SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V11I11P109

Volume 11 Issue 11, 95-109, November 2024

Original Article

Advanced Machine Learning Approach for Medicinal Plant Leaf Disease Detection: Combining Modified Sigmoid-Hyperbolic Functions with Logistic Regression

A. Rajeshkanna[1], Thalari Chandrasekhar[2], N. Subramanyan[3], S. Kiran[4]

[1,3]*Department of Computer Science, Sri S.R.N.M College, Tamil Nadu, India.*
[2]*Department of Electronics, Government Science College, Karnataka, India.*
[4]*Department of Computer Science and Engineering, YSREC of YVU Proddatur, Andhra Pradesh, India.*

[2]*Corresponding Author : tcsekhar2011@gmail.com*

Received: 10 September 2024 Revised: 09 October 2024 Accepted: 08 November 2024 Published: 03 December 2024

***Abstract* - *Detecting diseases in the medicinal plant leaves, especially on pepper and citrus, is very important in maintaining plant health and supporting food production. This research work, therefore, suggests using a more sophisticated machine learning algorithm that can diagnose and categorize the diseases affecting the leaves of the aforementioned plants with a high-efficiency level, especially for pepper and citrus crops. The method involves the addition of sigmoid, and hyperbolic sine functions with the efficiency of logistic regression to improve the accuracy of diagnosing disease. When combined with the hyperbolic sine function, the sigmoid function promotes increased activation flexibility in capturing disease patterns for interpretation by the model. The logistic regression model is adopted as the principal classification algorithm; it outperforms other classifiers in binary classification problems. Hyperparameter tuning optimizes the model's performance in terms of disease prediction, adding further layers of complexity and shouldering the added responsibility of good accuracy. The proposed approach is further validated through experiments conducted on a set of pepper and citrus leaf images, which show that the current approach yields better accuracy, sensitivity, and specificity than typical ML algorithms. The results evince the effectiveness of this hybrid approach in enhancing early disease diagnosis of medicinal plants that, in turn, can improve disease control of plants.*

***Keywords* - *Agriculture productivity, Disease detection, Hyperbolic sine sigmoid function, Hyperparameter tuning, Logistic regression.*

1. Introduction

Better and faster disease detection systems are revolutionising agriculture through automated systems. These systems utilize advanced imaging and analysis integrated with data to identify diseases before there are recognizably visible symptoms to the human eye. It allows farmers to act quickly and apply treatments exactly where they need to, meaning they no longer need to spray as far as their fields. For this reason, automated disease detection increases crop yields, lowers costs, and helps with more sustainable, environmentally friendly farming practices while conserving resources and protecting biodiversity. The ethical issues around automated disease detection systems in agriculture with particular attention on the role of such systems in promoting sustainability and support to farmers. These systems enable early, accurate detection of broad pests, decreasing reliance on generalized pesticide use, making crops healthier and lowering the environmental impact. Through their work, they help farmers make data-driven decisions, conserve resources, cut farming costs, and improve agriculture as a means of production. However, these tools can be affordable to enhance productivity and improve food security for smallholder farmers, thereby improving economic equity among people within agricultural communities. Therefore, the ultimate idea of such technology is to deliver a responsible but effective way of farming for man and the environment.

Citrus and pepper are known to have medical properties that play vital roles in the existing traditional and modern medicine. The leaf diseases of citrus and pepper plants can frequently impair crop yield and quality. This is why it becomes extremely important to diagnose these diseases at a very early stage when they are still manageable so that they do not have a devastating effect on the large-scale farming industries. Conventional methods of diagnosis of leaf diseases depend on visual assessments, which are cumbersome, tiresomely involving and inaccurate because of the human factor involved [1]. That is why it is necessary to develop better, more automatized methodologies of disease identification [2].

1.1. Healthy Versus Diseased Leaves

The condition is when the leaves of the pepper and citrus plants are green, fresh, smooth surfaces, and well-shaped, showing the plant's health status. Figure 1 shows the healthy and diseased leaves of Pepper and Citrus. In pepper plants, the healthy leaves should have vibrant green colors and a smooth texture of the leaves and should be rigid and firm to the touch and have no signs of developing blotches or any other colour. These are elliptical with pointed apices and without any abrasion to their margins. They always produce new leaves and retain mature ones. Conversely, the healthy leaves of the citrus plant have dark green colouration, which is evidence of adequate nutrient uptake; the leaves are somewhat leathery and have a glossy surface [3]. They are thick, tapered at the end, and of equal size, avoiding any curvature hindering the plant's growth or affecting fruit formation.

On the other hand, the diseased leaves of the pepper and the citrus are shown to have one or more stress and or damage indicators. In pepper leaves, there are signs of disease that cut across discolouration, which may include yellowing, browning or black spot-like lesions having changes in texture and may be rough, soft or brittle. This may culminate in an irregular shape of these leaves; there could be signs such as curling, wilting or twisting of the leaves. Stern or severe infections can bring about earlier abscission of the leaves, whereby the affected plant may develop very little foliage and grow in an erect manner. Consequently, the citrus leaves attacked by diseases may turn yellow or have brown or black spots or may have raised lesions depending on the disease kind. It can also change the texture of these leaves and make them rough, leathery or thickened and in shape, they may curl or crumple [4]. These leaves may fall prematurely depending on the severity of the condition, leading to low foliage cover, slow growth and low fruit yield in affected trees.

A few diseases include Bacterial Leaf Spot, Powdery Mildew, Anthracnose, and Phytophthora Blight, all of which form symptoms including water-soaked spots, white powdery patches, dark lesions and wilting. These diseases cause the dropping of the leaves, reduced vigour and the general ill health of the affected plants. Likewise, some diseases threaten citrus plants, such as Citrus Canker, Greasy Spot, Citrus Black Spot and Huanglongbing (Citrus Greening). All these diseases cause corky lesions, oily spots, sunken spots and mottled yellowing on fruits, leading to defoliation, twig dieback and poor fruit quality.

In recent years, Machine Learning (ML) has proven to be very powerful for identifying plant diseases. Diagnosing disease from leaf images can be done very accurately with machine learning, especially with those using Deep Learning (DL). Nevertheless, deep learning models create an enormous demand on computational resources and are often based on large, labeled datasets, which can not always be obtained. Further, they are more complex than classical methods, and the model generated from their application might be difficult to comprehend in real-life scenarios [5, 6].

To overcome these challenges, this research puts forward a novel machine learning technique suitable for detecting diseases in citrus and pepper medicinal plant leaves. The approach utilized a sigmoid expanded function fused with a hyperbolic sine fused function under logistic regression. The sigmoid-hyperbolic functions are enhanced to be more flexible in fitting temporal data with non-linear disease. In contrast, the model application of logistic regression is selected for its stability and interpretability. This approach aims to achieve the best sensitivity and specificity in disease detection while avoiding overfitting the model.

This position is measured on a citrus and pepper leaf images dataset, and yields improved performance and better accuracy, sensitivity, and specificity than classical approaches. This specific study contributes to developing the technology for disease identification of different kinds of medicinal plants. It provides tangible steps farmers and researchers can take to maintain plant health. The results are useful in extending the use of machine learning in the agricultural sector to enhance disease diagnostic tools that are more accurate and reduce the waste of resources.

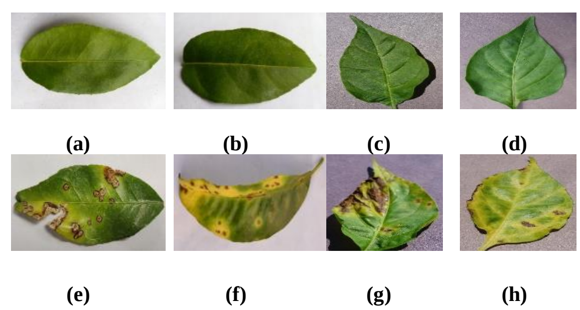


Fig. 1 (a-b) Citrus healthy, (c-d) Pepper healthy. (e) Citrus black spot, f) Citrus canker, and g-h) Pepper bacterial spot.

1.2. Research Challenges

Detecting diseases in citrus and pepper medicinal plant leaves has the following challenges. One of the main challenges of the process is the proper identification of symptoms that need treatment, which can be hardly detected and may also differ with the species of the plant and its surroundings. The conventional approach that mainly involves the physical assessment of fruits can be tiresome and, at the same time, involves some proneness to errors, especially where large-scale farming is involved. Also, for most of the models, especially those based on deep learning, the availability of large datasets and computational resources is paramount. The parameters of such models also make the models complicated and challenging to understand, let alone apply in events that happen in the real world. That is why they have exact, explainable, and resource-saving machine learning approaches that would still provide high diagnostic sensitivity at the current rates.

1.3. Motivation

The rationale for this research arises from the observation that correct early diagnosis of these diseases would go a long way in enhancing the health and yield of citrus and pepper medicinal plants. These types of plants are very important in addressing this disease, and any disease that affects them can lead to a reduced quality of these plants, hence having economic and health costs. Old traditional practices of detecting diseases are no longer adequate for the complex world of agriculture, where identifying the right disease at the right stage is crucial. Moreover, existing machine learning solutions initially work fine, but problems are associated with them, such as high complexity and a vast amount of resources needed for their usage. This research aims to fill this gap by establishing a machine learning model for accuracy, computational efficiency, and model interpretability, which is viable in different agricultural contexts.

