Plant Leaf Disease Detection Using Texture Features and XG-Boost Classifier

Dr. S. Satheesh[1], Dr. P. Ashok Babu[2]

[1]Professor, Department of ECE, RISE Krishna Sai Prakasam Group of Institutions, Ongole, Andhra Pradesh, India.

[2]Professor and Head, Department of ECE, Institute of Aeronautical Engineering, Dundigal, Hyderabad, Telangana, India.

Abstract: Identification and recognition of plant leaf disease require multiple computer vision-based procedures. Like image acquisition, image pre-processing, image segmentation, feature extraction, classification, and diagnosis of leaf diseases are examples of such measures. These processing methods move into and extract the finer details of the plant leaves that helps us to classify the type of the disease. However, the effectiveness of the processing methods relies on the data that has been collected and the samples that are involved in feature extraction. In the proposed approach, plant leaf detection is performed with enhanced texture features, which are used as features to discriminate different classes of diseases. Later, an XG-boost based classifier is used to perform the classification. The results are compared against traditional classifiers like Support Vector Machine (SVM) and conventional XG-boost classifiers. The classification process is validated with a standard Village Plat leaf Dataset with multiple class categorizations and compared with state-of-the-art classifiers. The experiments showed that when the enhanced features with the XG-booster classifier are employed, the accuracy is superior, with an improvement of 5% to 7%.

Keywords: *Plant Leaf Disease, Classification, Feature Extraction, XG-Boost Classifier*

1. INTRODUCTION

Agriculture is considered one of the chief sources of livelihood for people around the globe, and India is undoubtedly an agricultural generator globally. Agriculture in India started in the Indus Valley Civilization. India has been categorized second globally in agronomic products. According to [4], more than 50 percent of manpower was deployed in agriculture and committed to 17-18 percent of the GDP of India. Farming is a primary source of income for about 58 percent of the population of India. India stands as the sixth-largest food and packaged goods market and contributes 70 percent of the sales.

Diverse plants are harvested according to the country's requirements and the habitation circumstances and conditions, with agriculture being a vital source of economic growth. But there can be several issues that the farmers around the globe face, and that may comprise water shortage, bad weather conditions, natural disasters, and plant diseases. The problem of plant disease detection, for example, can be solved with the help of technical aids in the form of machine learning in general and image processing in particular. It may not be possible to remember the information related to each type of disease in plants. There are very few specialists involved in this area, so this type of technical aid will act as a boon for agriculturists and farmers around the globe.

In most developing countries globally, the vertebra of the economy is supported by agriculture. The plant sprouting determines the quality and quantity of a crop. Crop diseases frequently appear on the most delicate parts of the plant, namely the leaf. Peak care is required to reveal the diseases as timely as possible. A well-timed detection can inhibit the proliferation of diseases to the whole field or bunch of crops, which will improve yield. Observing the color and surface of the leaves is the traditional practice to detect the disease. Due to the time-consuming nature and requirement of regular endeavor and knowledge, the practice fails when it comes to disease detection in huge fields.

A specific plant can show contrasting symptoms for different diseases. It may be the diseased area's color, texture, or shape ranging from small-sized spots to large-sized patches. Due to these varying characteristics of a disease, there is a necessity to adopt an approach that is less time-consuming. Therefore, it will also boost the crop production rate and the economy. The disease can be detected on any part of a plant, be it leaves, stem, or the fruit itself, but a timely revelation is required. Many researchers have been devoting their time to this domain to help out farmers.

Dependency on crop production for economic growth cannot be ignored and hence constitutes an important part of it. The most important thing to consider to boost the production rate is monitoring the diseases related to different crops within the intended deadlines. Observation of plants' well-being based on the obvious or noticeable indications of disease on the leaves of plants is an indispensable part of agriculture as plants play a dominant role in our ecosystem, imperishable agriculture, and atmospheric conditions. Since, at the global scale, they are a dominant source of food and economy, it is both a societal and a phenomenal engagement. Significant operations in plants, such as photosynthesis, can be altered by a disease that can affect the quantity of agricultural production and the quality of the crop. This can further affect a country's economy like India, where agronomy is the dominant source of income for farmers working on a small scale. The disease symptoms in a plant generally appear on the leaves. It is a very challenging task to scrutinize the syndromes of leaves infected by diseases with the help of the naked eye [6]. The pathogens and weeds are the main causes of these infections and can be detected with the traditional optical method. However, it depends on the people's experience in this field which can otherwise be time consuming and cumbersome task.

