



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 8, August 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.524

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

Plant Leaf Disease Detection based on Image Processing and Machine Learning

Nisreen I. R. Yassin

National Research Centre, Dokki, Cairo, Egypt

ABSTRACT: In the field of agriculture, the precise identification of plant diseases holds significant importance. Various research endeavours have been made to automate the detection of diseases through the utilization of cutting-edge tools and methodologies in image processing and machine learning. Farmers often overlook a diverse range of plant diseases impacting agriculture. With no reliable treatment or diagnostic method available, a growing number of crops are lost due to these diseases on a daily basis. Detection of a disease within a plant cannot be effectively achieved through mere visual observation. Consequently, the misuse of pesticides which affects the human health through the development of chronic illnesses. Thus, the preservation of food security hinges on precise and automated disease identification. In this paper, a scheme for plant leaf disease classification based on image processing and machine learning is proposed. The proposed scheme comprises of five distinct stages, namely: image acquisition, pre-processing, segmentation, feature extraction, and ultimately classification. A combination of features are extracted using Grey Level Co-occurrence Matrix, Local Binary Patterns, and Histogram of Oriented Gradient. K Nearest Neighbours and Decision Tree are used for the classification process. The proposed scheme is applied on corn leaves dataset which consists of four classes Common Rust, Gray Leaf Spot, Blight, and Healthy. The experimental outcomes show the efficiency of the proposed algorithm to detect and classify corn leaf disease.

KEYWORDS: GLCM; HOG; LBP; Plant disease detection; Classification.

I. INTRODUCTION

Agriculture is the most extensive sector globally, accounts 18% of the Gross Domestic Product (GDP) [1]. Agriculture serves as the predominant revenue stream for a significant proportion of the global population. Consequently, the enhancement of crop yield holds paramount significance on a global scale. Current approaches for disease identification encompass visual examination, DNA-centric assays (e.g., PCR), and hyperspectral scanning. Despite their potential effectiveness, these methodologies are constrained by certain drawbacks. The process of visual inspection may consume considerable time and is subject to subjective interpretation, DNA-based testing could incur high costs and necessitate specialized apparatus, and the utility of hyperspectral imaging may be impeded by lighting circumstances. Incorrect identification of a plant disease may lead to the improper use of pesticides, posing a threat to human health through the development of chronic illnesses [2].

One method that is increasingly being adopted involves the utilization of machine learning algorithms to autonomously analyse images of crops and identify symptoms of diseases. Through the process of training these algorithms on extensive datasets of images, scientists are able to create highly precise tools for disease detection in agricultural settings. This particular strategy has the potential to offer a swifter, more impartial, and more expandable alternative to conventional disease identification techniques. However, it requires top-notch image data and meticulous calibration to guarantee precise outcomes [3]. Plant diseases can be classified using machine learning techniques based on a range of distinctive characteristics. Pre-processing steps, including image enhancement, color adjustment, and segmentation, are essential prerequisites for the efficient extraction of these features. Following feature extraction, a variety of classifiers, such as K-nearest neighbour (KNN) and support vector machine (SVM), decision tree, random forest (RF), naive Bayes (NB), and logistic regression (LR), can be employed [4].

Corn cultivation is prevalent on a global scale. The significance of corn farming lies in its capacity for exportation and its role in providing sustenance for a considerable number of agricultural workers. Various sectors such as food processing, poultry, and livestock husbandry rely heavily on corn. The yield of corn surpasses the combined productivity of wheat and rice, establishing it as a fundamental dietary component in numerous geographic areas. The corn's vulnerability to different diseases, along with its significantly reduced agricultural productivity, plays a major

role in the decreased corn yield [2]. Therefore, there is a necessity for a precise automated identification of diseases affecting corn plants. In this research, a scheme for identifying corn disease is proposed. The proposed scheme is based on collecting a feature vector using different feature descriptors algorithms. The Grey Level Co-occurrence Matrix (GLCM), Local Binary Patterns (LBP), and Histogram of Oriented Gradient (HOG) are used to collect discriminative features from images. Different classifiers are used to build a classification model such as K-nearest neighbor (KNN) and Decision Tree (DT). The rest of the paper is organized as follows. Section 2 presents a literature review, section 3 provides the applied methodology, section 4 depicts the experimental outcomes, and finally the paper is concluded in section 5.

