal data. [12] Random Forest is the algorithm employed in this case. Here, we first employed the Random Forest approach to suggest a crop, but we trained our model using KNN, SVC, and naive Bayesian (Gaussian) algorithms. However, the Random Forest approach outperformed the competition in terms of accuracy, therefore we decided to use it as the final algorithm to train our model.

The below graph represents the NPK values by comparing the Traditional approach and IGA recommendation as shown below:

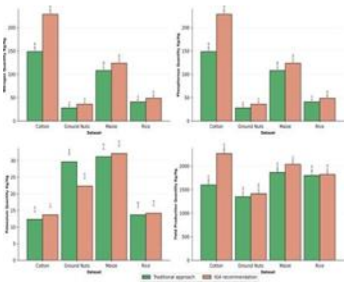


Fig.3 Bar graph showing NPK values

Algorithm Used:

Random Forest

- Due to its excellent accuracy, the supervised learning technique's Random Forest algorithm, a well-known machine learning method, is utilized for classification and regression applications.
- By averaging many decision trees, Random Forest lessens overfitting and is less susceptible to noise and outliers in the data.
- A random forest is a meta estimator that employs multiple sub-samples of the data and fits a number of decision trees to them.
- Higher accuracy results from a larger forest's tree population.
- Random Forest, a machine learning algorithm, has achieved outstanding results in a wide range of fields, which has greatly increased its popularity.

One of its most significant benefits is that it can manage large, multidimensional datasets with a range of properties. Since it uses a multitude of decision trees and their combined decision-making process, Random Forest can handle tens of thousands of characteristics and is good for tasks involving pictures, text, and other high-dimensional data types. In addition, Random Forest is flexible enough to handle classification and regression problems like regression and classification as well as unsupervised learning tasks like anomaly detection and clustering.

Another notable aspect of Random Forest is its ability to provide insights about feature importance. This is particularly helpful for feature selection because it allows us to identify the traits that most significantly improve the model's ability to forecast. This information can help us understand the underlying trends in the data and facilitate analysis. Additionally, Random Forest's natural capacity to handle missing values and outliers without requiring extensive data pre-processing makes the data preparation phase simpler, saving data scientists and practitioners time and effort.

Despite the fact that Random Forest is a powerful algorithm, it's crucial to keep in mind that it might not always outperform alternatives in every circumstance. For organized data with linear associations, for example, linear models like logistic regression may yield better results. Additionally, because Random Forest is an ensemble algorithm, it could use more resources than individual decision trees. Due to its versatility, resilience, and ability to handle many data sources, Random Forest is still a preferred option for many machine learning applications. As a result, it is a crucial weapon in the data scientist's toolbox.

Classification with Random Forest:

1. Probability Calculation for Classification:
The percentage of training samples in the leaf node (l) that belong to class (c) is used to determine the likelihood of a data point belonging to a specific class (c) for each decision tree in the forest.

Let:
- P (c|l) = Probability of class c given the samples in leaf node l.
- N (l) = Number of samples in leaf node l.
- N(c, l) = Number of samples in leaf node l that belong to class c.

By averaging over all the trees in the Random Forest, it is possible to determine the likelihood of class C for a given data point:

$$P(c|X)=(1/B)*\Sigma(P(c|l))$$

Where B is the total number of trees in the Random Forest.

Regression with Random Forest:

1. Prediction for Regression:
By averaging the target values of all the samples in the leaf node (l) of each decision tree, the prediction of the target variable (Y) for a given data point (X) in regression tasks is determined.

Let:
- Y (l) = Target values of samples in leaf node l.
- N (l) = Number of samples in leaf node l.

The prediction for the given data point is obtained by averaging over all the trees in the Random Forest:

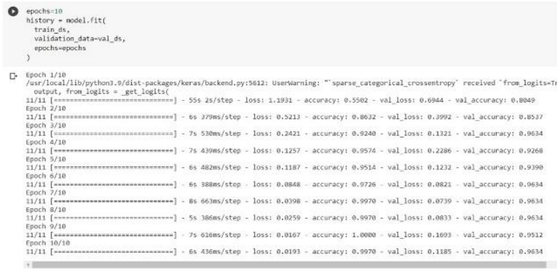
$$Y\_pred(X) = (1 / B) * \Sigma(Y (l)) \quad (7)$$

Where B is the total number of trees in the Random Forest.

