

CLASSIFICATION OF TOMATO DISEASES USING ENSEMBLE LEARNING

S. Jeyalakshmi and R. Radha

Department of Computer Science, Shrimathi Devkunvar Nanalal Bhatt Vaishnav College For Women, India

Abstract

A Plant disease is any dysfunction of a plant, caused by living organisms, which affects the quality and quantity of yield. These symptoms are visually shown on the plant leaves. This paper discusses classification of Tomato diseases such as Late Blight, Septoria Leaf Spot and Yellow leaf curl virus while distinguishing the healthy leaf at the same time. An experimental sample size of 1817 was considered in conducting this study. This work differentiates diseased tomato leaf images with healthy leaf images. The classifiers Random Forest, Multilayer Perceptron Neural Network and Support Vector Machines were trained and got a prediction accuracy of 88.74%, 89.84%, and 92.86% respectively in classifying diseases. Then, the prediction results of Random Forest, Multilayer Perceptron and Support Vector Machines were combined using Soft Voting classifier and obtained a highest accuracy of 93.13% in classifying tomato diseases.

Keywords:

Tomato diseases, Support Vector Machines, Multilayer Perceptron, Random Forest, Voting Classifier

1. INTRODUCTION

A disease is any abnormal growth of a plant. Disease may damage some or all parts of a plant thereby causing loss of production. Any visually noticeable change in color is an evidence of plant disease. If the disease is left unattended, the disease may spread to neighboring plants and will result in severe loss. Hence, early diagnosis of these diseases is essential to reduce the loss.

In this study tomato disease such as Late Blight, Septoria Leaf Spot and Yellow Leaf Curl Virus were considered while differentiating the healthy leaves at the same time. Late Blight is one of the serious tomato diseases caused by the fungus *Phytophthora infestans* [1]. The disease spots appear as dark and water soaked on young leaves. The old leaves show greasy and grayish indefinite patches. Septoria leaf spot is a fungal disease that appears as a dark brown edge with a white or gray center [1]. As the disease develops the surrounding spots turn yellow causing the leaf to die. Tomato yellow Leaf curl disease is caused by Tomato yellow Leaf curl virus. The symptoms of leaves include reduction in size, upward cupping and marginal chlorosis [1]. Healthy tomato leaf and diseased leaves are shown in Fig.1.

The suggested work extracts color, histogram and texture features from infected leaf tissues and classifies the disease using random forest classifier, Multilayer perceptron neural network with one hidden layer and Support Vector Machine. Further, the prediction results were improved by the ensemble learning model soft voting classifier.

Section 2 describes the related work in this field. Section 3 discusses the Materials and methods used in this paper. Section 4 discusses the proposed algorithm in detail, Section 5 details results obtained while Section 5 gives the conclusion and future work.

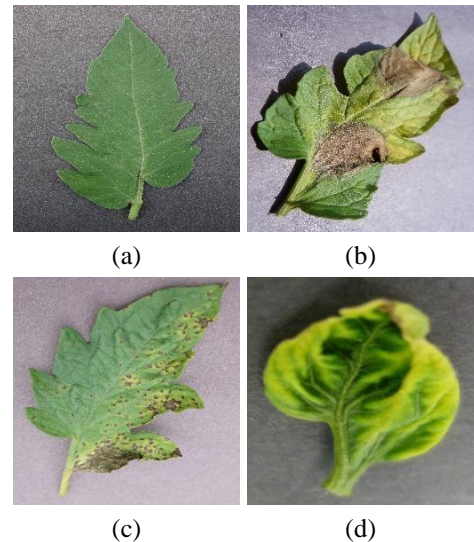


Fig.1. Types of Leaf Symptoms (a) Healthy (b) Late Blight (c) Septoria Spot (d) Yellow Leaf Curl Disease

2. LITERATURE SURVEY

Abed and Esmaeel [2] had extracted the features like Contrast, Correlation, Energy, Homogeneity, Mean, Standard Deviation, Entropy, Variance, Smoothness, Kurtosis, Skewness and Inverse Difference Moment (IDM) in detecting and classifying Bean diseases.

Jayamala Patil and Raj Kumar [3] had suggested the features namely Mean, Standard Deviation and Skewness of pixel values in R, G and B components to detect the diseases, Early blight, Late blight and Septoria leaf spot in tomato.

Juan-hua Zhu et al. [4] had extracted shape features such as area, perimeter, rectangularity, circularity and shape complexity of lesion region from chain code and segment Table. These parameters were used to identify the disease.