II. RELATED WORK

Lately, plant disease detection using image processing has become the most emerging and developed techniques for enhancing agriculture productivity. Patel et al. [5] proposed a semi-automatic technique to detect two soybean diseases. The technique depends on k-means clustering and KNN classifying algorithm. Abed and Esmaeel [6] presented a method to detect two kinds of bean leaf disease. The method includes image pre-processing, segmentation, feature extraction, and classification using public datasets. Pantazi et al. [7] proposed an automated model to detect four types of plant diseases. The model is based on Local Binary Patterns (LBPs) for feature extraction and applied on vine leaves.

Hossain et al. [8] introduced a classification model utilizing a support vector machine to identify infections in plants and during harvest. Alehegn [9] categorized three different types of corn leaf diseases in plants through the utilization of the K-nearest neighbor (KNN) and Artificial Neural Network (ANN) methodologies for identification, encompassing common rust, leaf blight, and leaf spot. Kanaka Durga and Anuradha [10] employed Support Vector Machine (SVM) and Artificial Neural Network (ANN) algorithms for the classification of leaf diseases in tomato and corn plants. The dataset comprised 200 images depicting both healthy and diseased leaves. A series of procedures were applied to identify the diseases, beginning with the conversion of the RGB image to grayscale, followed by division to calculate the intensity gradient at each pixel. The extraction of features was carried out using the Histogram of Oriented Gradients (HOG) technique. Zhang et al. [11] proposed a novel Support Vector Machine (SVM) incorporating a genetic algorithm approach, known as GA-SVM. The researchers obtained six distinct diseases that impact corn leaves and subsequently categorized them.

III. METHODOLOGY

The proposed scheme of plant disease detection and classification consists of five stages: image acquisition, pre-processing, segmentation, feature extraction, and finally the classification stage. Figure 1 depicts the proposed scheme.

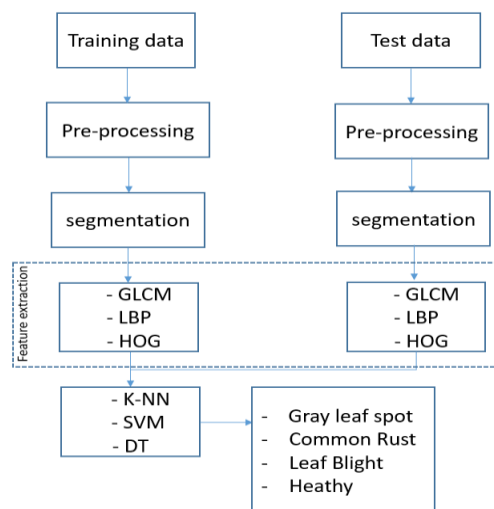


Figure 1. The proposed scheme for plant disease detection.

3-1 Image acquisition

The plant leaf image is taken by a camera as RGB color image. PlantVillage is a public dataset used for plant disease detection [12]. It consists of 87000 RGB images of healthy and unhealthy different plant leaves. The corn leaves were selected from the dataset to apply the proposed algorithm of plant disease detection. The corn leaves dataset consists of four classes which are 1192 Common Rust images, 513 Gray Leaf Spot images, 985 Blight images, and 1162 Healthy images. These images are color images of size 256x256x3. They are randomly partitioned into 80% for training and 20% for test. Samples of corn leaves are shown in Figure 2. Figure 3 shows the distribution of the training and test data classes.

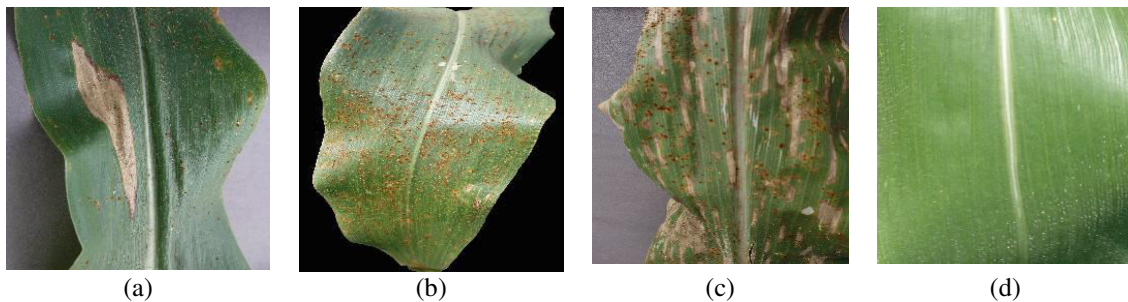


Figure 2. Corn leaves images: (a) Blight, (b) Common Rust, (c) Gray Leaf Spot, (d) Healthy.