Fig.4 Random forest structure

7. Results

In the first part of our project that is plant disease classifier we have trained our model with CNN where the data have been pre-processed with various pre-processing techniques like decoding the jpeg to RGB format where model can use them as the input to the next layers. By, following these techniques we have achieved the accuracy of 99%. The ML model has been trained for 11 epochs where it has given the accuracy of 99% while testing our model as shown:-



```
epochs=10
history = model.fit(
  train_ds,
  validation_data=val_ds,
  epochs=epochs
)

Epoch 1/10
/usr/local/lib/python3.9/dist-packages/keras/backend.py:5612: UserWarning: "`sparse_categorical_crossentropy` received `from_logits=Tr
  output, from_logits = _get_logits(
11/11 [==============================] - 55s 2s/step - loss: 1.1931 - accuracy: 0.5502 - val_loss: 0.6944 - val_accuracy: 0.8049
Epoch 2/10
11/11 [==============================] - 6s 379ms/step - loss: 0.5213 - accuracy: 0.8632 - val_loss: 0.3992 - val_accuracy: 0.8537
Epoch 3/10
11/11 [==============================] - 7s 530ms/step - loss: 0.2421 - accuracy: 0.9240 - val_loss: 0.1321 - val_accuracy: 0.9634
Epoch 4/10
11/11 [==============================] - 7s 439ms/step - loss: 0.1257 - accuracy: 0.9574 - val_loss: 0.2286 - val_accuracy: 0.9268
Epoch 5/10
11/11 [==============================] - 6s 482ms/step - loss: 0.1187 - accuracy: 0.9514 - val_loss: 0.1232 - val_accuracy: 0.9390
Epoch 6/10
11/11 [==============================] - 6s 388ms/step - loss: 0.0848 - accuracy: 0.9726 - val_loss: 0.0821 - val_accuracy: 0.9634
Epoch 7/10
11/11 [==============================] - 8s 663ms/step - loss: 0.0398 - accuracy: 0.9970 - val_loss: 0.0739 - val_accuracy: 0.9634
Epoch 8/10
11/11 [==============================] - 5s 386ms/step - loss: 0.0259 - accuracy: 0.9970 - val_loss: 0.0833 - val_accuracy: 0.9634
Epoch 9/10
11/11 [==============================] - 7s 616ms/step - loss: 0.0167 - accuracy: 1.0000 - val_loss: 0.1693 - val_accuracy: 0.9512
Epoch 10/10
11/11 [==============================] - 6s 436ms/step - loss: 0.0193 - accuracy: 0.9970 - val_loss: 0.1185 - val_accuracy: 0.9634
```

Fig.5 Showing the accuracy of of the model while train our model

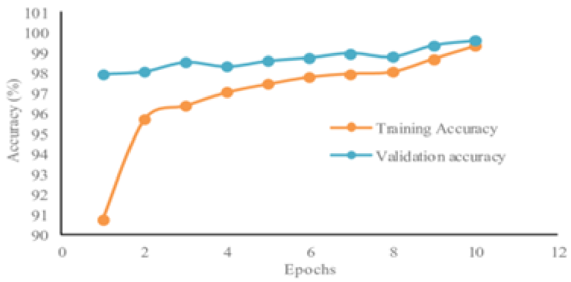


Fig.6 Training Vs Validation Accuracy

Our model has classified the plant disease correctly with the range of accuracy 96-99%.

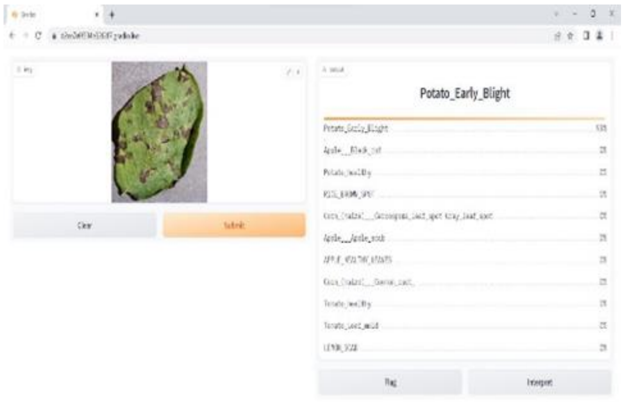


Fig.6 Plant disease classification

Now, in the second part of the project we have integrated crop recommendation system into our platform where it works based on the NPK values which have been stored in the database. Here, we have used the Random forest algorithm to recommend the crop as well as this system can recommended the NPK values for the particular crop from the database whether they have sufficient nutrients or , not if not it will also recommend the appropriate values to grow that crop.

The model has been trained with various algorithms such as SVM which has given accuracy of 98%, Linear regression with accuracy of 96%, KNN with accuracy of 97%, Decision Trees with accuracy of 98.899% and the Random forest which has given the accuracy of 99%, So we have selected our algorithm as Random Forest.