In agronomic productivity, the key requirement is a well-timed prognosis of the disease. If this timely prognosis of the disease is not considered seriously, then crop yield loss is an obvious adverse effect. This is why farmers worldwide face problems identifying disease symptoms in the diseased plants properly in the initial stage of the disease.

In order to support these people in identifying the correct diseases in plant leaves, Image processing techniques are applied. The advantage of using image processing in agriculture is the high degree of precision in forecasting and diagnosis and cost savings compared to conventional methods. Several image processing computer-based tools have been built in agriculture to help farmers monitor their crop growth process. On the other hand, image processing with a machine learning algorithm plays a significant role in classifying different diseases in plants. Therefore, these reasons in research have motivated us to work on bringing up a system that can identify leaf diseases on time to avoid any loss in the agronomic productivity, which can be increased in the best probable way. Hence, the farmers can take reasonable steps to treat the diseases involved in the leaves of the plants.

2. RELATED WORK

The following figure 1 depicts the various stages and approaches of image processing involved in plant leaf disease recognition.

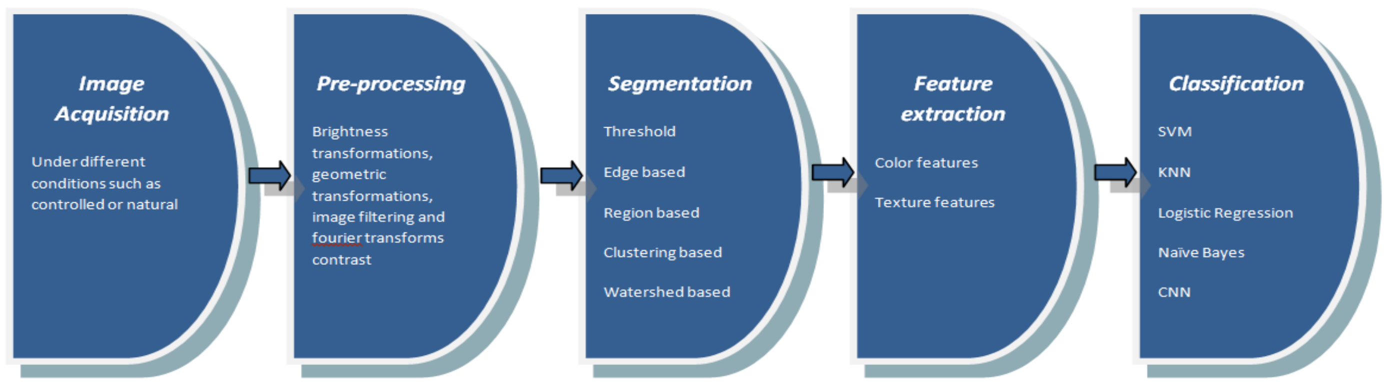

Figure 1: Phases of plant leaf disease detection

In [1], a unique approach hybridizes machine learning and deep learning techniques to detect leaf diseases in peanuts. A total of 6029 images of diseased leaves of peanut plants have been observed. Five categories of diseases have been taken into account. The device used was a mobile phone to capture the images. Rotation, scaling, and flipping are the three augmentations approaches adopted in this paper. Better accuracy has been obtained with the deep learning model. The same deep learning model even performed better after using it and stacking ensemble and augmentation. The machine learning technique for stacking ensemble is logistic regression, RF, and SVM. 97.59% of accuracy is achieved. ResNet50 and DenseNet121 have shown the maximum accuracy with logistic regression and RF, respectively, with data augmentation. The setup considered in the paper is a laboratory setup which is potential enough to obtain an accuracy of 97.59%. However, the natural environment is generally complex, and hence detection of disease will experience complications. The authors have suggested future work related to the same.