Table 1 Dataset distribution

	Train data	Test data
Gray Leaf Spot	410	103
Common Rust	954	238
Northern Leaf Blight	788	197
Healthy	930	232

3-2 Image pre-processing

Pre-processing stage includes noise removal, and contrast enhancement. The images of corn dataset are clear and have no complex background. Contrast enhancement is applied by identifying the upper and lower limits that can be used to stretch the contrast of the image.

3-3 Image segmentation

In this stage, region of interest (ROI) is extracted from the pre-processed image in order to extract different discriminative features that can differentiate between healthy and unhealthy images. The proposed algorithm performs image segmentation based on colour features using K-means clustering. The K-means clustering technique is applied to allocate a set of data points into distinct K-clusters that do not overlap. Its purpose is to segment the data points by reducing the within-cluster distances through the minimization of squared error, which is characterized as the total sum of squared distances of points from the centroid of the clusters. Figure 3. Shows the k-means clustering output.

3-4 Feature extraction

Feature extraction is an essential stage in image processing and computer vision undertakings. Images hold a huge volume of data therefore, extracting significant information from images is crucial for a multitude of applications. In this stage, different discriminative features that can differentiate between healthy and unhealthy images are extracted from the pre-processed images. The proposed algorithm performs the feature extraction stage based on three feature extraction algorithms: the Grey Level Co-occurrence Matrix (GLCM), the Local Binary Patterns (LBP), and Histogram of Oriented Gradient (HOG). GLCM is utilized to depict texture by examining the spatial associations of pixel intensities. The fundamental principle of GLCM involves the examination of the frequency at which gray levels co-occur in an image, with a particular focus on adjacent pixels. LBP characterizes texture by evaluating localized pixel configurations, while HOG delineates object morphology via histograms of gradient orientations.

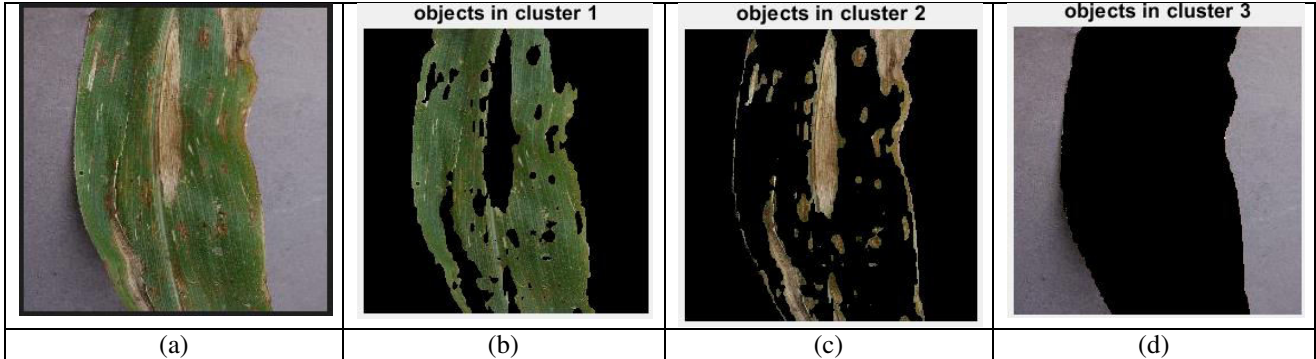


Figure 3. The K-means clustering output: (a) corn leaf, (b) cluster 1, (c) cluster 2, (d) cluster 3.

3-4-1 Grey Level Co-occurrence Matrix (GLCM)

The concept of GLCM is to account how often the neighboring pixels occur together inside the image. Figure 4 illustrates the GLCM generation method. Calculating GLCM depends on three factors: the number of intensity levels, the distance between neighboring pixels, and the direction of pixel pairing. The extracted features using GLCM are contrast, correlation, energy and homogeneity and their mathematical formulations are summarized in Table 2.

