

Research Article

Performance of Machine Learning and Image Processing in Plant Leaf Disease Detection

Abu Sarwar Zamani ¹, **L. Anand** ², **Kantilal Pitambar Rane** ³, **P. Prabhu** ⁴,
Ahmed Mateen Buttar ⁵, **Harikumar Pallathadka** ⁶, **Abhishek Raghuvanshi** ⁷,
and Betty Nokobi Dugbakie ⁸

¹Department of Computer and Self Development, Preparatory Year Deanship, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia

²Department of Networking and Communications, SRM Institute of Science and Technology, Chennai, India

³Department of Electronics and Communication, Koneru Lakshmaiah Education Foundation, Guntur, Andra Pradesh, India

⁴Alagappa University, Karaikudi. 630003, Tamilnadu, India

⁵Department of Computer Science, University of Agriculture Faisalabad, Faisalabad, Pakistan

⁶Manipal International University, Manipal, India

⁷Mahakal Institute of Technology, Ujjain, India

⁸Department of Chemical Engineering, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

Correspondence should be addressed to Betty Nokobi Dugbakie; bdnokobi@st.knust.edu.gh

Received 14 February 2022; Accepted 31 March 2022; Published 26 April 2022

Academic Editor: Rijwan Khan

Copyright © 2022 Abu Sarwar Zamani et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The aim of this study is to evaluate infected leaf disease images. Precision agriculture's automatic leaf disease detection system employs image acquisition, image processing, image segmentation, feature extraction, and machine learning techniques. An automated disease detection system offers the farmer with a fast and accurate diagnosis of the plant disease. Automation of plant leaf disease detection system is essential for accelerating crop diagnosis. Using machine learning and image processing, this paper describes a framework for detecting leaf illness. An image of a leaf can be used as an input for this framework. To begin, leaf photographs are preprocessed in order to remove noise from their images. The mean filter is used to filter out background noise. Histogram equalization is used to enhance the quality of the image. The division of a single image into multiple portions or segments is referred to as segmentation in photography. It assists in establishing the boundaries of the image. Segmenting the image is accomplished using the K-Means approach. Feature extraction is carried by using the principal component analysis. Following that, images are categorized using techniques such as RBF-SVM, SVM, random forest, and ID3.

1. Introduction

Due to dwindling natural resources, one of the biggest concerns in agriculture is that crop yields would not be able to keep pace with the growing global population. Increased productivity, regardless of unfavorable environmental factors, is the key problem here. Modern precision agriculture leverages the most modern advances in agricultural technology to improve productivity. Precision agriculture's automatic leaf disease detection system employs image

acquisition, image processing, image segmentation, feature extraction, and machine learning techniques. An automated disease detection system provides the farmer with an immediate and accurate diagnosis of the plant disease, speeding up the diagnostic process. Automation of the disease detection system is critical for expediting crop diagnosis [1, 2].

Image processing is a collection of tools and techniques to remove noise from images and improve their quality. The field of image processing is one that is expanding at a quick pace these days. Enhancement, segmentation, feature

extraction, classification, and other techniques used in image processing are all examples of image processing. The process of improving an image involves making adjustments to its brightness, color temperature, noise reduction, and sharpness [3].

Splitting an image into smaller, more manageable chunks is what image segmentation entails. In most cases, this technique is employed to recognize objects in digital photographs. Image segmentation can be done in a variety of ways, including thresholding, color based, transform, and texture-based methods. A form of dimensionality reduction known as “feature extraction” reduces the number of pixels in an image by extracting just the most important and visually appealing elements. Image matching and retrieval can be expedited by using a reduced feature representation and a high image size with this strategy. The labeling of photographs into one of a number of specified categories is known as “image classification.” supervised and unsupervised are the two subcategories in the classification [4, 5].

Agriculture image processing is a core application of image processing and the fastest-growing study topic in the field. A wide range of industries, including agriculture, have found that image processing can be a useful tool for data analysis. Photographs are taken with cameras, planes, or satellites, and then processed. Computers use image processing algorithms to process and analyze these images. Solving a wide range of issues in agriculture has never been easier, thanks to the recent developments in picture capture and data processing technologies. Images can be utilized in agricultural applications to extract sick leaves, stems, and fruits; quantify the affected area by disease; and determine the illness’s color, shape, and size [6, 7].

