Plants Diseases Prediction Framework: A Image-Based System Using Deep Learning

Conference Paper · August 2022

CITATIONS
0

READS
40

5 authors, including:

Madhu Kirola
Uttaranchal University
8 PUBLICATIONS **10** CITATIONS
SEE PROFILE

Neha Singh
Ningxia University
10 PUBLICATIONS **16** CITATIONS
SEE PROFILE

Kapil Joshi
Uttaranchal University
45 PUBLICATIONS **111** CITATIONS
SEE PROFILE

Sumit Chaudhary
Chandiragh University
20 PUBLICATIONS **33** CITATIONS
SEE PROFILE

Plants Diseases Prediction Framework: A Image-Based System Using Deep Learning

Madhu Kirola
Dept. of Computer Engineering
Uttaranchal University
Dehradun, India
madhukirola@gmail.com

Kapil Joshi
Dept. of Computer Engineering
Uttaranchal University
Dehradun, India
kapilengg0509@gmail.com

Sumit Chaudhary
Dept. of Computer Engineering
Indrashil University
Rajpur,Kadi, India
Sumit.chaudhary@indrashiluniversity.edu.in

Neha Singh
Dept. of Computer Engineering
Indrashil University
Rajpur,Kadi, India
Neha.singh@indrashiluniversity.edu.in

Harishchander Anandaram
Centre for Excellence in Computational Engineering and Networking
Amrita Vishwa Vidyapeetham
Coimbatore, Tamil Nadu, India
harishchander.a@gmail.com

Ashulekha Gupta
Dept. of Management studies
Graphic Era University
Dehradun, India
ashulekha26@gmail.com

***Abstract*— Plant diseases mostly harm the leaves, resulting in a loss in agricultural output's quality and quantity. Plant disease is the most common cause of large-scale crop mortality. India is a country where people's livelihoods are heavily reliant on agriculture. The disease has caused chaos in the agricultural industry. The human eye's perception is not quite as sharp as it needs to be to notice minute variations in the sick leaf region. It needs a complex process that requires both plant expertise and a large amount of processing time. As a result, plant diseases can be detected using machine learning. The disease detection method includes image acquisition, image pre-processing, image segmentation, feature extraction, and classification. To prevent crops at the initial stage from diseases, it is essential to develop an automatic system to diagnose plant diseases and identify its category. The goal of the proposed research is to examine several machine algorithms for plant disease prediction. The paper proposed a framework for disease and healthiness detection in plants and the classification of diseases based on symptoms appearing on a leaf. The diseases are grouped into three categories in the paper: bacterial, viral, and fungal. To conclude, the research paper investigates all of these factors and uses several machine learning(DL) techniques and deep learning(DL) techniques. The machine learning(ML) techniques used in the research work are SVM, KNN, RF(Random Forest), LR (Logistic Regression), and the deep learning(DL) technique used is-Convolutional Neural Network(CNN) for disease prediction in the plants. Following that, a comparison of machine learning and deep learning methodologies was conducted. The RF(Random forest) has the highest accuracy of 97.12 % among machine learning classifiers, however, in comparison to the deep learning model mentioned in the study, the CNN classifier has the highest accuracy of 98.43 %.**

Keywords: Machine Learning, Deep learning, Image segmentation, SVM, RF, LR, CNN.

I. INTRODUCTION

Machine learning can be used for any task that a human performs on a regular basis. Supervised learning, Unsupervised Learning, and Reinforcement learning are the three main learning algorithms used in machine learning. Plant disease prediction using machine learning and deep learning is the subject of this study In India, agriculture is the primary source of income. India is the world's second-largest producer of agricultural products, which accounts for a significant portion of the country's GDP. For over 61.5 % of the Indian population, farming is their principal source of income. Many crops are lost each year as a result of infections such as bacteria, viruses, and fungi. The agriculture sector has improved significantly as a result of technological advancements, but there is still a subject that needs to work on, namely the automatic diagnosis of plant infections by assessing their visually visible signs. The human visual system is an excellent detector, and it is frequently used in practice to determine whether a plant leaf is healthy or unhealthy. It's a good method, but it takes time and costs It works effectively in small plantations since it necessitates continual crop monitoring, which is labour-intensive in large cultivations. There are a thousand plant illnesses, thus continual expert monitoring is required to inspect crops for the highly expensive sickness. Because it is a very cost and labour-effective strategy, an automatic detection technique to identify the disease in the plant is necessary for this reason. A computing system will automatically detect whether the plant leaf is diseased or healthy in this case. If a leaf is unhealthy, it will be classified to determine the sort of disease that the plant has contracted so that proper measures can be taken to protect the entire crop.

