SUPPORT VECTOR MACHINE BASED CLASSIFICATION FOR TOMATO LEAVES DISEASES

Rakesh Kumar Ray[1], Manjeet Bhardwaj[2], Rohit Kumar[3], Sujata Chakravarty[4]

[1,2,3,4] Dept. of CSE, Centurion University of Technology and Management, Odisha, India
Email: [1]rakesh.ray@cutm.ac.in, [2]manjeet184jeet@gmail.com, [3]7165551rohitkumar@gmail.com, [4]sujata.chakravarty@cutm.ac.in

Abstract

Tomato is a very common vegetable used by Indians. It is widely harvested by farmers all over India including Odisha. But often, the yield quality and quantity are affected due to various diseases. In this study, the focus is to design an automated leaf disease detection technique based on image processing and machine learning for the early diagnosis of infection in tomato leaf. The model is designed based on a dataset containing 18,200 images of tomato leaves those classified in 10 classes (9 diseases and a healthy class) collected from the Plant Village. For feature extraction, methods like shape-based features, color-based features, and texture-based features are used, then different machine learning algorithms like linear regression, Decision tree, Random forest, and Support Vector Machine (SVM) are used for tanning and testing the model. Among all SVM is fitting very well with the underlying dataset and producing a classification accuracy of 98.2%.

Key words: Feature Extraction; Linear Regression; Decision Tree; Random Forest; Support Vector Machine (SVM)

Introduction

Agriculture is the backbone of India where a majority of the population's livelihood completely depends on it. All around the world, Tomato is the most produced crop and India stands second in the world next to China in tomato production[1]. It consists of three antioxidants those are vitamin C, vitamin E, and beta-carotene. The tomato crop farming area in India is around 3,50,000 hectares approximately and 53,00,000 tons of Tomato is produced every year. Due to disease affected plants, 10% to 30% of total crops are loosed every year[2].

Early recognition of such diseases on the Tomato plant is very essential so that losses of the crop can be minimized and the quality & quantity of yield can be maximized. Fungus, bacteria, virus and nematodes are the causes of some common tomato leaf diseases. Therefore, farmer need to gain knowledge on disease affect the tomato plant at early stage or hire export to detect and identify plant diseases with naked eye, which might be costlier and time consuming in case of large fields[3]. It is the main motivation of this research is to find fast, less expensive, automatic, timely and accurate model to detect tomato plant disease. A fast and more accurate way for disease detection can be possible by using image processing and machine learning techniques.

A number of researchers have attempted to make plant disease detection fully automatic. Despite many models that have been suggested and tested, still those models have their conditions and limitations. The main focus of this paper was to propose a classification model that can help detect common diseases of Tomato plant leaf-like Septorial leaf spot, Bacterial leaf spot, leaf mold, target spot, etc. The rest of the paper is organized into 4 sections. Section 2 presents the related work proposed by different researchers in tomato leaf disease detection. Section 3 presents the stages of model building. Section 4 presents the result and analysis. finally, section 5 presents the conclusion of the paper.

Related Work

There are research works that have proposed different models to identify and detect leaf diseases of plants, which guides research to progress further. Mokhtar et al[4] has proposed a model to identify and detect infected leaves of tomato with upward curling and yellowing. The model used SVM for classification using different kernel functions on 200 infected tomato leaves images, which gives 90% accuracy. In their work K-Fold cross-validation is used to validate the model. Machine Learning has been a tool to classify different leaf diseases. The author[5] has introduced KNN, Decision Tree, SVM, and Neural Network to detect plant leaf diseases. The objective of the proposed work was to identify the healthy and affected portion of leaves. The paperwork has focused on Early Blight, Black Rot and Late Blight are plant disease. In another paper, the author[3] has used SVM, Random Forest, and Logistic Regression to detect plant disease detection, where SVM gives the better result as compare to the other two methods. Image processing techniques are a powerful tool to detect plant leaves disease[3], [5],[6]. The proposed [6] work was on image analysis and classification to detect plant leaves disease. The model has four-part, which are image pre-processing; K-Means to detect the area of disease for segmentation of leaves; feature extraction using GLCM (Gray Level Co-occurrence Matrix); finally SVM (Support Vector Machine) to classification. Wu et al.[7] has proposed an automated leaf recognition for plant classification model using the Probabilistic Nural Network technique, which has given 90% accuracy. The model is been trained by 1800 leaves that can classify 32 kinds of plants. In the work, 5 features have been orthogonalized out of 12 extracted leaf features like physiological length, physiological width, diameter, leaf area, and leaf perimeter.