1.4. Objectives

More specifically, the conclusions reached aim to propose an innovative machine-learning methodology for diagnosing diseases in citrus and pepper medicinal plant leaves. In particular, this paper intends to improve the chosen model by using logistic regression in association with the modified sigmoid and hyperbolic sine functions to improve the model's capacity to detect intricate disease profiles. Another important goal is to fine-tune model parameters to increase performance while neglecting factors such as increasing the number of computations. The study also aims to determine the efficiency of this approach in practical applications with special reference to diseases shown by citrus and pepper leaves.

1.5. Contributions

The following are the major findings of this research: To begin with, it brings together a new set of modified sigmoid and hyperbolic sine functions with a logistic regression model, there being no previous studies on feature extraction and classification that are quite useful in medicinal plant leaf diseases. The study also shows that this hybrid model is as effective as conventional deep learning models in diagnosing diseases in citrus and pepper plants. Furthermore, the study enriches the general use of machine learning for agricultural purposes. It indicates how these sophisticated models might be applied to the benefit of various plants requiring different levels of attention. Last, these insights have implications for enhancing plant disease control and may contribute to the development of increased yields and quality when cultivating these medicinal plants.

The paper's content is planned as follows: Section 2 presents a literature review. Section 3 explains the proposed methodology. Section 4 results and analysis of the proposed approach using standard metrics over the dataset. Section 5 conclusion and future research directions.

2. Literature Review

The paper selected for review, titled “Classification of Plant Leaf Disease Using Deep Learning,” by K. Indira and H. [7], focuses on applying CNNs to identify plant diseases. The authors propose a deep learning model that can diagnose diseases based on the characteristics like colors, textures, and shape of the images of the leaves. This research proves that the architecture of the CNN model can accurately differentiate different plant leaf diseases and highlights its applicability in precision farming to create timely disease diagnosis and management strategies. The study might fail to capture some issues, such as the protracted process of gathering massive, diverse data or the computational power necessary for generating deep learning models, which is a constraint when implementing the proposed approach in scenarios with limited resources.

The original paper titled “Hybrid System for Detection and Classification of Plant Disease Using Qualitative Texture Features Analysis”; in this paper, analyzing the qualitative texture features of plant leaf images; the authors Anjana, Meenakshi Sood, et al. [8] proposed a hybrid system for both detection and classification of plant diseases. To achieve high diagnostic accuracy, the authors propose a multi-step plant leaf texture analysis system that utilizes both traditional methods and machine learning and deep learning approaches. Specifically, the texture-based features are extracted as primary indicators of plant diseases, and the classifier is trained to detect between healthy and diseased leaves. It is shown that the proposed hybrid system successfully classifies and identifies plant diseases, and thus, such a system can be employed in agricultural areas. The study may have had shortcomings that stem from the reliability of the quality and flexibility of the texture features acquired, which may not be comprehensive of all the disease manifestations, especially when the disease is abarker or exhibits overlapping symptoms. Besides that, the hybrid approach could be time-consuming.

The research paper “Identification of Plant Leaf Diseases Using a Nine-Layer Deep Convolutional Neural Network” by Geetharamani G. and Arun Pandian J. [9]. The system presented in this research paper should be an efficient deep learning system based on plant leaf disease identification. The authors designed a three-layer CNN to filter out pixels from leaf images to understand and extract features while training the model to differentiate between healthy and unhealthy leaves. It is shown that CNN can accurately discriminate plants from diseases with reasonable predictive accuracy, making it a viable addition in agricultural systems for integration to support timely and effective disease management. However, the model has some limitations; it requires a particular dataset to be available to apply and may not be suitable for other plant species or diseases. Furthermore, the study finds that large and diverse datasets are necessary to make the model robust under different conditions and disease stages.

The paper authored by Youyao Fu, Linsheng Guo, and Fang Huang [10] entitled “A Lightweight CNN Model for Pepper Leaf Disease Recognition in a Human Palm Background” introduces a novel Convolutional Neural Network (CNN) model for the identification of pepper leaf diseases, regardless of the leave position on a human palm. The authors pay particular attention to achieving high accuracy of disease detection while keeping the model’s weight low so it can be used in real-life applications, for example, on mobile devices. This study confirms that the proposed model can identify diseases for pepper leaves even under practical agricultural scenarios with a background containing hand images. Some possible shortcomings of the paper may be related to the decrease in the accuracy that can occur when using a lightweight architecture of the model if its design will imply the sacrifice of response speed.

The paper “Citrus Plant Disease Identification using Deep Learning with Multiple Transfer Learning Approaches” written by Talha Anwar and Hassan Anwar [11] developed a deep learning model for disease identification of citrus plants and focused on transfer learning. Here, the authors use various transfer learning techniques to improve the model's performance in disease detection using pre-trained models. It makes sense because, this way, the model is able to re-purpose features learned from large data sets to solve the current problem of deciding whether a given citrus plant is infected or not. The study proves that by achieving up to 96% accuracy and reducing false-negative rates, transfer learning enhances the model's effectiveness in diagnosing diseases affecting citrus crops early enough. Some strengths of the paper may include the fact that it may have some limitations, such as pre-trained models that may not be optimally developed to capture the unique characteristics of citrus plant diseases due to the transfer learning process that is followed. Also, the approach might need significant computational power to adjust the models, which might prove inconceivable, especially in less developed economies.

This paper was presented by Pranajit Kumar Das [12] with the title “Leaf Disease Classification in Bell Pepper Plant using VGGNet”. The paper details the use of VGGNet in classifying diseases in the bell pepper plant. VGGNet architecture, which is very effective in image classification given that it offers depth, is used by the author to forecast and categorize several diseases affecting bell pepper plant leaves. The study shows that VGGNet can distinguish the diseases that affect the bell pepper crops from the images of the affected leaves and, therefore, can be used as a tool in precision agriculture and early detection of diseases through image analysis. The paper may have some drawbacks; for example, the used VGGNet can have high computation complexity, and it is not feasible for real-time implementation in many cases or on low-power devices. Moreover, how the model’s accuracy could be affected by the type of training dataset or if it even applies to other plant species or other disease types. While accurate, the usage of VGGNet may also result in longer training times and higher resource consumption compared to lightweight models, which may be a concern in some agricultural settings.

‘Plant Leaf Disease Detection and Classification using Conventional Machine Learning and Deep Learning’ paper by Hardikkumar S. Jayswal and Jitendra P. Chaudhari [13] compares between conventional machine learning algorithms and deep learning models for detecting and classifying plant leaf diseases. Based on leaf image, it evaluates traditional approaches like SVM and Random Forest against deep learning solutions, e.g. CNNs for disease identification. The study shows that deep learning models are better than regular methods in accuracy and robustness, leading to their effectiveness in agricultural image classification. The paper also points out its limitations, but. In resource-constrained environments, deep learning models can be complex, so there is a need to train and test the process in a time-consuming and resource-intensive manner, which can be difficult. Moreover, the study fails to review how these models might be used in actual agricultural situations and how environmental noise will affect their performance because of these varying conditions.

The work presented in this paper is an effective research paper called “A Multiclass Plant Leaf Disease Detection using Image Processing and Machine Learning Techniques” by Nilay Ganatra and Atul Patel [14]; their research is interested in detecting and classifying multiple plant leaf diseases. In the proposed studies, computer vision techniques are used to automatically identify the color, texture, and shape features of leaf images for disease identification using machine learning classifiers. As evidenced through the study, the proposed method is indeed useful in differentiating between different types of plant leaf diseases.

Hence, it can be used for disease management among plants in agriculture. The limitations of the paper are as follows. The paper may also depend on the quality of the extracted features that may not provide for all the important features of diseases, especially those presenting mild symptoms or those with symptoms that mimic those of other diseases. The study may not comprehensively capture issues that could affect the utilization of the method in real-life agricultural conditions, including but not limited to changes in weather conditions and the presence of noise in the dataset.

S. Jana, A. Rijuvana Begum, S. Selvaganesan, & P. Suresh [15] wrote a paper titled “Deep Belief Network Based Disease Detection in Pepper Leaf for Farming Sector”, which focuses on the task of identifying diseases in plant leaves using DBNs. The authors create a DBN–based model that will use the hierarchical feature learning potential of the algorithm to diagnose and categorise selected images of the pepper plant leaves with different diseases. This paper illustrates that DBN based approach is efficient in terms of accurate distinctions of diseases that the farming segment can benefit from, hence offering a solution to incorporating complicated ML algorithms in farming to improve crop management and control diseases. There is evidence that the output of the model could result from the quality and variety of the training dataset and might have weaknesses in terms of its ability to recognize different species of plants or different types of diseases without having to conduct additional training again. The study may not fully capture some of the variability of the environment in which the DBN-based approach is likely to be practiced, noise and the need for intensive data pre-processing to use the proposed method in real farming environments.

