Global Mainstream Journal of Innovation, Engineering & Emerging Technology

Vol. 03, Issue 05, November, 2024
e-ISSN:2998-3967
Page:11-18

AN IMAGE RECOGNITION APPROACH FOR CROP DISEASE DETECTION IN AGRO-FIELD FROM INFECTED PLANT AREA

Md. Mehedi Hasan Rabbi[1], Mohammad Jamal Hossain [2], Md. Abdul Masud [3], Md. Atikqur Rahman[4]

[1,2,3,4] Department of Computer Science and Information Technology (CSIT)
Faculty of Computer Science and Engineering, *Patuakhali Science and Technology University, Patuakhali, Bangladesh*

[1]mehedirabbi16@gmail.com , [2]jamalpstu07@gmail.com,[3]masud@pstu.ac.bd ,[4]atikqurrahaman@gmail.com

Keywords

Image Processing
Crop Disease
Pixel
Sobel Edge
Ensemble Classifier

ABSTRACT

A plant pathogen is an abnormal metabolic disorder that inhibits a plant's typical structure, development and function. Disease lowers the quality of crops and the yield they produce, which harms the economies of countries like Bangladesh, where agriculture is the primary economic industry. Crop disease identification considers visually observable patterns and colors of the diseased region. Manually observing patterns and colors to categorize diseases is time-consuming and appears not to be as effective when dealing with diseases not indigenous to the area. Image processing has become the most widely used method for identifying and categorizing diseases. Despite this, it is still thought that implementing image processing is a difficult and complex undertaking. The first difficult assignment to complete is an entirely distinct feature extraction task. The authors of this paper present an ensemble-based model that applies unique approaches to extracting numerous features from plant leaves, such as color, shape, and texture to identify various diseases. The mean value of the pixels in the affected area is determined to extract the color feature. We employed the Sobel edge detector to exhibit edges, which primarily include shape and texture information about an image. The results showed that the proposed method was effective in identifying the infected areas of the leaves, with an accuracy rate of 94%. This method could potentially be used for early detection and prevention of plant diseases in agriculture.

Article Information

Received: *09, October, 2024*

Accepted: *16, November, 2024*

Published: *18, November, 2024*

Doi: 10.62304/jieet.v3i05.219

1 Introduction

Plant diseases can lead to food shortages and price increases, which can have a negative effect on the ability of producers and customers to maintain their standard of living. Disease Management is a broad term that involves recognizing the health condition, categorizing it, and providing treatment for it. Recognizing and categorizing the disease must first be completed, which can be an exhausting endeavor. On the basis of discolored spots or streaks that appear anywhere on the stalk or leaves, diseases can be recognized. The various forms of the disease each have their own distinct coloration and pattern. Even though traditional observation can be used to determine diseases, it is no longer practical when dealing with vast regions or diseases that are not endemic to the area. Image processing can be used in this situation rather than human observation, making disease identification quicker and more accurate. Due to the fine results it produces and the fact that it requires less human effort, image processing is helpful in detecting plant diseases, which can save time and resources for farmers while also improving crop yields.

Yet, it is still believed that implementing image processing is a challenging and complex task. The first challenging objective is to perform a unique feature extraction task. Effective feature extraction is needed to identify and classify various plant diseases. A dependable feature extraction technique can increase a classifier's effectiveness. In this paper, we propose a novel approach for feature extraction from plant leaves to aid in disease classification. The extracted attribute chosen for classifying plant diseases is the color, texture and shape of the leaves. The second challenge is selecting the appropriate algorithm to process the extracted features and achieve the desired outcome. The suggested approach demonstrates that groups of the selected classifiers can perform significantly better than single classifiers. Furthermore, the ensemble approach can also help to reduce overfitting and improve generalization performance.