To achieve two goals, the research study proposes a strategy based on image retrieval techniques and a machine learning model. The first is to determine whether a leaf is healthy or unhealthy, and the second is to determine which disease the leaf is suffering from if it is an unhealthy-bacterial disease, Fungal Disease, or Virus generated disease. Plant disease classification can be obtained by examining the plant's leaf stem or root, however, the focus of this study is on disease patterns visible on plant leaves. Such as Leaf colour, pale yellow or brownish denotes healthiness, leaf size, and shape coil leaf, roll leaf, crumble leaf indicate viral disease in leaves, and black specks on the leaf indicate fungal disease in leaves. This method employs image processing, machine

learning, and deep learning approaches. To detect leaf color, spot shape, leaf, texture, and size, an image processing technique is employed with machine learning and deep learning technology to do so across a large number of leaves. Machine learning techniques like Support vector machine(SVM), Nave Bayes Classifier(NB), Logistic regression(LR), Random Forest(RF), K Nearest Neighbor(KNN), and deep learning techniques like Convolutional Neural Network(CNN) will be used. The classification of plant diseases is represented in Figure 1. Pathogens are disease-causing biological organisms that infect plants. Viruses, bacteria, fungi, and nematodes are only a few of the most common plant pathogens. Several non-pathogenic disorders (in plants) can emerge when the pH value, moisture, humidity, soil, and other elements in the soil change.

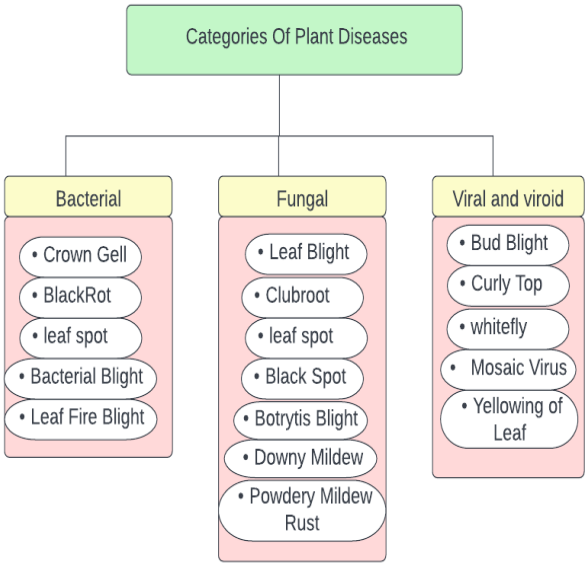

Figure 1. Plant disease classification

1.1 Bacterial Plant Diseases:

Bacterial infections in plants can harm the stems, leaves, and roots, as well as sometimes be transmitted internally without causing external symptoms. Bacterial diseases spread rapidly from plant to plant. Bacterial diseases are caused by bacteria. The most common bacterial diseases are crown gell, Black-Rot, bacterial blight, leaf fire blight, andleaf spot. Bacterial black rot is the most serious disease of crucifier crops like cabbage, cauliflower, etc. It becomes more destructive in warm, hot, humid weather.Figure[2]

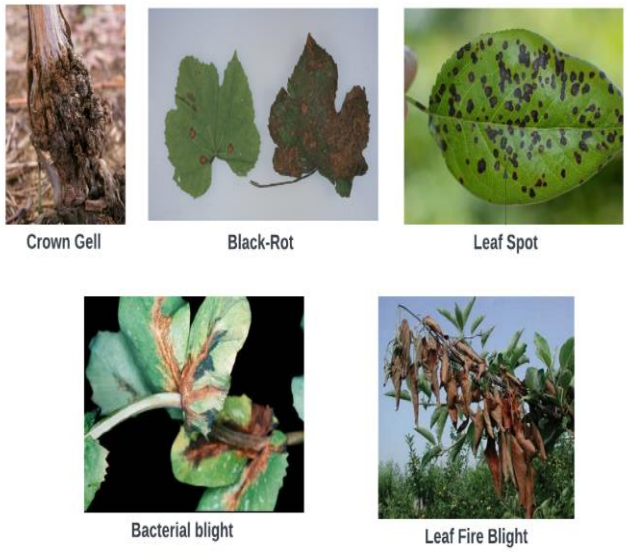

Figure 2. Bacterial Disease Categories

1.2 Fungal Plant Diseases

Fungal infections are more widespread and severe than other plant diseases, affecting crops and crop quality while also shortening plant growth and productivity. Phyto-pathogenic fungi (parasites and semi-parasitic) cause fungal plant diseases. The most common fungal diseases that are hazardous for plants are powdery mildew rust, downy mildew, botrytis blight, fusarium wilt, black spot, and leaf spot clubroot.Figure[3].