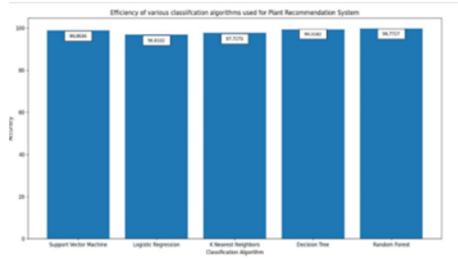


Fig.7 Accuracy analysis of ML algorithms

We chose the Random Forest method, which has an accuracy of 99%, as the model with the highest accuracy based on this graph.

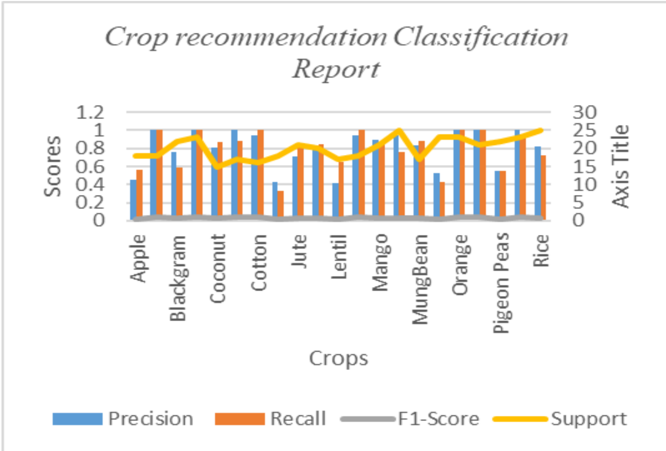


Fig.8 Classification report of crops

8. Conclusions

To sum up, plant phenotyping is extremely important for comprehending and examining numerous facets of plant growth, development, and response to environmental influences. In this document, the needs, user characteristics, limitations, presumptions, and particular functional requirements for a plant phenotyping system have been outlined. A system's technological and financial viability has also been described, and test cases have been offered to guarantee the system's correct operation. With the use of plant phenotyping, the properties of plants may be precisely measured and analyzed, diseases can be found, growth stages can be evaluated, and correlations can be examined. By leveraging technologies such as image analysis, machine learning, and data integration, these systems can accurately detect and classify plant diseases. Crop suggestions can assist farmers in making well-informed selections regarding the kinds of crops to produce by taking into account variables including climate, soil qualities, market demand, and farming practices.

References

[1] Tardieu F., Cabrera-Bosquet L., Pridmore T. et al. (2017). Plant Phenomics, From Sensors to Knowledge. Current Biology, 27(15): R770-R783. doi 10.1016/j.cub.2017.05.055.

[2] Sunder Reddy, K. S. ., Lakshmi, P. R. ., Kumar, D. M. ., Naresh, P. ., Gholap, Y. N. ., & Gupta, K. G. . (2024). A Method for Unsupervised Ensemble Clustering to

Examine Student Behavioral Patterns. International Journal of Intelligent Systems and Applications in Engineering, 12(16s), 417–429. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/4854.

[3] Ravindra Changala, "Brain Tumor Detection and Classification Using Deep Learning Models on MRI Scans", EAI Endorsed Transactions on Pervasive Health and Technology, Volume 10, March 2024.

[4] Neumann K., Klukas C., Friedel S. et al. (2015). Phenotyping Root Growth Dynamics of Arabidopsis Seedlings in Response to Hypoxia. PLoS ONE, 10(4): e0123942. doi 10.1371/journal.pone.0123942.

[5] Das Choudhury S., Bashyam S., Hu H. et al. (2021). Automated Hyperspectral Plant Phenotyping: A Review of Sensor Systems, Techniques, and Applications. Journal of Imaging, 7(8): 141. doi: 10.3390/jimaging7080141.

[6] Nagesh, C., Chaganti, K.R. , Chaganti, S. , Khaleelullah, S., Naresh, P. and Hussan, M. 2023. Leveraging Machine Learning based Ensemble Time Series Prediction Model for Rainfall Using SVM, KNN and Advanced ARIMA+ E-GARCH. International Journal on Recent and Innovation Trends in Computing and Communication. 11, 7s (Jul. 2023), 353–358. DOI:https://doi.org/10.17762/ijritcc.v11i7s.7010.

[7] Ravindra Changala, "Sentiment Analysis in Social Media Using Deep Learning Techniques", International Journal of Intelligent Systems and Applications in Engineering, 2024, 12(3), 1588–1597.

