Enhanced Plant Leaf Disease Detection Using Modified Logistic Regression for Sustainable Agriculture Practices

S. Kiran
Y.S.R.E.C of YVU
N. Subramanyan
S.R.N.M College
M. Bhavsingh
bhavsinghit@gmail.com

Ashoka Womens Engineering College
K Samunnisa
Ashoka Womens Engineering College
B. Swathika
Y.S.R.E.C of YVU
Lavanya A
Universitat Politécnica De Valencia
Jaime Lloret
Universitat Politécnica De Valencia

Research Article

Keywords: Accuracy, Dull intensity images, Disease detection, Plant leaf diseases, Logistic Regression, FPR

Posted Date: April 1st, 2025

DOI: https://doi.org/10.21203/rs.3.rs-6241975/v1

Additional Declarations: No competing interests reported.

Enhanced Plant Leaf Disease Detection Using Modified Logistic Regression for Sustainable Agriculture Practices

[1*]S. Kiran, [2]N. Subramanyan, [3*]M. Bhavsingh, [4]K Samunnisa,[5]B. Swathika, [6]Lavanya. A, [7]Jaime Lloret

[1*]Associate Professor, Department of CSE, Y.S.R.E.C of YVU, Proddatur, AP, India
e-mail: rkirans125@gmail.com
[2]Research Scholar, Department of Computer Science, S.R.N.M College, Sattur, TN, India
e-mail: nsm2517@gmail.com
[3*]Associate Professor, Department of CSE, Ashoka Women's Engineering College, Kurnool, AP, India. e-mail: bhavsinghit@gmail.com
[4]Associate Professor, Department of Computer Science and Engineering Ashoka Womens Engineering College , Kurnool, Andhra Pradesh, India. Email ID: samunnisa14@gmail.com
[5]U.G Student, Department of CSE, Y.S.R.E.C of YVU, Proddatur, AP, India,
e-mail: swathikareddy67@gmail.com
[6]Universitat Politécnica De Valencia, Valencia, Spain, e-mail: phani.lav@gmail.com
[7]Universitat Politécnica De Valencia, Valencia, Spain,e-mail: jlloret@dcom.upv.es

Abstract –Pepper, Hibiscus, and Basil are medicinal plants with a rich history in traditional medicine and health benefits. They are essential in culinary and medicinal applications, contributing to natural health solutions. Disease detection is crucial to protect their agricultural, economic, and medicinal value. Early detection minimizes crop losses, maintains plant health, and ensures plant availability for traditional medicine and culinary uses. This promotes sustainable and eco-friendly agricultural practices. Traditional logistic regression for plant leaf disease detection struggles with imbalanced data and a fixed linear decision boundary, making it less effective in capturing complex disease patterns. The modified logistic regression model with the One Half Constant improves performance metrics and handling intricate features of leaf images by addressing class imbalance more effectively. It adjusts the decision boundary to handle imbalanced datasets, enhancing classification accuracy for minority classes while maintaining simplicity and interpretability. This study uses a dataset collected from Kaggle and surrounding of Kadapa district AP, India. For the evaluation of the proposed model in disease detection, the traditional logistic regression and other machine learning algorithms were used, and the corresponding key metrics of accuracy, precision, recall, false positive rate (FPR) and F-Measure were assessed. A comparison with existing methods show overwhelming improvement of 32.94% in accuracy, 17.64% in precision, 33.6% in recall, 86.91% in FPR improvement, 34.6% in F-Measure. The proposed approach seeks to improve overall diagnostic accuracy, thereby providing a reliable tool for early detection and treatment planning in clinical sectors.

Keywords-Accuracy, Dull intensity images, Disease detection, Plant leaf diseases, Logistic Regression, FPR.

1. Introduction

The agricultural domain is imperative for underneath global health and food security, medicinal plants had a significant role in pharmaceuticals, traditional medicine, and diverse industries[1]. Ensuring the health and yield of medicinal plants are crucial to meet the growing demand for natural medications [2]. Plant diseases pose a considerable threat to crop yield and quality, making the appropriate detection and supervision of these diseases important [3].

1.1 Healthy Versus Diseased Leaves

Healthy and diseased leaf images of pepper, hibiscus and basil are shown in Fig.1 and Fig.2. Pepper leaves are typically vibrant green, with a smooth texture and uniform shape. They are firm and free from any spots, discoloration, or distortion. Diseased pepper leaves may show symptoms such as brown or black spots (indicative of bacterial leaf spot), yellowing, curling, or a mosaic pattern of light and dark green areas (caused by viral infections). Severely affected leaves may become brittle, wilt, or drop prematurely.

(a) (b) (c)

Fig. 1. Healthy Leaves a) Pepper b) Hibiscus c) Basil

(a) (b) (c)

Fig. 2. a) Pepper Bacterial Leaf spot b) Basil Leaf spot c) Hibiscus turning yellow.

Hibiscus leaves are usually dark green, glossy, and slightly serrated along the edges. They are firm and resilient, with no signs of yellowing or spots. When diseased, hibiscus leaves can turn yellow (chlorosis) due to nutrient deficiencies or root problems. They may also exhibit spots, necrotic patches, or a generally wilted appearance if suffering from infections or environmental stress. Diseased leaves may become limp, drop off, or develop distorted shapes. Healthy basil leaves are bright green, smooth, and fragrant, with an even texture and a slightly oily feel. They should be free from spots, discoloration, or any visible damage. Diseased basil leaves might show dark brown or black spots, often with a yellow halo (due to fungal or bacterial infections). Leaves can also yellow and become brittle or wilted. In severe cases, leaves may curl, dry out, or drop off, significantly reducing the plant's vigor.