Convolution Neural Network (CNN) based on a MobileNet to detect and classify plant leaf diseases is presented in the study by S. Ashwinkumar, S. Rajagopal, V. Manimaran, and B. Jegajothi [16]. Given its lightweight architecture, tailor-designed for mobile and embedded platforms, MobileNet is a highly suitable choice for real-time agricultural applications. From the results, it is evident that the best MobileNet model classification accurately distinguishes various plant leaf diseases; hence, it is useful and feasible for farmers in their farming practices via their mobile devices. The only disadvantage of using MobileNet is that it tends to be less accurate than deeper and heavier models for deeper and heavier image classification challenging scenarios that involve diseases with similar or vague symptoms. The study might also not elaborate on some drawbacks of implementing the model in actual field environments, especially due to light changes, noise interferences and other real field conditions that may force the model to undergo significant pre-processing to give accurate results.

From the summarized limitations of the various plant disease detection models above, one can note some of the prevalent issues: This is because most of these models rely on some given datasets and, hence, might not perform well when applied to other plant species or diseases. The mentioned DNNs, conventional statistical models, and most hybrid models, in general, require a larger and more diverse data set and are more time-consuming to train and then apply in real-time in environments with limited resources. MobileNet, for instance, can be deemed small and possibly have lesser accuracy ratios than other models. VGGNET, deep belief networks, and many more are some of the most complex and time-consuming. Also, some factors that could be regarded as drawbacks include the increased degree of tuning of the parameters in the algorithm and the features and other changes in the environment of the applications limiting these models from being implemented into real practice in agriculture.

3. Methodology

An approach combining Hyperbolic Sine Sigmoid Function and Logistic Regression is proposed. This approach improves the disease detection rate by implementing high-end feature extraction and classification techniques. The approach starts from data collection and data pre-processing, where a rich data set of medicinal plant leaves, specifically Pepper and Citrus, is acquired and pre-processed through image resizing, noise elimination, and image data augmentation.

3.1. GLCM

Relative to feature extraction, the applied methodology uses the GLCM for texture-based features of the leaf images [17]. Four basic features termed contrast, correlation, energy, and homogeneity are calculated and furnish valuable data pertaining to the spatial distribution of pixel densities [18].

$$Energy = \sum_{i,j}^{N-1} (P_{ij})^2 \tag{1}$$

The uniformity of texture is evaluated in terms of the periodicity of pixel pairs.

$$Contrast = \sum_{i,j=0}^{N-1} P_{ij}(i-j)^2 \tag{2}$$

Gives information about the spatial frequency of an image.

$$Correlation = \sum_{i,j=0}^{N-1} P_{ij} \frac{(i-\mu)(j-\mu)}{\sigma^2} \tag{3}$$

Determines the linear relationship between its neighbouring pixels’ gray levels.

$$Homogeneity = \sum_{i,j=0}^{N-1} \frac{P_{ij}}{1+(i-j)^2} \tag{4}$$

Determines the fluctuations in image intensity.

Nine additional statistical metrics are computed based on these GLCM features: Root Mean Square (RMS) error, mean, standard deviation, variance, skewness, kurtosis, smoothness, and Inverse Difference Moment (IDM) [19]. These features give an approximate model of the leaf's surface, thus boosting the model’s performance when mapping disease patterns.

1. Mean

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} \tag{5}$$

2. Standard Deviation $$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}} \quad (6)$$
3. Entropy $$E(F) = \sum_{i=0}^{G-1} p(i)\, log_2\, p\,(i) \quad (7)$$
4. RMS $$RMS = \sqrt{\frac{\sum_{i=1}^{N} x_i^2}{N}} \quad (8)$$
5. Variance $$\sigma^2 = \frac{\sum(X - \mu)^2}{N} \quad (9)$$
6. Smoothness $$SI = \sqrt{\sum^{\Sigma}(TSi^2{}_{max}} \quad (10)$$
7. Kurtosis $$Kurtosis = \frac{1}{n}\frac{\sum_{i=1}^{n}(X_i - \overline{X})^4}{S^4} \quad (11)$$
8. Skewness $$Skewness = \frac{1}{n}\frac{\sum_{i=1}^{n}(X_i - \overline{X})^3}{S^3} \quad (12)$$
9. IDM $$IDM = \sum_{i,j} \frac{P_{i,j}}{1 + (i - j)^2} \quad (13)$$

These extracted features are then used to feed a logistic regression model, as it is simple but proved quite efficient for the classification models. The training and fine-tuning of the model are accomplished using cross-validation and hyperparameter tuning procedures. The trained model is tested on validation sets and possesses an assessment that includes accuracy, precision, recall, and F1-score [20]. This method fuses a GLCM-based feature extraction tool and modified logistic regression with hyperbolic sigmoid function to provide a potent medical plant leaf disease recognition device.

3.2. Hyperbolic Sine Function

The logistic regression can be further demonstrated using the hyperbolic sine function (sinh) instead of the usual sigmoid function. Compared to standard logistic regression, this substitution is designed to increase performance on all typical statistical measures.

Hyperbolic functions are non-oscillatory, not being periodic like trigonometric functions, with periodic oscillations in both value and slope. Hyperbolic functions basically constitute a trig in terms of a hyperbola instead of a circumference. Both Vincenzo Riccati and Johann Heinrich Lambert independently introduced them in the 1760s. Hyperbolic functions can be expressed in terms of the side lengths of a right triangle within a hyperbolic sector.

For a given hyperbolic angle, the area of the associated hyperbolic sector is twice the length of that angle. In that case, we assign a hyperbolic angle to be a real variable. The most common real examples of hyperbolic functions include sinh, cosh, tanh, sech, cosech, and coth, common in hyperbolic geometry when we must find relationships involving hyperbolic angles or distance.

3.2.1. Sinh Function

The sinh function is used because, like its sine counterpart, the sinh function is derived from the properties of a hyperbola. In this case, the exponential function is expressed as defined by Equation (14), which defines sinh(x).

$$sinh\, x = \frac{e^x - e^{-(x)}}{2} \quad (14)$$

For hyperbolic functions, when considering a circle, the parameterization is given by $x^2+y^2=1$. For a hyperbola, it is defined as $x^2–y^2 =1$, as illustrated in Figures 2 and 3.

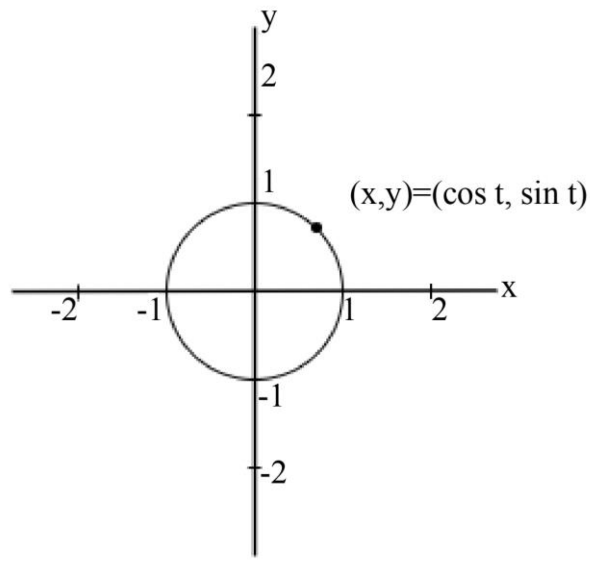


Fig. 2 Points on the circle $x^2+y^2=1$

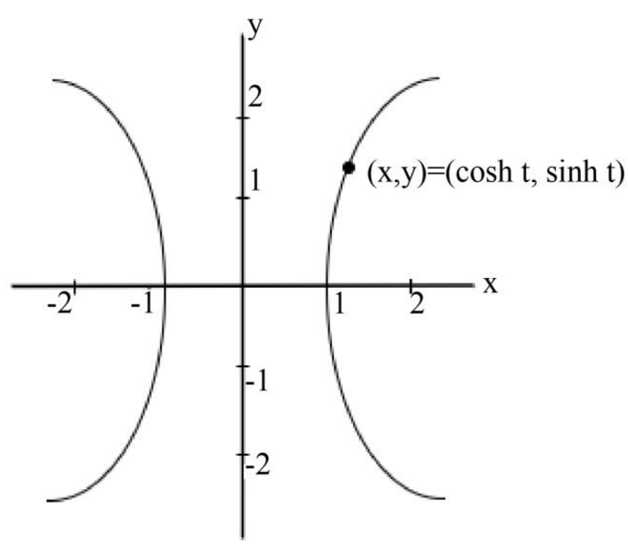


Fig. 3 Points on the hyperbola x^2-y^2=1

In logistic regression, both z and x in sinh(x) are real numbers, making the substitution of sinh in logistic regression yield improved results. The parameters and outcomes for the proposed objective and a statistical evaluation are detailed in section 4, results and analysis. The overall steps involved in the proposed scheme are in the block diagram in Figure 4. Algorithm 1 details the sequence of steps of the current disease detection system.