SVM, ANN, and Random Forest are been used to analyze potato images to detect disease symptoms and classify healthy and unhealthy leaves in plants[8]. The paper has proposed an automated disease management tool to identify late blight and early blight affected part in leaves of potato. Genetic algorithm [9] is been used for image segmentation to detect plant leaf disease for building an automatic detection and classification of plant leaf disease. The author [10] has proposed an automatic model to classify and recognize thirteen types of plant disease using Deep CNN.

The paper [11]proposed a model to identify tomato leaves disease by using the Gabor wavelet transformation technique for feature extraction. the SVM classifier is been trained with the extracted features as input to train the model which then identifies the type of disease in tomato leaf. In the pre-processing stage, Image resizing, noise illumination, and background removal have been carried out. The paper has used Gabor transformation to extract appropriate features by identifying the textual patterns of the affected leaf. SVM with different kernel functions is used for Disease classification of tomato leaves and cross-validation technique has been used for performance evaluation. The experimented result shows 99.5% accuracy using the proposed model. The main disadvantage of Gabor transformation is computationally exhaustive for feature extraction.

In the paperwork [12], Artificial Nural Network is been used to predict late blight infection on tomatoes. The gradient-descent learning algorithm is been used in Back-propagation Nural Network to develop the model. The model is been trained with field data and remote sensing image data which gives the result(coefficient of determination) as 0.62 and 0.66 respectively. Deep learning has enormous potential to fulfill the research requirement. In the proposed work [13] deep learning is been used to detect various diseases on a tomato plant leaf. In the model, two deep learning architectures are been tested named AlexNet and SqueezeNet. PlantVillage dataset has been used for training the model with ten different classes including healthy images. Deep convolutional Nural Network is been used to design a model to detect 3 desiseas namely Leaf Miner, Target Spot and Phoma Rot[14]. The model is been trained using 4923 healthy and infected images of tomato leaf, which is collected throuth automated image capuring box. The model gives a accuracy of 91.67%.

Methodology

Four important stages are proposed in this model: Data acquisition, Image processing, and Image segmentation, Feature extraction, and the last Classification as shown in [Figure-1].

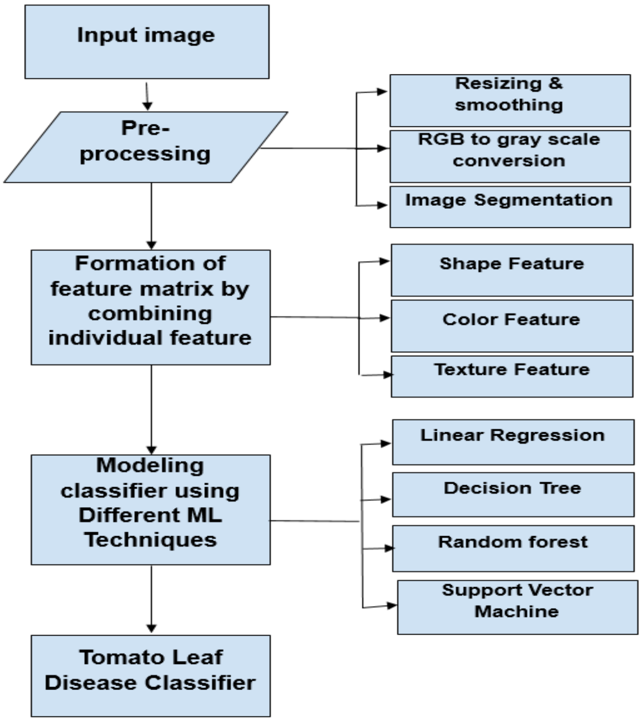

Fig. 1. Flow Diagram of the Model

3.1 Data Acquisition

The images of tomato leaf disease have been collected from the Plant Village repository. The collected dataset has almost 18200 images that belong to 10 various classes [Figure-2]. It includes images of all dominant types of leaf diseases that occur in the tomato plant. Each downloaded images are in the RGB colour space and was saved in the unzipped JPG format. The details of the dataset provided in [Table-1]