Conventional methods of disease recognition in plants often entail manual examination, which is labor-intensive, sustained, and lying face down to human miscalculation [4]. The combination of machine learning techniques in plant disease recognition has come up as a capable solution, offering the possible for automation, precision, and effectiveness[5] . In this circumstances, logistic regression, a widely-used method, has shown accomplishments in various fields, including medical diagnostics. However, utilizing logistic regression for the meticulous and complicated patterns there in plant diseases needs adaption to enhance its usefulness.

The dataset worn in this study comprises a assorted collection of medicinal plant leaves images, each thoroughly labeled with the equivalent disease class[6]. Leveraging image processing methods and refined feature extraction, the proposed model aims to recognize subtle visual cues investigative of diseases, such as stain, lesions, and other signs[7]. As there is a need for medicinal plants continues to rise, the proposed model holds the possible to transform disease detection in a manner that aligns with supportable and accurate agriculture systems. This research contributes to the extensive field of plant pathology by contributing an novel and effective methods for plant health observation, eventually facilitates the production of quality medicinal plants and promising the continuous availability of natural remedies for various infections.

In this research, the focus is on directing the challenges in identifying and classifying diseases in medicinal plant leaves. The physical examination methods are cost effective and error-prone, prioritize the need for an automated, precise, and proficient solution. The purpose of the research is to enhance the logistic regression model, and conventional machine learning technique, to more successfully address the minutiae associated with plant diseases. The proposed model will be intended to identify assorted disease patterns, with discoloration and lesions, in various medicinal plant kinds and ecological conditions.

The motivation for the research on medicinal plant leaves disease recognition using modified logistic regression (MLR) deceit in the vital substance of medicinal plants for pharmaceutical and conventional medicine industries.

The key motivations include improving exactness and competence in disease recognition, enabling automated and timely mediation for disease management, optimizing resource consumption through meticulousness agriculture, contributing to agricultural novelty, addressing global health needs by establishing the quality of medicinal plants, and accommodating well-versed decision-making for farmers.

The key challenges contain adapting the model, improve feature engineering, obliging data multiplicity, ensuring competence and extensibility, and validating the model's relevancy. Productively get control of these challenges is estimated to result in a convenient and consistent tool for early disease identification in medicinal plant leaves, contributing to exactitude agriculture and legitimate agriculture practices.

Highlighting the key contributions made in current study as, the study focuses on implementing modifications to the traditional logistic regression algorithm, specifically tailored for images with uneven illumination. To address the challenges posed by irregular intensity, a pre-processing technique is employed to reduce noise, enhance contrast, and improve feature detection within these images. An effective disease detection framework is also integrated into the system, enabling the accurate identification of various disease types. Experimental validations are presented in the results and discussion section, where the performance of the proposed method is evaluated using specific parameters. Additionally, the proposed approach is compared with baseline methods to highlight its advantages and effectiveness.

The subsequent section of this paper investigates background study in section-2, proposed methodology in section-3, experimental results in section-4, and finally conclusion and future scope in section-5.

2. Literature review

The inclusion of a wide-ranging literature study in this research is necessary for several crucial reasons. This section is essential in guiding the appreciative of the background study of plant disease detection methods. Firstly, the literature study aims in the recognition of established methods in the area. By reviewing past work, the pros and cons of a variety of methods become apparent, contributes a initial knowledge base for further implementations. Secondly, it provides benchmarking analysis of existing methods allows researchers to review the proposed model against other learning algorithms or conventional methods, enhances its originality and latent advancements. Furthermore, the literature review helps researchers realize the challenges and confines faced by current disease detection methods. This understanding is crucial in identifying gaps in existing approaches.

The research by Mathew, Mahesh, et al.[8] uses deep learning, specifically the YOLOv5 model, to identify bacterial spot disease in bell pepper plants. This method, which collects images from various parts of the farming area, uses a mobile phone to detect the disease by identifying visible symptoms on the plant leaves. The model is quick and accurate, identifying even minute disease spots. The study's significance lies in its practical application for farmers, enabling timely and effective measures to stop the grow of the disease and mitigate potential yield losses. The model's utility is highlighted, enabling farmers to identify plant diseases as early as they appear, contributing to improved plant health and agricultural production. The research aims to provide an effective method for identifying bacterial spot infection in plant leaves of bell pepper.

Meenakshi et al. [9] have developed a method for automatically identifying diseases in medicinal plant leaves using image processing, DL and ML. They use the Modified Logistic Regression(MLR) ML algorithm, adaptive gamma adjustment, k-mean clustering, and GLCM for quality extraction and classification. The dataset is categorized into different ratios for training and testing, and performance assessment compares the proposed method to existing methods. Results show the proposed method outperforms the existing approach, indicating its efficacy in disease recognition and classification. This research provides a potential solution for farmers for early disease detection.

The study by S. Shreya, P. Likitha et al. [10] highlights the importance of agriculture in India, which employs a large portion of the rural population and contributes significantly to the GDP. The research focuses on identifying and organizing diseases affecting various types of leaves, such as Brown Spot, Late Blight, Bacterial Spot, Septoria Leaf Spot, and yellow-curled. The authors propose a novel approach using image processing methods and a convolution algorithm, employing the K-Means clustering method to partition the dataset into non-overlapping clusters based on attributes. The study also emphasizes the consistency and competence of deep learning models, particularly CNN, compared to statistical models during computation.

The automated approach to plant disease detection presented by V. Sathiya, M. S. Josephine, and colleagues [11] will be particularly helpful for large agricultural operations in reducing demands for continuous monitoring. The rapid and reliable diagnosis of leaf diseases is highly important, since the leaves are important for plant nutrition. In this research, a convolutional neural network (CNN) based model is introduced for automating the detection of diseases affecting Basil and Mint leaves. Advancements in CNN technology allow the study to leverage better image processing than that afforded of traditional methods. In specific, the Adam Optimizer is used for model training on the Istance V3 in order to classify disease affecting Basil and Mint plants. The results show promise for the performance with an accuracy of 77.55% for Basil leaves and 70.89% for Mint leaves.