With the help of artificial intelligence and image processing, this paper proposes a system for identifying leaf disease on trees. As an input, this framework accepts a picture of a leaf. To begin, leaf photographs are preprocessed in order to remove background noise. The mean filter is used to filter out background noise. Histogram equalization is used to enhance the quality of the image. Division of a single image into multiple portions or segments is referred to as segmentation in photography. It assists in establishing the boundaries of the image. Segmenting the image is accomplished using the K-Means approach. Feature extraction is carried by using the principal component analysis. Following that, images are categorized using techniques such as RBF-SVM, SVM, random forest, and ID3.

2. Literature Survey

Plant and fruit diseases can be identified and classified using a variety of methods. Anthracnose and Downy mildew, watermelon leaf diseases, were classified by Suhaili Kutty et al. [8]. In order to do this, the region of interest must be identified using RGB color components in an infected leaf sample. Authors have used mean filters to remove noise from the input data.

Scab, apple rot, and apple blotch are among the many diseases of apples that Dubey and R. Jalal [9] investigated. In this scenario, K-means clustering is utilized to segment the

data. The extracted characteristics are then applied to the segmented image. Classification is accomplished via the usage of the multiclass support vector machine (SVM).

Image processing and artificial intelligence are being used by Sanjiv Sannakki et al. [10] in an effort to diagnose the condition. Downy mildew and powdery mildew of the grape leaf are the two most common causes of this disease. Using masking, you may remove the backdrop and achieve more precise results. This information is maintained via anisotropic diffusion in the damaged leaf region. A technique known as K-means clustering is used to separate the data into manageable chunks. In order to complete the feature extraction, the gray level co-occurrence matrix must be calculated. Feed forward back propagation networks are used to classify the data. To obtain a more realistic result, they simply employed the Hue option.

For illness identification and fruit grading, Monika Jhuria et al. [11] used an image processing approach. Disease classification has been aided using an artificial neural network (ANN). Color, texture, and morphology are all factors that they take into account. Morphological features are the best of the bunch. Apple scab and rot can be detected in grapes, as well as black rot and powdery mildew in grapes. Fruit grading is done using two methods: the spread of disease and an automatic weight computation.

Sachin Khirade [12] described how image processing methods may be used to diagnose and categorize plant diseases. Images are collected, preprocessed, segmented, and features are extracted before being classified. Segmentation approaches include Otsu’s approach, which involves transforming RGB photos into HIS models, and K-means clustering. Of all the algorithms, K-means clustering produces the most exact results possible. This is followed by the extraction of attributes such as color and texture as well as morphology, edges, and more. Motif extraction is a better choice than the other methods available. Features are classified using an artificial neural network (ANN) and back propagation neural network (BPN).

Computer vision and image processing techniques were used by Kaiyi Wang et al. [13] to develop a new method for diagnosing vegetable diseases and insect pests. Images collected by smartphones are used in the study of vegetable disease and insect pest status. To identify leaves in these photos, we employed a brand-new extraction and classification technique. Then, a region-labeling technique was used to determine the number of insects and sick areas in the images segmented. For the separation of the objects, a mathematical morphology technique was utilized to deal with the areas of adhesion. The proposed strategy was tested in the field using mobile smart devices. A high level of efficiency and accuracy were found in the experimental results.

According to Dipali Majumder et al. [14], BTH (benzothiadiazole) provided systemic protection for wheat against powdery mildew infection by interfering with numerous stages of the pathogen’s life cycle. For sturdiness, we use the support vector machine (SVM) machine learning technology. Information on wheat plants and disease preventive strategies is the primary topic of this article. The support vector machine (SVM) can be used to diagnose and

treat any illness that may be present in wheat leaves. Support vector machine delivers a wealth of information that makes it simple to identify and complete the procedure early on. There is also a comparison of the various leaf disease detection methods.