Table 2 Mathematical formula of GLCM extracted features

Feature	Mathematical formulation
Contrast	$\sum_i \sum_j (i - j)^2 g_{ij}$
Correlation	$\frac{\sum_i \sum_j (ij) g_{ij} - \mu_x \mu_y}{\sigma_x \sigma_y}$
Energy	$\sum_i \sum_j g_{ij}^2$
Homogeneity	$\sum_i \sum_j \frac{1}{1 + (i - j)^2} g_{ij}$



Figure 4 Example of GLCM calculation in the horizontal direction and using offset 1 pixel.

3-4-2 Local Binary Patterns (LBP)

LBP [13] is a popular feature extraction method that used in many computer vision tasks. It illustrates the local construction of an image that remains constant despite variations in lighting conditions. LBP operates through the comparison of the luminance level of a central pixel within a limited vicinity with the luminance levels of the neighboring pixels. A binary value is allocated to each pixel in the vicinity, depending on whether its luminance level exceed or falls below the luminance level of the central pixel. These binary digits are subsequently merged into a binary numeral, effectively depicting the textural characteristics of that vicinity. Figure 5 shows an example of LBP method.

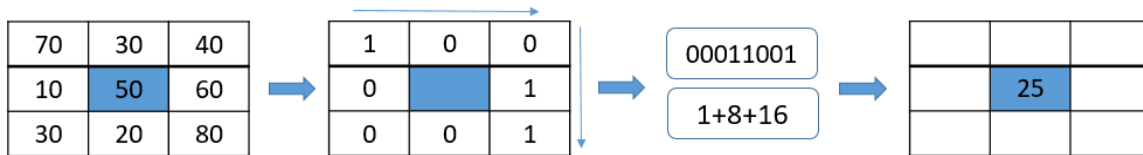


Figure 5 Example of LBP method.

Steps for computing LBP are as follows:

1. For a grayscale image, consider a window of a centre pixel and eight neighboring pixels in circular circle of radius 1.
2. The central pixel is compared to its eight neighbors. If the difference between the neighboring pixel and the central pixel is positive then the neighboring pixel is changed to 1 otherwise, changed to 0.
3. The resulted 8-binary number is converted to a decimal value which represents the LBP value of the central pixel.
4. Repeat the pervious steps for each pixel in the image to obtain the corresponding LBP image. Construct the histogram of the LBP image. This histogram is used as a feature vector for the classification process.

3-4-3 Histogram of Oriented Gradient (HOG)

HOG [14] is a feature extraction method that extracts the significant shape and edge information from an image. It is invariant to scaling, rotation, and illumination. Steps for applying HOG method is as follows:

1. For each pixel in the image, compute the magnitude and direction of the gradients.
2. Divide the image into small cells. The histogram of gradients of each cell is computed.
3. Normalize histograms over multiple cells called blocks.
4. HOG features are collected from image blocks and concatenated to obtain The HOG feature vector of the image.

3-5 Machine Learning Techniques (MLTs)

Multiple MLTs are utilized in the field of plant leaf disease detection and classification. In the realm of classification algorithms, the dataset undergoes a partitioning process into training and test sets. The construction of the model is conducted utilizing the training dataset, followed by the validation of said model using the test dataset. In the proposed scheme, KNN and DT used to classify corn leaf images.

The KNN [15] classifier is a supervised classification method wherein an unknown sample is categorized through an initial computation of its distance from each of the training samples. This method identifies the k smallest distances, and subsequently assigns the output class label of the unknown sample based on the most frequently occurring class within these k classes. Different distance measures are used but the most common is the Euclidean distance.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_{ik} - y_{ik})^2} \quad (1)$$

Where x_{ik} and y_{ik} are the i -th points of the k -th variables from the training and test data respectively.

DT formulates a series of meticulously devised inquiries regarding the characteristics of the test record within a hierarchical tree framework. Upon receiving each response, a subsequent query is posed until a determination is made regarding the classification of the record. A decision tree classifier comprises a singular root node, numerous internal nodes, and multiple terminal nodes. The root and internal nodes encompass the test conditions for the attributes to differentiate among records with varying characteristics. Each terminal node is allocated a specific class label. In the process of constructing decision trees, the fundamental algorithm utilizes entropy to assess the level of homogeneity within the sample. When perfect homogeneity is achieved, the entropy is reduced to zero; conversely, if the sample is evenly divided, the entropy is one. The creation of decision trees revolves around identifying an attribute that yields the greatest information gain, essentially leading to the most homogeneous branch.