The study by Gong X, Zhang S. et al. [12] introduces a dataset featuring apple leaf diseases to address the deficiencies in existing datasets. They train Faster R-CNN and YOLOv3 techniques on the dataset to identify apple leaf diseases. Comparative studies and evaluation metrics were conducted to evaluate the performance of the deep learning algorithms. The experimental results show the potential of these algorithms, particularly YOLOv3 and Faster R-CNN, for plant disease exposure. To overcome weaknesses, these algorithms may be combined with parameter tunings.

S. Banerjee and A. C. Mondal et al. [13] proposed a novel technique to minimize the loses in the agriculture and balance the ecological process using deep learning algorithms. In. recent years computer vision and object recognition had a potential progress. CNN based models used to identify the leaf diseases based on visual data. The experiments are evaluated with different statistical features finally it is concluded that CNN models outperforming when compare with other state of art algorithms.

Dai M, Sun W, et al. [14] proposed CNN based disease identification for pepper leaves. In this work embedding the CNN based computation is a challenging task therefore this work proposes light weight GoogleNet architecture. The proposed architecture considered a real time dataset with 9183 images which contains 6 types of diseases. The experimental results are cross validated with other algorithms thre exist a 6 percent of improvement in accuracy evaluation.

The study by Rakgadi Grace Malapela, Gloria Thupayagale et al. [15] suggests that home remedies can help prevent COVID-19 without the need for professional supervision. After collecting 526 articles from 2019 to 2022, only 40 were screened, with 4 articles excluded due to inappropriate abstracts. The study suggests that medicinal plants, such as Ayurveda, homeopathy, and yoga, can be used to improve immunity without prescription. Additionally, leaves like lemon, guava, and honey can be used to treat COVID-19. The study concludes that plant-based products can boost human immunity when used properly.

The SVM and logistic regression are belongs to supervised learning algorithms commonly used for classification tasks. The aim of the proposed work is to improve the probability of classification by introducing half constant as a specific parameter in logistic regression. It often favored for its simplicity and interpretability with reference to SVM while dealing with the binary classification tasks. Introducing half constant in logistic regression could be a way of fine tune the model to extract specific characteristics of the leaf disease detection.

Table. 1. Summary of literature review

Ref. no	Proposed	Findings	Limitations
[8]	Leaf-based disease detection in bell pepper plant using YOLO v5	• The labelled files are configured using YAML. • YOLO v5 is considered for training, testing and classification. • Even small spot is able to identify using YOLO v5.	• Need to extend the disease detection for other diseases effects on bell pepper plant.
[9]	Identification of diseases in medicinal plant leaves using MLR	• Pre-processing using AGCWD • Segmentation using K-Mean • GLCM Feature extraction • Classification modified logistic regression with hyperbolic sine function.	• Need to attain best performance in the parameters named accuracy, recall, precision, F1score and FPR. • Faded intensity noisy and obscure images need to consider for future extension.
[10]	Suggested CNN based method for early leaf disease detection for Tomato, Hibiscus, Spinach, Mango and Bitter gourd	• Pre-processing • K-Mean for segmentation • Kaggle Dataset used • GLCM based feature extraction • CNN classifier	• Colour, shape and texture considered for disease evaluation. • Large dataset need to be considered for future work.

[11]	Classification of plant diseases affecting basil and mint leaves through the use of CNN.	• CNN based approach proposed • Inception V3 model with Adam optimizer for classification	• Model need to improve with hybrid algorithms to attain higher accuracy • The methodology can be applied different plant spices • Need to consider multiclass classification with more than two spices.
[12]	Analysing the identification of plant diseases using deep learning methods.	• Studied significance of deep learning methods like CNN, Fast R-CNN and YOLO v3 • Dataset considered to identify leaf diseases for apple leafs. • Five diseases are considered among them four diseases data is collected from Plant Pathology 2021-FGVC8. • Remaining mosaic disease data is collected from AI Studio. • The study has been conducted with different algorithms, finally they identify YOLO v3 and faster R-CNN are suitable for this dataset.	• Background information has been removed from the dataset. • In future need to accommodate the model for portable devices.
[13]	An Intelligent Approach towards Plant Leaf Disease Detection through CNN model	• Gaussian filter for pre-processing • Colour to grayscale conversion for extracting features • CNN based approach for classification.	• Dull intensity images are not consider for evaluation. • Need to develop Mobile based model with IoT.
[14]	Recognition of pepper leaf disease using improved lightweight convolutional neural networks.	• Lightweight based model GoogleNet is introduced. • Inception structure is reduced in this method. • Experiment is divided into two parts, • First part is training and second one is testing. • SGDM optimizer is introduced over adam. • Dataset is collected from Yangzhou city.	Further it is essential to investigate on dull intensity images and large field environment.
[15]	Utilization of domestic remedies for the management and prevention of coronavirus disease.	• Review has been done to prevent the COVID-19 infections and boost the immunity with medicinal leaves without any side effects. • Various herbal based medicines considered including ginger, clove, black pepper, coriander. • In case of African communities they considered tobacco leaves or powder to enhance the respiratory system.	• The studies finds that how far this traditional home remedies are effective to prevent the COVID-19. • Various home remedies are showing their inference in improving the respiratory system of COVID-19 patients. • WHO also given a clear guidelines, home remedies are alternative treatments to fight with COVID-19.

3. Proposed method

Timely detection of diseases in medicinal leaves and make them available for remedies is a challenging task with uneven illuminated images. Traditional techniques are non-standardized due to image intensity distribution and not able to achieving higher accuracy. In response to overcome above limitations this work proposes an innovative approach in disease detection of pepper, basil and hibiscus leafs with modified logistic regression model. It is also incorporating feature engineering techniques and refinement of algorithm, this approach aims to provide robustness in early disease detection in challenging scenarios. Ultimately the proposed work outcomes large yielding of crop for above leafs to enhance the clinical system of health care. The proposed method consisting five phases as follows.