Arivazhagan et al. [5] proposed an algorithm for detecting bacterial, fungal and viral diseases on plants from plant leaves. In this work, the RGB image was converted into HSI (Hue Saturation and Intensity) color space and the Healthy region was masked from the plant leaf by thresholding the Hue component value of the leaf. Texture statistics namely, Contrast, Energy, Local homogeneity, Cluster shade and Cluster Prominence were calculated for useful segments by obtaining Grey level co-occurrence matrix. Tomato leaf was classified into types of diseases using Minimum distance criterion and Support Vector Machines and had obtained an accuracy of 95.24 and 92.5% with a training set size of 10 images and test set size of 27 images.

Patil and Raj Kumar [6] had proposed a method for detecting Grey spot and common rust diseases. The input RGB image was converted into Grey scale and Grey level co-occurrence matrix (GLCM) for different offset values specified by distance 'd' and

angle ' α ' were generated. Texture properties like correlation, inertia (contrast), energy and homogeneity were obtained from GLCM.

Tigadi and Sharma [7] had classified and graded banana plant diseases namely Banana top Virus, Black Sigatoka Yellow Sigatoka, Panama Wilt and Banana Streak Virus using feed-forward back propagation neural networks. Color features such as mean and standard deviation of HSV color space and Histogram Of Templates (HOT) texture and gradient magnitude features were used in classifying the diseases.

Muthukannan et al. [8] classified diseases in bean and bitter gourd plant leaves using Feed Forward Neural Network Algorithm, Learning Vector Quantization Algorithm and Radial Basis Function Algorithm from texture Contrast, Homogeneity, Energy, Correlation and shape feature area. Among the three methods, Feed Forward Neural Network Algorithm gave a best accuracy of 90.7%.

Mokhtar et al. [9] have proposed a method for identifying tomato virus diseases tomato yellow leaf curl Virus and tomato spotted wilt orthospovirus using Support Vector Machines with data set size 200 images. The SVM model was trained with different kernel functions such as linear, RBF, QP, MLP and Polynomial kernels. The study found that QP kernel gave a highest accuracy of 91.5% in identifying the disease.

Sabrol and Kumar [10] have suggested intensity-based feature extraction for classification of tomato diseases Bacterial Leaf Spot, Septoria Leaf Spot, Fungal Late Blight, Bacterial Canker and Leaf Curl along with Healthy leaf. The data set sized used for this study was 520 images. Mean, Standard Deviation and Skewness of X, Y and Z components of CIEXYZ color space were considered as feature set and the diseased images were classified using decision tree and obtained an accuracy of 72%.

James and Punitha [11] had proposed ensemble learning approach for classification of tomato diseases Anthracnose, Bacterial cranker, Bacterial spot, Bacterial speck, Early blight and Late blight. The shape features such as centroid, eccentricity, orientation, diameter, solidity, extent and perimeter; color co-occurrence features such as energy, correlation, contrast, entropy and homogeneity; color statistical features like hue, saturation, red, green, blue, meanR, meanG and meanB were extracted. Tomato diseases were classified using Multi-class ensemble methods like LogitBoost, TotalBoost and AdaBoost, algorithms and obtained an accuracy of 82.00%, 86.83% and 92.33% respectively.

3. MATERIALS AND METHODS

3.1 MATERIALS

The images used in this research work were taken from Plant Village dataset [12]. A dataset size of 1334 diseased leaf images and 483 healthy leaf images were considered to pursue this work.

3.2 METHODS

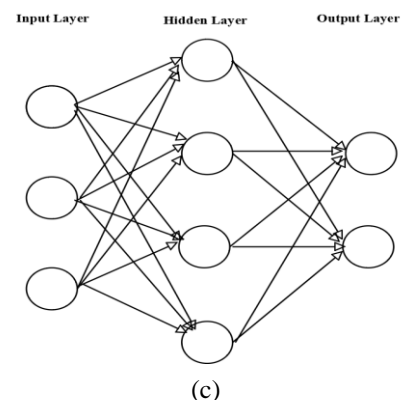
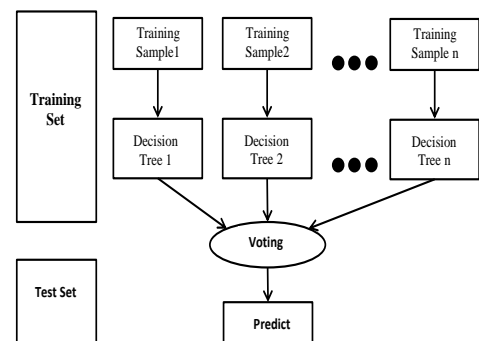
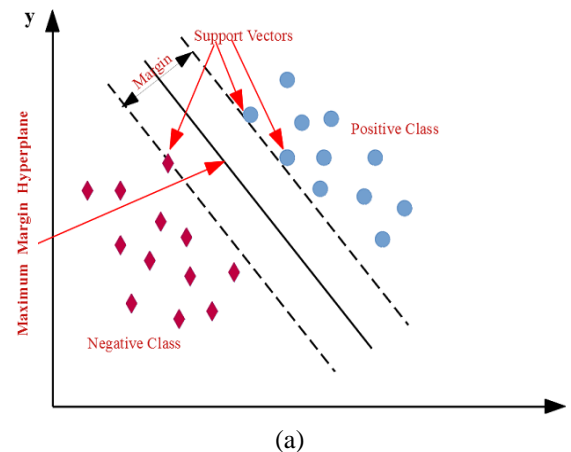
3.2.1 Texture Features:

Texture features are extracted from Gray Level Co-occurrence matrix (GLCM). GLCM tabulates the frequency of combination of gray levels that occur in an image. Grey Level Co-occurrence

matrix can be expressed as probability using the following normalization Eq.(1):

$$P_{i,j} = \frac{V_{i,j}}{\sum_{i,j=0}^{N-1} V_{i,j}} \quad (1)$$

where i and j are row and column numbers respectively. V is the value in cell i, j of the image window. $P_{i,j}$ is the probability value for the cell i, j . N is the total number of rows and columns. The following texture measures can be calculated from GLCM and can be used in analysing the Grey level intensities of neighbouring pixels.



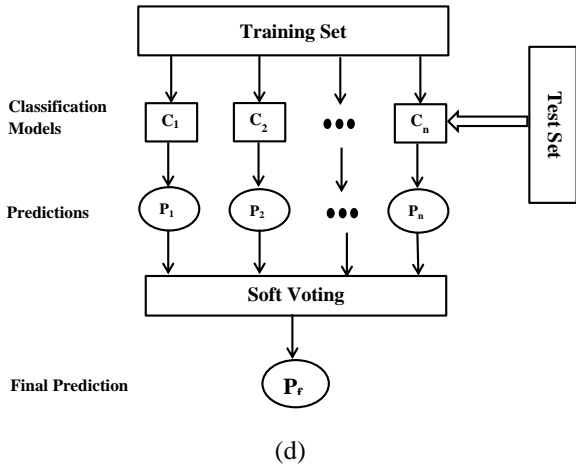


Fig.2. Classifiers (a) SVM (b) Multilayer Perceptron with one hidden layer (c) Random Forest (d) Soft Voting

Contrast and Dissimilarity: Contrast and dissimilarity are calculated with weights assigned to the distance from the GLCM diagonal. Exponential (0,1, 4, 9,16 and so on) and linear (0,1,2,3,4 and so on). Weights are assigned to Contrast and Dissimilarity [13]. Contrast and Dissimilarity can be calculated using Eq.(2) and Eq.(3) respectively.

$$Contrast = \sum_{i,j=0}^{N-1} P_{i,j} (i - j)^2 \quad (2)$$

$$Dissimilarity = \sum_{i,j=0}^{N-1} P_{i,j} |i - j|^2 \quad (3)$$

Homogeneity: Homogeneity is a measure that gives the value of how close the elements are distributed to that of the diagonal of the matrix [13] and it is given by Eq.(4).

$$Homogeneity = \sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1 + (i - j)^2} \quad (4)$$

Angular Second Moment (ASM): Angular Second Moment, given by Eq.(5), is the square of the elements added together in the matrix [13].

$$ASM = \sum_{i,j=0}^{N-1} P_{i,j}^2 \quad (5)$$

Energy: Energy is the square root of ASM [13]

$$Energy = \sqrt{ASM} \quad (6)$$

Correlation: The Correlation texture measures the linear dependency of grey levels on those of neighboring pixels [13] given by Eq.(7).

$$Correlation = \sum_{i,j=0}^{N-1} P_{i,j} \left[\frac{(1 - \mu_i)(j - \mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}} \right] \quad (7)$$

Entropy: Entropy is a statistical measure of randomness [13], calculated using Eq.(8).

$$Entropy = \sum_{i,j=0}^{n-1} P_{i,j} (-\ln P_{i,j}) \quad (8)$$

3.2.2 Classifiers:

This section discusses machine learning models Support Vector Machines, Multilayer Perceptron, Random Forest and Soft Voting Classifiers used in this research work.

Support Vector Machines: Support Vector Machines are supervised machine learning algorithms that generate hyperplane that separates classes with large margin. Linear Kernel SVM was used in this work. SVM classifier is illustrated in Fig.2(a).