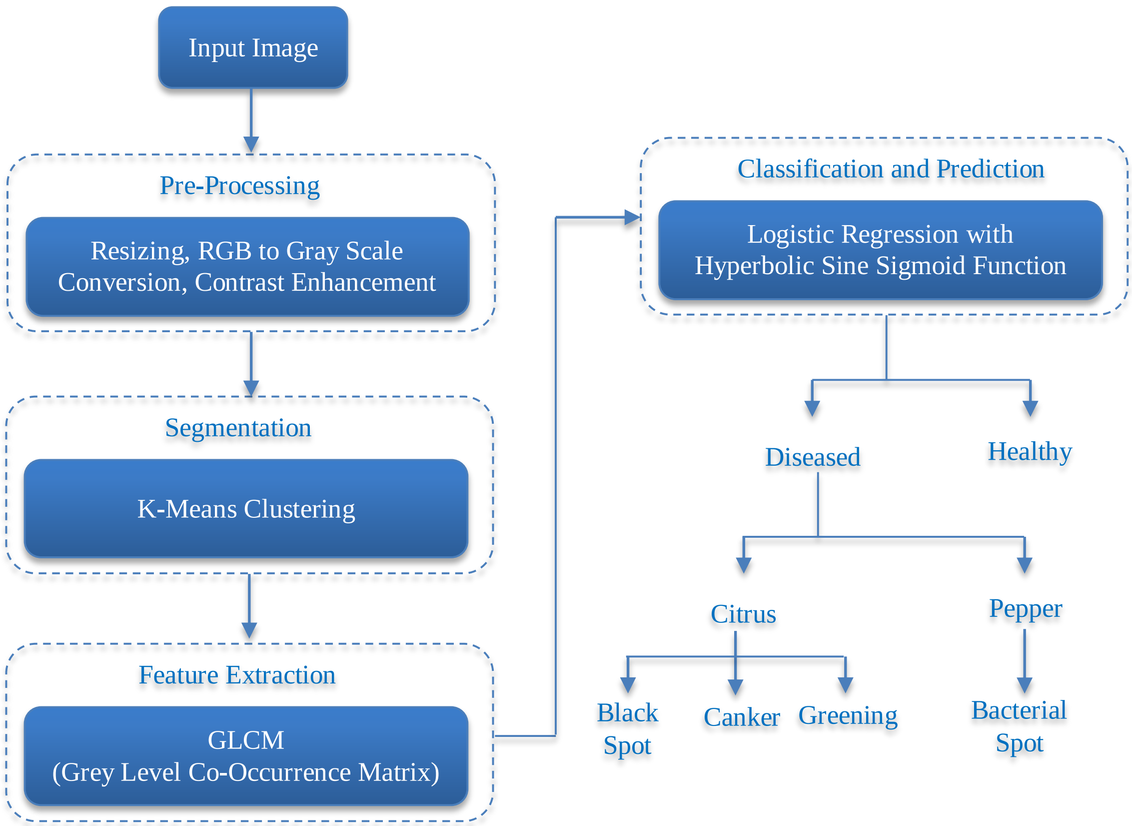


Fig. 4 Block diagram of proposed model

Algorithm 1. Medicinal Plant Leaf Disease Detection
Input:
- Leaf Image Dataset: Set of medicinal plant leaf images with labels
- Initial Hyper parameters: Initial values for logistic regression

Output:
- Predicted Labels: Predicted labels (healthy or diseased) for test images

Begin
  # Step 1: Image Pre-processing
  For each image in the Leaf Image Dataset, do
    Convert image to grayscale
    Normalize pixel values
    Apply noise reduction using a Gaussian filter
    Resize the image to consistent dimensions
  End For

  # Step 2: Feature Extraction
  For each pre-processed image in the Leaf Image Dataset, do
    Extract GLCM features
    Compute other relevant features from the GLCM features

  End For
  # Step 3: Model Initialization
  Initialize the Logistic Regression Model by setting the following hyperparameters
  • Learning rate
  • Number of epochs
  • Regularization rate
  • Batch size
  Define the sigmoid function with the hyperbolic sine function.

  # Step 4: Model Training
  Train the Logistic Regression Model with optimal hyper parameters on full training data

Step 5: Model Evaluation
Evaluate the Logistic Regression Model on test data
Compute performance metrics: accuracy, precision, recall, F1-score
End

3.3. Logistic Regression

Logistic regression is among the most basic algorithms of supervised learning used in binary classification tasks that target estimating the probability of an input belonging to a specific class out of two [21, 22]. As is well known, logistic regression transforms the linear combination of input features to the probability scale using the sigmoid function. However, to increase the model's features and differentiate multiple relations in the data, the more complex one that includes the hyperbolic sine function can be applied instead of the simple sigmoid function.

Modified Sigmoid Function with Hyperbolic Sine

The standard sigmoid function is defined as:

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{15}$$

This is a type of function, a continuous function that generalizes any real x into the [0,1] interval. To improve the model's capability to handle non-linearities and more complex data distributions, we can modify the sigmoid function by incorporating the hyperbolic sine function (sinh(x)), which is defined as:

$$sinh(x) = \frac{e^{x}-e^{-x}}{2} \tag{16}$$

The modified sigmoid function can then be expressed as:

$$\sigma(sinh(x)) = \frac{1}{1+e^{-sinh(x)}} \tag{17}$$

Here, the sigmoid function is used as a nonlinear transformation of the input features, and the hyperbolic sine function maps these features to the rest of the network. This transformation helps the model consider more general patterns since it brings more non-linearity into the logistic regression model and increases its flexibility regarding data distribution.

The function has been modified using the hyperbolic sine function, and it has some benefits when used in logistic regression models. It increases non-linear features, making it suitable for capturing the non-linear trends and relationship between features and target variables. This modification also enhances the flexibility, improving the model's ability to fit more complicated data. Furthermore, another function used is the hyperbolic sine function, and its purpose is to maintain the equality of statistics due to the fact that this function helps in normalizing values; this work also prevents the skewing of data by the presence of outgrowths that are likely to affect the outcome of a model. All these additions provide an increased reliability and flexibility of the given model.

Another important aspect that can improve the performance of logistic regression models is the selection of hyperparameters, especially when using the modified sigmoid function [23]. Amongst these hyperparameters is the learning rate (α), which determines the size of the step to be made during the weights update and the strength of the regularization (λ which is used to overcome the overfitting of the model. Another factor that greatly influences model convergence and overfitting is the batch size and the number of epochs.

Algorithm 2. Logistic Regression Model with Hyperbolic Sine Sigmoid Function
Input:
- Training data: $\mathrm{X} = \{\mathrm{x}_1, \mathrm{x}_2, \ldots\ldots\ldots, \mathrm{x}_\mathrm{m}\}$(features) where *X* is of size *m x n*
- Labels: $y = \{y_1, y_2, \ldots\ldots\ldots, y_m\}$ (size *m x 1*)
- Learning rate: α=0.05
- Number of epochs: n_epochs=100
- Regularization rate: λ=0.04
- Batch size: batch_size=64

Output:
- Optimized model parameters θ (size $n\times1$)

Step 1: Initialize Parameters
1. Initialize θ to a vector of zeros (size n×1) or small random values.

Step 2: Training with Mini-Batch Gradient Descent
2. For epoch=1 to n_epochs, do the following:
   2.1. Shuffle the training data *(X, y)*.
   - Randomly reorder the rows of X and y to ensure each epoch uses a different order of data.

   2.2. For each mini-batch in the shuffled data:
   - Divide the shuffled training data into mini-batches of size batch_size=64.

   2.3. For each mini-batch X_{batch}, y_{batch}:
   - Compute the Hypothesis for the Mini-Batch:
   - For each example x_j in the mini-batch, compute: $h\theta(x_j) = \frac{1}{1+exp(-sinh(\theta^T x_j))}$
   - Compute the Regularized Cost Function for the Mini-Batch:
   - Calculate the regularized cost function *J(θ)* for the mini-batch:

$$J(\theta) = -\frac{1}{batch\_size} X \sum_{j=1}^{batch\_size}[y_j \, log(\, h\theta(x_j)) + (1-y_j)\, log(\,1-h\theta(x_j))]$$

   - Perform Gradient Descent for the Mini-Batch:

- For each parameter θ_k :

$$\theta_k = \theta_k - \alpha[\frac{1}{batch\_size}(h\theta(x_j) - y_j)x_{jk} + \frac{\lambda}{m}\theta_k]$$

Step 3: Return Final Model

3. Return the optimized parameters θ.

Also, feature scaling methods like standardization help make features more uniform and, in turn, improve the model's effectiveness. The point of division for binarization usually equals 0.5 in the clinical setting, and it can be optimized to achieve higher sensitivity for early diagnosis or higher specificity for accurate diagnosis. In models that employ a hyperbolic sine function with parameters inside, these parameters may also need to be optimized to achieve the best performance.

4. Results and Analysis

4.1. Dataset Overview

In the present research, a data set containing both normal and disease-affected citrus and pepper leaves has been used. The citrus leaf images were sourced from Kaggle [24], consisting of 607. PNG-type images (each image is 256x256 in size) are categorized into five classes: Blackspot, Canker, Greening, Melanose, and Healthy options. The total size of the citrus dataset is 43.29MB.

The identified images of pepper leaves to which the used bell peppers belong were sourced from Mendeley Data [9]. The applied dataset consists of 2475. PNG-type images (each image is 256x256) in total, where 997 images are related to bacterial spots and the remaining 1478 are classified as healthy images. The total size of the pepper dataset is 158.4MB. To assess the performance of the proposed model, the dataset was split into multiple training and testing ratios: Perfect ratio for excellent successful planning. 5 fold cross-validation was done to improve the model's reliability.