1. Image Acquisition
2. Pre-Processing
3. Segmentation
4. Feature Extraction
5. Classification and Prediction

The above phases are represented in the following block diagram shown in Fig.3

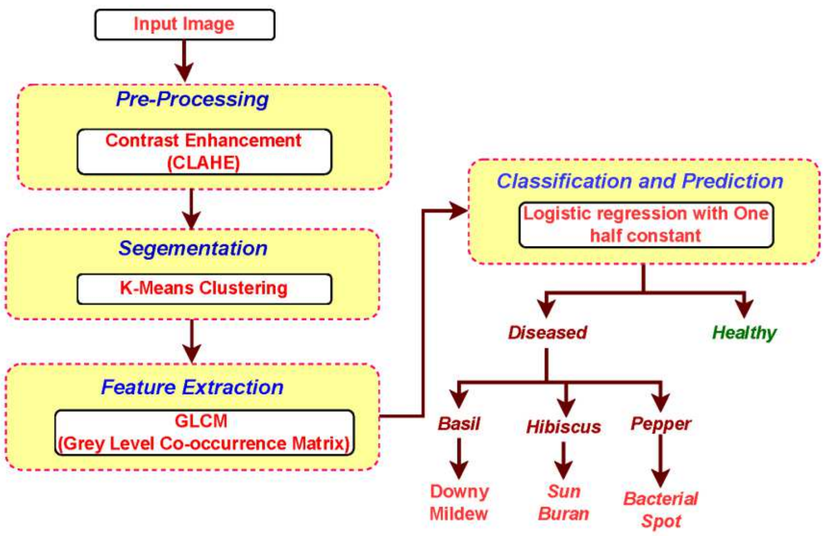


Fig. 3. Block diagram of proposed architecture

3.1 Image Acquisition

Image acquisition is the procedure of collecting visual information from various sources. The proposed method collected the required images from the web sources like Kaggle and UCI machine learning for pepper, hibiscus and basil plants, while preprocessing involves noise elimination and image enrichment. The acquired images are stored with associated metadata, and they serve as the base for applications in computer vision, medicinal imaging, and remote sensing.

3.2 Pre-Processing

This procedure confines of histogram equalization in cases where image gray levels are erratically distributed. It introduces Contrast Limited Adaptive Histogram Equalization (CLAHE) as a clarification, particularly effective for images with erratic intensity circulation[16]. The method redistributes gray levels across adjacent regions and prevents disproportionate intensification over noise, a major advantage over traditional histogram equalization. CLAHE employs a clip limit to restrict intensification by assessing the slope of the transformation function, which is proportional to the cumulative allocation function in histogram equalization [17]. The anticipation of intensification over noise is illustrated in Fig.4 conspicuously, CLAHE was at first implemented in aircraft cockpits.

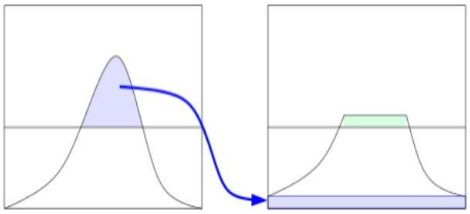


Fig. 4. Intensification limitation (shaded green region)

This algorithm is intended to enhance local dissimilarity in images by tuning the contrast of each pixel within its immediate region and capitalize contrast over the complete image, CLAHE focuses on enhancing the contrast at local level [18]. Fig.5 shows the effect of CLAHE method on sample leaf image. The important steps of the CLAHE algorithm are as follows:

1. Segregate the image into unbroken and non-overlapping blocks.
2. Snip the histogram of each block above a specified threshold.
3. Dispense the snipped pixels to all gray levels consistently, avoiding unnecessary intensification.
4. Apply histogram equalization separately to each block.
5. Interpolate the intensity delineate between adjacent blocks.
6. The mapping at any pixel is built from the intensity mappings of four adjacent tiles.

(a) (b)

Fig. 5. a) Hibiscus original leaf image b) Contrast enhanced leaf image using CLAHE

3.3 Segmentation

In this process K- mean clustering algorithm is used. The procedure of K-Mean for segmentation is described as follows. K-Means clustering is a admired unsupervised machine learning algorithm used for dividing a dataset into K different, non-overlapping subsets (clusters)[19]. The goal of K-Means is to assemble related data points collectively and determine fundamental patterns or structure in the data [20]. The key steps of K-Mean as follows.

Initialization:

1. Select the number of clusters (K)
2. Arbitrarily initialize K cluster centroids. Each centroid represents the center of one cluster.

Assignment Step:

3. For each data point, compute the distance to each centroid.
4. Assign the data point to the cluster whose centroid is the adjoining (usually using Euclidean distance).

Update Step:

5. Recalculate the centroids of the clusters build on the data points assigned to each cluster in the earlier step.

Repeat:

6. Replicate the assignment and revise steps repeatedly until convergence.
7. Convergence occurs when the centroids no longer change considerably, or a particular number of repetitions is reached.

Result:

8. The concluding result is K clusters, and each data point be affiliated to one cluster.

The result of K-Mean segmentation (with K=3) on Hibiscus leaf image is shown in Fig.6

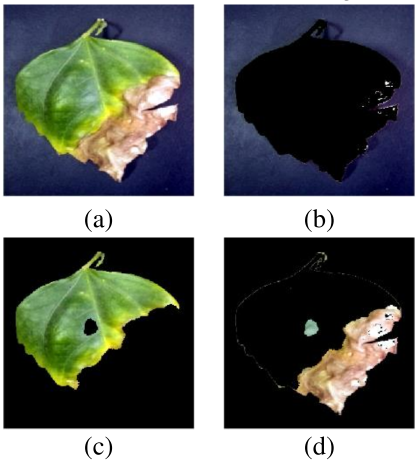


Fig. 6. a) Hibiscus original leaf image b)-d) Cluster images

3.4 Feature Extraction

There are two categories of Haralick's texture feature extraction techniques[21]. The co-occurrence matrices [22]must first be calculated, and then texture features must be determined using the co-occurrence matrices that have been determined. The Gray-Level Co-occurrence Matrix (GLCM) matrix calculation is shown with an example. The GLCM is a computational technique in image processing used for texture investigation[23][24]. It concerns the construction of a matrix that stores the co-occurrence of any two pixel intensities at a particular displacement in a gray scale image.