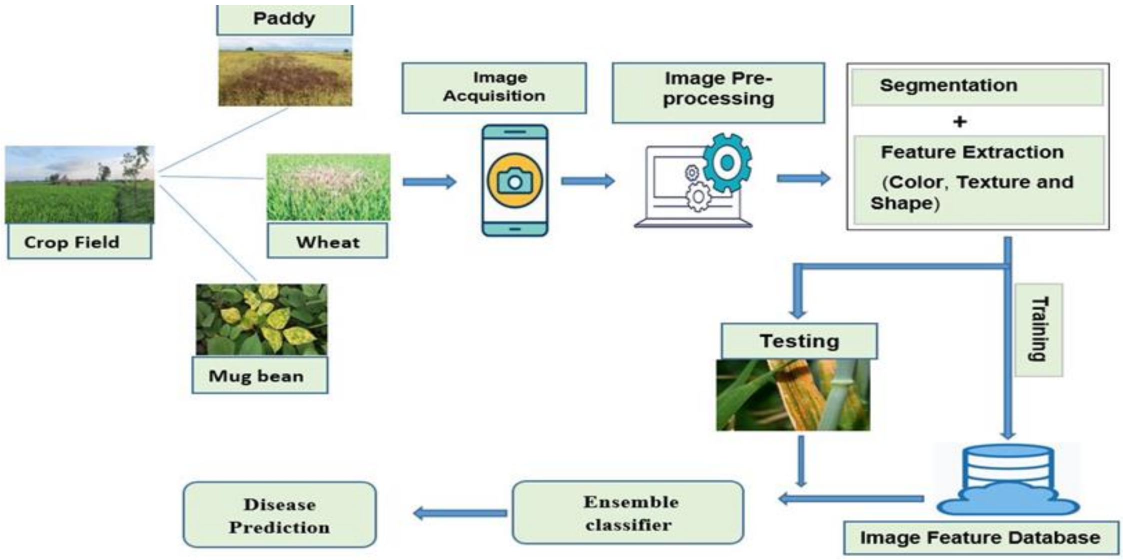
2 Literature Review

For the purpose of detecting plant diseases with machine learning, a model was introduced by S. Ramesh, R. Hebbar, Niveditha M. Pooja R.; Prasad B. N., Shashank N., Vinod P.V. consists of the following steps, preprocessing, extracting features, training a classifier, and classifying [1]. Hu moments, the color histogram, and the haralick texture are the three feature descriptors used in this research. The Histogram of an Oriented Gradient (HOG) can be used to extract features from an image, and the Random Forest classifier can be used to classify the features with an accuracy rate of 70%.

S. R. Shinde, S. Sabale, S. Kulkarni, and D. Bhatia stated a color feature extraction-based image categorization system [2]. In this research, the author uses data pre-processing to color attributes of images to create a feature vector, and then they combine this feature vector with machine learning classifiers to categorize the images.

Rice brown spot and blast infections are automatically classified using a segmentation method based on vegetation indices by S. Phadikar and J. Goswami [3]. The segmentation approach employed by the author is Otsu's. Then, using GLCM, five distinct texture features homogeneity, correlation, contrast, energy, and entropy of the infected spots are calculated. Classifiers are then applied to these feature values in order to categorize the diseases. The accuracy of the proposed approach is 84%.

Figure 1: Proposed methodology diagram.



A disease recognition and categorization approach using machine learning mechanisms and image processing techniques was put out by V. Pooja, R. Das, and V. Kanchana [4]. Using K-means clustering, converting RGB to HIS, and Ostu's detection, segmentation is carried out using this method. For feature extraction, factors such colour co-occurance, skewness, contrast, and correlation are applied. SVM is utilized for classification.

Technique for the detection of wheat fungal disease was proposed by Varsha P. Gaikwad and Dr. Vijaya Musande [5]. The use of histograms allowed for the differentiation of healthy wheat from unhealthy wheat. They were able to recognize diseases with the usage of Neural Networks (NN) and Support Vector Machines (SVM) from their extracted feature vector. 89.23%

A corn leaf disease recognition system using image processing was developed by T. Ren, Y. Zhang, and C. Wang [6]. Hu invariant moment, shape factor F, roundness R, equivalent area radius Re, and radius of inscribed circle Ri were chosen for feature extraction. The classification was then done using a support vector machine.

A K-mean clustering-based segmentation technique was developed by A. K. Singh, A. Rubiya, and B. Raja for detecting rice blast disease using SVM [7]. Their methodology yielded 82% accuracy.

A method for detecting rice diseases named rice sheath blight, rice blast and brown spot was proposed by Anthonys G. and N. Wickramarachchi [8]. Shape, texture, and color characteristics considered for categorisation. However, they only managed to classify objects with an accuracy of 70%, which required the inclusion of three features.