Here, we split the dataset into 5 parts of equal sizes, and then, in each iteration, we used one part for testing and remaining four parts for training. These split were performed five times with each fold form as the test set; this was in order to gain a general idea of how the model would perform across different splits. After completing the process, the averages of the results acquired were taken from the five-fold validation to estimate the model’s accuracy and how well it could generalize.

4.2. System Specification

The process was carried out in MATLAB 2021b environment on the system, and the system is powered with the Intel Core i5 8th Generation, the processor that comes with 4 cores and 8 threads and has a base clock rate of 1.6 GHz the bandwidth of which can also be increased to 3. 4 GHz. MATLAB 2021b edition on Window 10 Working Platform professional 64, has been used to analyze its tool boxes, Statistical and Machine learning, Image Processing, Deep Learning, Optimization toolbox. These toolboxes helped investigate and apply logistic regression, combine sigmoid-hyperbolic function, and tune hyperparameters. The Lenovo ThinkPad T480, with its average performance and strong computation power, proved to be suitable for the implementation of the proposed model.

4.3. Performance Evaluation Parameters

A performance metric analysis is necessary to evaluate the approach's success and the proposed model's accuracy. The proposed model’s key metrics are evaluated against accuracy, precision, recall, F1-score, confusion matrix and the area under the ROC curve [25]. Based on these metrics, the modified sigmoid hyperbolic function was coupled with the logistic regression algorithm to assess the method's ability to classify and identify diseases in medicinal plant leaves. Figure 5 represents the structure of the confusion matrix. The confusion matrices obtained for Citrus and Pepper datasets are represented in Figures 6(a) and 6(b). Figure 7 depicts the flow from raw values and/or labeled data to the F1-Score of the metrics hierarchy.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{18}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{19}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{20}$$

$$\text{FPR} = \frac{FP}{TN+FP} \tag{21}$$

$$\text{F1Score}=\frac{2\times\text{Precison}\times\text{Recall}}{\text{Precision}+\text{Recall}} \tag{22}$$

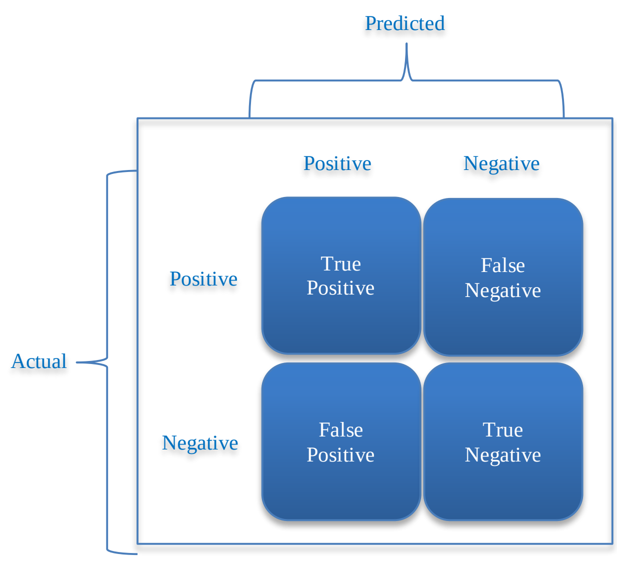


Fig. 5 Confusion matrix

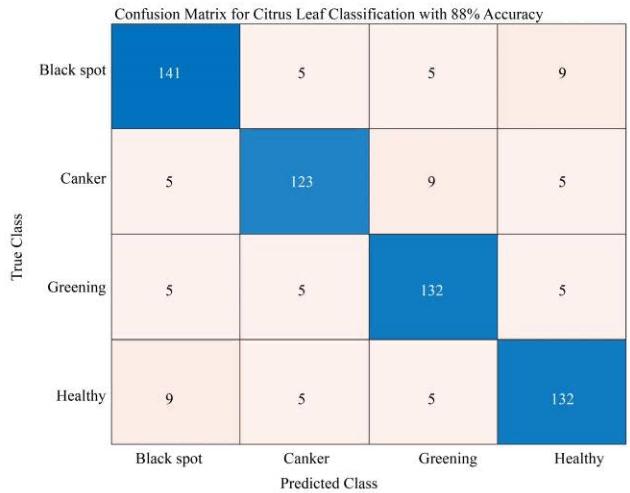


Fig. 6(a) Confusion matrix for citrus dataset

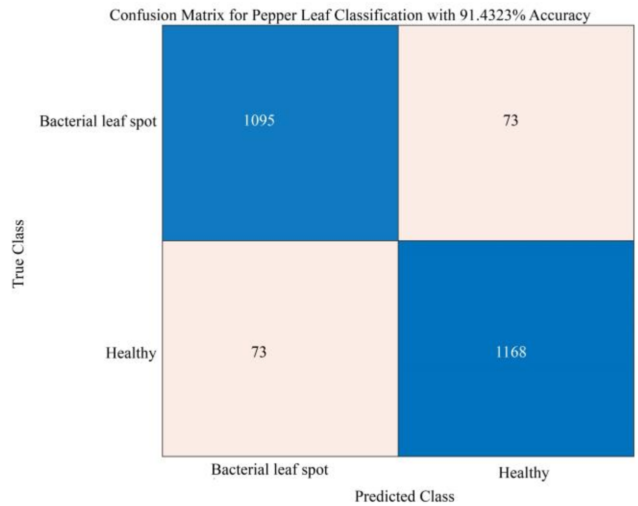


Fig. 6(b) Confusion matrix for pepper dataset

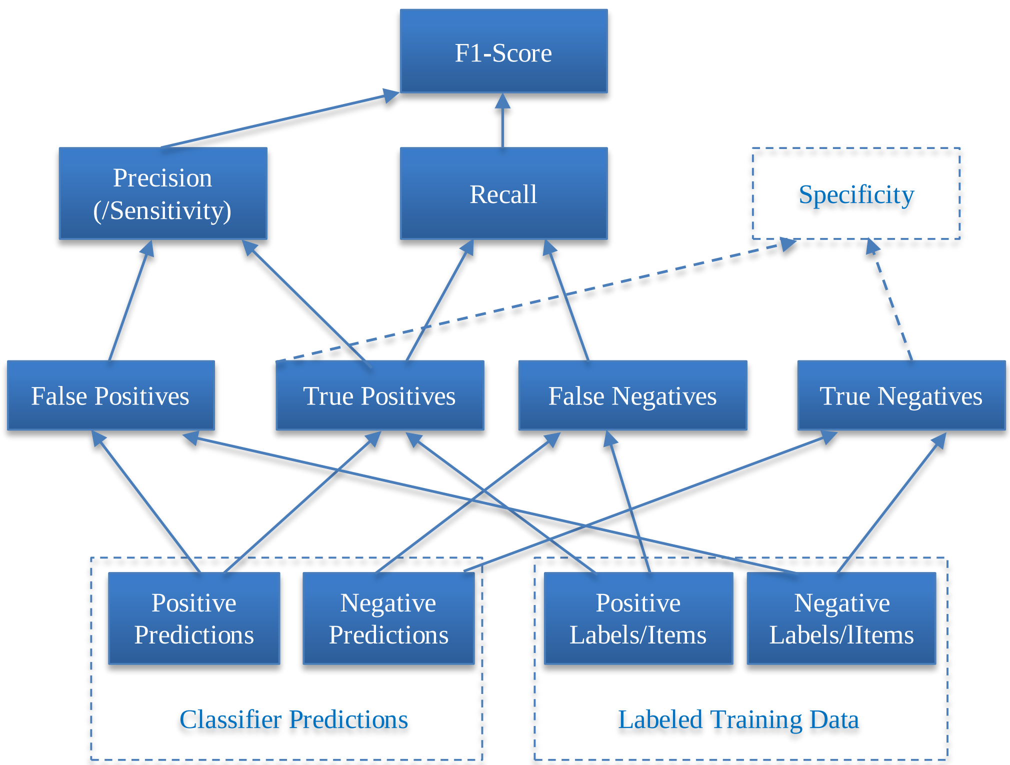


Fig. 7 Impact of confusion matrix in parameter evaluation

Table 1 presents the performance evaluation metrics for the Citrus dataset, comparing the results between the logistic regression (LR) model and the proposed model. Similarly, Table 2 displays the parametric comparison results between the modified SVM classifier [26] and the proposed model. Figures 8(a) to 12(a) illustrate the graphical comparison of metrics between LR and the proposed method for the Citrus dataset. On average, the proposed model demonstrates improvements of 64.44% in accuracy, 53.16% in precision, 64.44% in recall, 54.02% in FPR, and 150.91% in F-Measure compared to the logistic regression model. Figures 8(b) to 12(b) present the graphical comparison of metrics—accuracy, precision, recall, FPR, and F-Measure—between the modified SVM classifier [26] and the proposed modified logistic regression with the hyperbolic sine sigmoid function for the Pepper dataset. The proposed model shows average improvements of 38.52% accuracy, 14.51% in precision, 37.39% in recall, 74.81% in FPR, and 31.74% in F-Measure over the modified SVM classifier. Figure 13 shows the ROC curves for the Citrus and Pepper datasets.