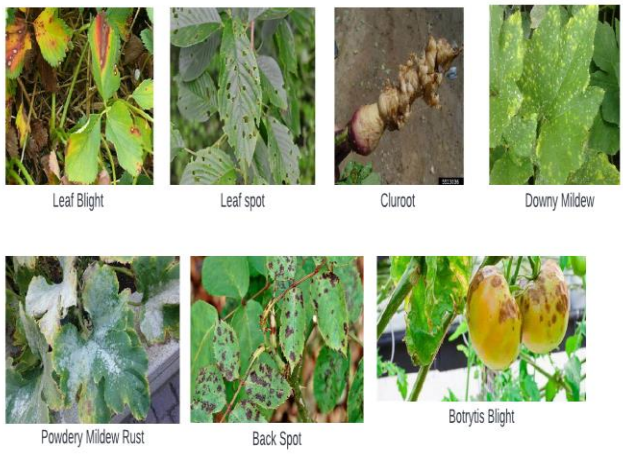

Figure 3. Fungal Disease Categories

1.3 Viral or Viroid Plant Diseases

Viruses are infectious, sub-microscopic particles that only reproduce inside living host cells. Plant viruses consist of two parts: a protein covering and a nucleic acid core. Unlike fungi, viral infections do not produce spores or fruiting bodies. Unlike bacterial diseases, they also do not create ooze. They can't even be cultured in a lab because they only develop in living cells. Viral diseases are caused by the pathogen virus. yellowing of leaf, Leaf curl, curly top, whitefly, bud blight, and the mosaic virus is the most common viral diseases.Figure[4]

Figure 4. Viral Disease Categories

This article is divided into six sections: Section II- showcases related research from a number of researchers along with a comparative analysis of various machine learning and deep learning techniques. In Section III -Our proposed work is discussed. Section IV -The experiment's outcomes are discussed here and lastly in Section V- The Conclusion and Future Scope are discussed. This article is divided into six sections: Section II- showcases related research from a number of researchers along with a comparative analysis of various machine learning and deep learning techniques. In Section III -Our proposed work is discussed. Section IV -The experiment's outcomes are discussed here and lastly in Section V- The Conclusion and Future Scope are discussed.

II Literature Review

[1] Md. Asif et.al."presents an approach for detecting potato leaf disease using image segmentation and machine learning techniques. Colour information, colour intensity, and brightness of the HSV colour space model are used for generating masks. The region of interest is segmented by thresholding HSV images for green and brown colors. Seven classifiers namely Random Forest (RF), Logistic Regression (LR), k-Nearest Neighbours (KNN), Decision Trees (DT), Naive Bayes (NB), Linear Discriminant Analysis (LDA), Support Vector Machine (SVM) to classify potato leaf disease but Random forest gives the highest accuracy of 97%".

[7]Vyshnavi. G.K.P "proposes an approach for plant disease recognition and identification. Mainly focuses on fungal diseases. Image Segmentation is done by hierarchical clustering. Method. Support Vector Machine is usedto fulfill classification purposes". [10] S. Arivazhagan "proposed a methodology in which the Hue Saturation Intensity (HSI) colour model is applied for the transformation of the input image and the infected region of the leaves is then segmented into equal patches. The colour, texture, and feature analysis are done using Spatial Grey Level Dependence Metrics (SGDM). It first uses the Minimum Distance Criterion (MDC) approach for the classification of disease and then improved the accuracy of the second classifier Support Vector Machine (SVM)". [3] Mr. Ashish Nage et.al. "proposed an approach for plant disease detection using python which captures an image of the diseased leaf and identifies the disease. The convolutional neural network has been trained for the classification of the disease. It worked on common plant diseases like bacterial, black-spotted, rust, and Mildew. It first performs the down sampling of the document image and pixel value normalization as pre-processing and then feeds thenormalized image to the CNN to predict the class label". [2] Mohan Sai Singamsetti et.al. "used Convolutional Neural Network Layers are used as a low light enhancement technique. Then the RGB components of the image are transformed into a YCbCr colour space model then the segmentation into k clusters based on a genetic algorithm thathelps search large space and gets optimized in every phase. For texture feature extraction it uses a mathematical method using Grey Level Co-occurrence Matrix. For the classification of the disease into various categories, it uses SVM with different". [8] Sachin D. Khirade contains "a Study on plant disease detection using image processing such as image clipping, image smoothing, and image enhancement for increasing the contrastof the image. The colour Co-occurrence method is discussed for feature extraction which generates an SGDM matrix for the computation of texture statistics then the GLCM function is used for feature calculation. A Self Organization feature map with a backpropagation neural network is implemented. For the classification of diseases, it discussed Artificial Neural Network (ANN) and Backpropagation Algorithm". [9] P. R. Rothe et.al. "suggested an approach for cotton leaf disease detection. The diseased leaves are classified using a backpropagation neural network. Training is performed by extracting seven invariant moments from three types of images of diseased cotton leaves.The active contour model is used for segmentation. The classification model gives average accuracy of 85.52%". [4] Merlin Francis et.al "created and developed a network model to accomplish plant disease prediction and classification on tomato and apple leaves. An accuracy of 88.7% has been achieved by performing a binary classification using CNN to classify whether the leaf is diseased or healthy". [6] Melike Sardogan et.al." presents a method by combining a convolutional neural network model with LVQ stands for learning vector quantization-based methodfor tomato leaf disease detection and classification. Learning Vector Quantization algorithm is used as a classifier because of its topology and adaptability the algorithm". [5] Kshyanaprava Panda Panigrahi et.al. "Proposes the study of the various supervised machine learning algorithms for maize plant disease detection. The

