IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, August 2023

Prediction of Disease in Tomato Leaves with use of Machine Learning Technique

Sandeep K H[1] and Rakesh B S[2]
Assistant Professor, Department of CSE, PES Institute of Technology & Management, Shivamogga, India [1]
Assistant Professor, Department of ISE, Vemana Institute of Technology, Bengaluru, India[2]

Abstract: *India's sizable agricultural market provides the perfect conditions for cultivating a variety of products, including the tomato harvest. Detecting the transmission of diseases from unhealthy to healthy plants poses a severe threat to the agricultural industry because, if caught early enough, they can quickly spread and perhaps infest the entire farm. In terms of profit in good forming, early stage crop disease identification and severity monitoring is quite important. K-means clustering with fuzzy logic is used in the proposed study to evaluate the disease-affected region of the leaf and, as a result, assess the severity of the diseases. In this thesis, illnesses are detected using machine learning models for convolutional neural networks (CNN) and K-nearest neighbours (KNN).*

Keywords: Convolutional neural networks (CNN), K-nearest neighbours (KNN), K-means Clustering, Fuzzy logic, Severity.

I. INTRODUCTION

In India, tomatoes are a common crop, and approximately 90% of farmers decide to plant them in soil that drains well. In the country, they grow on more than 3,500,000 hectares of land and produce about 54,000 tons each year. While tomatoes are also grown in gardens, it can be difficult for both farmers and gardeners to achieve ideal growth and yield. The impact of plant disease, which not only results in major losses of products intended for human consumption but also adversely impacts the livelihoods of farmers who depend on healthy crops for income, is one ofthe biggest dangers to the agriculture business. Therefore, whenever a crop has a disease or a deficiency, automatically detecting diseases and gauging their severity will have a positive impact.

An important study field is the automatic detection of crop leaf diseases and severity measurement since it has the potential to significantly improve farmers' yields and food security. If diseases are not addressed in a timely manner, serious loss results. Pesticide use should be kept to a minimum because it can burden farmers and pollute the environment even when used excessively. The risk of toxicity in agricultural products directly harming human health can result from excessive consumption. If we can determine the severity of the disease, we can efficiently decide how much pesticide should be used, and it can be targeted to the affected area. We can't rely on visual observation with our unaided eyes for precise outcomes. On the other hand, a computerized system that employs the machine learning method may be extensively employed to monitor the health of the plant, its growth, and even the early detection of different diseases.

Each type of disease typically leaves behind patterns on plant leaves that can be used to identify it. With the aid of these patterns, machine learning ought to be able to offer a productive and affordable solution for the growth of tomato plants. This paper's main goal is to accurately identify illnesses that damage tomato leaves. In this work, K-means clustering using fuzzy logic is used to analyse the area of the leaf that is impacted by the disease and, as a result, to determine how severe the disease is. The implemented CNN and KNN work on the collected data to classify the kind of sickness.

II. RELATED WORK

A deep convolution neural network model was utilized in a study on the identification of plant diseases by Madhulatha and Ramadevi (2020), and the results showed that the proposed work produced an accuracy of 96.50% [1]. The study uses the renowned AlexNet architecture to categorise the many plant diseases. Most image classification use case

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, August 2023

scenarios use the AlexNet architecture, a Neural Network with eight layers of learnable features. All of the photos from the Plant Village dataset, which includes 54,323 photographs of plant diseases and 38 different disease categories, were used to create the dataset for this study.

A study on the detection of plant diseases was conducted by Hatuwal, Shakya, and Joshi [2] using a variety of machine learning models, including Support Vector Machine (SVM), K-nearest Neighbour (KNN), Random Forest Classifier (RFC), and Convolution Neural Network (CNN). The CNN model had the best accuracy among all machine learning models, scoring 97.89%, followed by the RFC model with 87.436%, the SVM model with 78.61%, and the KNN model with 76.969%. Contrary to earlier studies, this one used precision, recall, and f1 scores to assess its models; however, when the models were ultimately compared, accuracy was only taken into account when determining the top-performing model.