3 PROPOSED METHOD

The systematic, theoretical examination of the techniques used in an area of study is known as methodology. The strategy can be broken down into two components. The first portion discusses methods for the learning phase, and the second section discusses leaf disease recognition and classification methodologies. The approach taken in constructing the leaf disease monitoring system is illustrated in figure 1. To start with, samples of the crop fields and the internet's leaves are collected. Following the collection of the samples, we developed a model for extracting characteristics

and developing a detection technique specifically geared toward plant (mungbean) disease. The design has been simulated by using a variety of classification techniques, and its performance has been evaluated until it was deemed satisfactory.

3.1 *Image Acquisition*

To start with, high-quality images are mandatory for effective disease identification. We use these high-quality pictures for training and evaluating our method to detect the most common illnesses. We trained and tested our classifier on a vast collection of high-quality photos to recognize and categorize diseases. Most of the photographs we collected came from the field, although some also came from the internet. Digital cameras were used to capture the imagery collection needed for the purpose of the work.

3.2 *Image Preprocessing*

To enhance the image's quality, preprocessing is applied to the collected images. So that it won't influence the outcome, it must remove the background that contains various types of noise. Additionally, we downsized the pictures while keeping the aspect ratio, which keeps the shape of the disease spot constant. When we crop an image, rather than processing the entire leaf, only a tiny section of it is processed using an image processing technique that takes a significantly shorter amount of time than processing the entire leaf. This results in a reduction in the amount of time required to process the image. Overall, these techniques help to improve the efficiency and accuracy of disease detection in plants.

3.3 *Feature extraction of color*

Color is one of those aspects of a picture that contributes many essential features to the whole. As a consequence of this, it is much simpler to classify different diseases using color characteristics. This is because, in general terms, every sickness has its own particular shade. To retrieve color features from a picture, the impacted region of the image is utilized to compute the mean value of the three fundamental RGB color bands, namely red, green, and blue. The following are the phases that are necessary for the separation of color features.

3.3.1 Masking of Image:

The process of masking an image involves selectively hiding or revealing certain portions of the image based on specific criteria. To ascertain the mean value of the affected diseased segment, it is unnecessary to take into account the healthy unaffected portion that refers to green pixels. The phenomenon of healthy green pixels being masked by black pixels can be attributed to the RGB color space value of [0, 0, 0] possessed by the latter, which has no effect on the calculation of the mean value. Figure 2. shows how masking work.

3.3.2 Separation of Color the bands:

Afterward the masking process, the resultant RGB image undergoes decomposition into its constituent primary colors, namely red, green, and blue. Here we have three images in the most fundamental three bands. The computation of the mean value of the three bands is straightforward, given that each band contains a single value per pixel.

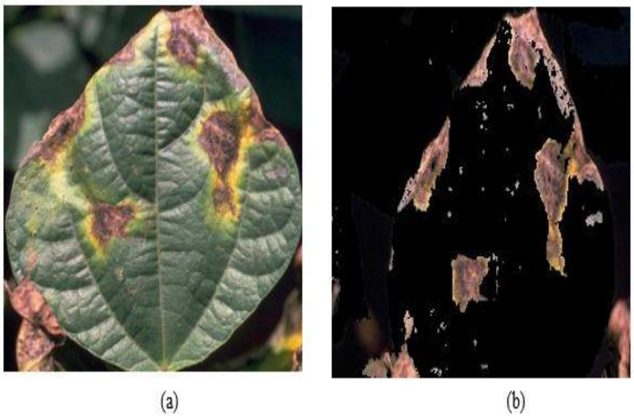
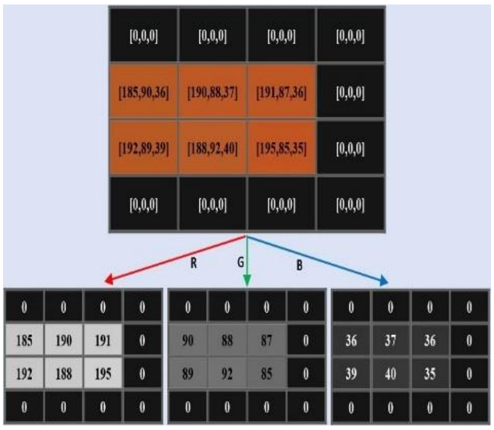