Rong et al. [15] were able to identify early cercospora leaf spot in sugar beet by combining template matching and support vector machine approaches (SVM). To ensure accuracy, they used a three-stage methodology. Plant disease may be detected and qualified on-site using continuous quantification under daylight conditions.

It was proposed by Revathi et al. [16] to use fuzzy curves and fuzzy surfaces to pick image features for cotton leaves disease diagnosis (FS). This inquiry is divided into two phases. The extraction of a small selection of relevant characteristics from a large number of original features is automated and quick. Fuzzy curves are a method used to eliminate irrelevant information. Another technique is to isolate just the most significant components of a given characteristic using fuzzy surfaces. To make the feature space smaller, an approach like this may be used in practical classification applications.

To identify and classify diseases, a neural network-based effort was done by Sanjeev S Sannakki [17]. Under the umbrella of an intelligent system, the author outlined a diagnostic strategy for isolating the ailment. Grapes have been a focus of the author's research. It is proposed that this system be broken down into two distinct phases. The object is identified from the image in the first stage. In order to carry out this object detection, the segmentation method is defined. In the second stage, the image masking might be done under the prediction of disease. The author uses the K-means clustering approach for disease identification and classification.

A family of statistical learning algorithms based on biological neural networks are known as artificial neural networks (ANNs). A neural network is a synthetic network of neurons that may be used to recognize patterns. The way neural networks learn is by iteratively rearranging the weights of their connections. The accurate estimation of functions that depend on several unknown variables can be achieved using this method. Interconnected "neurons" in an artificial neural network may compute input/output values and perform machine learning and pattern recognition [18].

In the K-nearest neighbor classifier (k-NN) nearest neighbor classification, a test tuple and training tuples that are comparable are used to make comparisons. This is an n-dimensional tuple that represents a single point in that space. An n-dimensional pattern space is used to hold all training tuples. Simply by finding the nearest neighbor in tuple space, the classifier is able to categorize the unknown tuple as belonging to the same class as its known neighbor. Pattern space is searched by the k-nearest neighbor classifier for the nearest training tuple to the unknown tuple. The unknown tuple's k-nearest neighbor classifier is constructed using these training tuples. Any distance metric, such as Euclidean distance, can be used to measure closeness. Classifiers based on distance comparisons that allocate equal weight to all attributes are known as nearest neighbor

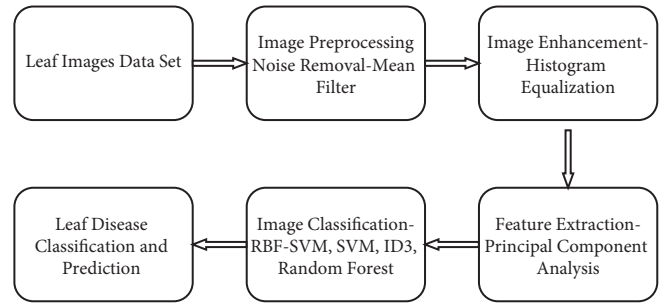


FIGURE 1: Block diagram of leaf disease detection by image processing and machine learning.

classifiers. If there is a lot of noise or an unimportant attribute, they may have a lower level of accuracy.

Support vector machines (SVMs) are a new kind of statistical learning algorithm based on modern statistical learning theories (SLT). This method works for both linear and non-linear data. Data are transformed from its original form into a higher dimension using support vectors, important training tuples, in order to locate a hyperplane for the separation of the data. Support vector machines may be formalized by separating hyperplanes (SVMs). Either an alternative, training may be referred to as a single hyperplane or a set of hyperplanes generated by a support vector machine may represent an infinitely complex space. Any class of hyperplane with a so-called functional margin that has the furthest distance to the closest training data point, intuitively, obtains a fair separation. Because a classifier's generalization error decreases as the size of the margin increases, this is an important consideration.

A simple Bayesian classifier assumes class-independence. As a result, the effect of a specific attribute value on a given class is independent of the other attribute values. In order to save money, this assumption is made, and it is deemed naive. For huge datasets, a naive Bayesian model is ideal because it does not require complex iterative parameter estimates.