This measure is also the most familiar and commonly used at the present for texture features extraction and is known as the co-occurrence matrix. It estimates the image attributes using second order statistical information. A co-occurrence matrix sums the chances of finding a pair of pixel of a certain separation in an image. On this basis, a method called GLCM algorithm is employed. The matrix reflects the chance for pixels to be matched with another pixel $P_{d,\theta}(i,j)$ of the i^{th} and j^{th} values, and arranged at a distance of d and position angle θ. The typical values of angular direction θ which can be used are 0^o, 45^o, 90^o, and 135^o. Therefore, there are four pixel neighbourhood relations required for matrix computations as it is depicted in the Fig.7. The matrices produced for each of the four strategies are as follows: The following formulae were used to derive the matrices in the context of the four strategic directions as shown in equations 1 to 4.

$$P(i,j,d,0) = \left|\left\{\left((k,l),(m,n)\right)\epsilon(N \times M)^2 while\ (k-m=0, |l-n|=d, I_{k,l}=i, I_{m,n}=j)\right\}\right| \quad (1)$$

$$P(i,j,d,45) = \left|\left\{\begin{matrix} ((k,l),(m,n)\epsilon(N \times M)^2 while \\ (k-m=d, l-n=-d)\vee(k-m=-d, l-n=d), I_{k,l}=i, I_{m,n}=j \end{matrix}\right\}\right| \quad (2)$$

$$P(i,j,d,90) = \left|\left\{\left((k,l),(m,n)\right)\epsilon(N \times M)^2 while\ (|k-m|=d, l-n=0, I_{k,l}=i, I_{m,n}=j)\right\}\right| \quad (3)$$

$$P(i,j,d,135) = \left|\left\{\begin{matrix} ((k,l),(m,n)\epsilon(N \times M)^2 while \\ (k-m=d, l-n=d)\vee(k-m=-d, l-n=-d), I_{k,l}=i, I_{m,n}=j \end{matrix}\right\}\right| \quad (4)$$

Where, (k,l) representing the grayscale pixel at $i\epsilon[0, nmax-1]$
(m,n) representing the grayscale pixel at $j\epsilon[0, nmax-1]$

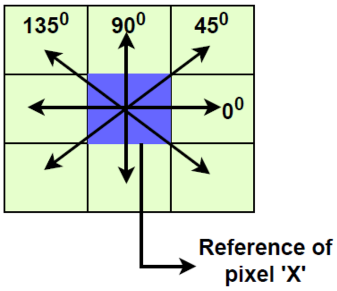


Fig. 7. Pixel 'x' closest neighbors in all directions

Fig. 8. showing an example of how to calculate P(i, j) from an image of size 4 × 4 with grayscale values 0, 1, 2 and 3. The only case considered in the example is d=1 and θ= 0.

The matrix element value is P(1,0)=4 when element (2,3) is considered, which means that there are four instances in the image where a pixel with grayscale value 2 is horizontally separated by a distance of 1 from a pixel with grayscale value 3. These instances are highlighted with gray lines in Fig.8.

The GLCM matrix is derived by including synchronized pixel intensity pairs, and various texture characteristics, such as energy, contrast, correlation, and homogeneity [25], can be extracted from it.

$$Energy = \sum_{i,j=0}^{N-1} \left(P_{ij}\right)^2 \tag{5}$$

Measures the texture uniformity that is pixel pair repetitions.

$$Contrast = \sum_{i,j=0}^{N-1} P_{ij}(i-j)^2 \tag{6}$$

Measures the spatial frequency of an image.

$$Correlation = \sum_{i,j=0}^{N-1} P_{ij}\frac{(i-\mu)(j-\mu)}{\sigma^2} \tag{7}$$

Measures the linear dependency of gray levels of neighboring pixels.

$$Homogenetiy = \sum_{i,j=0}^{N-1} \frac{P_{ij}}{1+(i-j)^2} \tag{8}$$

Measures the variations in image intensity.

0	0	1	2
0	1	3	2
0	2	3	2
1	2	3	0

$$\text{P(i, j,1, 0°)=} \begin{pmatrix} 2 & 2 & 1 & 1 \\ 2 & 0 & 2 & 1 \\ 1 & 2 & 0 & 4 \\ 1 & 1 & 4 & 0 \end{pmatrix} \qquad \text{P(i ,j,1, 45°)=} \begin{pmatrix} 2 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 5 & 0 \end{pmatrix}$$

$$\text{P(i, j,1, 90°)=} \begin{pmatrix} 4 & 2 & 0 & 1 \\ 2 & 0 & 1 & 1 \\ 0 & 1 & 6 & 0 \\ 1 & 1 & 0 & 4 \end{pmatrix} \qquad \text{P(i ,j,1,135°)=} \begin{pmatrix} 0 & 1 & 2 & 2 \\ 1 & 0 & 1 & 1 \\ 2 & 1 & 0 & 2 \\ 2 & 1 & 2 & 0 \end{pmatrix}$$

Fig. 8. Co-occurrence matrices got from 4x4 image with 4 greyscale values.

Using these four primary parameters, an evaluation was conducted on nine additional features: mean, standard deviation, variance, entropy, root-mean-square (RMS), smoothness, kurtosis, skewness, and inverse difference moment (IDM).

