TITLE:

Automated Plant Leaf Disease Detection with K-Means Clustering and Support Vector Machine

Anilkumar,Dr.ponmaniraj

Department of computer science and engineering ,Saveetha School of Engineering SIMATS ,CHENNAI,

Tamilnadu .

Email of student : [na9918588@gmail.com](mailto:na9918588@gmail.com)

Email of guide :

Abstract :-

Aim:- To develop an automated system for the detection and classification of plant leaf diseases using a combination of K-Means clustering and Support Vector Machine (SVM) algorithms.

Materials and methods:- The K-means clustering is used to classify objects based on a set of features into K number of classes [1]. Machine Learning, Computer vision and IOT (Internet of Things) approaches are helpful in identifying plant diseases [2] [3]. Feature Extraction method is applied to take out necessary features from segmented clusters. It helps in accurate classification and recognition of images. Texture, colour and shape features are extracted for categorization of diseases [4]

Result and decision:- The study employed feature selection, including the Gray-level co-occurrence matrix (GLCM) features (98.71% accuracy) and Principal Component Analysis (PCA) for dimension reduction (98.97% accuracy). Additionally, two deep learning methods, CNN (86.82% accuracy) and Google Net (94.05% accuracy), were compared, with the proposed method demonstrating significantly shorter processing times than both deep learning models [5].

Conclusion:-according to experimental data ,the proposed technology can correctly detect and diagnose plant sickness with a 97.2 percent accuracy

Key Words:-machine learning GLCM algorithm, k-means clustering, LPB,SVM

INTRODUCTION

According to the Food and Agricultural Organization of the United Nations [6], global crop losses due to pests range from 20% to 40%. Plant diseases cost the global economy $220 billion annually, with $70 billion attributed to invasive pests [6]. Traditional methods struggle to detect leaf diseases efficiently.

Environmental conditions, as noted in [7], significantly influence plant diseases. The rate of disease spread depends on crop health and susceptibility to infections [8]. Infections, whether bacterial, parasitic, viral, or fungal, result in various symptoms like leaf spots, blights, wilts, scabs, cankers, and root rot [9]. These symptoms range from discoloration to plant decay.

Plant diseases are a significant concern as they can harm crop quality and reduce yield. Detecting these diseases has been greatly improved through advanced machine learning and deep learning methods, leveraging the power of computer-aided engineering (CAE) networks to enhance accuracy [10].

Plant disease detection employs a range of techniques, including Convolutional Neural Networks (CNNs), hyperspectral imaging, artificial neural networks (ANNs), visible and infrared spectroscopy, and various imaging methods. These methods are essential for identifying diseases caused by emerging pathogenic species like huanglongbing citrus, chestnut blight fungus, as well as bacterial, fungal, viral infections, and plant damage [11].

SVM is employed for automated object recognition and characterization, particularly in efficiently identifying plant disease backgrounds with minimal training data. This section's related work outlines diverse methods utilized in plant disease detection.

Literature Survey:-

In [10], an automatic detection method was employed to pinpoint damaged portions in leaves and crops. They initially used SVM to acquire and process the original image. In the segmentation process, the background and black pixels were separated into defective area hue and saturation components. This approach successfully identified diseased and healthy features within the designated area, revealing disease presence in 5.56% of the region. However, it's important to note that Support Vector Machine (SVM) exhibited lower accuracy in disease detection and became somewhat outdated as technology advanced.

Zia Ur Rehman et al. introduced a new way to classify citrus diseases in [12]. They used feature fusion and transfer learning and enhanced image quality with hybrid contrast stretching as a preprocessing step. They retrained Densely Connected Convolutional Networks (DenseNet201) and Convolutional Neural Networks for Mobile Vision Applications (MobileNet-v2) to generate feature vectors. Their results showed improved classification accuracy compared to previous work in this field.

Hyperspectral imaging detects plant diseases using optical sensors and pathogen-induced measurements. It operates at various scales, from canopy to tissue levels, based on RGB cameras following a mosaic principle. However, it's expensive, intricate, and has sensitive detectors, resulting in lower accuracy. [13]

The CNN method, as described in [14], automates plant disease detection by analyzing leaf images with bacterial spot disease. It offers improved disease prediction and faster image processing but faces challenges in handling image rotation and losing compositional information. Additionally, it demands a sizable dataset for neural network training.

Algorithm:-

***Image Pre-Processing***  (1)

***Image Segmentation***

(2)

where X represents group data, Q shows the object in space, and is the centroid of the cluster

***Feature Extraction***

(3)

where (a,b)  are the pixel values. (u,v) are the spatial position of the image. (Δu,Δv)(are the spatial relation for matrix calculation. I(u,v) represents the pixel value of u and v.