An ensemble of decision-tree-based classifiers, the random forest is an example. Each tree is built using a bootstrap sample of the data and a candidate set of characteristics chosen at random. Trees are built using both bagging and random selection. Class predictions are made by the trees when a forest is developed. Because of the strong association between any two trees in a random forest, its error rate is highly variable. Regression and classification problems can be ranked in a natural way using this method [19].

3. Methodology

This section contains a machine learning and image processing framework for leaf disease detection. In this framework, a leaf image is used as the input. First of all, leaf images are preprocessed to remove noise. Noise removal is performed using the mean filter. Image enhancement is achieved by histogram equalization. Image segmentation

divides a single image into multiple parts or segments. It helps in identification of image boundaries. Image segmentation is achieved by the K-Means algorithm. Feature extraction is performed by principal component analysis. Then, image classification is performed by RBF-SVM, SVM, random forest and ID3 algorithms. The block diagram is shown in Figure 1.

There is a great deal of reliance on the adaptive median filter (AMF) algorithms [20] for the removal of unwanted noise from images. Spatial processing like this is used by the AMF method to identify which pixels in a picture are affected by impulse noise. When a high number of pixels are not spatially aligned, it is called “impulse noise.” Thus, noise pixels are masked by utilizing the median value of pixels in their immediate vicinity that have been labeled as being free of the noise.

To improve contrast, the histogram equalization communicates pixel intensity values to create a consistent intensity distribution and a continuous histogram in the output image. In situations when the picture’s practical data are characterized by very high contrast values, this approach is used often to boost the overall contrast of the image. An equal distribution of intensities may be achieved by using this technique [21]. The upshot of this is that certain sections of the image may benefit from a boost in contrast. Using histogram equalization, the most common intensity values are distributed more evenly over the whole histogram.

K-means clustering assigns each observation to a cluster according to the local mean, enabling the formation of a pattern of groupings. When looking for clusters, this approach uses the entire number of groups provided by k to find them. Squared distances are used to identify the most important data points. According to the stated qualities, each data point is allocated one of the k groups and analyzed. Grouping data points based on feature similarity is common [22].

Data parameters such as time (the length of the connection) and SRC bytes (the size of the data) are standardized using z-score normalization. In this article, the principle component analysis (PCA) method is used to extract feature information. PCA’s linear method to dimensionality reduction may help in data analysis and compression [23]. According to this method, it is possible to combine a large number of uncorrelated traits by finding orthogonal linear combinations of the original characteristics.

Support vector machines (SVMs) are a new type of statistical learning algorithm that is based on new ideas about how to learn (SVM). This method works with both linear and non-linear data, so it can be used with both types of data. People use support vectors, important training pairs, to move data into a higher dimension so that they can find a hyperplane where the data can be separated. Support vector machines can be formalized by separating hyper planes from each other (SVMs). Training, on the other hand, could be called that. An infinitely complex space can be shown by a single line or a group of lines made by a support vector machine. This means that any class of hyperplane with a “functional margin” that has the furthest distance from the

closest training data point gets a fair amount of space. Because a classifier’s generalization error decreases as the size of the margin grows, this is an important thing to think about. SVM performs better in the RBF kernel mode. Radial basis function is best suited for SVM.

Random forest is an example of a group of decision-tree-based classifiers. Each tree is made with a bootstrap sample of the data and a set of characteristics that are chosen at random for each one. There are two ways that trees are made: bagging and random selection. Class predictions are made by the trees when a forest grows up, and they do this because there is a strong connection between any two trees in a random forest, its error rate can be very high or very low at any given time. If you use this method, you can rank regression and classification problems in a natural way [19].

A decision-tree-based approach known as ID3 was the first to evolve. Entropy and information gain metrics are the foundation of this strategy. There is an initial nodule, and each subsequent iteration computes the entropy of the functional features. Datasets are divided into subsets depending on the characteristic with the lowest error rate (entropy) and the largest information gain, and these subsets are referred to as split attributes. It is recursively performed on all subsets of data if the procedure is not properly classified to its target classes. The branch’s final subset defines the terminal nodes of the decision tree, which is formed using a nonterminal node. The split property specifies the nonterminal node, while the terminal node represents the class labels.