4.3.1. Baseline Model Comparison

Performance metrics used in the assessment of the different models have been captured in Table 3. For the purpose of extracting the features for the baseline models, DarkNET-19 [27] was used, whereas, for the implementation of the proposed method, GLCM was used. An average of 94 was realized using the proposed method, indicating that it had the highest accuracy in terms of student performance. 45% across all ratios. Linear SVM, along with Linear SVM+PCA, came second and third in efficiency. This proves that in the aspect of potential performance, GLCM features are far superior compared to DarkNET-19 features. Furthermore, compared with the other methods, the method with modified logistic regression through the hyperbolic sine function has made the maximum accuracy, precision, sensitivity, and the F-Measure possible.

Table 4 presents a performance assessment comparing the accuracy of the baseline model, which uses a transfer learning approach combined with an attention-based pre-trained VGG16 model, to the proposed model. In the proposed model, features are extracted using GLCM, which are then utilized for disease detection in citrus leaves through MLR with a hyperbolic sine sigmoid function. The proposed model shows a significant performance improvement compared to the baseline model, achieving a 2.25% increase in accuracy.

4.3.2. Limitation of the Current Study

The study uses GLCM for disease detection in Citrus and Pepper plants; however, the work has challenges regarding image quality since variations in lighting and noise might be obtained, negatively impacting the GLCM feature extraction process. The above-mentioned model is computationally intensive and requires the right choice of parameters, which may limit its applicability to other plant species.

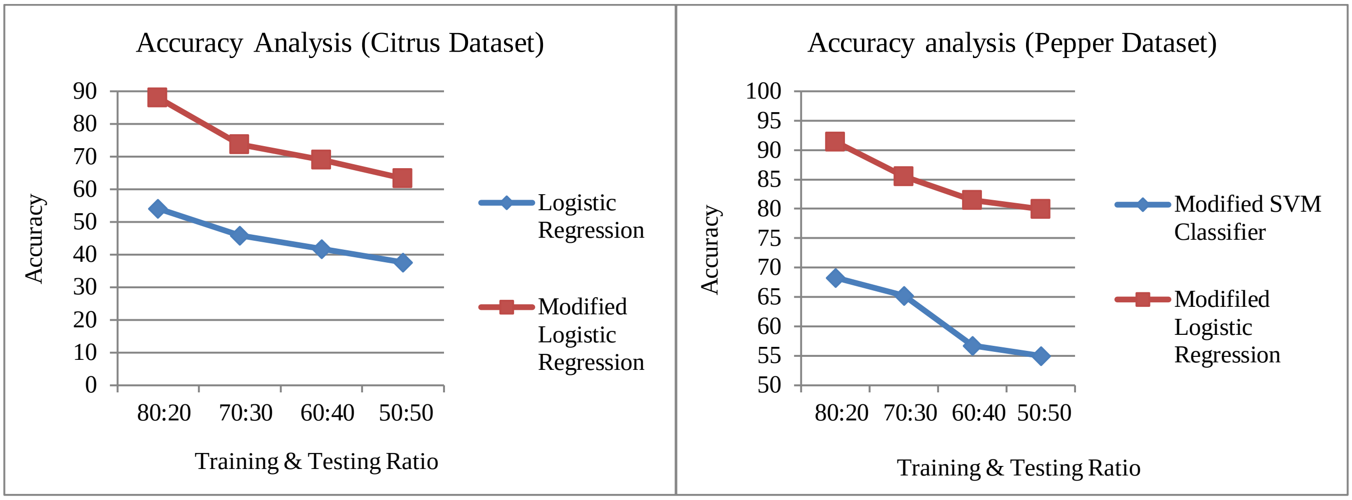
Increasing performance comes at the cost of elaborated hyperparameter optimization trade-off, which in turn leads to overfitting the results, and the methods applied are rather computational-demanding, which may not allow for their application in real-world problems, particularly in environments where resources are limited.

Table 1. Model performance evaluation for Citrus dataset

Training & Testing Ratio	Model	Accuracy	Precision	Recall	FPR	F-Measure
80:20	LR	54.1667	0.5321	0.5417	0.1528	0.5368
	MLR	88.0101	0.8868	0.8801	0.0783	0.8774
70:30	LR	45.8333	0.3929	0.4583	0.4231	0.1806
	MLR	73.6227	0.7941	0.7362	0.1221	0.7138
60:40	LR	41.6667	0.5389	0.4167	0.1944	0.4700
	MLR	69.0910	0.6542	0.6909	0.1434	0.7010
50:50	LR	37.5000	0.5071	0.3750	0.4311	0.2083
	MLR	63.3233	0.6211	0.6332	0.1294	0.6162

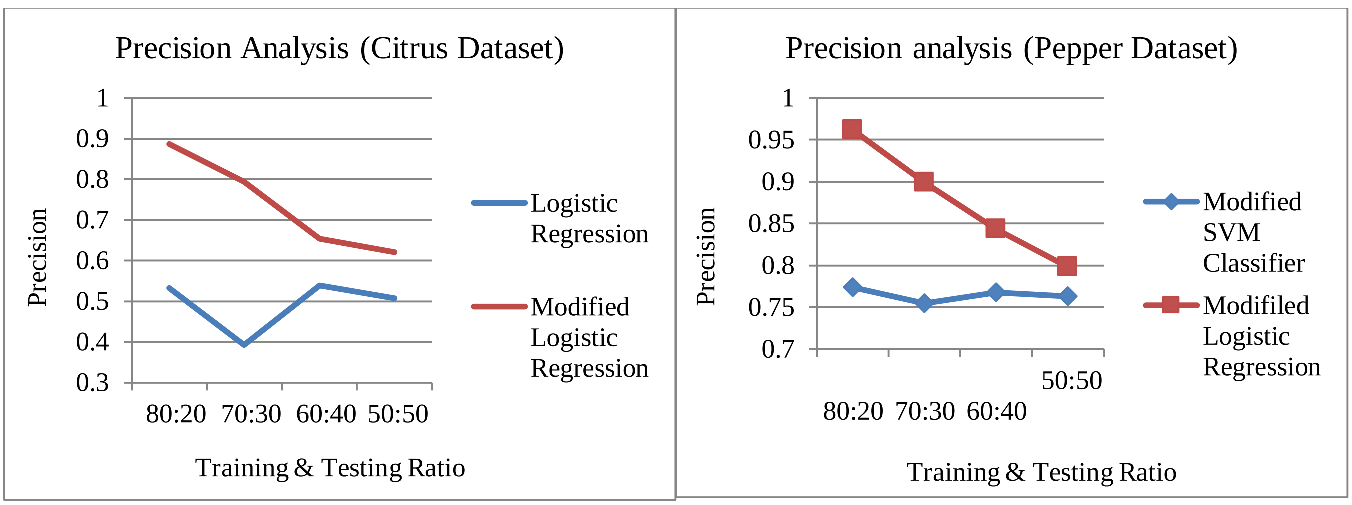
Table 2. Model performance evaluation for Pepper dataset

Training & Testing Ratio	Model	Accuracy	Precision	Recall	FPR	F-Measure
80:20	SVM	68.2351	0.7734	0.6833	0.6822	0.6739
	MLR	91.4323	0.9616	0.9143	0.1008	0.9170
70:30	SVM	65.2523	0.7543	0.6745	0.6599	0.6534
	MLR	85.4612	0.8992	0.8546	0.1243	0.8496
60:40	SVM	56.6672	0.7679	0.5667	0.4333	0.6521
	MLR	81.4754	0.8436	0.8147	0.1932	0.8163
50:50	SVM	55.0010	0.7632	0.5500	0.4500	0.6393
	MLR	79.9342	0.7986	0.7993	0.1014	0.8674



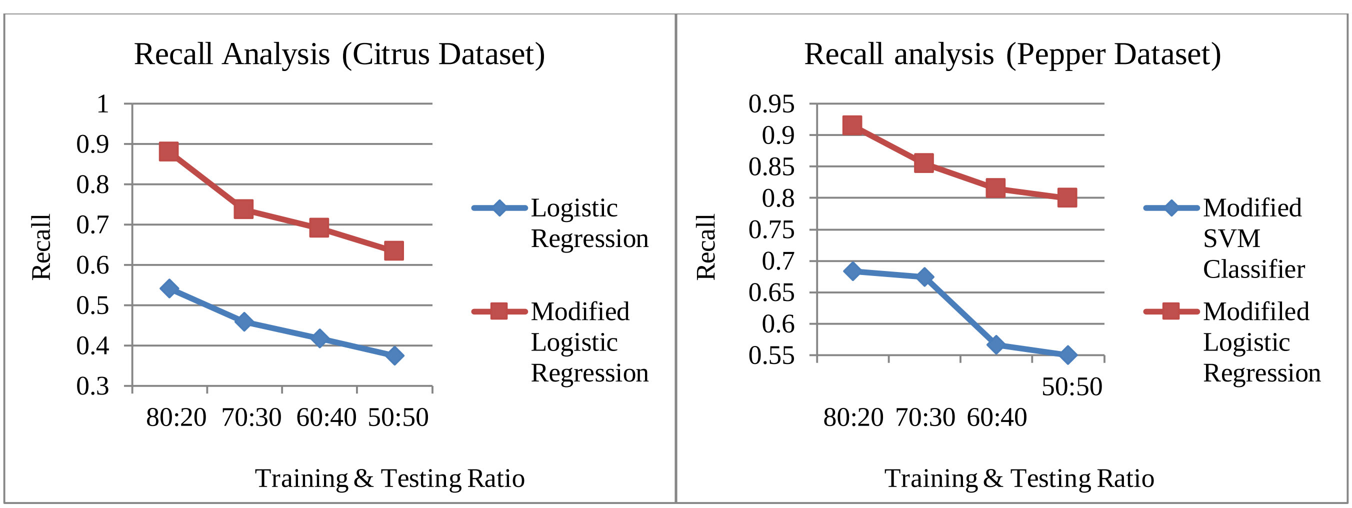
(a) **(b)**

Fig. 8 Accuracy analysis a) Citrus dataset, and b) Pepper dataset.