***Energy***

Energy is also known as uniformity; It provides the sum of square elements of homogeneous and non-homogeneous regions in the GLCM matrix. It provides high frequency and high image pixels as well; the energy formulation is given in [Eq. (4)](https://www.techscience.com/cmc/v76n1/53063/html#eqn-4).

Energy (4)

***Entropy***

Entropy is used to calculate the image randomness. Sometimes the homogeneous image will reflect the lower entropy and the equation derived in [Eq. (5)](https://www.techscience.com/cmc/v76n1/53063/html#eqn-5).

Entropy (5)

***Contrast***

Contrast measures the link between the neighbor image and the pixel, and it is formulated in [Eq. (6)](https://www.techscience.com/cmc/v76n1/53063/html#eqn-6).

Contrast (6)

***Correlation***

Correlation is used to measure the linear grey tone dependencies of the image. It defines the relationship between a pixel and its neighbor and the equation given in [Eq. (7)](https://www.techscience.com/cmc/v76n1/53063/html#eqn-7).

Correlation= (7)

***Homogeneity***

Homogeneity is a term that describes pixel similarity. The value of 1 in the GLCM matrix of a homogeneous image, if just little changes are required to the image texture, the formulation part is relatively slow. The formulation part is given in [Eq. (8)](https://www.techscience.com/cmc/v76n1/53063/html#eqn-8).

Homogeneity= (8)

where,

(9)

(10)

(11)

is the variation in pixel brightness.

This binary pattern system contains 0–1 and 1–0 uniform transactions. The method reduced the length of the texture features 3 × 3 window from 4200 to 2340. The concatenate histograms are applied in all the cells, and the disease is classified with the help of texture features [15].

**Materials and Methods:-**

This paper presents a plant disease detection framework employing SVM. Figure 1 illustrates the key steps in SVM techniques, including image acquisition, pre-processing with DCT and colour space conversion, image segmentation using K-means clustering, feature extraction with LBP and GLCM, classification with radial basis and polynomial kernels, and final evaluation using SVM classifiers.

Image Acquisition

Image pre-processing using DCT domain and colour space conversion

Image Segmentation using K-means Clustering

Feature Extraction using GLCM and LBP feature

Identify disease type and total affection area

Image classification using radial basic kernel & polynomial kernel methods

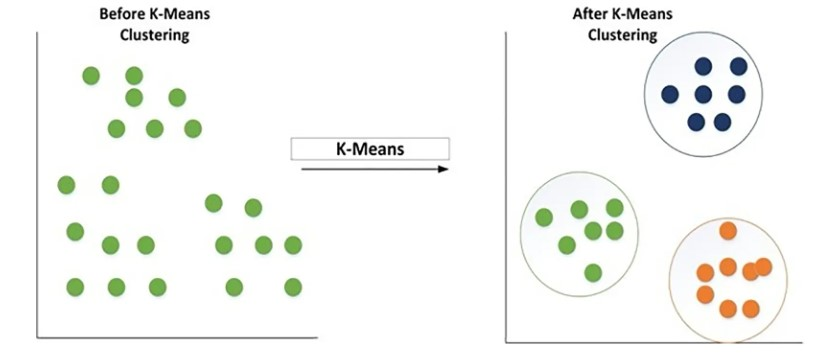
**Figure 1:** Flowchart of plant disease detection using SVM techniques

Image pre-processing

The pre-processing step enhances image data by eliminating background and noise, which is crucial for improving disease analysis. Initially, RGB images are resized and converted to the HSV format. A median filter is then applied to reduce noise and enhance image smoothness. Additionally, image enhancement techniques are employed to boost image contrast. RGB images are subsequently transformed into grayscale using the DCT color conversion method outlined in Eq. (1) for plant disease detection.[16].

Image Segmentation

K-Means clustering is favored over hierarchical clustering. This method groups images based on specific attributes, encompassing both diseased and healthy pictures. In the process, the background is removed from HSV images, reducing attributes like brightness and darkness, ultimately leading to segmented images for analysis. For instance, if four clusters are identified, the image is sorted into these clusters, yielding effective image segmentation [17]. Eq. (2) outlines the K-Means clustering algorithm, while Fig. 2 illustrates how it operates.