In [2], Tyagi stressed that the ability of humans to identify and analyze plant diseases is limited because it is dependent on Nano scale activity. Computer-assisted images rearrangement approaches are used in precise plant disease classification and identification. A k-mean clustering procedure has been used to determine disease on an actual plant leaf image. After detection has been completed, it's time to go on to the next step. The GLCM filter draws out features. The spatial frequency components, or how frequently a grouping of images illuminates contrast levels in pixels appearances, are bundled as GLCM. Feature extraction is the process of transforming contribution information into a group of spatial and texture statistic features. SVM-based technologies are often used for classification. However, it has a low accuracy level when it concerns textural characteristics. A Back Propagated ANN technique based on innovative artificial intelligence is used for classification to achieve feature-formulated comparison. The proposed method has been tested in MATLAB software, and its accuracy is far superior to that of standard methods. This analysis includes applying a deep machine learning methodology to establish a generic disease classification for diverse infections.

In [3], Setiawan et al. adduced a technique to project timely identification of diseases in maize crops. Almost all traditional classifiers have been compared based on accuracy and F1 score. Feature extraction on the grounds of color is performed. The appearance of the diseases leads to the variation in diseases. The color feature differentiates the diseases from one other. RGB feature extraction is conducted to filter out the color information from the images. 3823 images were examined for the experiment, with four types of classes being considered. 90% of the images are utilized for training, and the rest 10% for testing. Conventional classifiers such as SVM, NB, KNN, DT, and RF have been compared. RF is found to obtain a maximum accuracy of 80.68%. As a part of future work, the paper researchers have recommended the utilization of high-dimensional datasets.

In [4], Pethybridge et al. adduced a CBIR method with the aid of Ranklet Transform as a step in pre-processing and color features. K means clustering was utilized for this purpose. For making the image invariable towards rotation, Ranklet Transform has been utilized. As a part of future work, the authors have suggested utilizing shape and texture features accompanying color features.

In [5], Bhusri and Jain adduced techniques based on localization to categorize diseases into three crop types for distinct diseases. Initially, localization in the leaf zone is performed with the aid of color features and subsequently utilization of mixture model-based region expansion. Different patterns are shown in the leaf images. These features of differentiating characteristics impact the classification of images. Characteristics of infected leaves in images, such as spots and damaged leaf areas in localized images, reveal how healthy images can be differentiated from infected leaves, which can be done easily with Fisher Vector (FV) extraction. Gaussian distribution's differentiation of various orders aids the FV. FV is obtained by Fisher Kernel (FK). FK illustrates a feature by a gradient vector with the help of SIFT. The conduct of classification is determined on the grounds of AUC and accuracy. The analysis is done on bell pepper, tomato, and potato images of the Plant Village dataset. Two bell pepper categories, three potato categories, and ten tomato categories have been considered. SVM and Multi-layer perceptron is utilized to test the performance. 94.35% is the maximum accuracy achieved, and AUC is found to be 94.7%. In contrast with the existing techniques, the new approach performs better. The authors have suggested using other crops and future work in this research.

In [6], Prasad et al. put forward a fuzzy classification approach to automatically recognize the diseases in apple plant leaves. A fuzzy association function is applied that views the association between the pixels on the grounds of the degree. 20 samples are stored in the database and captured using digital cameras. As a significant part of image processing, the pre-processing step is also employed in edge detection and thresholding. Otsu's approach is applied for thresholding the image with the utilization of a histogram. Otsu's approach splits the pixels into two categories – the forepart and the backdrop. Various mathematical approaches are adopted for detecting edges. The edges are identified by taking into account the sharp variations in the image's brightness, which ultimately aids in identifying the borders of objects in images. Color features and LBP features are extracted later on after employing the contrast enhancement technique to make the features in the image emerge prominently. Following this, K means the approach is utilized for segmentation purposes. Eventually, the fuzzy classification is employed to obtain the right disease class. The adduced model offers an accuracy of 93% that more in contrast to state-of-the-art techniques. The authors have suggested future work as well. In the future, modern background detachment approaches can be employed to split up the objects from the backdrop in the image.

In [7], Johannes et al. have presented a color feature-based rice plant disease categorization technique. One of the most important characteristics of rice plant diseases is color. 14 distinct color spaces were investigated, and four were extracted. There are 172 features in total, with features from each color channel. Four statistical parameters were recovered and utilized as attributes in this machine learning architecture. The extracted color features were input to classifiers to classify rice plant diseased images. Operating a 10-fold cross-validation process, the dataset was partitioned into training and test sets. Furthermore, the performance of seven different classifiers was evaluated, revealing that the SVM classifier achieved the greatest classification accuracy of 94.65%. There are four classes in the dataset. The dataset containing 619 images was used to train and test the models. The optimistic findings of this article reveal that color features can play an essential role in developing a rice plant disease diagnostic model, allowing farmers to adopt proactive steps and improve product type and effectiveness. The authors hope to collect more datasets of rice plant illnesses with a larger number of labels in the future and use deep learning approaches to characterize rice plant diseases.