Figure 2. Pixel representation of (a) source image in RGB (b) masked image



[0,0,0]	[0,0,0]	[0,0,0]	[0,0,0]
[185,90,36]	[190,88,37]	[191,87,36]	[0,0,0]
[192,89,39]	[188,92,40]	[195,85,35]	[0,0,0]
[0,0,0]	[0,0,0]	[0,0,0]	[0,0,0]

0	0	0	0
185	190	191	0
192	188	195	0
0	0	0	0

0	0	0	0
90	88	87	0
89	92	85	0
0	0	0	0

0	0	0	0
36	37	36	0
39	40	35	0
0	0	0	0

3.3.3 Estimating the Mean Value:

Throughout this process stage, the mean value of each masked images of three principal color bands is determined and calculated. For the determination of the mean, the following equation is utilized:

$$\text{Mean } (\mu) = \frac{P}{N} \qquad (1)$$

Where P is the sum of all pixel values in the affected region and N is the overall number of affected pixels.

The equation can be rewritten as follows:

$$\mu = \frac{1}{n_d} \sum_{i=1}^{r} \sum_{j=1}^{c} p_{i,j} \qquad (2)$$

Where n_d represents the number of pixels affected by the disease, r represents the number of rows, c represents the number of columns, and $p_{i,j}$ represents the pixel value located at i×j position.

Subtracting the number of black pixels from the total number of pixels in the masked picture results in the total number of pixels in the infected region.

$$n_d = n - n_b \qquad (3)$$

The overall number of pixels in the masked image is denoted by n, while the number of black pixels is denoted by n_b .

To determine the total value of the pixels in the infected area, we added the values of each pixel in the masked image. All three color channels have a value of zero when black pixels are substituted for regular green pixels. The total number of pixels in the masked image equals the total number of pixels in the infected region. This is how the mathematical formula appears:

$$Sum = \sum_{i=1}^{r}\sum_{j=1}^{c} p_{i,j} = p_{1,1} + p_{1,2} + \cdots .. + p_{r,c} \quad (4)$$

With the help of Figure 3. the following formula can be used to determine the mean value for the first picture of the green color band.

$$n_d = n - n_b = 16 - 6 = 10$$

I.

$$Sum = \sum_{i=1}^{r}\sum_{j=1}^{c} p_{i,j} = p_{1,1} + p_{1,2} + \cdots + p_{44}$$

$$= 0 + 0 + +0 + 0 + 90 + 88 + 87 + 0 + 89 + 92 + 85 + 0 + 0 + 0 + 0 + 0 = 531$$

$$mean, \mu = \frac{1}{n_d}\sum_{i=1}^{r}\sum_{j=1}^{c} p_{i,j} = \frac{531}{10} = 53$$

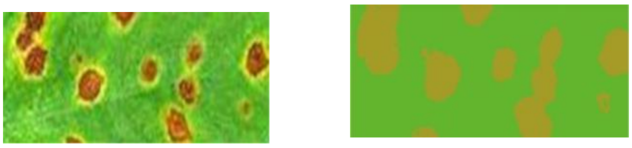
We obtained 3 feature values as a color feature from a single image after completing the calculation for each down sampled image from each color band. This feature values are required in next section for classification.

3.3.4 Extraction of Shape and Textural Features:

Disease identification requires considering additional characteristics like shape and texture, as leaf color alone is insufficient. The Sobel edge identification algorithm, combined with other tools, extracts shape and texture characteristics from images. The following procedures have to be followed in order to acquire these features.

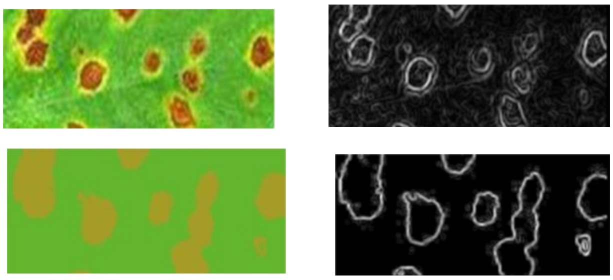
The Sobel edge detector mainly was employed in this paper to depict the edges of the infected region of a leaf. This edge detector must correctly recognize the boundaries from the diseased outer border to determine the shape and texture of a disease. However, under typical lighting conditions, a diseased area's center is deeper while its periphery is lighter. This indicates that the color of a diseased area fades from its darkest core to its lightest periphery. Therefore, it is difficult to detect edges from an infected region, as the Sobel edge detector requires rapid changes in color to identify edges. To fix this, images are segmented into color groups using k-means clustering [9][10], and their color values are rapidly modified. Following this approach, the Sobel edge detector will be able to recognize the infected boundaries as edges.