**Figure 2:** K-means clustering algorithm

Feature Extraction:

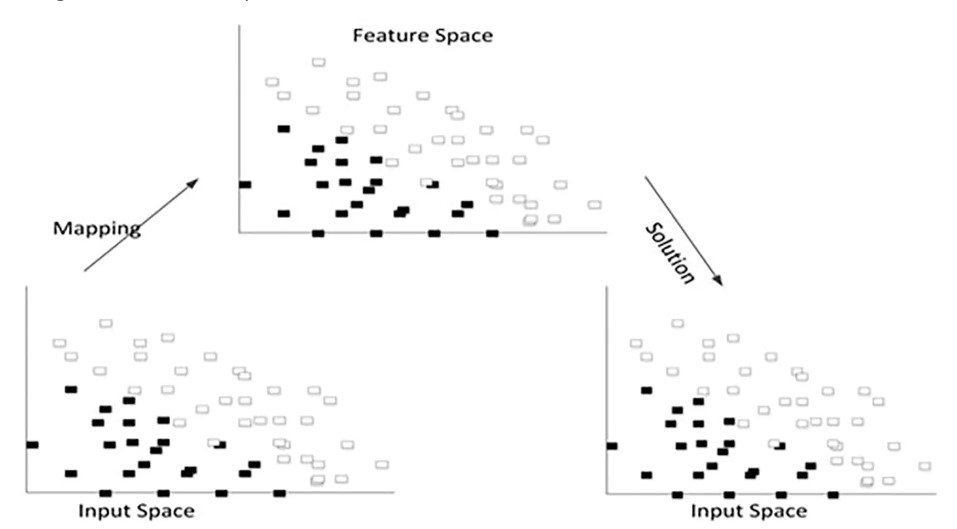
The GLCM feature values such as standard deviation, energy, entropy, correlation, smoothness, contrast, variant, Integrated Disease Management (IDM), and homogeneity [[28](https://www.techscience.com/cmc/v76n1/53063/html#ref-28)] are also used to calculate texture features. Here [Eq. (3)](https://www.techscience.com/cmc/v76n1/53063/html#eqn-3) is denoted by the algorithm used in GLCM for plant disease detection

Classification:

In the final stage, images are classified using the radial basis kernel method within the Support Vector Machine (SVM). SVM is preferred for its ability to excel when there's a substantial gap between classes. It's memory-efficient and particularly effective in high-dimensional spaces, making it valuable when dealing with more features than data points. Unlike other classifiers, SVMs determine boundaries that optimize separation between classes, often referred to as the maximal margin hyperplane.

The SVM constructs a hyperplane in an infinite-dimensional space for both image classification and regression. The radial basis kernel algorithm is utilized for data analysis and pattern recognition in disease classification. It categorizes binary sets and transforms them into non-probabilistic binary linear classifiers.

In the final major step, a training and validation process using SVM is executed. The dataset is divided into training and testing sections. The training feature is employed to train the SVM model, while the testing feature evaluates its accuracy. Finally, image accuracy is assessed to determine the percentage of the diseased region, which is estimated through the ratio of leaf data to diseased data [5]. Fig. 3 illustrates the SVM algorithm, and it is complemented by Fig. 4, which presents the flow diagram of SVM technique



**Figure 3:** SVM algorithm

Disease prediction

Collecting results and classify image using SVM

Extract image applying GLCM and LBP feature

Image pre-processing applying DCT domain and colour space conversion

3500 for traiiining

1806 for testing

5306 images captured

**Figure 4:** Flowchart for SVM techniques

**Results and Discussion**

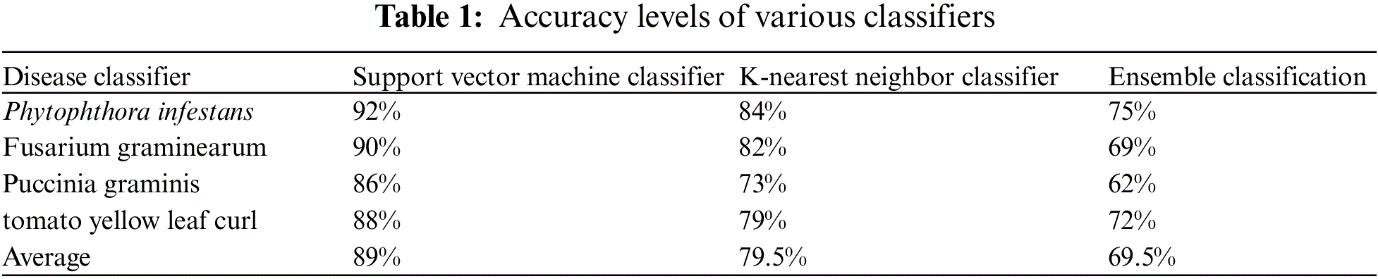
**Data analysis**

A dataset of 5,306 images has been gathered for analysis. This process involves categorizing data classifiers and parameters into distinct classes. Using the "Load image" function, the first infected segment is loaded. In this initial step, the image is enhanced, noise is removed, and segmentation is carried out using K-means clustering. The infected region is then extracted using GLCM and LBP methods, and the extent of the infection is calculated.