[8] PlantSelect. (2021). Plant Recommendations for a Water-Wise Landscape. Retrieved from https://plantselect.org/plant-recommendations/

[9] Naresh, P., & Suguna, R. (2021). Implementation of dynamic and fast mining algorithms on incremental datasets to discover qualitative rules. Applied Computer Science, 17(3), 82-91. https://doi.org/10.23743/acs-2021-23.

[10] Ravindra Changala, Development of CNN Model to Avoid Food Spoiling Level, International Journal of Scientific Research in Computer Science, Engineering and Information Technology, ISSN: 2456-3307, Volume 9, Issue 5, September-October-2023, Page Number 261-268.

[11] M. I. Thariq Hussan, D. Saidulu, P. T. Anitha, A. Manikandan and P. Naresh (2022), Object Detection and Recognition in Real Time Using Deep Learning for Visually Impaired People. IJEER 10(2), 80-86. DOI: 10.37391/IJEER.100205..

[12] American Horticultural Society Plant Propagation: The Fully Illustrated Plant-by-Plant Manual of Practical Techniques. (1999). DK Publishing.

[13] Ravindra Changala, "Optimization of Irrigation and Herbicides Using Artificial Intelligence in Agriculture", International Journal of Intelligent Systems and Applications in Engineering, 2023, 11(3), pp. 503–518.

[14] Hussan, M.I. & Reddy, G. & Anitha, P. & Kanagaraj, A. & Pannangi, Naresh. (2023). DDoS attack detection in IoT environment using optimized Elman recurrent neural networks based on chaotic bacterial colony optimization. Cluster Computing. 1-22. 10.1007/s10586-023-04187-4.

[15] Mahlein, A.-K. (2016). Plant Disease Detection by Imaging Sensors - Parallels and Specific Demands for Precision Agriculture and Plant Phenotyping. Plant Disease, 100(2): 241-251. doi 10.1094/PDIS-03-15-0340-FE.

[16] B. Narsimha, Ch V Raghavendran, Pannangi Rajyalakshmi, G Kasi Reddy, M. Bhargavi and P. Naresh (2022), Cyber Defense in the Age of Artificial Intelligence and Machine Learning for Financial Fraud Detection Application. IJEER 10(2), 87-92. DOI: 10.37391/IJEER.100206.

[17] Ravindra Changala, Development of Predictive Model for Medical Domains to Predict Chronic Diseases (Diabetes) Using Machine Learning Algorithms And Classification Techniques, ARPN Journal of Engineering and Applied Sciences, Volume 14, Issue 6, 2019.

[18] Ravindra Changala, "Evaluation and Analysis of Discovered Patterns Using Pattern Classification Methods in Text Mining" in ARPN Journal of Engineering and Applied Sciences, Volume 13, Issue 11, Pages 3706-3717 with ISSN:1819-6608 in June 2018.

[19] Singh, A., Ganapathy subramanian, B., Singh, A.K., Sarkar, S., & Sarkar, S. (2016). Machine Learning for High-Throughput Stress Phenotyping in Plants. Trends in Plant Science, 21(2): 110-124. doi: 10.1016/j.tplants.2015.10.01

[20] P. Naresh, S. V. N. Pavan, A. R. Mohammed, N. Chanti and M. Tharun, "Comparative Study of Machine Learning Algorithms for Fake Review Detection with Emphasis on SVM," 2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS), Coimbatore, India, 2023, pp. 170-176, doi: 10.1109/ICSCSS57650.2023.10169190.

[21] Ravindra Changala, AIML and Remote Sensing System Developing the Marketing Strategy of Organic Food by Choosing Healthy Food, International Journal of Scientific Research in Engineering and Management (IJSREM), Volume 07 Issue 09, ISSN: 2582-3930, September 2023.

[22] Rath, M., & Maheshwari, A. (2019). A Review on Plant Leaf Diseases Detection and Classification using Image Processing Techniques. Computers and Electronics in Agriculture, 161: 272-287. doi: 10.1016/j.compag.2019.03.016.

[23] Mahlein, A.-K., Steiner, U., Hillnhütter, C., Dehne, H.-W., & Oerke, E.-C. (2012). Hyperspectral Imaging for Small-Scale Analysis of Symptoms Caused by Different Sugar Beet Diseases. Plant Methods, 8: 3. doi: 10.1186/1746-4811-8-3.

[24] Armitage, A. M., & Lachman, J. M. (2011). Specialty Cut Flowers: The Production of Annuals, Perennials, Bulbs, and Woody Plants for Fresh and Dried Cut Flowers. Timber Press.