Figure 3: In this case, we have an input image (left) and a clustered output image (right).



The state both before and after the clustering is depicted in Figure 4. Also, before clustering, there are various noises in this leaf, such as veins, undesired colour and so on. However, after clustering, these are removed. As a result, it is evident that clustering is also required for removing various sorts of noise from a leaf and then we goes for edge detection.

Figure 4 : Edge detection from (a) unclustered and (b) clustered images



The Sobel edge detector cannot recognize edges properly from the affected area without clustering the image, as seen in Fig 5 (a). Edge detection of the non-clustered image is also affected by noise. The clustered image in Fig 5 (b) doesn't have this issue. After clustering, Sobel edge detector can efficiently identify the infected edges. This is why we undertake clustering as a preliminary process. Generally an affected leaf is divided into two parts: the green healthy section and the diseased portion. So, in most circumstances, two clusters (k=2) produce the best results.

3.3.4.1 Texture Feature extraction:

The distances between the various edges are determined at this stage. Simply put, the distances between white pixels are the necessary distances because we're treating white pixels as edges and black pixels as non-edges. The Manhattan distance [11] has been computed using pixels as our measuring unit. So, for example, two white pixels might be located at (xn, yn) and (xn+1, yn+1) and Manhattan can be defined as,

$$d = |x_{n+1} - x_n| + |y_{n+1} - y_n| \quad (5)$$

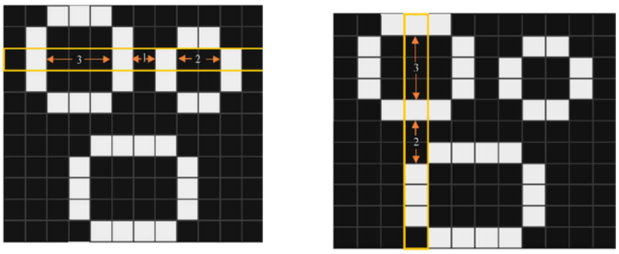
An image is a two-dimensional array of pixels, consisting of rows and columns. Thus, we have determined the distances between each row and column using two distinct techniques such as row scan and column scan. In an empty array, we recorded the amount of black pixels (distance) between each row and column, which we used to calculate the distance between pairs of white pixels.

Distances between adjacent columns (white pixels) have been measured, and their X-coordinates have been calculated in order to do a row scan. In this situation, all distances along the Y axis equal 0, hence the corresponding equation (5) is

$$d_{row} = |x_{n+1} - x_n| \quad (6)$$

The number of horizontal (from left to right) pixels in a row is denoted by the variable n and similarly, along with the Y plane we get the column scan.

Figure 5: One image has a row scan, whereas the other has a column scan



Using the above technique, we were able to independently calculate distances for each row and column in an image, ultimately arriving at the two arrays [23, 1, 11, 2, 17] and [12, 7, 20, 10, 3], respectively. The extracted texture features are the numbers in these two arrays.

3.3.4.2 Shape Feature extraction:

The two frequency arrays are used to extract shape features : one from row scanned distances and the other from column scanned distances. Positional values from the first array are divided by the corresponding value from the second array, resulting in a recursive process.

$$r = \frac{f_i}{S_i} \quad (7)$$

Here, fi and Si stand for the ith item in the first collection and second array, respectively.

The two sets of frequencies we measured are as follows:

X = [21, 3, 13, 6, 19]

Y = [15, 8, 18, 11, 5]

When equation (7) is applied, the following value is generated.

Z = [1.4, 0.375, 0.72222, 0.54545, 3.8]

The indexed of the array stands for the calculated value of the shape component.