IV. EXPERIMENTAL RESULTS

The experiments are conducted utilizing MATLAB R2017b on a computing system running the Windows 10 operating platform, equipped with 4GB of RAM. The corn leaf images were collected from Plant Village dataset. They have been divided into 80% for training and 20% for test. A feature vector combining features extracted using

GLCM, LBP, and HOG is obtained. Two classification algorithms are employed which are KNN and DT. They are used to classify the corn leaf into four classes Common Rust, Gray Leaf Spot, Blight, and Healthy. To evaluate the proposed algorithm, accuracy, precision, recall, and confusion matrix are used.

$$Accuracy = (T_p + T_N) / (T_p + F_p + T_N + F_N) \% \tag{2}$$

$$Recall = T_p / (T_p + F_N) \tag{3}$$

$$Precision = T_p / (T_p + F_p) \tag{4}$$

$$F1 - score = 2T_p / (2T_p + F_p + F_N) \tag{5}$$

Where, T_p is true-positive value, F_p is false-positive value, T_N is true negative value and F_N is false-negative value.

The Confusion Matrix is a tabular representation employed in the evaluation of performance, serving to enhance the process of visualizing results as shown in table 3. Accuracy represents the proportion of accurately classified instances among the total quantity of available data. Precision, also known as the positive predictive value (PPV), denotes the correlation existing between true positive outcomes and all positive predictions. Recall, or otherwise referred to as the true positive value (TPV), signifies the connection between true positive results and all positive elements. The F1-Score, a metric utilized in statistical analysis, is calculated as twice the product of the precision value and the recall value, divided by the sum of the precision value and the recall value.

Table 3 Confusion matrix

Actual Class	Predicted Class	
	Positive	Negative
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

For KNN classifier, different K values are tested. The achieved accuracy is approximately 84% with K equals 3. Also, the achieved F1-score is 80%, the precision is 80%, and the recall is 80%. Almost the same results are achieved for k equals 5 and 10. According to DT, the achieved accuracy is 80% while the F1-score is 74% with precision of 74% and recall of 74%. Table 4 collects the performance evaluation of both KNN and DT for corn leaf disease detection and classification using the proposed scheme. Figure 6 is a bar chart depicts the performance of KNN and DT using the proposed scheme. It is shown that KNN outperforms DT in the classification of corn leaves using the proposed scheme. The confusion matrix of KNN classifier is illustrated in table 5.

Table 4 KNN performance using different values of K

	Accuracy	F1-Score	Precision	Recall
KNN	0.8415	0.8144	0.8105	0.8183
DT	0.8000	0.7424	0.7428	0.7420

Table 5 Confusion matrix using KNN (K=3)

	Gray Leaf Spot	Common Rust	Northern Leaf Blight	Healthy
Gray Leaf Spot	61	0	39	3
Common Rust	0	236	1	1
Northern Leaf Blight	33	0	157	7
Healthy	26	0	14	192

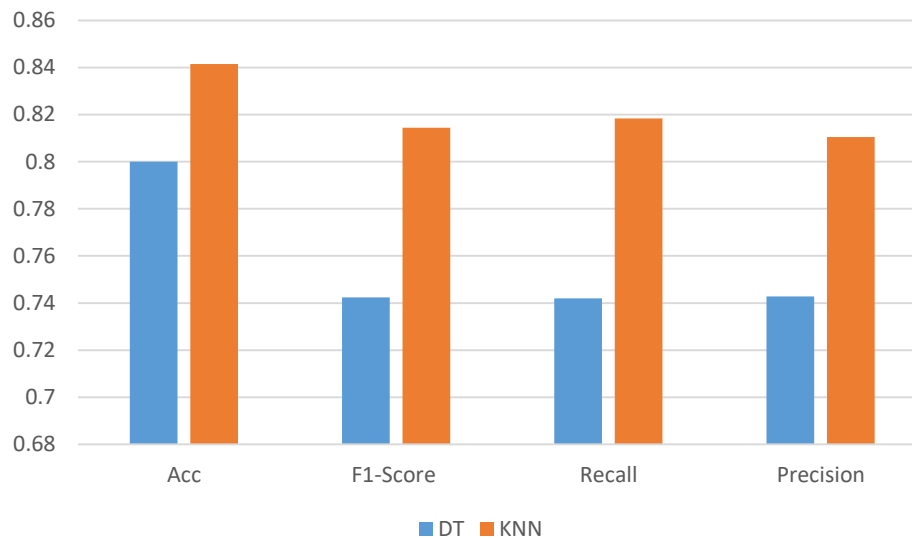


Figure 6 Bar chart of the performance of KNN and DT for corn leaf disease classification.