algorithms considered for this study are Naive Bayes, Decision Tree, K-Nearest Neighbour, SVM, and Random Forest. Image segmentation is done by the Label edge detection method. Compared to all the discussed classification techniques, Random forest achieves the highest accuracy of 79.23%".

III Proposed Methodology

The system of detection, classification, and prediction of plant disease is proposed with the following steps. Figure[5].

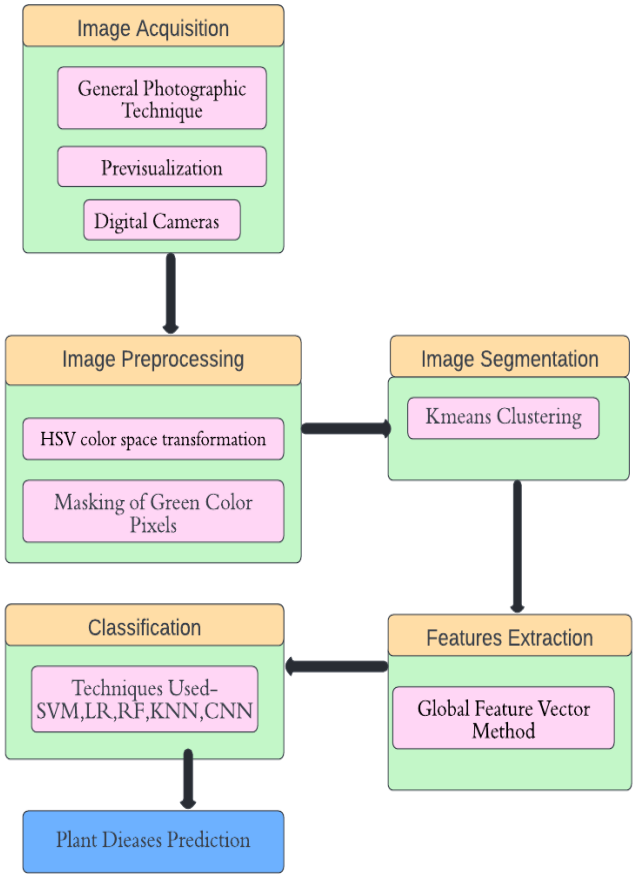

Figure 5. Proposed Referenced Framework for Plant Diseases Prediction

- ***Image Acquisition***

The first step in the detection of disease in plants is image acquisition which includes acquiring various images of plants either through the digital camera and making a dataset out of them or acquiring a plant disease dataset available online. The images of some kind of disease should be contained in one folder which is considered a class labelfor classification.

- ***Image Preprocessing***

In this step, all the images are first resized to an equal fixed image size before using training and validation. The stepis comprised of two parts: In the first part, HSV colour space transformation which is a device-independent colour space transformation is applied to all images. Hue, Saturation, and Value are the three properties of the HSV colour space transformation model. Saturation is the percentage of the purity of the hue, i.e. the amount of white light combined with hue. Hue is the measure of the pure colour received by the observer. The value represents the intensity/brightness of the colour. The second part of pre-processing module consists of masking green colour pixels as green colour pixels contain no extra information about the disease because the green part of the leaf assumes to behealthy part. So we need to mask the green pixels so that the green part can be removed so that better results can be acquired figure shows the BGR, RGB, and HSV format of the sample image.[13][16].