3.3.5 Classification

This is where a classifier comes in, as it identifies and categorizes diseases. As a means of categorizing plant diseases, we opted for the use of bagging ensemble learning (Random Forest) for our research. Ensemble learning approaches have performed significantly better than single classifiers. By combining the results of multiple weak classifications, ensemble learning can produce predictions that are more accurate. Random Forest is a popular bagging-based ensemble learning algorithm that constructs multiple decision trees and aggregates their predictions to make the final prediction. This approach has been successfully applied in various fields, including image classification, natural language processing, and medical diagnosis. The process of classification can be broken down into two distinct phases: the first, known as the learning phase, and the second, known as the testing stage. During the learning phase, the classifier is taught by using feature values (color, shape and texture) derived from previously known objects. And in this section where we test it, we use the same kinds of feature values from an unknown object to figure out how to classify this one.

Feature values taken as a whole are known as feature vectors. Our classifier takes these feature vectors as input in order to recognize diseases. We used seventy-five percent of the images that we acquired from both the field and the internet to train the classifier, and we used the remaining twenty-five percent of the images for testing.

4 Result and Discussion

In this segment, we presented the results of our analysis of the accomplishments of various classifiers. Performance measures that are frequently used for any categorization problem include accuracy, precision, recall, F1 score. In table I. showed the accuracy rate for crops (mugbean) disease detection. We also work with rice and wheat plant but for result discussion mungbean diseased named Tan spot, Frogeye, Powdry Mildew and Yellow Mosaic is showed.

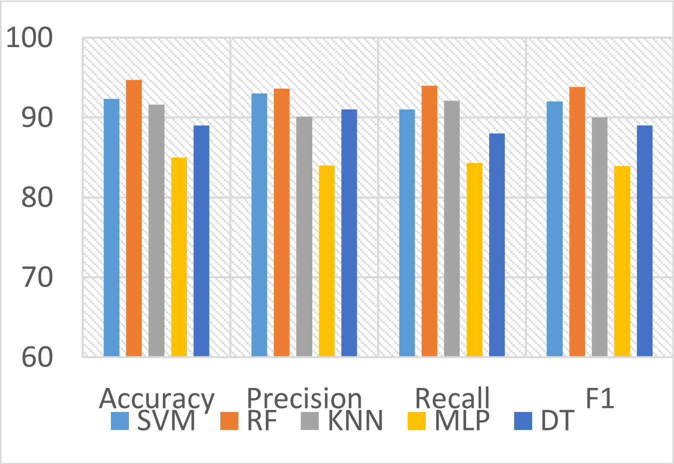
Table 1. ACCURACY OF CLASSIFICATION

Diseased / Healthy Leaf	**Total Test Sample**	**Correctly Classified**	**Accuracy**
Tan Spot	80	76	95%
Frogeye	65	61	93%
Powdry Mildew	54	49	92%
Yellow Mosaic	100	94	94%
Healthy leaf	50	48	96%
Combined	349	328	94%

Table 2: CLASSIFICATION REPORTS OF DIFFERENT CLASSIFIERS

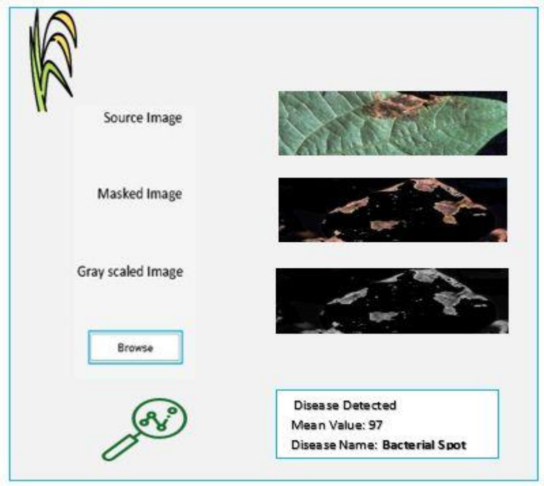
Classifiers	**Precision**	**Recall**	**F1 Score**	**Accuracy (%)**
KNN	0.901	0.907	0.915	91.6
SVM	0.917	0.912	0.927	92.32
Decision Tree	0.897	0.899	0.901	90.21
MLP	0.843	0.848	0..853	85..47
Random Forest	0.945	0.939	0.945	94.61

Figure 6: Illustration of different classifier performance.