Multilayer Perceptron Classifier: A Multilayer Perceptron classifier is a feed forward Neural Network [14]. An MLP has at least 3 types of layers, an input, an output and a hidden layer. The number of hidden layers may be increased based on the complexity of the problem. The input data is fed into the first layer of the network the input layer and outputs or predictions are taken from the output layer. The hidden layers are located between input and output layers. They perform computations on the weighted inputs and produce net input which is then applied with activation functions to produce the actual output. The number of hidden layers and number of hidden neurons is chosen in such a way that it results in high accuracy and low error. The Fig.2(b) shows a Multilayer Perceptron with one hidden layer.

Random Forest: Random Forest Classifier is a supervised learning algorithm [15]. This algorithm creates decision trees on a dataset, gets the prediction from each of them and selects the best one by voting. The Fig.2(c) demonstrates the Radom Forest classifier.

Soft Voting Classifier: A Voting Classifier is a meta-classifier for combining prediction results of various machine learning classifiers by majority voting [16]. A soft voting classifier predicts the output based on highest probability of chosen class as the output. Fig.2(d) details the Soft Voting Classifier.

3.2.3 Performance Metrics:

The following performance analysis are performed in this study.

Confusion Matrix: A confusion matrix is an $N \times N$ matrix where N is the number of classes. A confusion matrix for a multiclass classifier is shown in Fig.3.

In Confusion matrix, True Positive denoted by TP is the Actual Positives predicted to be Positive. False Negative denoted by FN is the Actual Positives predicted to be Negative. True Negative indicated by TN is the Actual Negatives predicted as Negative. False Positive denoted by FP is the Actual Negatives predicted as Positive.

		Predicted Label			
		Class 1	Class 2	...	Class N
True Label	Class 1	TP	FN	...	FN
	Class 2	FP	TN	...	TN
	⋮	⋮	⋮	...	⋮
	Class N	FP	TN	...	TN

Fig.3. Confusion matrix for multiclass classifier

Accuracy: The classification accuracy gives the overall correctness of the classifier and is calculated by Eq.(9) [17]

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (9)$$

Accuracy is measured as a percentage. A high accuracy specifies that the classification model performs well.

Precision: Precision is the number of true positives divided by the number of true positives and false positives, represented in percentage [18]. A high precision shows less number of False Positives. It can be calculated using Eq.(10).

$$Precision = TP / (TP + FP) \quad (10)$$

positives and false negatives [18]. A high recall indicates a small number of False Negatives.

$$Recall = TP / (TP + FN) \quad (11)$$

A high recall and a low precision indicate most of the positive examples are recognized correctly but there are more false positives. On the other hand, a low recall and a high precision means that a lot of positive examples are wrongly predicted and false positives are high.

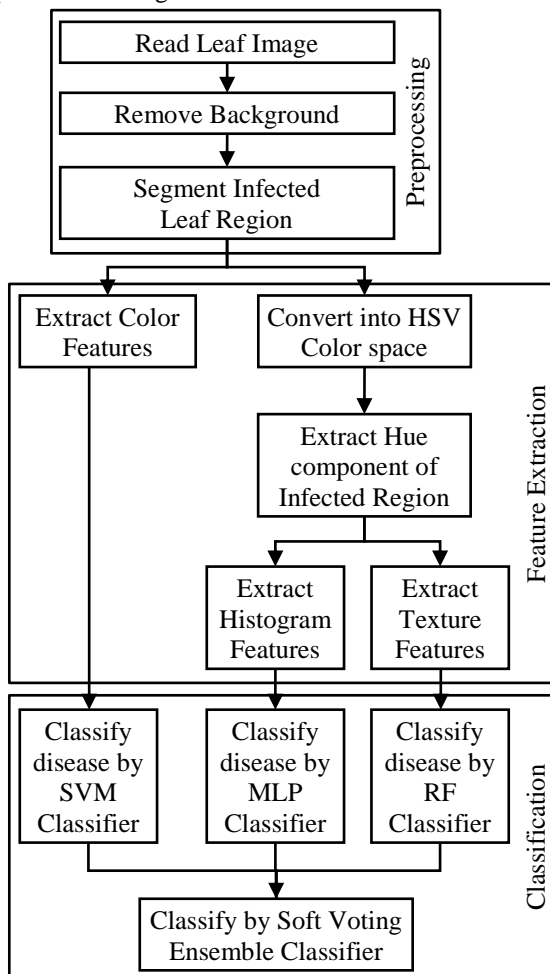


Fig.4. Overall workflow

F1 Score: F1 score tells the balance between precision and recall, represented as a percentage [18]. The value of F1 score will be closer to the value of precision or recall whichever is less. F1 score can be calculated using Eq.(12).

$$F1\text{-Score} = 2 * ((Precision * Recall) / (Precision + Recall)) \quad (12)$$

Sensitivity. Sensitivity is the percentage of actual positives that were predicted correctly and is given by Eq.(13) [19].

$$Sensitivity = TP / (TP + TN) \quad (13)$$

Specificity: Specificity is the ratio of actual negatives that were predicted correctly and is given by Eq.(14) [19].

$$Specificity = TN / (FP + TN) \quad (14)$$

4. PROPOSED METHODOLOGY

The major steps involved in this work include image pre-processing, feature extraction and classification of tomato diseases from healthy one. The overall workflow of the research study is shown in Fig.4.