Table 1. Dataset

DATASET	INFORMATION
Collected from	**Plant village Repository**
Format	**JPG**
Total images size	**18200**
Training size	**14528**
Testing size	**3674**

Fig.2. (a) Healthy (b) Late blight (C) Leaf mold (d) Two-spotted Spider mite attack (e) Target spot (f) Tomato mosaic virus disease (g) Tomato yellow leaf curl virus disease (h) Spider mites Two-spotted spider mite(i) Early blight (j) Bacterial Spot

3.2 Data Pre-Processing

The collected dataset contains images having minimum noise and therefore there is no need to remove any noise. Images of datasets were resized into 512*512 resolution so that the training process of the model should be increased and beneficial for the computation. The importance of normalizing either input or output variables lean or needed to increase the training process speed. It is done by improving the numerical conditions of the normalization problems. Normalization also helps to get all pixels values of images in a particular range by using the standard deviation and mean value. Smoothing of images by using Gaussian Filter. A Gaussian filter is a linear filter. It's usually used to reduce noise. It is also used for edge detection. The Gaussian filter is helpful to blur reduce contrast and edges. Adaptive image thresholding using Otsu's thresholding method. Image thresholding is a simple, effective way of partitioning an image into a background and foreground. This image analysis method is a type of image segmentation that segregates objects by transforming grayscale images into binary images. Closing of holes using Morphological Transformation. Were morphological transformations are some simple operations based on the image shape. It is normally performed on binary images. It needs two inputs, one is our original image, and the second one is called the kernel or structuring element which decides the nature of the operation.

3.3 Feature Extraction

In this section Image processing, algorithms are used to detect and isolate various desired portions or shapes (features) of digitized images. In this method, various features of plant leaves were used like shape features, colour features, texture features. Shape features include features like Surface area, Surface Perimeter, Disfigurement. Shape features are essential as they provide different methods to define an image, using its most essential characters and diminish the amount of data stored. Shape features can be extracted using the contour method. Colour feature includes the variance of red, green, and blue. Using colour features, the model can classify images easily. The number of pixels that have the same colour value is computed for each image. Therefore, it is common to use colour features for image classification. Colour features can be extracted by using red, green, and blue channels of the image respectively. Texture features might be useful to depict certain repeated local patterns and arrangement regularity in specified regions of images, could be used to describe local characteristics of images, and could provide characteristics metrics such as contrast, correlation, and homogeneity. Where texture features can be extracted by using mahotas method. Feature extraction is particularly important in the area of optical character recognition or classification.

3.4 Designing the Classifier

To design the classifier different ML techniques are used like linear regression, Decision tree, Random Forest, and Support Vector Machine (SVM). Where among all SVM is a powerful method for the construction of the model. SVM [3], [4], [11] is used as a classifier. A Support Vector Machine (SVM) is a supervised machine learning algorithm and it can be used for both regression purposes and classification. SVMs are more commonly used in classification problems and as such, this is what we are focusing on in this research. SVM divides a dataset into two or more classes by finding a hyperplane as shown in [Figure-3]. Support vectors are the data points adjacent to the hyperplane, the points of a data set that, if removed, would alter the position of the dividing hyperplane. That's why they are treated as the critical elements of a dataset. Where hyperplane is decision boundaries that help classify the data points. Data points laying on either side of the hyperplane can be associated with different classes. In simple terms, it is the ability of your selected machine learning model to correctly classify different groups of data. To find the right hyperplane model select a hyperplane having the greatest possible margin between the hyperplane and any point belonging to the training set, giving a greater chance of new data being classified correctly.