In a study by Agarwal et al. [3] on the detection of diseases in tomato leaves using convolutional neural networks, the proposed CNN model outperformed pre-trained CNN models like VGG16, Mobilenet, and Inception, scoring 91.2% accuracy, 77.2% accuracy, and 63.4% accuracy, respectively. Three Convolution layers and three Max pooling layers make up the suggested CNN model in this study. This study by Agarwal et al. (2020) further illuminates the advantages of not employing a pre-trained model, revealing that the suggested model used far less storage space—1.5 MB—than pre-trained models, which required 100 MB.

The proposed work obtained a test accuracy of 76.59%, according to research on the detection of illnesses in paddy leaves using the KNN classifier conducted by Suresha, Shreekanth, and Thirumalesh [4] on a database including 330 photos of paddy leaves. The study did not consider metrics like precision, recall, or f1-score, which would be the subject of this work, while evaluating the KNN classifier; instead, it used accuracy as the only parameter.

In a study [5] by SujanSarkar et al. (2018) the primary issue of many sorts of crop leaf diseases has been addressed in this paper. They effectively separated the leaf pictures and divided them into the defective and normal regions using image processing. This segmentation makes it simpler to determine whether crop leaf diseases exist simply by using a digital image. They implemented the k-means clustering algorithm using the unsupervised learning technique, which is significantly more accurate than the current techniques for segmenting the defective region in leaves. They are attempting to increase the accuracy and establish a ratio between the healthy and defective zones.

III. METHODOLOGY

Plant leaf disease detection and severity measurement involves a few fundamental image processing steps in order to identify and categorize plant leaf disease. These processes include leaf disease detection, image acquisition, image pre-processing, image segmentation, feature extraction, and classification. Below is a description of these steps.

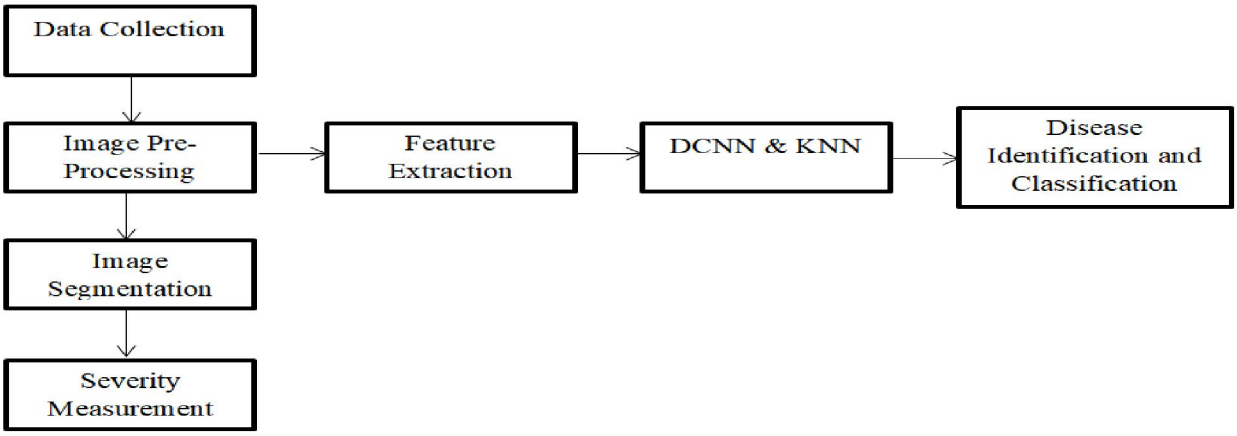

Fig 1: Methodology

Data Collection: There are eleven main classifications in the dataset that was used for this study. One leaf class stands in for the healthy leaf class, while the other ten represent the unhealthy. For a total of about 7000 leaf images, there are more than 1500 examples in each class.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, August 2023