4.1 IMAGE PRE-PROCESSING

Background of the leaf image was removed by the enhanced GrabCut algorithm [20]. Segmentation of infected leaf region from the foreground leaf image is done by threshold method applied on hue component of HSV color space for removing green (healthy) region from the image. The results obtained from background removal and segmentation is shown in Fig.5(a) and Fig.5(b) respectively.

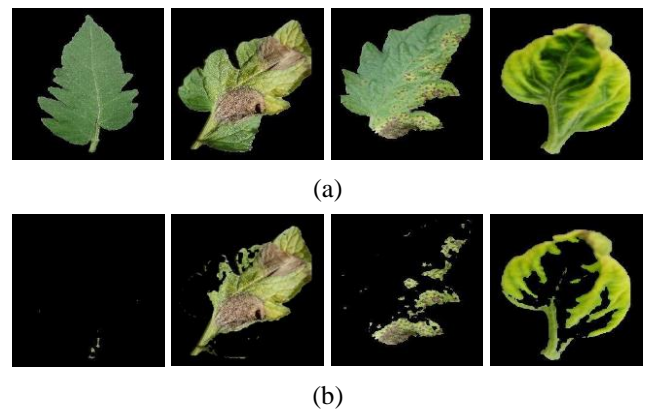


Fig.5. For Images shown in Fig.1 (a) After removing background (b) Infected regions obtained from segmentation

4.2 FEATURE EXTRACTION

Color Features were extracted from Red, Green and Blue components of RGB Color space and Hue components of HSV Color space. Histogram features and textures features were generated from Hue component of HSV Color space because color plays a vital role in distinguishing the diseases and pure color can be considered using Hue.

4.2.1 Color Features:

Average Red, Green and Blue intensities of infected pixels in RGB image and average Hue intensity of HSV image were calculated using Eq.(15)

$$I_c = \frac{\sum_{i=1}^M \sum_{j=1}^N infected_c(i, j)}{M \times N} \quad (15)$$

where

$infected_c$ is the color component Red, Green, Blue or Hue. I_c is the average intensity and $M \times N$ is the image size. The Table.1 tabulates color features obtained from infected/healthy leaves.

Table.1. Color features

Disease	Hue Intensity	Red Intensity	Green Intensity	Blue Intensity
0: Healthy	35.7202	103.2755	108.7719	81.4995
1: Late Blight	26.9843	135.4986	130.0422	82.2398
2: Septoria Leaf Spot	33.2087	76.2231	80.0866	48.9944
3: Yellow Leaf Curl Virus	25.8946	80.3977	80.2316	41.2456

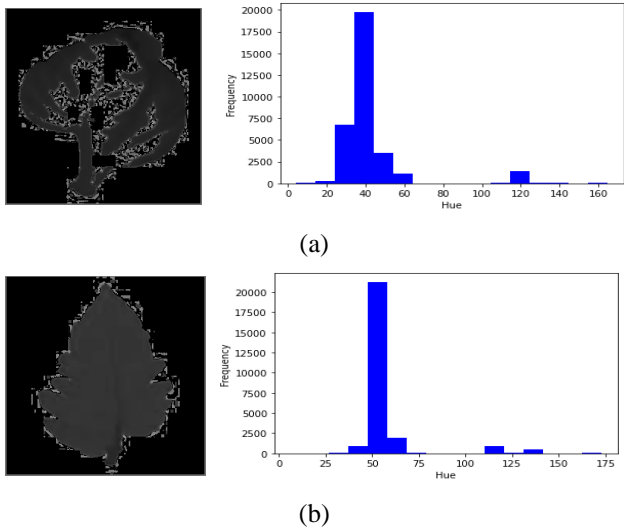


Fig.6. (a) Hue component of Infected Leaf and its Histogram (b) Hue component of Healthy leaf and its Histogram

4.2.2 Histogram Features:

Histogram of hue component of infected leaf region plotted and mean and standard deviation values were extracted from the histogram. The Fig.6 shows the infected leaf and histogram of the same. The Table.2 shows mean and standard deviation for a healthy/ infected leaf.