In [8], Kamal et al. developed a new approach for diagnosing tomato plant illness. Pre-processing, segmentation, feature extraction, and classification are the four stages of the proposed technique. The obtained images were downsized, and the noise was suppressed using the Weiner filtering technique in the pre-processing step. For that improved K-Means image segmentation algorithm, segmentation was proposed as one of the crucial processes. Following the segmentation of the image, a Region of Interest (ROI) was identified, and relevant characters were collected using the feature extraction approach from this ROI. The authors employed the GLCM algorithm to extract significant factors from ROI in the suggested technique. After the segmentation stage, important characteristics are recovered from the segmented image utilizing the GLCM feature extraction approach. Eventually, classification techniques such as Support Vector Machine (SVM) and Adaptive Neuro-Fuzzy Inference System (ANFIS) are used to categorize sick leaves. The tests are based on images of tomato leaves found in the plant village database. The proposed scheme is tested in tomato plants for five different illnesses. The adduced advanced K-Means with ANFIS classification algorithm achieved the best accuracy of 98.60 percent.

3. PROPOSED METHODOLOGY

The Block diagram of the proposed approach is depicted in figure 2. The approach requires a sequence of steps where the input plant leaf image is segmented initially to extract the leaf portion accurately since it is known that features in the leaf help to categorize the disease class. This segmentation process is performed with traditional K-means clustering. Despite many, K-means was chosen as it is faster in partitioning the regions of the objects in the image.