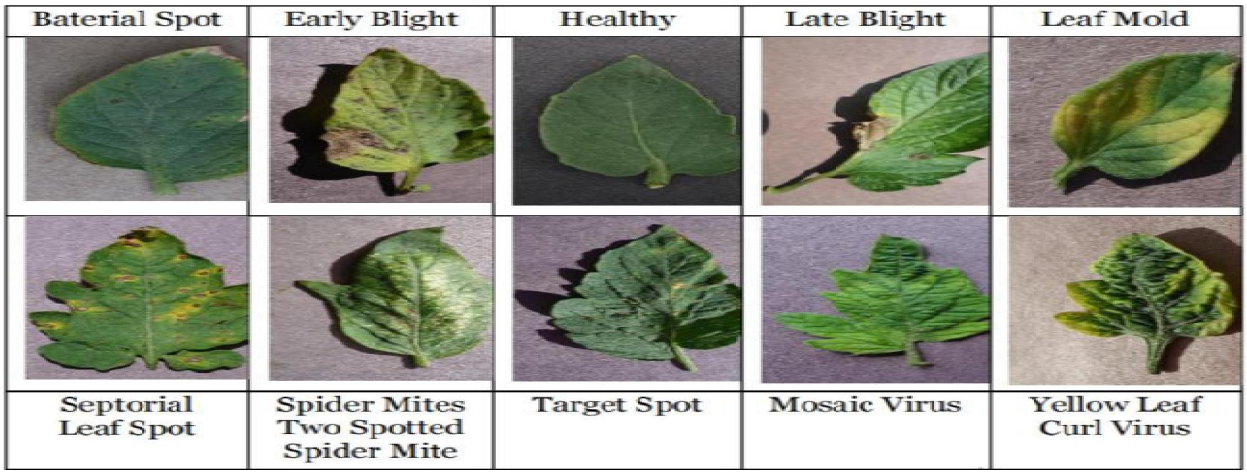

Fig 2: Various tomato leaf diseases

Image Pre-Processing: The basic goal of image pre-processing is to strengthen certain image features or enhance the image information that contains undesired distortions in preparation for any processing. Pre-processing techniques include dynamic picture size and form, noise filtering, image conversion, image enhancement. Dewdrops, dust, and insect waste on the plants make up the visual noise of the tomato leaf. To address these issues, the RGB input image is converted to a gray scale image for precise results. The image's background is first eliminated, but because the image size is so huge in this case, it needs to be resized. The image is shrunk to a size of 256 × 256 pixels. Denoising of the image must be done if there is noise.

Image Segmentation: Image segmentation is crucial for the identification and classification of plant diseases. The image is simply segmented into different objects or parts. It evaluates visual input to glean information that could be processed further. To locate illness regions, identify the type of disease, and classify the conda critical phase in which the area of interest is chosen. One should be cautious enough to select an appropriate method because it has such a vital part to play. The approaches include thresholding, binarization, and K-means clustering, among others. In the suggested study, K-means clustering has been used to pick the region of interest from five (k=5) clusters.

In terms of image processing, it is among the most popular segmentation algorithms. Centroids are computed repeatedly in this approach until the best result is obtained. The number of clusters to be created should be known from the outset, and it is indicated by K. The ideal way for choosing K value is the elbow method. To allocate the data points, the Euclidian distance between the centroids is determined.

Severity Measurement: Segmentation produces regions of interest in the photos. One should carefully pick photographs with the afflicted part to determine the severity. For improved visibility and accuracy, five clusters are being constructed in the proposed work. The majority of the time, no single cluster has an unhealthy section because the colour and surface of the leaf vary depending on the condition. To determine the overall area of the diseased leaf, we must choose multiple clusters, and their associated binary images are made. The area is then expressed in pixels. The same pixels collected will then be used to compute severity. The disease-affected area discovered in a particular region of interest is divided by the total leaf area discovered during pre-processing to determine the percentage of affected area.

One can quickly determine the severity based on the area that is impacted. The user will be guided by severity to take the proper step to stop future crop loss. The suggested work employs fuzzy logic to determine the severity and to estimate the overall area affected. The output of fuzzy logic is determined based on the sets and assumptions. Horsfall and Heuberger values are used to determine the severity.