Table.2. Histogram features

Disease	Mean	SD
0: Healthy	9.833908	3.286402
1: Late Blight	38.5049	52.3845
2: Septoria Leaf Spot	12.39678	15.24848
3: Yellow Leaf Curl Virus	22.6472	30.9545

4.2.3 Texture features:

The hue component of infected leaf image was partitioned into 16x16 blocks and texture features were extracted for blocks with 10% of information. Otherwise, the block was discarded. To reduce the calculation complexity and time the number of gray levels reduced to 32 by dividing the hue component by 8. Gray Level Co-occurrence matrices for each useful block (GLCM) were generated in the directions 00, 450, 900, 1350 degrees for neighboring and alternate pixel pairs. This results in eight GLCM matrices and hence eight set of texture features. Average of these features was calculated to arrive at one texture feature set for one block. Texture feature set for one leaf image was calculated by averaging out texture features of all useful blocks. Texture features such as Contrast, Dissimilarity, Homogeneity, Energy, Correlation, Angular Second Moment, and Entropy were extracted for each block. Algorithm 1 elaborates the steps involved in extracting features Sample texture features obtained or one single infected or healthy leaf are given in Table.3.

The final set of features generated from the infected leaf regions include Red Intensity, Green Intensity, Blue Intensity, Hue Intensity, Mean and Standard Deviation, Contrast, Dissimilarity, Homogeneity, Energy, Correlation, Angular Second Moment, and Entropy.

Algorithm 1

Let $infected_{hue}$ be the hue component of infected region of leaf image. Texture features were extracted by the following step by step procedure.

Step 1: The hue (gray) levels in $infected_{hue}$ image was reduced from 256 to 32 by dividing $infected_{hue} / 8$. Let the resultant image be $infected_{hue32}$

Step 2: The $infected_{hue32}$ image was divided into equal sized blocks of 16 x 16. Let n be the number of blocks.

Step 3: Calculate the total area of one block in pixels,

$$area = 16 \times 16 = 256 \text{ pixels.}$$

Step 4: For $i = 1$ to n do the following:

Calculate number of pixels in $block_i$ with information, hue value > 0. Let it be $area_i$ in pixels.

Calculate % of information in $block_i$.

$$info_i = \frac{area_i}{area} * 100$$

If $info_i > 10\%$ then

Generate Gray Level Co-occurrence matrix (GLCM) for $block_i$ in the directions $0^0, 45^0, 90^0, 135^0$ for neighboring pixels and alternate pixels. This generates 8 GLCM matrices

Table.3. Texture features

Disease	Contrast	Dissimilarity	Homogeneity	Energy	Correlation	Angular Second Moment	Entropy
0: Healthy	1.0934	0.2897	0.9264	0.7417	0.8082	0.5557	0.3095
1: Late Blight	0.3488	0.1574	0.9394	0.6610	0.8489	0.4473	1.2722
2: Septoria Leaf Spot	0.9857	0.2949	0.9148	0.8078	0.7013	0.6541	0.1693
3: Yellow Leaf Curl Virus	0.6322	0.2218	0.9258	0.6730	0.8441	0.4658	1.1144

Extract texture features such as Contrast, Dissimilarity, Homogeneity, Energy, Correlation, Angular Second Moment (ASM) and Entropy from each of the eight GLCM.

The feature sets were averaged out to generate one feature set for $block_i$.

Else

Discard $block_i$.

Step 5: Calculate feature set for infected leaf $infected_{hue32}$ by finding average of features, Contrast, Dissimilarity, Homogeneity, Energy, Correlation, Angular Second Moment and Entropy, of all blocks $block_1, block_2, \dots, block_n$

The features Hue Intensity, Red Intensity, Green Intensity, Blue Intensity, Mean and Standard Deviation, Contrast, Dissimilarity, Entropy were normalized to the range of values 0 and 1 using min-max scaling as they were not in normal form.

4.3 CLASSIFICATION

The study was done with a dataset size of 1817 tomato leaf images. The Machine learning models were fed in with Texture features such as Contrast, Dissimilarity, Homogeneity, Energy, Correlation, Angular Second Moment and Entropy; Color features such as Red Intensity, Green Intensity, Blue Intensity and Hue Intensity; Histogram features like Mean and Standard Deviation. The classifiers Random Forest, Multilayer Perceptron and Linear Support Vector machines were trained with 80% of dataset and remaining 20% dataset was used to test the performance of the classifiers. Multilayer Perceptron with one hidden layer and 12 hidden neurons was implemented as it gave high accuracy.

An accuracy of 88.74%, 89.84%, and 92.86% was obtained from Random Forest, MLP Classifier and SVM classifier respectively. To improve the prediction results, Soft Voting Classifier was used to combine the results of the above classifiers and got an accuracy of 93.13% in classifying tomato. diseases while differentiating healthy leaves at the same time.