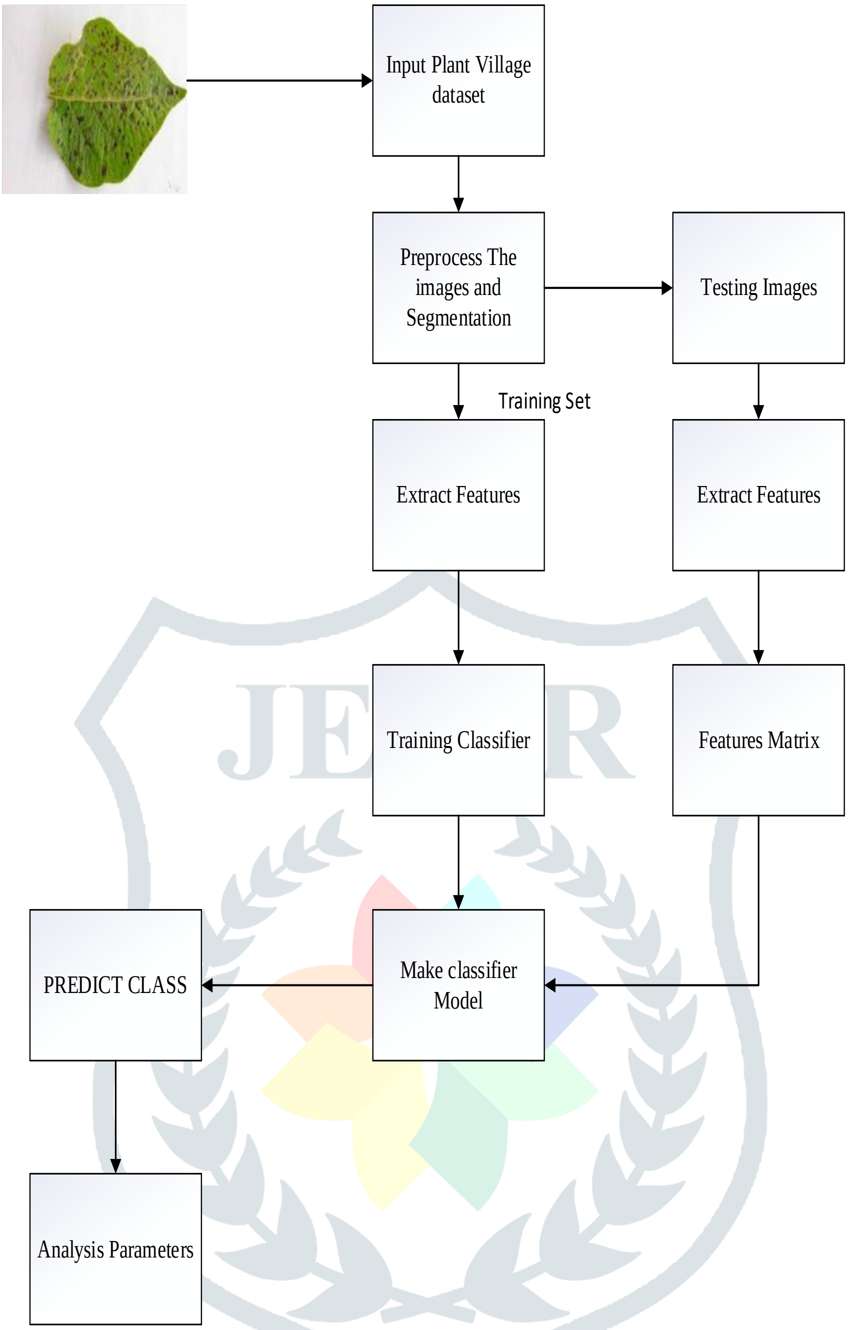

Figure 2: Block diagram of the proposed leaf disease detection approach

Initially, to improve the segmentation efficiency and extract the leaf regions more accurately, the images are pre-processed with Gaussian distribution that aims to normalize the pixel values, which minimizes the images' noise content. Later, the clustering method is applied for extracting the leaf regions.

In the next step, the features are improved by finding the specific location on the image where the most sickness possibilities exist. Following clustering, fundamental segmentation recovers various characteristics, followed by feature extraction through partitioned areas to find Law Mask GLCM and LBP features using the convolution kernel.

Let the textural energy assessment be carried out using K. I. Laws, which are used in various applications. Such parameters are defined by applying certain convolutional layers to a digital image and then performing a non-linear windowing procedure. The kernels of the convolution will then be initiated, to which we may respond afterward. L2-D convolution masks are commonly employed in texture-based differentiation and are made up of a set of 1D convolution masks with the following lengths: L3 = 1 2 1, E3 = 1 0 1, S3 = 1 2 1 L5 = 1 4 6 4 1, E5 = 1 2 0 2 1, S5 = 1 0 2 0 1, W5 = 1 2 0 2 1, R5 = 1 4 6 4 1. Spot denotes the extraction of spots, Edge denotes the extraction of edge features, Level denotes the average grey level, Ripple denotes the extraction of ripples, and Wave denotes the extraction of wave characteristics. Texture-based convolution using Law masks and energy stats yields a texture overview that can be used to distinguish between textures.

Finally, classifications are used to pick features in which the learning is done in several classes employing XG-Boost, and testing is conducted using accuracy, recall, and precision.

Algorithm for classification with XG-Boost Classifier

Input: *Input Images*
Output: *Segmented Images*
Begin
N $\leftarrow$ *no. of Images*
P $\leftarrow$ *i x j Pixels*
While (N>0)
Start

$$G(P) = \frac{1}{\sqrt{2\pi\sigma}} C^{-\frac{x^2}{2\sigma^2}} \dots\dots\dots\dots\dots\dots\dots\dots \quad (1)$$

σ = *difference* $(P_{i,J-1} - P_{i-1,J})$
End
Define centroid $\{X_1, X_2, \dots\dots\dots\dots, X_n\}$
While (Centroid >0)
Start
Define the population of grey wolves
GW $\leftarrow$ *Centroid*
$G_\propto \leftarrow G(P)$
$G_\beta \leftarrow N$
$G_\delta \leftarrow P$
Update weights

$$w^{n+1} = \frac{w_0\, G_\propto + \sum GW}{G_\propto + G_\beta + G_\delta} \dots\dots\dots\dots\dots\dots\dots\dots.. \quad (2)$$

End
End

4. EXPERIMENTAL RESULTS

Plant Village Dataset is used to validate the performance of the proposed approach with enhanced features and an XG-boost classifier. For the experimental analysis, multiple classes are considered, like Pepper, Potato, and others. Multiple metrics like Accuracy, precision, and Recall are employed to validate the performance. Table 1 depicts the experimental setup of the approach.