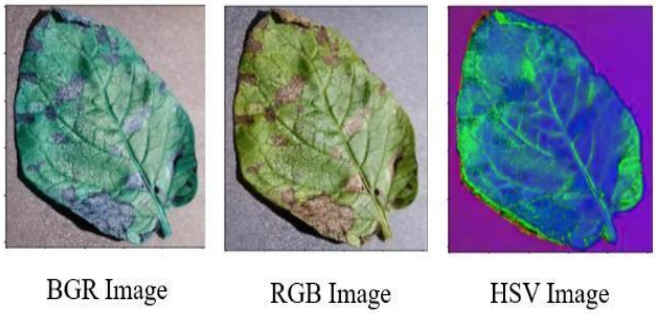

Figure 6: Conversion of BGR image to RGB image and HSV image Representation

- ***Image Segmentation***

The technique of separating an object from its surroundings is known as image segmentation. We have applied a threshold-based method for extracting green and brown colour part from the image and then masked all green colour pixels because green colour pixels does not provide extra information about the disease. Figure7 shows the segmented part of the image, the part other than the leaf becomes black.[11][12]Figure[7].

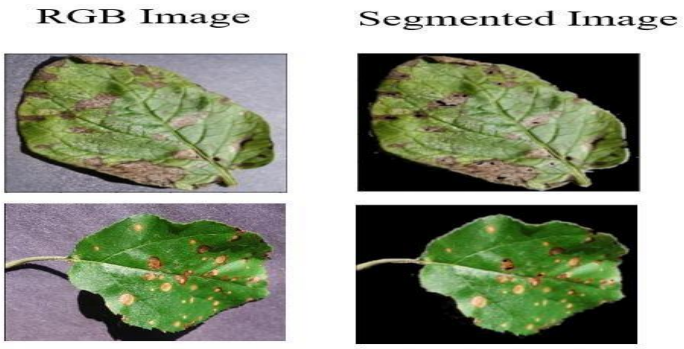

Figure 7. Leaf region Segmentation Process

- ***Feature Extraction***

For extracting the features we are using a global feature vector which is generated by stacking Hu moments, Haralick texture feature, and Colour histogram.

- ***Global Feature Vector***

Global feature vectors are the set of values that describes the entire image in the form of a vector. In this work, we are mainly focusing on three features namely, the shape, colour, and texture features of plant leaves, and stacked all three features to make a single global feature vector for every image. We are using Hu moments as a shape descriptor, Haralick features for texture analysis, and for colour-related information, we are using the colour histogram. All these three feature descriptors are described as:

Hu moments. Hu moments are computed from central moments which are unvarying in size, position, and orientation. There are seven Hu moments in which six are invariant to translation, size, and orientation while the seventh one is skew invariant which means significant changes for image reflection

Haralick Texture. Haralick texture feature is a method of representing the texture information of an image. A grey-level co-occurrence matrix, abbreviated as GLCM, is used to determine Haralick texture features. GLCM depicts the relationship between neighbouring pixels. It is a kind of adjacency matrix. GLCM matrix quantifies how many times the pair of adjacent pixels having specific values occur.

Colour Histogram. Colour histogram is the colour descriptor that represents the colour feature of an image. It represents the continuous distribution of colours in an image. It is the representation of the number of pixels belonging to a particular colour range.

- ***Image Classification***

The classification of images in different labels is fulfilled by 5 classifiers so that we can compare the performance of every model on the same dataset and get better results. The five classifiers are explained as follows:[15]

SVM. Support Vector Machine is a supervised machine learning algorithm used for classification and regression purposes. It can handle both linearly separable as well as non-linearly separable data points. A hyperplane drawn between two classes acts as a boundary to decide the class of the object. A line parallel to the hyperplane is drawn on both sides of the hyperplane and the complete distance from one line to another is known as the margin. The points nearest to the hyperplane on both sides are known as support vectors. The maximum margin hyperplane is chosen for classifying instances.

Logistic Regression. Logistic regression is the statistical classification algorithm for the categorical class variables. Our goal is to predict and categorize plant diseases into target classes. A simple Logistic regression algorithm islimited to only binary classification but here we have more than two classes so we will use multiclass logistic regression. The multiclass logistic regression trains n binary classifiers for n target classes for each class *I* to determine the probability of the target class.