Mean (9)

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

Standard deviation (10)

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

Entropy (11)

$$E(F) = \sum_{i=0}^{G-1} p(i) \, log_2 \, p(i)$$

RMS (12)

$$RMS = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{N}}$$

Variance (13)

$$\sigma^2 = \frac{\sum (X - \mu)^2}{N}$$

Smoothness (14)

$$SI = \sqrt{\sum (TS_{max} - TS_i)^2}$$

Kurtosis (15)

$$Kurtosis = \frac{1}{n} \frac{\sum_{i=1}^{n} (X_i - \bar{X})^4}{S^4}$$

Skewness (16)

$$Skewness = \frac{1}{n} \frac{\sum_{i=1}^{n} (X_i - \bar{X})^3}{S^3}$$

IDM (17)

$$IDM = \sum_{i,j} \frac{P_{i,j}}{1 + (i - j)^2}$$

GLCM is important for tasks like texture classification, segmentation, and object recognition, as it gathers information about the spatial allocation of pixel intensities in an image. The method's parameters, including spatial offset and gray levels, can affect the computed texture features. Extracted GLCM features for Basil leaf are shown in Fig.9.

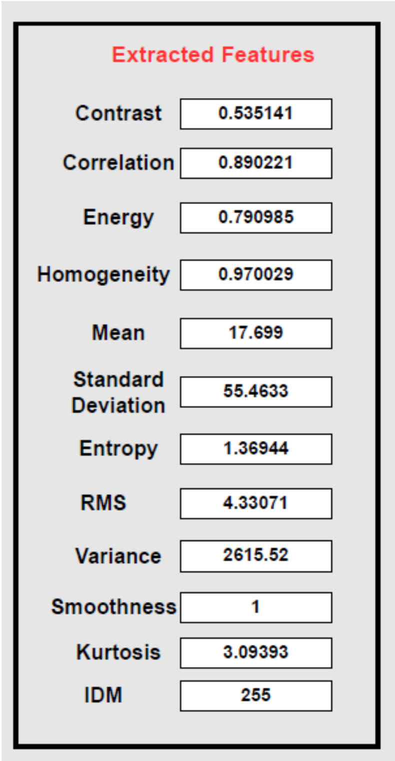


Fig. 9. Extracted features for Basil leaf input image

3.5 Classification and Prediction

In implementing this procedure modified logistic regression with One Half Constant is used. The detailed description of the methods as follows.

3.6 Multinomial Logistic Regression

Logistic Regression is a flexible and interpretable algorithm for binary and multiclass classification, extensively employed for its effortlessness and efficiency in various fields[26][27]. In case of binary classification, it is suitable for problems where the reliant variable is binary, meaning it has only two possible solutions (usually 0 or1), where in multi class classification it can be extended to hold multiclass classification with one-vs-all (OvA) or one-vs-one (OvO)[28]. Algorithm.1 shows the multinomial logistic regression. In the equation (18), the model uses the sigmoid (logistic) function to apply to the linear combination of input features and generates a probability 0 to 1 final output.

$$g(z) = \frac{1}{1+e^{-(z)}} \tag{18}$$

For the above equation One Half Constant is introduced to attain better classification and prediction.

One Half Constant (OHC):

One-half, depicted as the fraction$\frac{1}{2}$, is a fundamental mathematical concept. It represents the division of a conception into two equal parts. The relationship between multiplication by one-half and division by two, as well as the thought of halving and doubling, is a basic arithmetic rule.

In several contexts, one-half is normally used in mathematical equations, measurements, and other concern where a division into two equal parts is applicable. The concept of one-half is fundamental in perceive fractions and proportions. Fig.10 describing a geometric structure involving a square with a side length of one, which is dissected into rectangles. The areas of these rectangles come out to be successive powers of one-half.

Let's consider the analysis of the square into rectangles where the areas follow a structure related to powers of one-half. It start with a square of side length one, the area of the square is $1 \times 1 = 1$.

Dissect the square into two rectangles, each with an area of $\frac{1}{2}$, it would represent $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$ of the original. Continue this process and dissect one of the leftover squares into two rectangles, each with an area of $\frac{1}{4}$ will get rectangles with areas $\frac{1}{8}$ each. This structure continues, and at each step, essentially dividing the area of the leftover square by two, creating rectangles with areas that are successive powers of one-half: $\frac{1}{2^2}, \frac{1}{2^3}, \frac{1}{2^4}$ and so on.This geometric content is often used to exemplify the concept of infinite geometric series and the sum of an infinite geometric progression where the common ratio is one-half. The sum of such a series bound to 1, reflecting the total area of the original square.

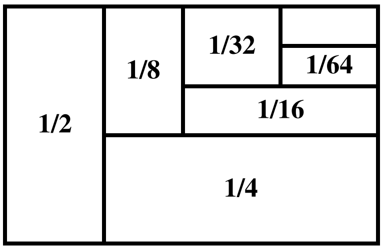


Fig. 10. One half constant

For example, the area S of a triangle can be calculated.

$$S = \frac{base}{2} \times perpendicular\ height \tag{19}$$

One half is also involved in the formulas used to compute figurate numbers, including triangular and pentagonal numbers.

$$\frac{n}{2}[(S-2)n-(4-S)] \tag{20}$$

and the formula for computing magic constant for magic squares

$$M_2(n) = \frac{n}{2}(n^2 + 1) \tag{21}$$

The Riemann hypothesis states that every nontrivial complex root of the Riemann Zeta function has a real part equal to $\frac{1}{2}$.

Algorithm 1. Multinomial logistic regression with modified sigmoid function

Input: Input consists of data (features and labels) and hyper parameters.
Output: Output consists of trained model parameters and predictions.