Table 1: Experimental Setup

Database	**Plant Village**
Leaves	Bell pepper, Potato, Tomato
Number of classes	2,3,10
Classifier	XG-BOOST
Features	Texture, Law’s mask
2-class	Bell Pepper
3-class	Potato
10-class	Tomato
Clustering	K means with GWO
Evaluation Metrics	Accuracy, Precision, Recall

The experimental configuration for the proposed methodology is shown in Table 1. The research employs a Plant Village dataset with three cases: two classes, three classes, and ten classes, with accuracy, precision, and recall used in the study.

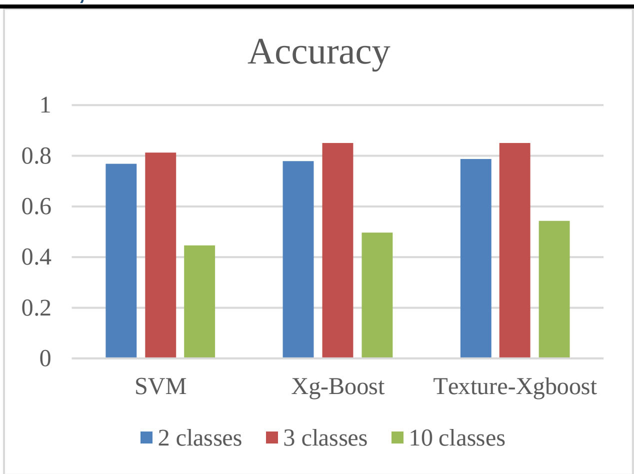

Figure 3: Accuracy comparison of the classifiers for multiple classes

Table 2: Accuracy Comparison

	2 Classes	3 Classes	10 Classes
SVM	0.768506057	0.812693498	0.44576
XG-Boost	0.779273217	0.849845201	0.49719
Texture-XG-boost	0.787348587	0.851393189	0.54383

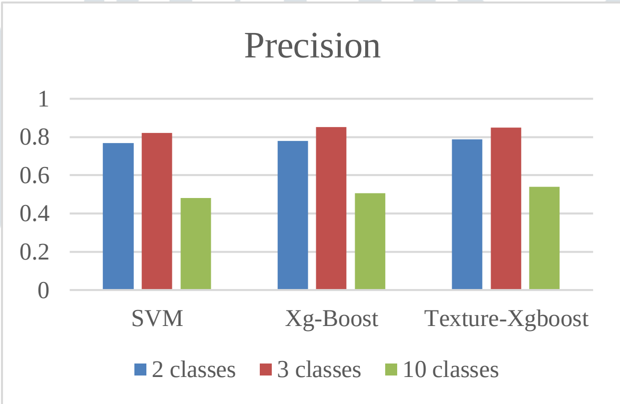

Figure 4: Precision comparison of the classifiers for multiple classes

Table 3: Precision Comparison

	2 Classes	3 Classes	10 Classes
SVM	0.749054905	0.578799602	0.37045
XG-Boost	0.760714821	0.691485597	0.42856
Texture-XG-boost	0.764963996	0.758774814	0.47614

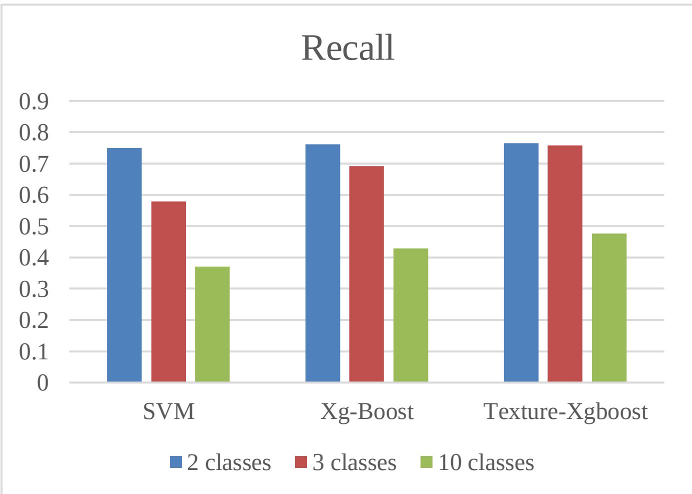

Figure 5: Recall comparison of the classifiers for multiple classes

Table 4: Recall Comparison

	2 Classes	3 Classes	10 Classes
SVM	0.767086938	0.821973748	0.48009
XG-Boost	0.778141335	0.850679989	0.50684
Texture-XG-boost	0.788346674	0.84957792	0.53949

Numerous observations are found in the data, and the first one is to increase the recommended approach's accuracy, precision, and recall. However, sophisticated capability and segmentation are used to improve measures.