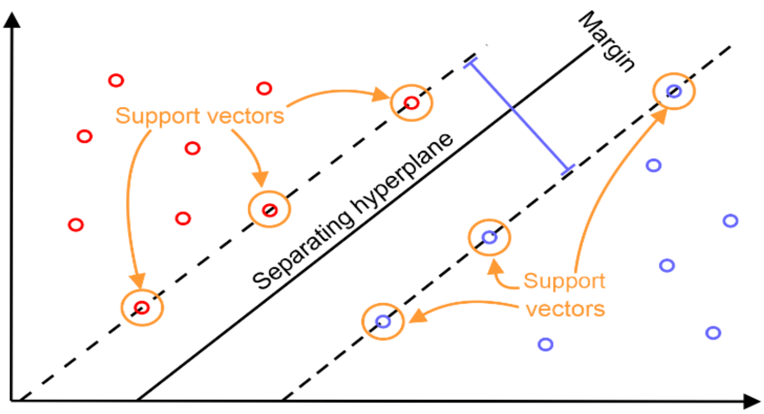

Fig.3. Support Vector Machine (SVM)

The Support Vector Machine is a preferable machine learning model having more accuracy in pattern recognition. SVM is exclusively found beneficial for the classification of high-dimensional datasets and has been found better in comparison to other machine learning algorithms.

Result and Discussion

To analyze the performance of the defined models based on the input extracted feature set, Overall accuracy has been considered and listed in [Table-2].

Table-2: Overall accuracy of different models

Sl No	Model	Accuracy
1	**Linear Regression**	**75.6%**
2	**Decision Tree**	**60.1%**
3	**Random Forest**	**82.4%**
4	**SVM**	**95.2%**

The best classification accuracy of 98.2% is achieved by the 4th model using SVM that is shown by red color in the bar graph [Figure-4]. The other models like Linear Regression, Decision Tree, and Random Forest that are used here to compare classification accuracy are also performing well.

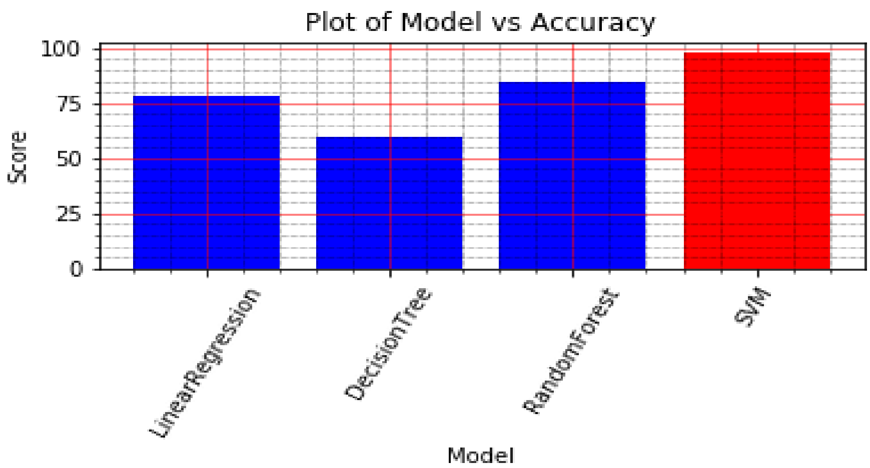

Fig.4. Accuracy Comparison plot.

Conclusion:

An Automatic model for the classification of Tomato leaf diseases is elucidated in this paper. A set of machine learning techniques are used to achieve the same, out of which SVM fits better with the set of extracted features in comparison to others with a classification accuracy of 98.2%. The model is robust and accurate as it trained and tasted with a large size dataset, which includes almost 18,200 images. The accuracy of the classifier can further be improved by using deep learning techniques. In the future, this method also can be used to detect the diseases of other plant leaves like potato.

References

[1] N. Madhumurthy and M. Sundaramoorthy, "A Report on The Study of Tomato Value Chain and its Financing," 2018. https://cab.rbi.org.in.