4. Result Analysis

The rice data collection [24] comprises of three illness categories that have been identified. Leaf Smut, Leaf Blight, and Brown Spot are the three classifications. There are 40 photos in each disease group. It is possible to find a total of 120 photos in the data collection. 96 images were used for training of machine learning algorithms and remaining 24 images were used for the testing of the machine learning algorithms. The images are preprocessed with the use of a mean filter and a histogram. The K-means technique is used to segment the data into several groups. PCA is used to extract the features from the data. Image classification techniques such as RBF-SVM, SVM, random forest, and ID3 are used in the following phase to classify photos based on their contents. Performance of several algorithms is evaluated using three parameters: accuracy, sensitivity, and specificity, which are all measured in this study.

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$$

$$\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{Specificity} = \text{TN} / (\text{TN} + \text{FP})$$

$$\text{TP} = \text{True Positive}$$

$$\text{TN} = \text{True Negative}$$

$$\text{FP} = \text{False Positive}$$

$$\text{FN} = \text{False Negative}$$

Performance comparison of different algorithms is shown in Figure 2, Figure 3, Figure 4, and Figure 5.

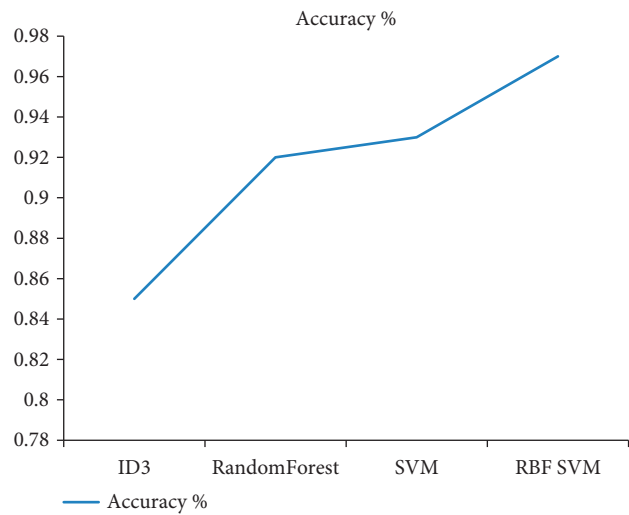


FIGURE 2: Accuracy results of machine learning algorithms.

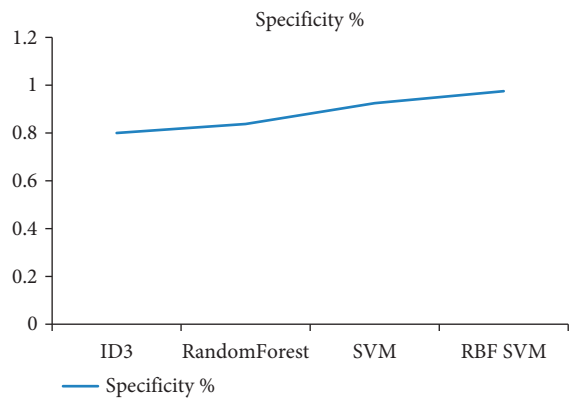


FIGURE 3: Specificity results of machine learning algorithms.

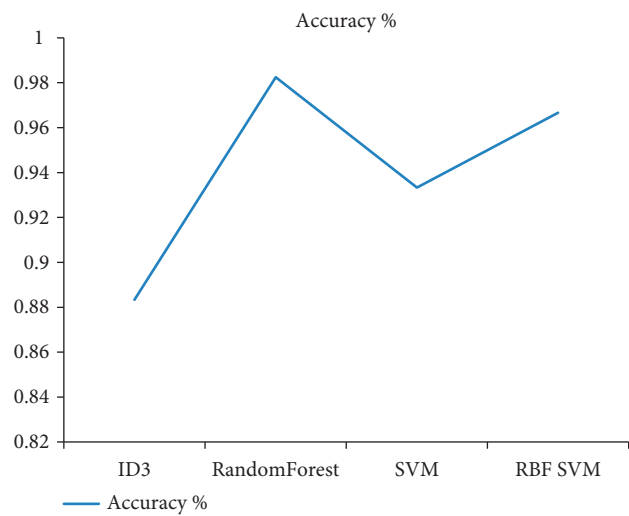


FIGURE 4: Sensitivity results of Machine Learning algorithms.