 # Initialize parameters
 # Parameter for modified sigmoid function
alpha = 1
learning_rate = 0.01
iterations = 250
 # Initialize weights randomly or with zeros
num_features = nf # Number of features
num_classes = nc   # Number of classes
weights = initialize_weights(num_classes, num_features)
 # Training loop
for iter from 1 to iterations:
    # Forward pass
    for each training example:
        # Calculate weighted sum of inputs
        z = dot_product(weights, features)

```
        # Apply modified sigmoid function
        activation = modified_sigmoid(z, alpha)

        # Apply softmax to get probabilities for
each class
        probabilities = softmax(activation)

    # Backward pass
    for each training example:
        # Calculate error
        error = predicted_probability -
actual_probability

        # Update weights using gradient descent
        weights = update_weights(weights, error,
learning_rate, features)

# Predictions
for each test example:
    z = dot_product(weights, features)
    activation = modified_sigmoid(z, alpha)
    probabilities = softmax(activation)
    predicted_class = argmax(probabilities)
    # output predicted class and probabilities
```

The definition of function dot_product is given below.

```
# Dot product function
function dot_product(weights, features):
    result = 0
    for i from 0 to length(weights) - 1:
        result += weights[i] * features[i]
    return result
```

The definition of function exp is given below.

```
# Exponential function
function exp(x):
    result = 1
    for i from 1 to some_large_number:
        term = power(x, i) / factorial(i)
        result += term
    return result
```

The definition of function sum is given below.

```
# Sum function
function sum(x):
    result = 0
    for i from 0 to length(x) - 1:
        result += x[i]
    return result
```

The definition of function modified_sigmoid is given below.

```
# Modified sigmoid function with one half
constant
function modified_sigmoid(z, alpha):
    return 1 / 2* (1 + exp(-alpha*z))
```

The definition of function softmax is given below.

```
# Softmax function for multinomial logistic
regression
function softmax(z):
    exp_z = exp(z)
    sum_exp_z = sum(exp_z)
    result = []
    for i from 0 to length(z) - 1:
        result[i] = exp_z[i] / sum_exp_z
    return result
```

The definition of function initialize_weights

```
# initialize_weights function
function initialize_weights(num_classes,
num_features):
    weights = new
Array[num_classes][num_features]
    for i from 0 to num_classes - 1:
        for j from 0 to num_features - 1:
            weights[i][j] = random_initialization()
    return weights
```

4. Result and discussion

4.1 Dataset

The objective of this research is to early detection of plant leaf diseases. To implement this different kinds of plant diseases are considered. The basil leaf data is collected from Railway koduru region, Annammaiah District, Andhra Pradesh, India. Healthy leaf images 500 and 960 diseased leafs. In case of hibiscus 400 healthy images and 950 unhealthy sunburn images are collected from Korrapadu and Gopavarm villages belongs to Proddatur city, YSR Kadapa district, Andhra Pradesh, India. For pepper leafs dataset is taken from Kaggle[29] with 1470 healthy leafs and bacterial leaf spot 990 leaves. Further dataset images are classified into two types training and testing under healthy and diseased categories. Training and testing sets are considered into different ratios 80:20, 70:30, 60:40 and 50:50. In order to analyze the performance of algorithms, it is necessary to define a confusion matrix. Fig.11 represents the impact of the confusion matrix in various parameter evaluations.

4.2 System Specification

All the models were developed and deployed using MATLAB, leveraging its Machine Learning Toolbox for implementing and evaluating various machine learning models. Computation was accelerated through experiments using a high-performance GPU.For disease detection, a LR model was carried out on a desktop machine with platform running MATLAB R2021b on Windows 10. The system configuration utilized in conjunction with a GeForce RTX 3080_{Ti} version with 32 GB graphics processing unit on an Intel Core i7 processor. At the time of experimenting, the Logistic Regression with one half constant was used to train the model, with the training process set epochs=250 epochs and learning rate λ=0.01. The dataset images are resized to 256x256 resolution. The dataset was split into different ratios like 80:20, 70:30, 60:40 and 50:50 for training and validation. We used a custom Logistic Regression model tailored to handle the specific characteristics of the dataset, which includes leaf images categorized as healthy and diseased. An experimental study was conducted on the pepper, basil and hibiscus datasets to verify the effectiveness of the proposed Logistic Regression model within MATLAB. The results were compared against the performance of previous classification models discussed in our study, demonstrating the Logistic Regression with one half constant model's capability in handling the classification tasks on this datasets.