[2] P. Tm, A. Pranathi, K. Saiashritha, N. B. Chittaragi, and S. G. Koolagudi, "Tomato Leaf Disease Detection Using Convolutional Neural Networks," *2018 11th Int. Conf. Contemp. Comput. IC3 2018*, pp. 2–4, 2018, doi: 10.1109/IC3.2018.8530532.

[3] D. Das, M. Singh, S. S. Mohanty, and S. Chakravarty, "Leaf Disease Detection using Support Vector Machine," *Proc. 2020 IEEE Int. Conf. Commun. Signal Process. ICCSP 2020*, pp. 1036–1040, 2020, doi: 10.1109/ICCSP48568.2020.9182128.

[4] U. Mokhtar *et al.*, "SVM-Based detection of tomato leaves diseases," *Adv. Intell. Syst. Comput.*, vol. 323, pp. 641–652, 2015, doi: 10.1007/978-3-319-11310-4_55.

[5] N. G. Kurale and M. V. Vaidya, "Classification of Leaf Disease Using Texture Feature and Neural Network Classifier," *Proc. Int. Conf. Inven. Res. Comput. Appl. ICIRCA 2018*, pp. 1–6, 2018, doi: 10.1109/ICIRCA.2018.8597434.

[6] M. P. Vaishnnave, K. Suganya Devi, P. Srinivasan, and G. Arutperumjothi, "Detection and classification of groundnut leaf diseases using KNN classifier," *2019 IEEE Int. Conf. Syst. Comput. Autom. Networking, ICSCAN 2019*, pp. 1–5, 2019, doi: 10.1109/ICSCAN.2019.8878733.

[7] S. G. Wu, F. S. Bao, E. Y. Xu, Y. X. Wang, Y. F. Chang, and Q. L. Xiang, "A leaf recognition algorithm for plant classification using probabilistic neural network," *ISSPIT 2007 - 2007 IEEE Int. Symp. Signal Process. Inf. Technol.*, pp. 11–16, 2007, doi: 10.1109/ISSPIT.2007.4458016.

[8] P. Patil, N. Yaligar, and S. Meena, "Comparision of Performance of Classifiers - SVM, RF and ANN in Potato Blight Disease Detection Using Leaf Images," *2017 IEEE Int. Conf. Comput. Intell. Comput. Res. ICCIC 2017*, pp. 1–5, 2018, doi: 10.1109/ICCIC.2017.8524301.

[9] V. Singh and A. K. Misra, "Detection of plant leaf diseases using image segmentation and soft computing techniques," *Inf. Process. Agric.*, vol. 4, no. 1, pp. 41–49, 2017, doi: 10.1016/j.inpa.2016.10.005.

[10] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, "Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification," *Comput. Intell. Neurosci.*, vol. 2016, 2016, doi: 10.1155/2016/3289801.

[11] U. Mokhtar, M. A. S. Ali, A. E. Hassenian, and H. Hefny, "Tomato leaves diseases detection approach based on Support Vector Machines," *2015 11th Int. Comput. Eng. Conf. Today Inf. Soc. What's Next?, ICENCO 2015*, pp. 246–250, 2016, doi: 10.1109/ICENCO.2015.7416356.

[12] X. Wang, M. Zhang, J. Zhu, and S. Geng, "Spectral prediction of Phytophthora infestans infection on tomatoes using artificial neural network (ANN)," *Int. J. Remote Sens.*, vol. 29, no. 6, pp. 1693–1706, 2008, doi: 10.1080/01431160701281007.

[13] H. Durmus, E. O. Gunes, and M. Kirci, "Disease detection on the leaves of the tomato plants by using deep learning," *2017 6th Int. Conf. Agro-Geoinformatics, Agro-Geoinformatics 2017*, 2017, doi: 10.1109/Agro-Geoinformatics.2017.8047016.

[14] R. G. De Luna, E. P. Dadios, and A. A. Bandala, "Automated Image Capturing System for Deep Learning-based Tomato Plant Leaf Disease Detection and Recognition," *TENCON 2018 - 2018 IEEE Reg. 10 Conf.*, no. October, pp. 1414–1419, 2018.