KNN. The supervised machine learning method K-Nearest Neighbour categorizes data points based on their similarity notion. It assumes that there is some closeness between similar things. The distance between the new data point and the data points in the training set is used to determine the similarity notion. K minimum distant points are assumed to be the most similar points. It extracts the most similar k data points from the training set and categorizesthe new data point to the highest occurring category of K similar data points.

Random Forest. Random forest is built on the notion of supervised methods, which combines numerous weak classifiers to create a strong classifier and improve the model's performance. Random Forest works on both training labels and validation labels of the dataset. Several decision trees are formed on the subsets of the given dataset. Each decision tree is performed on its own and the class of the instance is predicted based on majority voting of the prediction class predicted by each decision tree.

CNN. Convolutional Neural Network, or ConvNet, is a type of artificial neural network with multiple hidden layers that is often used for image data training. The Convolution layer, ReLU layer, and Pooling layer are the three layers that make up the hidden layer. There are numerous convolution layers, and for each one, we must define the number of filters that the layer should have. A filter is a NxN matrix with arbitrary numeric values that slides across the image and conducts the convolutional operation on every NxN block. The feature maps obtained from the convolving procedure are fed into the ReLU layer as input. To obtain a corrected feature map of the input picture, the ReLU activation function is used to the convolution layer. ReLU layer makes all the negative values 0. The ReLU layer introduces non-linearity to the network. It performs the element-wise operation. . Pooling layer reduces the dimensionality of the image after having convolving and ReLU operations.Is it simply down sampling the image by performing a pooling operation? The Fully connected layer consists of an artificial neural network that takes a flattened feature vector as input and adjusts some weights in given iterations.The iterations are commonly known as epochs. This layer provides classification results as output. It uses the soft-maxfunction to predict the class of a given leaf. An image is given as input to the number of hidden layers. The hidden layer performs feature extraction by performing some calculations and the output of the hidden layer is flattened ten by the flattening process and given as input to the feed-forwarded neural network in the form of a single feature vector. The feed-forward network outputs the result. Figure8 shows the

working of CNN model.Figure[8].[14]

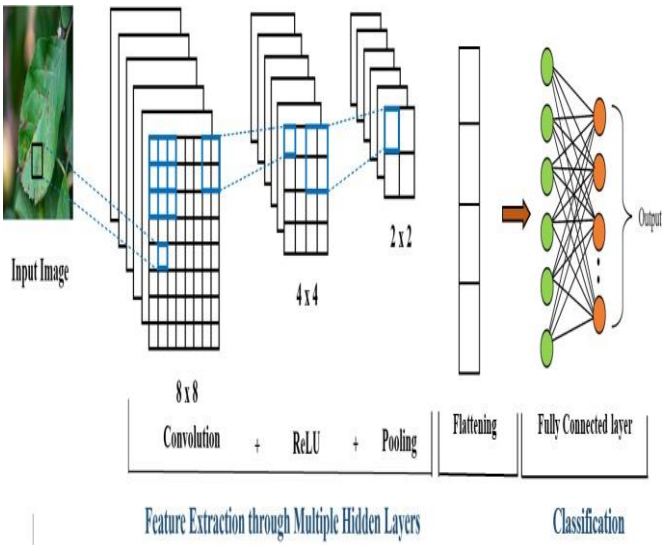

Figure 8. CNN model Working Architecture

IV Results Analysis

In this experiment, 53200 image samples were used, which contain the most frequent fungal, bacterial, and viral pathogens occurring in the plants. Initially, all of the sample images are reduced to a fixed size of (256 x 256) pixels. The classification was performed on five supervised machine learning algorithms namely SVM(Support Vector Machine), Logistic Regression(LR), K Nearest Neighbors(KNN), Random Forest(RF), and Naïve Bayes (NB), and Deep learning algorithm Convolutional Neural network(CNN).Figure 9: shows the classification accuracy achieved by all the classifiers, where the Random Forest algorithm achieves the percentage of accuracy i.e 97.12% among all the machine learning algorithms used in this work along with graphical representation [Figure 9].