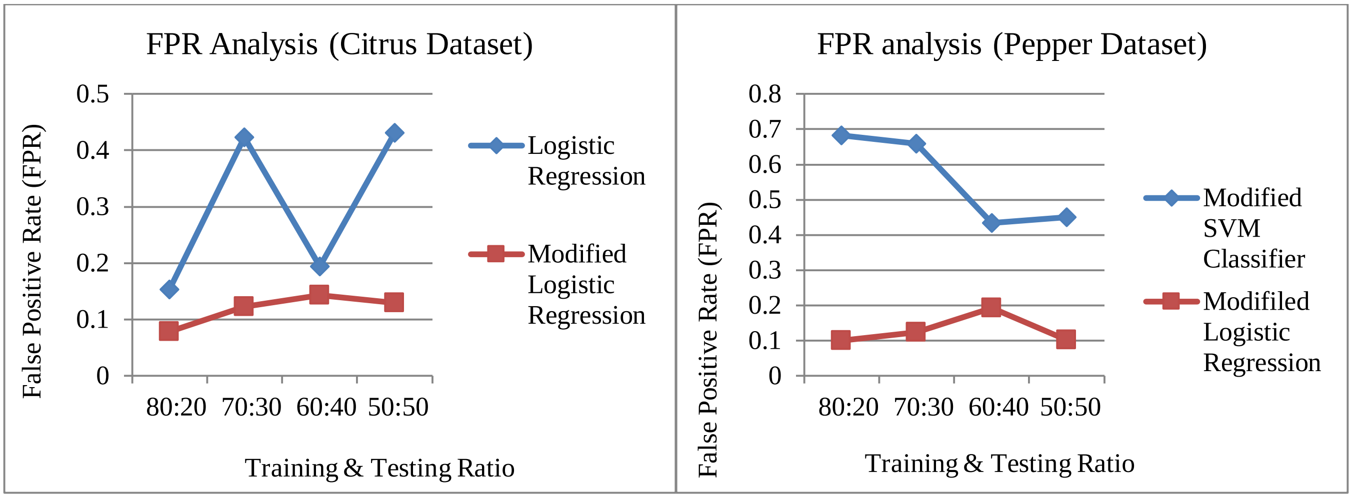
(a) **(b)**

Fig. 9 Precision analysis a) Citrus dataset, and b) Pepper dataset.



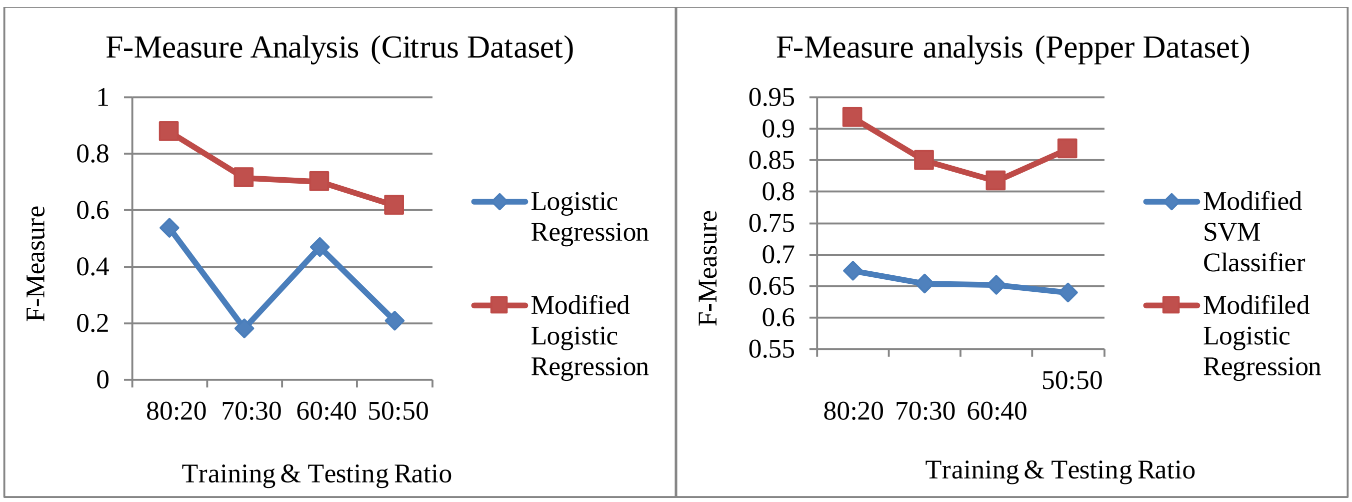
(a) **(b)**

Fig. 10 Recall analysis a) Citrus dataset, and b) Pepper dataset.



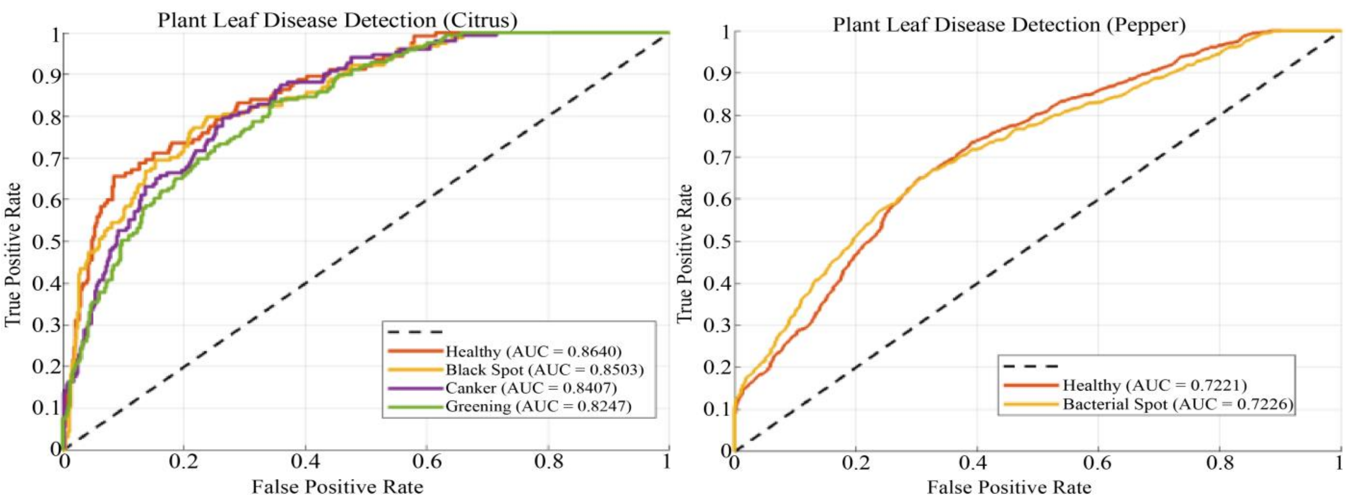
(a) (b)

Fig. 11 FPR analysis a) Citrus dataset, and b) Pepper dataset.



(a) (b)

Fig. 12 F-Measure analysis (a) Citrus dataset, and b) Pepper dataset.



(a) (b)

Fig. 13 ROC curves (a) Citrus dataset, and b) Pepper dataset.

Table 3. Pepper leaf disease detection for different models

Model	Accuracy	Precision	Recall	F-Measure
DarkNET-19 Features[27]				
Gaussian NB	85.3%	0.88	0.82	0.82
Gaussian NB+PCA	78.4%	0.80	0.75	0.72
Linear SVM	93.5%	0.94	0.94	0.92
Linear SVM+PCA	93.1%	0.92	0.94	0.91
K-NN.	90.5%	0.87	0.97	0.87
K-NN. +PCA	73.3%	0.69	1.00	0.50
DT	82.5%	0.84	0.80	0.78
DT+PCA	81.3%	0.82	0.81	0.75
GLCM Features				
Logistic Regression[28]	79.59%	0.79	0.79	0.78
SVM Classifier[26]	62.59%	0.76	0.63	0.64
SVM classifier with RBF kernel[26]	68.23%	0.77	0.68	0.67
The proposed method (MLR)	**91.43%**	**0.96**	**0.91**	**0.91**

Table 4. Citrus leaf disease detection between VGG16 and the proposed model

Model	Accuracy	Precision	Recall	F-Measure
VGG16 [29]	86.07%	90.25	91.05	97.23
Proposed LR model	**88.01%**	88.68	88.01	87.74

5. Conclusion

The proposed study presented a new method of disease identification and diagnosis from citrus and pepper plant leaves. Thus, the new combination of a modified sigmoid function with the hyperbolic sine enhances the logistic regression model to achieve a better fit compared to the previous nonlinear association of various disease conditions. The enhancement done with the hyperparameter tuning has enhanced the model even more, resulting in increased accuracy and standardized performance for various data sets.