Table.4. Dataset Size

Class Label	Class	Training Set Size	Test Set Size	Data Set Size
0	Healthy	387	96	483
1	Late Blight	360	96	456
2	Septoria Leaf Spot	350	81	431
3	Yellow Leaf Curl Virus	356	91	447
Total		1453	364	1817

5. RESULTS AND DISCUSSION

The dataset size considered in this work was 1817. The dataset was split into 80% and 20% as training and test set respectively. The size of dataset, training and test set considered in this work is tabulated in Table.4.

Table.5. Confusion Matrix

True Label	Predicted Label			
	Class	0	1	2
0	92	0	4	0
1	2	83	8	3
2	0	8	70	3
3	3	1	9	78

(a) Random Forest

True Label	Predicted Label			
	Class	0	1	2
0	92	2	2	0
1	2	85	7	2
2	1	4	71	5
3	2	3	7	79

(b) Multilayer Perceptron

True Label	Predicted Label			
	Class	0	1	2
0	93	1	2	0
1	1	86	6	3
2	0	1	74	6
3	1	0	5	85

(c) Support Vector Machines

True Label	Predicted Label			
	Class	0	1	2
0	93	0	3	0
1	2	89	3	2
2	0	2	73	6
3	1	0	6	84

(d) Soft Voting Classifier

The Table.5(a)-Table.5(d) shows the Confusion Matrix for Random Forest, Multilayer Perceptron Classifier, Support Vector Machines and Soft Voting Classifiers respectively. Out of 96 healthy leaves tested, 92, 92, 93 and 93 were correctly classified by the classifiers RF, MLP, SVM and Soft Voting. 83, 85, 86 and 89 out of 96 Late Blight infected leaf were classified correctly by RF, MLP, SVM and Soft Voting Classifiers respectively. 81 Septoria Leaf Spot diseased images were tested and 70, 71, 74 and 73 were classified correctly by RF, MLP, SVM and Soft Voting Classifiers respectively. 91 images of Yellow Leaf Curl Virus infected leaf images were considered for testing and 78, 79, 85 and 84 images were correctly classified by classifiers RF, MLP, SVM and Soft Voting Classifiers respectively.

Performance metrics such as Precision, Recall, F1-Score, Sensitivity and Specificity of the classifiers Random Forest, Multilayer Perceptron, Support Vector Machines and Soft Voting Classifiers are tabulated in Table.6. An accuracy of 88.74%, 89.84%, 92.86% and 93.13% in predicting the tomato diseases were obtained by the classifiers Random Forest, Multilayer Perceptron, Support Vector Machines and Soft Voting Classifiers respectively.

Table.6. Performance of Classifiers

Classifier	Class	Precision	Recall	F1-Score	Sensitivity	Specificity	Support	Accuracy
RF	0	94.85%	95.83%	95.34%	95.83%	98.13%	96	88.74%
	1	90.22%	86.46%	88.30%	86.46%	96.64%	96	
	2	76.92%	86.42%	81.40%	86.42%	92.58%	81	
	3	92.86%	85.71%	89.14%	85.71%	97.80%	91	
ANN	0	94.85%	95.83%	95.34%	95.83%	98.13%	96	89.84%
	1	90.43%	88.54%	89.47%	88.54%	96.64%	96	
	2	81.61%	87.65%	84.52%	87.65%	94.35%	81	
	3	91.86%	86.81%	89.27%	86.81%	97.44%	91	
SVM	0	97.89%	96.88%	97.38%	96.88%	99.25%	96	92.86%
	1	97.73%	89.58%	93.48%	89.58%	99.25%	96	
	2	85.06%	91.36%	88.10%	91.36%	95.41%	81	
	3	90.43%	93.41%	91.89%	93.41%	96.70%	91	
VC	0	96.88%	96.88%	96.88%	96.88%	98.88%	96	93.13%
	1	97.80%	92.71%	95.19%	92.71%	99.25%	96	
	2	85.88%	90.12%	87.95%	90.12%	95.76%	81	
	3	91.30%	92.31%	91.80%	92.31%	97.07%	91	

Table 7. Classifier Performance

Disease	Number of Images Tested	Number of images classified correctly			
		RF	ANN	SVM	VC
Healthy	96	92	92	93	93
Late Blight	96	83	85	86	89
Septoria Leaf Spot	81	70	71	74	73
Yellow Leaf Curl Virus	91	78	79	85	84
Total	364	323	327	338	339

6. CONCLUSION AND FUTURE WORK

In this research work, tomato diseases were classified using the classifiers Random Forest, Multilayer Perceptron and Support Vector Machines. Support Vector Machine gave a highest accuracy of 92.86%. To improve the accuracy of prediction ensemble learning was applied by using Soft Voting Classifier and obtained an accuracy of 93.13% in classifying the tomato diseases. Table 7 tabulates the total number of images test in each disease and number of images correctly classified by the classifiers Random Forest, Multilayer Perceptron, Support Vector Machines and Soft Voting classifier. The study can be extended in future to other plants such as Corn, Soyabean, Grape and Potato.