Feature Extraction: The majority of current and historical feature extraction strategies exclusively focus on the leaf's visual features and attempt to extract the essential data from them. The system takes into account the leaf image as well as sensor data. Between visual features and environmental factors like humidity, temperature, and soil moisture, it maps the semantic representation.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, August 2023

DCNN: The data can be used to learn feature representations using deep learning algorithms. A CNN model uses images as its input, performing one or more convolution operations before extracting the image's features and reducing the input's dimensionality. These retrieved features have a direct impact on the accuracy of image processing. Convolution, ReLU Layer, Pooling, Fully Connected, Flatten, and Normalization are just a few of the layers that make up the CNN model. The photographs would be compared piece by piece using CNN. A feature or filter is the name for each component. CNN employs the weight matrix to analyse the input image and extract the precise characteristics while preserving the spatial arrangement information.

KNN: Based on the vote of the majority labels that the data points closer to the new sample are located in, the K-nearest Neighbour machine learning algorithm calculates how likely it is that a new sample belongs to a label. In order to sort the calculated values in ascending order, the algorithm first calculates the distance between the new sample and each of the other samples. The majority label is selected as the forecast based on the 'k' value. The integer "k" here denotes how many of the closest neighbours will be considered for choosing the majority label. The algorithm's predictions are made or broken by the choice of 'k'in the KNN. The classification is particularly susceptible to noise if the value of 'k' is too tiny.Disease Identification and Classification: Finally, the datasets are trained and tested using classifiers. These classifiers could be based on fuzzy logic, neural networks, k-nearest neighbour, etc. These techniques are employed to categorize and find leaf diseases.

IV RESULTS

The following results show the percentage of disease affected area estimated using K-means algorithm.

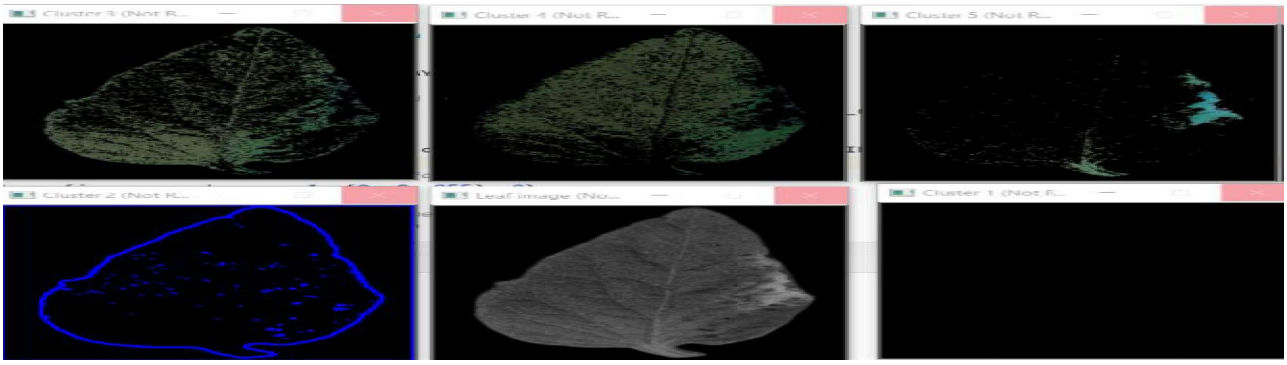

Fig. 3: Clustering

Fig. 4: Original

Fig. 5: Clustered

```
"C:\Program Files\Python37\python.exe" D:/severity/severity.py
Total number of pixels in leaf:  98881.5
Enter the cluster number to choose (1-5): 5
Number of pixels in ROI from cluster 5 is 1841
Percentage of Affected Area :  1.8618245071120483
```

Fig 6: Output for severity of disease in percentage

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, August 2023

The following results present the Accuracy, Precision, Recall and F1-Score of both the CNN and KNN model.