The random forest ensemble-based model is chosen as the final model for identifying crop disease following the above outcomes. A software prototype given bellow can identify diseases from a input image also the affected area of leaf.

Figure 7: Mungbean Bacterial Spot is identified from the affected portion of the sample.



Disease Detected
Mean Value: 97
Disease Name: Bacterial Spot

5 Conclusion

Image processing and machine learning face significant difficulties in the area of plant disease identification. The main contribution of this research is the development of an ensemble classification strategy that makes use of multiple feature extraction algorithms. An effective ensemble classifier has been made with the help of SVM, MLP, KNN, and decision trees, among others. The ensemble classification that uses various features has been implemented, and their effectiveness has been evaluated. It has come to our attention that our ensemble classifier (Random Forest) produced the highest levels of accuracy, precision, and recall when combined with the suggested work. The experiments used three disease groups from rice, four from wheat, and four from mungbean. Furthermore, the research findings provide valuable insights into developing targeted and sustainable strategies to combat diseases in these crops, ultimately improving their overall productivity and sustainability.

6 Acknowledgements

We express our gratefulness to ICT Division Bangladesh for providing us with ICT fellowship grants for this research work for the financial year 2022-2023. This thesis has been funded by the ICT Division Bangladesh (Government of the People's Republic of Bangladesh) so we are thankful to ICT Division Bangladesh.

References

[1] M. R. H. N. M. P. R. P. B. N. S. N. P. V. V. Shima Ramesh, "Plant Disease Detection Using Machine Learning," in *International Conference on Design Innovations for 3Cs Compute Communicate Control*, 2018.

[2] M. S. S. M. S. K. M. D. B. Ms. Sandhya R. Shinde, "Experiments on Content Based Image Classification using Color Feature Extraction," in *International Conference on Communication, Information & Computing Technology (ICCICT)*, 2015.

[3] S.Phadikar and J. Goswami, "Vegetation indices based segmentation for automatic classification of brown spot and blast diseases of rice," in 3rd International Conference on Recent Advances in Information Technology (RAIT), 2016.

[4] R. D. V. K. V Pooja, "Identification of plant leaf diseases using image processing techniques," in *IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR)*, 2017

[5] V. P. G. a. D. V. Musande, "Wheat Disease Detection Using Image Processing," in *1st International Conference on Intelligent Systems and Information Management (ICISIM)*, 2017.

[6] T. Ren, Y. Zhang and C. Wang, "Identification of Corn Leaf Disease Based on Image Processing," in *2nd International Conference on Information Systems and Computer Aided Education (ICISCAE)*, 2019.

[7] R. B. R. Amit Kumar Singh, "CLASSIFICATION OF RICE DISEASE USING DIGITAL IMAGE PROCESSING AND SVM CLASSIFIER," *International Journal of Electrical and Electronics Engineers*, vol. 07, no. 01, pp. 294-299, 2015.

[8] G. Anthonys and N. Wickramarachchi, "An image recognition system for crop disease identification of paddy fields in Sri Lanka," in *International Conference on Industrial and Information Systems (ICIIS)*, 2009.

[9] W. Gao, X. Zhang, L. Yang and H. Liu, "An improved Sobel edge detection," in *3rd International Conference on Computer Science and Information Technology*, 2010.

[10] . Kanopoulos, N. Vasanthavada and R. Baker, "Design of an image edge detection filter using the Sobel operator," *IEEE Journal of Solid-State Circuits* , vol. 23, no. 2, pp. 358 - 367, 1988.

[11] A. H. a. D. A. K. Charu C. Aggarwal, "On the Surprising Behavior of Distance Metrics in High Dimensional Space," in *International Conference on Database Theory*, 2001.

[12] L. J. A. K. R. a. P. K. S. anti Kumari Behera, "Disease Classification and Grading of Orange using Machine Learning and Fuzzy Logic," in International Conferance on Communication and Signal Processing, 2018.

[13] H. Manisha Bhangea, "Smart Farming: Pomegranate Disease Detection Using Image Processing," *Second International Symposium on Computer Vision and the Internet (VisionNet'15)*, vol. 58, p. 280 – 288, 2015.

[14] K. D. M.P.Vaishnnave, "Detection and Classification of Groundnut Leaf Diseases using KNN classifier," 2019.