REFERENCES

- [1] Plant Village, "Tomato", Available at https://plantvillage.psu.edu/topics/tomato/infos/diseases_and_pests_description_uses_propagation, Accessed at 2020.
- [2] S. Abed and A.A. Esmael. "A Novel Approach to Classify and Detect Bean Diseases based on Image Processing", *Proceedings of IEEE Symposium on Computer Applications and Industrial Electronics*, pp. 297-302, 2018.
- [3] Jayamala K. Patil and Raj Kumar, "Color Feature Extraction of Tomato Leaf Diseases", *International Journal of Engineering Trends and Technology*, Vol. 2, No. 2, pp. 72-74, 2011.
- [4] Juan Hua Zhu, Ang Wu and Peng Li, "Corn Leaf Diseases Techniques based on Image Recognition", *Proceedings of International Conference on Communications in Computer and Information Science*, pp. 334-341, 2012.
- [5] S. Arivazhagan, R. Newlin Shebiah, S. Anathi and S. Vishnu Varthini, "Detection of Unhealthy Region of Plant Leaves and Classification of Plant Diseases using Texture Features", *Agriculture Engineering*, Vol. 15, No. 1, pp. 211-217, 2013.
- [6] J.K. Patil and Raj Kumar, "Feature Extraction of Diseased Leaf Images", *Journal of Signal and Image Processing*, Vol. 3, No. 1, pp-60-63, 2012.
- [7] Basavaraj Tigadi and Bhavana Sharma, "Banana Plant Disease Detection and Grading using Image Processing", *International Journal of Engineering and Computing*, Vol. 6, No. 6, pp. 6512-6513, 2016.
- [8] K. Muthukannan, P. Latha, R. Pon Selvi and P. Nisha, "Classification of Diseased Plant Leaves using Neural Network Algorithms", *ARPN Journal of Engineering and Applied Sciences*, Vol. 10, No.4, pp. 1913-1919, 2015.
- [9] U. Mokhtar, M.A.S Ali and A.E. Hassanien, "Identifying Two of Tomatoes Leaf Viruses Using Support Vector Machine", *Advances in Intelligent Systems and Computing*, Vol 339, pp. 1-19, 2015.
- [10] H. Sabrol and S. Kumar, "Intensity Based Feature Extraction for Tomato Plant Disease Recognition by Classification using Decision Tree", *International Journal of Computer*

- Science and Information Security*, Vol. 14, No. 9, pp. 1-13, 2016.
- [11] Ginne M James, S.C. Punitha, "Tomato Disease Classification using Ensemble Learning Approach", *International Journal of Research in Engineering and Technology*, Vol. 5, No. 10, pp. 104-108, 2016.
- [12] Plant Village Data Set, Available at <https://github.com/spMohanty/PlantVillage-Dataset>, Accessed at 2020.
- [13] Hall-Beyer, "GLCM Texture: A Tutorial v. 3.0", Available at https://prism.ucalgary.ca/bitstream/handle/1880/51900/text%20tutorial%20v%203_0%20180206.pdf?sequence=11&isAllowed=y, Accessed at 2020.
- [14] A Beginner's Guide to Scikit-Learn's MLP Classifier, Available at <https://analyticsindiamag.com/a-beginners-guide-to-scikit-learns-mlpclassifier/>, Accessed at 2020.
- [15] Classification Algorithms - Random Forest, Available at https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_classification_algorithms_random_forest.htm, Accessed at 2020.
- [16] Ensemble Vote Classifier, Available at http://rasbt.github.io/mlxtend/user_guide/classifier/EnsembleVoteClassifier/, Accessed at 2020.
- [17] Classification Accuracy, Available at <https://developers.google.com/machine-learning/crash-course/classification/accuracy>, Available at 2020.
- [18] Accuracy Precision Recall, Available at <https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9>, Accessed at 2020.
- [19] Sensitivity vs Specificity, Available at <https://www.technologynetworks.com/analysis/articles/sensitivity-vs-specificity-318222>, Accessed at 2020.
- [20] S. Jeyalakshmi, R. Radha, "A Novel Approach to Segment Leaf Region from Plant Leaf Image using Automatic Enhanced Grabcut Algorithm", *International Journal of Advanced Computer Technology*, Vol. 8, No. 11, pp. 3485-3493, 2018.