Accuracy(%)	Precision(%)	Recall(%)	F1-Score(%)
98.5	92	91	93

TABLE 1- CNN EVALUATION METRICS

Accuracy(%)	Precision(%)	Recall(%)	F1-Score(%)
98.5	92	91	93

TABLE 2-KNN EVALUATION METRICS

For the CNN model, the training vs. validation accuracy and training vs. validation loss were also plotted. As shown in Figs. 5 and 6, as training accuracy rises, so does validation accuracy, and as training loss falls, so does validation loss, indicating that the model is neither under fitting nor over fitting and is continuously learning.

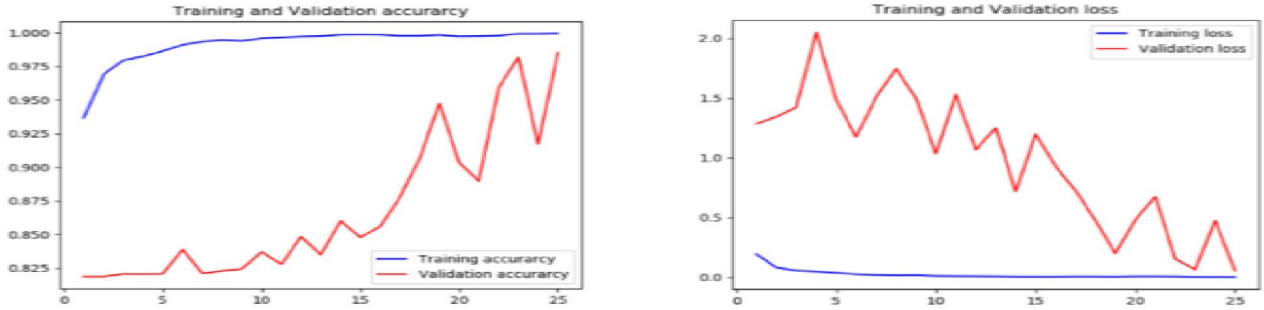

Fig 7: Training v/s Validation Accuracy and loss

V. CONCLUSION

Today, timely detection and identification of diseases that affect the leaves is crucial because they seriously harm both the quantity and quality of crop production. This study provides a model that uses a machine learning approach to pre-process a total of 7000 photos. The study demonstrates that the CNN model performs better than the KNN model in the plant disease detection of tomato leaves by outperforming the KNN model in all four evaluation metrics. The study implements two machine learning models, Convolutional Neural Networks (CNN) and K-nearest neighbors (KNN), on the disease detection of tomato leaves and also evaluates the aforementioned model using the following metrics: Accuracy, Precision, Recall, and F1-Score.

REFERENCES

[1]. Madhulatha, G. and Ramadevi, O. (2020) _Recognition of Plant Diseases using Convolutional Neural Network‘, in 2020 Fourth International Conference on I˘SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC). 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), pp. 738–743.

[2]. Agarwal, M. et al. (2020) ToLeD: Tomato Leaf Disease Detection using Convo˘lution Neural Network‘, Procedia Computer Science, 167, pp. 293–301.

[3]. Hatuwal, B. K., Shakya, A. and Joshi, B. (2020) _Plant Leaf Disease Recognition Using Random Forest, KNN, SVM and CNN‘, POLIBITS, 62, p. 7.

[4]. Suresha, M., Shreekanth, K. N., Thirumalesh, B. V. (2017) _Recognition of diseases in paddy leaves using knn classifier‘, in 2017 2nd International Conference for Convergence in Technology (I2CT). 2017 2nd International Conference for Convergence in Technology (I2CT), pp. 663–666.

[5]. SujanSarkar et al. —Fault Area Detection Using K-means Clustering‖, International Journal of Computer Applications, August 2018.

[6]. Vinutha B Hiremath, SandarshGowda MM —Disease Prediction of Tomato Leaf Using CNN, Deep Learning Techniques”, International Research Journal of Modernization in Engineering Technology and Science, June-2022.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, August 2023

[7]. Sachin D. Khirade and A. B. Patil. ―Plant Disease Detection Using Image Processing.‖ International Conference on Computing Communication Control and Automation (ICCUBEA), 2015 International Conference on,pp. 768-771. IEEE, 2015.