Models Used in the Study		**Diseases Prediction Accuracy (%)**
Machine Learning Models	Logistic Regression(LR)	71.89%
	Support Vector Machine(SVM)	75.76%
	K-Nearest neighbour (KNN)	82.17%
	Random Forest(RF)	97.12%
	Naïve Bayes Classifier(NB)	81.12%
Deep Learning Model	Convolutional Neural Network(CNN)	98.43%

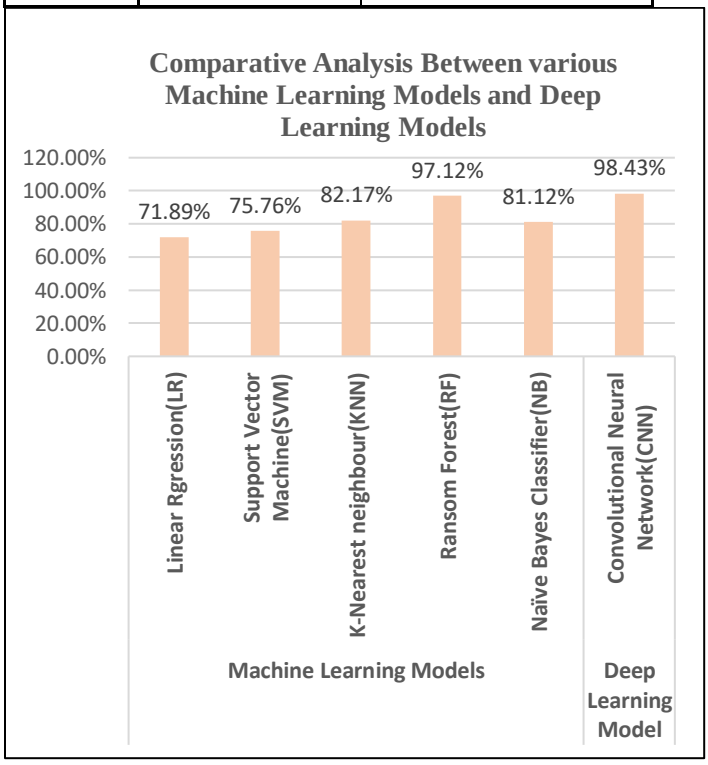

Figure 9. Graphical Representation of Machine learning models and Deep Learning Models Performance Classification Accuracy

Using a similar dataset, used to train a CNN model on all 5,3200 images of a dataset containing 30 classes. Our CNN model took 11,206,302 total parameters out of which 11,196,702 are trainable parameters and 9600 are non-

trainable parameters. We have trained our model on 60 epochs and achieved an accuracy of 98.43%. As we increase the number of epochs, the training and validation accuracy improves and training and validation loss reduce. Figure 10 shows the graph achieved on training and validation accuracy. Figure 11 shows a graph between training and validation loss.