- Collection of attributes is improved via segmentation because overlapped segmentation, such as fuzzy segmentation, reduces feature distortion.
- The new framework makes use of optimized segmentation to eliminate overlapped segments.
- Accuracy, precision, and recall all increased due to feature extraction. The properties employing the Law's mask characteristics convoluted kernels in the suggested approach.
- The suggested method enhances the feature extraction process, but the learning process does not boost 10-class categorization because it depends on the classifier's learning form.
- Consequently, the suggested method greatly improves 2-class, 3-class, and 10-class parameters.

Figure 3, 4, 5 depicts the performance comparison of the proposed approach in terms of metrics like Accuracy, Precision, and Recall for 2, 3, and 10 classes of diseases. It was observed that the proposed classification approach could retain higher metrical indexes when compared with state-of-the-art methods, with an average improvement of 5% to 7%.

The proposed method could improve accuracy by 10 percent, recall rate by 5 percent, and precision rate by 10 percent. This is a considerable achievement in classification when tested with thousands of images.

5. CONCLUSIONS

The proposed approach attempts to resolve concerns with leaf disease classification, with 2, 3, and 10 classes. This approach could detect and classify several critical diseases in the standard dataset. Such a strategy aims to collect image characteristics while employing texture features and a convolutional-based Law's mask and then create the metric space utilizing the principle of resemblance among distinct image characteristics. The performance of the XG-Boost classification is determined by the accumulation of supervised samples and the coherence of the metric space. Grey wolves' methodology is used to optimize fragmentation to reduce object blending, as can be seen. Multiple classes and state-of-the-art classifiers are used to classify diseased and non-diseased potato, tomato, and bell pepper types. It was observed that the proposed approach of enhanced feature set with an XG-boost classifier could retain high metrical values than the existing state of art classifiers. The suggested approach was able to classify many essential diseases in the dataset with an average accuracy of 72.3 %, which is a significant accomplishment in plant disease classification.

REFERENCES

[1] S. C. Mittal, "Role of Information Technology in Agriculture and its Scope in India," Fertiliser News, vol. 46, no. 12. pp. 83–88, 2001.

[2] V. Tyagi, "India's Agriculture: Challenges for Growth & Development in Present Scenario," *International Journal of Physical and Social Sciences*, vol. 2, no. 5. pp. 116–128, 2012.

[3] A. S. Setiawan, Elysia, J. Wesley, and Y. Purnama, "Mammogram Classification using Law's Texture Energy Measure and Neural Networks," Procedia Comput. Sci., vol. 59, no. Iccsci, pp. 92–97, 2015, doi: 10.1016/j.procs.2015.07.341.

[4] S. J. Pethybridge, I. Plant, P. Pathology, and P. Biology, "Leaf Doctor : A New Portable Application for Quantifying Plant Disease Severity," vol. 99, no. 10, pp. 1310–1316, 2015.

[5] S. Bhusri and S. Jain, "Analysis of breast lesions using laws' mask texture features," 2016 4th Int. Conf. Parallel, Distrib. Grid Comput. PDGC 2016, pp. 56–60, 2016, doi: 10.1109/PDGC.2016.7913115.

[6] S. Prasad, S. K. Peddoju, and D. Ghosh, "Multi-resolution mobile vision system for plant leaf disease diagnosis," Signal, Image Video Process., vol. 10, no. 2, pp. 379–388, 2016, doi: 10.1007/s11760-015-0751-y.

[7] Johannes et al., "Automatic plant disease diagnosis using mobile capture devices, applied on a wheat use case," Comput. Electron. Agric., vol. 138, pp. 200–209, 2017, doi: 10.1016/j.compag.2017.04.013.

[8] K. Kamal, R. Qayyum, S. Mathavan, and T. Zafar, "Wood defects classification using laws texture energy measures and supervised learning approach," Adv. Eng. Informatics, vol. 34, no. September, pp. 125–135, 2017, doi: 10.1016/j.aei.2017.09.007.