Following classification, the images are subjected to performance comparison with three different classifiers (SVM, K-Nearest Neighbor (KNN), and Ensemble classifier). Essential parameters such as accuracy, specificity, and sensitivity are estimated [18–21]. The employed method outperforms previous approaches in detecting leaf diseases and shows promise for disease detection and treatment. This method effectively addresses many of the limitations of older techniques.

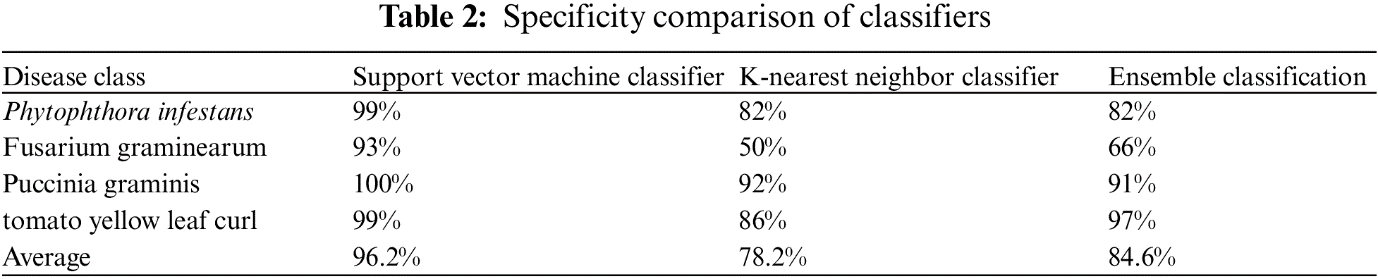
**Accuracy Calculation**

Accuracy is computed by multiplying the values of sensitivity, prevalence, and specificity. Table 1 provides a comparison of accuracy for the three different classifiers.



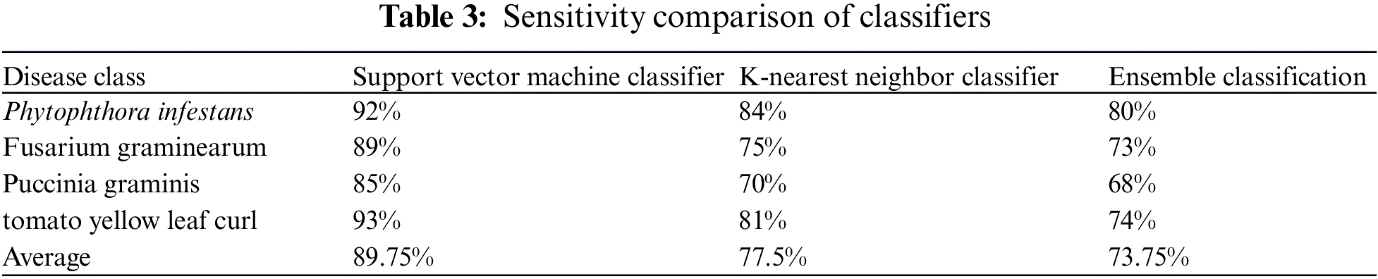
**Specificity Measurement**

Specificity, which quantifies the precise classification of non-diseased plants among all non-diseased plants, is a crucial parameter. Table 2 presents a comparison of specificity values for the three classifiers.



**Sensitivity Assessment**

Sensitivity evaluates the accurate classification of diseased plants among the total number of diseased plants. Table 3 displays a comparison of sensitivity for the three classifiers.



Conclusion

Recognizing the critical role of nature and plant life for human survival, it's imperative to develop effective solutions for safeguarding plants from diseases. The decline in crop production negatively impacts a country's economy. An automated plant leaf disease detection method is essential. This study aims to create an image processing system capable of identifying and classifying four different plant diseases: Phytophthora infestans, Fusarium graminearum, Puccinia graminis, and tomato yellow leaf curl.

Disease-affected areas of plant leaves are identified using GLCM and LBP features. Experimental results indicate that the proposed technology achieves a remarkable accuracy of 97.2% in detecting and diagnosing plant illnesses. Future plans involve expanding the database to encompass more plant diseases and using extensive data for classifier training. As the training data increases, system accuracy will continue to improve.