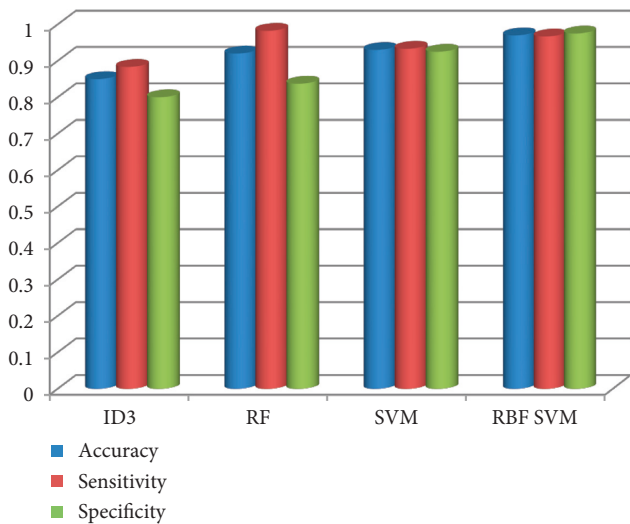


FIGURE 5: Performance of machine learning techniques for plant leaf disease detection.

5. Conclusion

An automated disease detection system gives the farmer a quick and accurate diagnosis of the plant disease, allowing the diagnostic process to be sped up, so the farmer can get more crops out of his fields. As a result, it is very important to make the disease detection system automated in order to speed up crop diagnosis. This paper talks about how to use machine learning and image processing to figure out if leaves are sick. As a starting point, this framework can be used with a picture of a leaf. To start with, leaf photos are cleaned up to remove any noise from them. In order to get rid of noise, the mean filter is used. Segmentation is the act of breaking up a single picture into parts or segments. It can help you figure out how big the picture is. The K-means algorithm is used to divide the image into parts. The principal component analysis is used to find features. In the next step, images are classified based on their content with help from algorithms like RBF-SVM, SVM, random forest, and ID3. RBF-SVM performs better in accurate leaf disease detection.