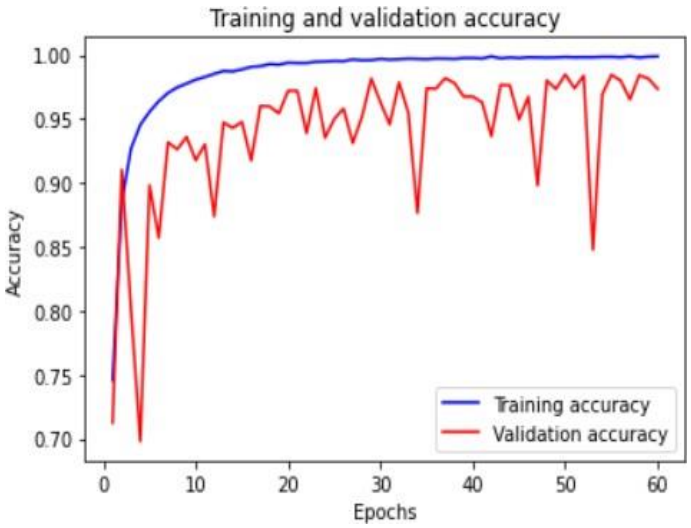

Figure 10. Training and Validation accuracy graph of CNN

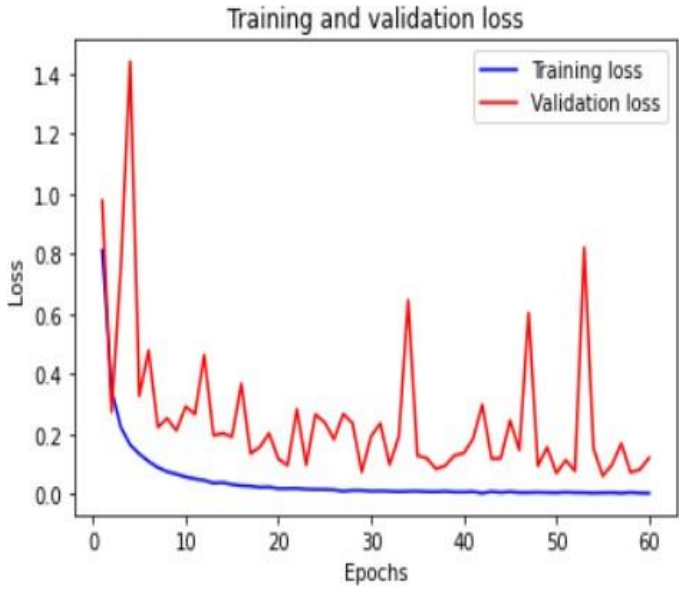

Figure 11. Training and Validation loss graph of CNN

V. Conclusion and Future Scope

In this research article, an approach for plant disease prediction and classification has been proposed using deep learning and machine learning classifiers. Classification is performed after the segmentation and feature extraction process. Diseases have been broadly classified into three categories namely Fungal, Bacterial and viral. A dataset consisting of 5,3200 images has been trained upon SVM, KNN, Logistic Regression, Random Forest, and CNN. Among all the classifiers, CNN gives the best prediction accuracy of 97.34%. In the future, image augmentation can also be applied to the dataset to increase its size and compare the results between all classifiers as well as The agricultural department seeks to automate the process of recognizing high-yield crops (real-time). This method can be automated by displaying the prediction result in a web or desktop application. To make the work easier to implement in an Artificial Intelligence context.

VI. References

[1].Iqbal, M. A., & Talukder, K. H. (2020, August). Detection of potato disease using image segmentation and machine learning. In 2020 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET) (pp. 43-47). I

[2]. Mohan Sai, S., Gopichand, G., Vikas Reddy, C., & Mona Teja, K. (2019). High Accurate Unhealthy Leaf Detection. arXive-prints, arXiv-1908.

[3].Raut, V. R., & Nage, M. A. (2019). Detection and Identification of Plant Leaf Diseases based on Python.

[4]. Francis, M., & Deisy, C. (2019, March). Disease detection and classification of inagricultural plants using convolutional neural networks—a visual understanding. In 2019 6th International Conference on Signal Processing and Integrated Networks(SPIN) (pp. 1063-1068). IEEE.

[5].Panigrahi, K. P., Das, H., Sahoo, A. K., & Moharana, S. C. (2020). Maize leaf disease detection and classification using machine learning algorithms. In Progress in Computing, Analytics, and Networking (pp. 659-669). Springer, Singapore.

[6].Sardogan, M., Tuncer, A., & Ozen, Y. (2018, September). Plant leaf disease detection and classification based on CNN with LVQ algorithm. In 2018 3rd International Conference on Computer Science and Engineering (UBMK) (pp. 382-385). IEEE.

[7].Vyshnavi, G. K. P., Sirpa, M. R., Chandramoorthy, M., & Padmapriya, B. (2016). Healthy and unhealthy plant leaf identification and classification using hierarchical clustering. Int. Res. J. Eng. Technol. (IRJET), 3, 448-453.

[8].Khirade, S. D., & Patil, A. B. (2015, February). Plant disease detection using image processing. In 2015 International conference on computing communication control and automation (pp. 768-771). IEEE.

[9].Rothe, P. R., & Kshirsagar, R. V. (2015, January). Cotton leaf disease identification using pattern recognition

techniques. In 2015 International conference on pervasive computing (ICPC) (pp. 1-6). IEEE

[10].Arivazhagan, S., Shebiah, R. N., Ananthi, S., & Varthini, S. V. (2013). Detection of the unhealthy regions of plant leaves and classification of plant leaf diseases using texture features. Agricultural Engineering International: CIGR Journal, 15(1), 211-217.

[11]. F. Ahmed, H. Abdullah Al-Mamun, A. S. M. Hossain Bari, E. Hossain and P. Kwan, Classification of crops and weeds from digital images: a support vector machine approach, *J. Crop Prot.* **40** (2014), 98–104.

[12]. J. Baghel and P. Jain, K-means segmentation method for automatic leaf disease detection, *Int. J. Eng. Res. Appl.* **6** (2016), 83–86.

[13]. Singh, V., & Misra, A. K. (2017). Detection of plant leaf diseases using image segmentation and soft computing techniques. Information Processing in Agriculture, 4(1), 41-49.

[14]. Kaur, L., & Laxmi, V. (2016). Detection of Unhealthy Region of plant leaves using Neural Network. Disease management, 1(05), 34-42.

[15] Ishak, S., Rahiman, M.H.F., Kanafiah, S.N.A.M., Saad, H.: Leaf disease classification using artificial neural network. J. Teknologi, 77(17) (2015).

[16]. Joshi, K., Diwakar, M., Joshi, N. K., & Lamba, S. (2021). A Concise Review on Latest Methods of Image Fusion. Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science), 14(7), 2046-2056.