Acknowledgement:

The authors thanks to the Saveetha school of engineering and SIMTS university for their support in providing the infrastructure needed to complete this work successfully .

References:-

[1]Kumar, A. and Tiwari, A., 2019. A comparative study of otsu thresholding and k-means algorithm of image segmentation. Int. J. Eng.

Technol. Res, 9, pp.2454-4698.

[2] R. Rout and P. Parida, "A Review on Leaf Disease Detection Using Computer Vision Approach," 2020.

[3] S. S. Shinde and M. Kulkarni, "Review Paper on Prediction of Crop Disease Using IoT and Machine Learning," IEEE, 2017.

[4] D. A.K., S. M. and M. M.R., "Image Processing Based Leaf Rot Disease,Detection of Betel vine(Piper BetleL.)," International Conference

on Computational Modeling and Security, vol. 85, pp.748-754.

[5]Nema, S. and Dixit, A., 2018, December. Wheat leaf detection and prevention using support vector machine. In 2018 International

Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET) (pp. 1-5). IEEE.

[6] FAO, "Food and Agricultural Organizations of the United Nations,"[Online]. Available:

http://www.fao.org/news/story/en/item/1187738/icode/. [Accessed 212 2020].

[7] K. A. Garett and A. H. Sparks, "Plant Pathogens as Indicators ofClimate Change," Science Direct, 2016.

[8] G. B. Lucas, C. L. Campbell and L. T. Lucas, Introduction to Plant Diseases: Identification and Management, Google Scholar, 2012.

[9] G. N. Agrios, Plant Pathology, 5th Edition ed., Google Scholar, 2005.

[10]R. Kaur and S. S. Kang, “An enhancement in classifier support vector machine to improve plant disease detection,” in IEEE 3rd Int. Conf. on MOOCs, Innovation and Technology in Education (MITE), Amritsar, India, pp. 135–140, 2015.

[11] S. Sankaran, A. Mishra, R. Ehsani and C. Davis, “A review of advanced techniques for detecting plant diseases,” Computers and Electronics in Agriculture, vol. 72, no. 1, pp. 1–13, 2010.

[12] M. Zia Ur Rehman, F. Ahmed, M. Attique Khan, U. Tariq, S. Shaukat Jamal et al., “Classification of citrus plant diseases using deep transfer learning,” Computers, Materials & Continua, vol. 70, no. 1, pp. 1401–1417, 2021.

[13]S. Thomas, K. T. Matheus, B. David, B. Anna, A. Elias et al., “Benefits of hyperspectral imaging for plant disease detection and plant protection: A technical perspective,” Journal of Plant Diseases Protection, vol. 125, no. 1, pp. 5–20, 2018.

[14]P. Bedi and P. Gole, “Plant disease detection using hybrid model based on convolutional autoencoder and convolutional neural network,” Artificial Intelligence in Agriculture, vol. 5, no. 1, pp. 90–101, 2021.

[15] D. Kumar, “Feature extraction and selection of kidney ultrasound images using GLCM and PCA,” Procedia Computer Science, vol. 167, no. 10, pp. 1722–1731, 2020.

[16] K. Elangovan and S. Nalini, “Plant disease classification using image segmentation and SVM techniques,” International Journal of Computational Intelligence Research, vol. 13, no. 7, pp. 1821–1828, 2017.

[17] D. Das, M. Singh, S. S. Mohanty and S. Chakravarty, “Leaf disease detection using support vectormachine,” in Int. Conf. on Communication and Signal Processing (ICCSP), Chennai, India, pp. 1036–1040, 2020.

no. 10, pp. 1722–1731, 2020.

[18]S. Ramesh, R. Hebbar, M. Niveditha, R. Pooja, N. Shashank et al., “Plant disease detection using machine learning,” in Int. Conf. on Design Innovations for 3Cs Compute Communicate Control (ICDI3C), Bangalore, India, pp. 41–45, 2018.

[19] L. V. Madden, G. Hughes and F. Van Den Bosch, “Spatial aspects of epidemics-III: patterns of plant diseas,” in Title of the Study of Plant Disease Epidemics, Third printing, vol. 1. St. Paul, USA: The American Phytopathiological Society, pp. 235–278, 2007.

[20] B. J. Staskawicz, F. M. Ausubel, B. J. Baker, J. G. Ellis and J. D. Jones, “Molecular genetics of plant disease resistance,” Science, vol. 268, no. 5211, pp. 661–667, 1995.

[21] M. A. Hussein and A. H. Abbas, “Plant leaf disease detection using support vector machine,” Al-Mustansiriyah Journal of Science, vol. 30, no. 1, pp. 105–110, 2019.