In this work, the authors provide evidence of the feasibility of the proposed method as a tool for early diagnostics and differentiation of healthy and affected leaves in citrus and pepper plants. Besides, this method is superior to traditional models and allows for further research to generalize the approach to other medicinal plants and implement it in current-time agriculture. The research helps refine precision agriculture instruments that will better help in efficient farming and enhance crop management.

Future Scope

More importantly, the dataset should be expanded to improve the model generalization. The inclusion of more advanced deep learning algorithms increases the model's accuracy. A promising direction is also real time deployment of the model on edge devices or mobile applications. The deployment of IoT data could also improve disease projection accuracy. The sigmoid hyperbolic function was refined, and transfer learning was applied to the model to make the model more adaptable to different agricultural applications. Furthermore, further improvement in robustness will be achieved by addressing challenges in dealing with uneven lighting and varying image capture angles. Structural similarities in image and disease patterns could lead to extending the methodology to other crops.

References

[1] Rahul Kundu, Usha Chauhan, and S.P.S Chauhan, "Plant Leaf Disease Detection Using Image Processing," *2022 2nd International Conference on Innovative Practices in Technology and Management*, Gautam Buddha Nagar, India, pp. 393-396, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[2] Sunil S. Harakannanavar et al., "Plant Leaf Disease Detection Using Computer Vision and Machine Learning Algorithms," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 305-310, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[3] Citrus: Diseases and Disorders of Leaves and Twigs, UC IPM. [Online]. Available: https://ipm.ucanr.edu/PMG/C107/m107bpleaftwigdis.html

[4] Caroline Bloomfield, Common Pepper Plant Problems – Pepper Plant Diseases and Pests, Gardening Know How, 2023. [Online]. Available: https://www.gardeningknowhow.com/edible/vegetables/pepper/common-pepper-plant-problems.htm
[5] Aliyu Muhammad Abdu, Musa Mohd Muhammad Mokji, and Usman Ullah Ullah Sheikh, "Machine Learning for Plant Disease Detection: An Investigative Comparison between Support Vector Machine and Deep Learning," *IAES International Journal of Artificial Intelligence*, vol. 9, no. 4, pp. 670-683, 2020. [CrossRef] [Google Scholar] [Publisher Link]
[6] Rehan Ullah Khan et al., "Image-Based Detection of Plant Diseases: From Classical Machine Learning to Deep Learning Journey," *Wireless Communications and Mobile Computing*, vol. 2021, no. 1, pp. 1-13, 2021. [CrossRef] [Google Scholar] [Publisher Link]
[7] K. Indira, and H. Mallika, "Classification of Plant Leaf Disease Using Deep Learning," *Journal of The Institution of Engineers (India): Series B*, vol. 105, pp. 609-620, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[8] Anjna, Meenakshi Sood, and Pradeep Kumar Singh, "Hybrid System for Detection and Classification of Plant Disease Using Qualitative Texture Features Analysis," *Procedia Computer Science*, vol. 167, pp. 1056-1065, 2020. [CrossRef] [Google Scholar] [Publisher Link]
[9] G. Geetharamani, and J. Arun Pandian, "Identification of Plant Leaf Diseases Using a Nine-Layer Deep Convolutional Neural Network," *Computers & Electrical Engineering*, vol. 76, pp. 323-338, 2019. [CrossRef] [Google Scholar] [Publisher Link]
[10] Youyao Fu, Linsheng Guo, and Fang Huang, "A Lightweight CNN Model for Pepper Leaf Disease Recognition in a Human Palm Background," *Heliyon*, vol. 10, no. 12, pp. 1-16, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[11] Talha Anwar, and Hassan Anwar, "Citrus Plant Disease Identification Using Deep Learning with Multiple Transfer Learning Approaches," *Pakistan Journal of Engineering and Technology*, vol. 3, no. 2, pp. 34-38, 2020. [CrossRef] [Google Scholar] [Publisher Link]
[12] Pranajit Kumar Das, "Leaf Disease Classification in Bell Pepper Plant Using VGGNet," *Journal of Innovative Image Processing*, vol. 5, no. 1, pp. 36-46, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[13] Hardikkumar S. Jayswal, and Jitendra P. Chaudhari, "Plant Leaf Disease Detection and Classification Using Conventional Machine Learning and Deep Learning," *International Journal on Emerging Technologies*, vol. 11, no. 3, pp. 1094-1102, 2020. [Google Scholar]
[14] Nilay Ganatra, and Atul Patel, "A Multiclass Plant Leaf Disease Detection Using Image Processing and Machine Learning Techniques," *International Journal on Emerging Technologies*, vol. 11, no. 2, pp. 1082-1086, 2020. [Google Scholar]
[15] S. Jana, A. Rijuvana Begum, and S. Selvaganesan, "Design and Analysis of Pepper Leaf Disease Detection Using Deep Belief Network," *European Journal of Molecular & Clinical Medicine*, vol. 7, no. 9, pp. 1724-1731, 2020. [Google Scholar]
[16] S. Ashwinkumar et al., "Automated Plant Leaf Disease Detection and Classification Using Optimal MobileNet Based Convolutional Neural Networks," *Materials Today: Proceedings*, vol. 51, no. 1, pp. 480-487, 2022. [CrossRef] [Google Scholar] [Publisher Link]
[17] M. Yogeshwari, and G. Thailambal, "Automatic Feature Extraction and Detection of Plant Leaf Disease Using GLCM Features and Convolutional Neural Networks," *Materials Today: Proceedings*, vol. 81, no. 2, pp. 530-536, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[18] T. Meenakshi, "Automatic Detection of Diseases in Leaves of Medicinal Plants Using Modified Logistic Regression Algorithm," *Wireless Personal Communications*, vol. 131, pp. 2573-2597, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[19] Aswathy K Cherian, E. Poovammal, and C. Malathy, "Automatic Feature Extraction for Breast Density Segmentation and Classification," *Asian Journal of Pharmaceutical and Clinical Research*, vol. 10, no. 12, pp. 111-115, 2017. [CrossRef] [Google Scholar] [Publisher Link]
[20] K. Saravanapriya, and J. Bagyamani, "Performance Analysis of Classification Algorithms on Diabetes Dataset," *International Journal of Computer Sciences and Engineering*, vol. 5, no. 9, pp. 15-20, 2017. [Google Scholar] [Publisher Link]
[21] Maher Maalouf, "Logistic Regression in Data Analysis: An Overview," *International Journal of Data Analysis Techniques and Strategies*, vol. 3, no. 3, pp. 281-299, 2011. [CrossRef] [Google Scholar] [Publisher Link]
[22] Chao-Ying Joanne Peng, Kuk Lida Lee, and Gary M. Ingersoll, "An Introduction to Logistic Regression Analysis and Reporting," *The Journal of Educational Research*, vol. 96, no. 1, pp. 3-14, 2002. [CrossRef] [Google Scholar] [Publisher Link]
[23] Ahmed Hamdy Ahmed Arafa et al., "Logistic Regression Hyperparameter Optimization for Cancer Classification," *Menoufia Journal of Electronic Engineering Research*, vol. 31, no. 1, pp. 1-8, 2022. [CrossRef] [Google Scholar] [Publisher Link]
[24] Citrus Leaf Disease Image, Kaggle, 2023. [Online]. Available: https://www.kaggle.com/datasets/myprojectdictionary/citrus-leaf-disease-image/data
[25] Andrew Churcher et al., "An Experimental Analysis of Attack Classification Using Machine Learning in IoT Networks," *Sensors*, vol. 21, no. 2, pp. 1-32, 2021. [CrossRef] [Google Scholar] [Publisher Link]
[26] T. Meenakshi, and Pushpa Rani, "Dull Intensity Medicinal Plant Leaf Disease Detection Using Machine Learning Algorithms," *NeuroQuantology*, vol. 20, no. 8, pp. 7228-7242, 2022. [Google Scholar] [Publisher Link]
[27] Alper Özcan, and Emrah Dönmez, "Bacterial Disease Detection for Pepper Plant by Utilizing Deep Features Acquired from DarkNet-19 CNN Model," *Dicle University Faculty of Engineering Engineering Journal*, vol. 12, no. 4, pp. 573-579, 2021. [CrossRef] [Google Scholar] [Publisher Link]
[28] Midhun P. Mathew, and Therese Yamuna Mahesh, "Leaf-Based Disease Detection in Bell Pepper Plant Using YOLO v5," *Signal, Image and Video Processing*, vol. 16, pp. 841-847, 2022. [CrossRef] [Google Scholar] [Publisher Link]
[29] Kalisetty Sai Kumar, "Citrus Leaf Disease Detection Using Attention Basedpre-Trained VGG16," pp. 1-11, 2023.