Data Availability

The data used to support the findings of this study are available upon request from the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] K. N. Bhanu, H. J. Jasmine, and H. S. Mahadevaswamy, "Machine learning implementation in IoT based intelligent system for agriculture," *International Conference for Emerging Technology (INCET)*, pp. 1–5, 2020.
- [2] A. Sharma, A. Jain, P. Gupta, and V. Chowdary, "Machine learning applications for precision agriculture: a comprehensive review," *IEEE Access*, vol. 9, pp. 4843–4873, 2021.
- [3] A. Muniasamy, "Machine learning for smart farming: a focus on desert agriculture," *2020 International Conference on Computing and Information Technology (ICCIT-1441)*, pp. 1–5, 2020.
- [4] S. Minaei, M. Jafari, and N. Safaie, "Design and development of a rose plant disease-detection and site-specific spraying system based on a combination of infrared and visible images," *Journal of Agricultural Science and Technology A*, vol. 20, no. 1, pp. 23–36, 2018.
- [5] A. Raghuvanshi, U. K. Singh, G. S. Sajja et al., "Intrusion detection using machine learning for risk mitigation in IoT-enabled smart irrigation in smart farming," *Journal of Food Quality*, vol. 2022, pp. 1–8, 2022.
- [6] V. Hemamalini, S. Rajarajeswari, S. Nachiyappan et al., "Food quality inspection and grading using efficient image segmentation and machine learning-based system," *Journal of Food Quality*, vol. 2022, pp. 1–6, 2022.
- [7] A. Raghuvanshi, U. K. Singh, and C. Joshi, "A review of various security and privacy innovations for IoT applications in healthcare," *Advanced Healthcare Systems*, vol. 4, pp. 43–58, 2022.
- [8] S. B. Kutty, N. E. Abdullah, D. H. Hashim et al., "Classification of Watermelon Leaf Diseases Using Neural Network Analysis," *IEEE, Business Engineering and Industrial Applications Colloquium (BEIAC)*, pp. 459–464, 2013.
- [9] S. R. Dubey and A. Singh Jalal, "Detection and Classification of Apple Fruit Diseases Using Complete Local Binary Patterns," *IEEE Computer and Communication Technology (ICCCCT)*, pp. 346–351, 2012.
- [10] S. S. Sannaki, V. S. Rajpurohit, V. B. Nargund, and Pallavi-Kulkarni, "Diagnosis and Classification of Grape Leaf Diseases Using Neural Network," *IEEE, Tiruchengode*, p. 1–5, July 2013.
- [11] M. Jhuria, A. Kumar, and R. Borse, "Image processing for smart farming: detection of disease and fruit grading," *IEEE, International Conference on Image Processing, Shimla*, pp. 521–526, 2013.
- [12] S. D. Khirade and A. B. Patil, "Plant disease detection using image processing," *IEEE, International Conference on Computing Communication Control and Automation, Pune, Feb*, pp. 768–771, 2015.
- [13] K. Wang, S. Zhang, Z. Wang, L. Z. Liu, and F. Yang, "Mobile smart device-based vegetable disease and insect pest recognition method," *Intelligent Automation & Soft Computing*, vol. 19, no. 3, pp. 263–273, 2013.
- [14] D. Majumder, T. Rajesh, E. G. Suting, and A. Debbarma, "Detection of seed borne pathogens in wheat: recent trends," *Meghalaya, AJCS*, vol. 7, no. Issue 4, pp. 500–507, 2013.
- [15] R. Zhou, S. Kaneko, F. Tanaka, M. Kayamori, and M. Shimizu, "Early Detection and Continuous Quantization of Plant Disease Using Template Matching and Support Vector Machine Algorithms," *IEEE, Sapporo*, pp. 300–304, 2013.
- [16] P. Revathi and M. Hemalatha, "Cotton leaf spot diseases detection utilizing feature selection with skew divergence method," *International Journal of Scientific Engineering and Technology*, vol. 3, no. 1, pp. 22–30, 2014.
- [17] S. S. Sanjeev, "Leaf disease grading by fuzzy logic & machine vision," *International Journal of Control Theory and Applications*, vol. 2, no. Issue.5, pp. 1709–1716, 2018.
- [18] A. Ingole, "Detection and classification of leaf disease using artificial neural," *Network International Journal of Technical Research and Applications*, vol. 3, no. Issue 3, pp. 331–333, 2015.
- [19] B. Sandika, S. Avil, S. Sanat, and P. Srinivasu, "Random forest based classification of diseases in grapes from images captured

- in uncontrolled environments,” *IEEE International Conference on Signal Processing (ICSP)*, pp. 1775–1780, 2016.
- [20] M. Rakhra, R. Singh, T. K. Lohani, and M. Shabaz, “Meta-heuristic and machine learning-based smart engine for renting and sharing of agriculture equipment,” *Metaheuristic Problems in Engineering*, vol. 2021, pp. 1–13, 2021.
- [21] W. Zhao and J. Wang, “A new method of the forest dynamic inspection color image sharpening process,” *International Conference on Advanced Computer Theory and Engineering (ICACTE)*, 2010.
- [22] M. N. Reza, I. S. Na, S. W. Baek, and K.-H. Lee, “Rice yield estimation based on K-means clustering with graph-cut segmentation using low-altitude UAV images,” *Biosystems Engineering*, vol. 177, pp. 109–121, 2019.
- [23] C. Dou, L. Zheng, W. Wang, and M. Shabaz, “Evaluation of urban environmental and economic coordination based on discrete mathematical model,” in *Mathematical Problems in Engineering*, Dr. D. Singh, Ed., vol. 2021, pp. 1–11, 2021.
- [24] <https://archive.ics.uci.edu/ml/datasets/Rice+Leaf+Diseases>.