V. CONCLUSION

The suggested research illustrated that the identification of plant infections necessitates a series of five crucial stages: image acquisition, pre-processing, segmentation, feature extraction, and categorization. The input images are randomly partitioned into training and test. The available approach employs the GLCM, LBP, and HOG algorithms to extract texture descriptors. GLCM examines the correlations among adjacent pixels, offering a comprehensive analysis of textural details. The LBP descriptor determines the local construction of an image that remains constant despite variations in lighting conditions. The HOG descriptor extracts the significant shape and edge information from an image. KNN and DT are used for the classification process of corn plant. The experimental results showed that the proposed scheme attains good performance in detecting and classifying corn leaf disease.

REFERENCES

- [1]A. Waheed, M. Goyal, D. Gupta, A. Khanna, A. E. Hassanien, and H. M. Pandey, "An optimized dense convolutional neural network model for disease recognition and classification in corn leaf," *Computers and Electronics in Agriculture*, vol. 175, p. 105456, 2020.
- [2]F. Rajeena PP, A. SU, M. A. Moustafa, and M. A. Ali, "Detecting plant disease in corn leaf Using efficientNet architecture—An analytical approach," *Electronics*, vol. 12, p. 1938, 2023.
- [3]P. K. Sethy, N. K. Barpanda, A. K. Rath, and S. K. Behera, "Deep feature based rice leaf disease identification using support vector machine," *Computers and Electronics in Agriculture*, vol. 175, p. 105527, 2020.
- [4]M. Aamir, M. Irfan, T. Ali, G. Ali, A. Shaf, A. Al-Beshri, et al., "An adoptive threshold-based multi-level deep convolutional neural network for glaucoma eye disease detection and classification," *Diagnostics*, vol. 10, p. 602, 2020.
- [5]P. Kulkarni, A. Karwande, T. Kolhe, S. Kamble, A. Joshi, and M. Wyawahare, "Plant disease detection using image processing and machine learning," *arXiv preprint arXiv:2106.10698*, 2021.
- [6]A. A. Esmael, "A novel approach to classify and detect bean diseases based on image processing," in *2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, 2018, pp. 297-302.
- [7]X. E. Pantazi, D. Moshou, and A. A. Tamouridou, "Automated leaf disease detection in different crop species through image features analysis and One Class Classifiers," *Computers and electronics in agriculture*, vol. 156, pp. 96-104, 2019.
- [8]S. Hossain, R. M. Mou, M. M. Hasan, S. Chakraborty, and M. A. Razzak, "Recognition and detection of tea leaf's diseases using support vector machine," in *2018 IEEE 14th International Colloquium on Signal Processing & Its Applications (CSPA)*, 2018, pp. 150-154.

- [9]E. Alehegn, "Maize Leaf Diseases Recognition and Classification Based on Imaging and Machine Learning Techniques," 2020.
- [10]N. K. Durga and G. Anuradha, "Plant disease identification using SVM and ANN algorithms," International Journal of Recent Technology and Engineering (IJRTE), vol. 7, 2019.
- [11]Z. Zhang, X. He, X. Sun, L. Guo, J. Wang, and F. Wang, "Image recognition of maize leaf disease based on GA-SVM," Chemical Engineering Transactions, vol. 46, pp. 199-204, 2015.
- [12]S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," Frontiers in plant science, vol. 7, p. 1419, 2016.
- [13]T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE Transactions on pattern analysis and machine intelligence, vol. 24, pp. 971-987, 2002.
- [14]N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), 2005, pp. 886-893.
- [15]Z. Zhang, "Introduction to machine learning: k-nearest neighbors," Annals of translational medicine, vol. 4, 2016.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details