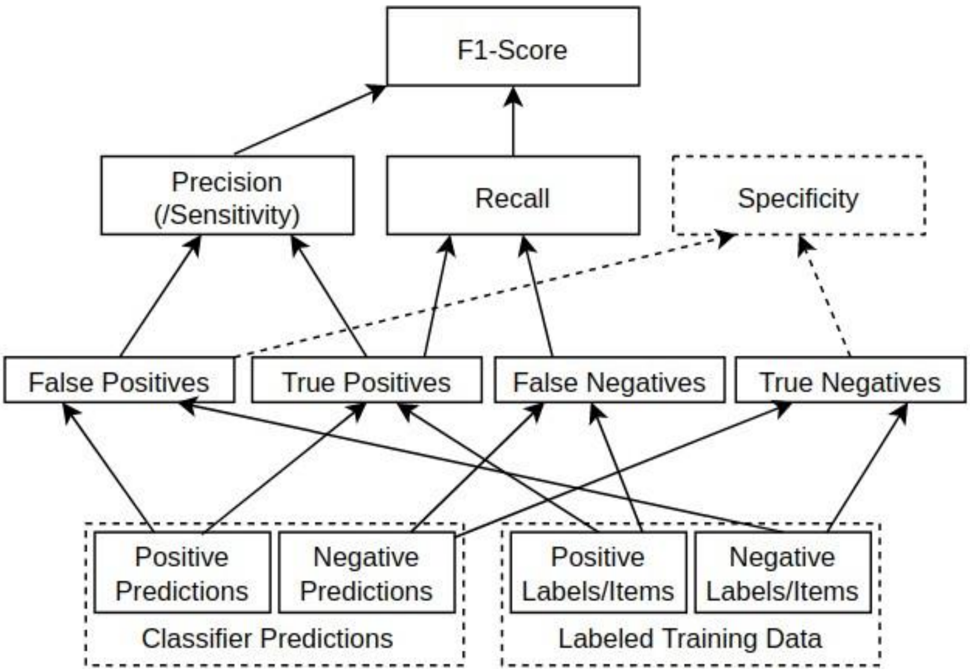


Fig. 11. Impact of confusion matrix in parameter evaluation

This kind of plot representation against classification or prediction is termed a confusion matrix. In the representation of the confusion matrix, two kinds of values are used: actual and predicted values. Further, these two values are classified based on binary classification representation. Fig.12 shows the confusion matrix. The confusion matrices of Basil, Hibiscus and Pepper datasets are shown in Fig. 13.

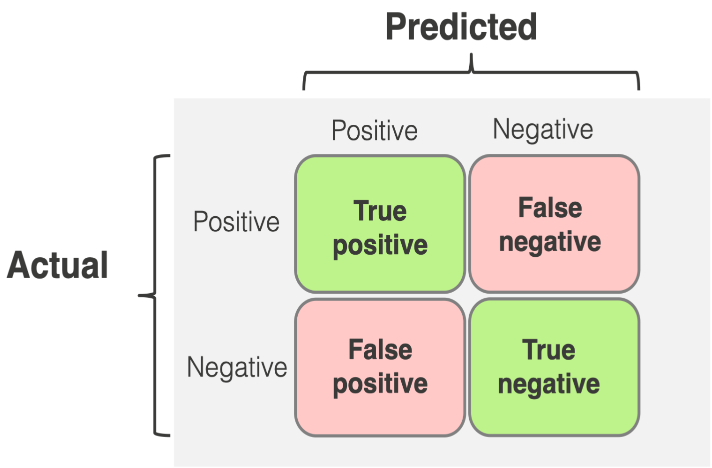


Fig. 12. Confusion matrix

Based on the above four parameters, the following parameters Accuracy, Precision, Recall, FPR (False positive rate) and F-Measure or F1-Score [30] are derived.

$$Accuracy = \frac{\text{No.of correct predictions}}{\text{Total Predictions}} \tag{22}$$

$$\text{Precision} = \frac{\text{No.of correctly predicted possitive instances}}{\text{Total No of possitive and negative predictions observed}} \tag{23}$$

$$\text{Recall} = \frac{\text{No.of correctly predicted possitive instances}}{\text{Total No of possitive instances}} \tag{24}$$

$$\text{FPR} = \frac{\text{FP}}{\text{TN+FP}} \tag{25}$$

$$\mathrm{F-Measure} = 2 * \frac{\mathrm{Precision} * \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \quad (26)$$

Table.2 Model performance evaluation for Basil Training dataset (dull)

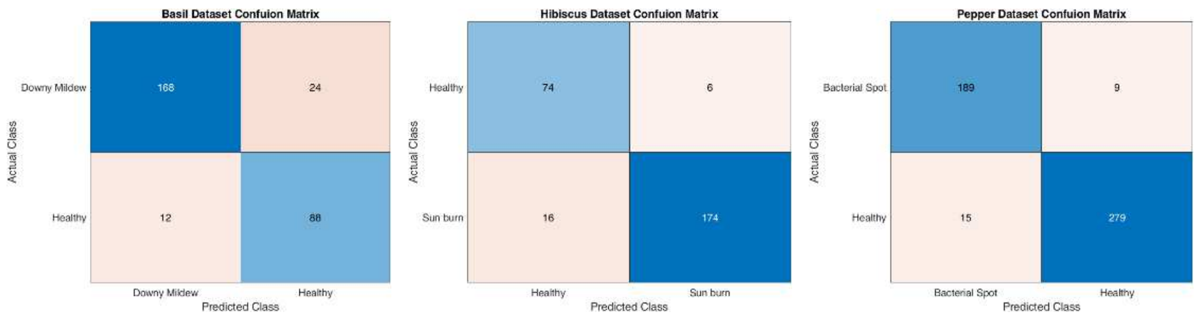
Training and Testing Ratios	Modified SVM with RBF Kernel					Proposed MLR				
	Accuracy	Precision	Recall	FPR	F-Measure	Accuracy	Precision	Recall	FPR	F-Measure
80:20	68.5621	0.7931	0.6889	0.6858	0.6989	87.5000	0.8667	0.8667	0.1333	0.8667
70:30	63.3333	0.7827	0.6332	0.5167	0.6748	87.5000	0.8643	0.8778	0.1222	0.8710
60:40	56.6667	0.7779	0.5667	0.4333	0.6521	84.3750	0.8373	0.8583	0.1417	0.8477
50:50	55.0000	0.7632	0.5500	0.4500	0.6393	85.0000	0.8446	0.8667	0.1333	0.8555

Table.3 Model performance evaluation for Hibiscus Training dataset (dull)

Training and Testing Ratios	Modified SVM with RBF Kernel					Proposed MLR				
	Accuracy	Precision	Recall	FPR	F-Measure	Testing Accuracy	Precision	Recall	FPR	F-Measure
80:20	69.4533	0.7723	0.6824	0.6679	0.6662	91.6667	0.9286	0.9167	0.0833	0.9226
70:30	66.1121	0.7692	0.6432	0.5223	0.6613	94.4444	0.9500	0.9444	0.0556	0.9472
60:40	56.6667	0.7693	0.5873	0.4452	0.6533	87.5000	0.8776	0.8750	0.1250	0.8763
50:50	55.4500	0.7602	0.5544	0.4322	0.6433	90.0000	0.9018	0.9000	0.1000	0.9009

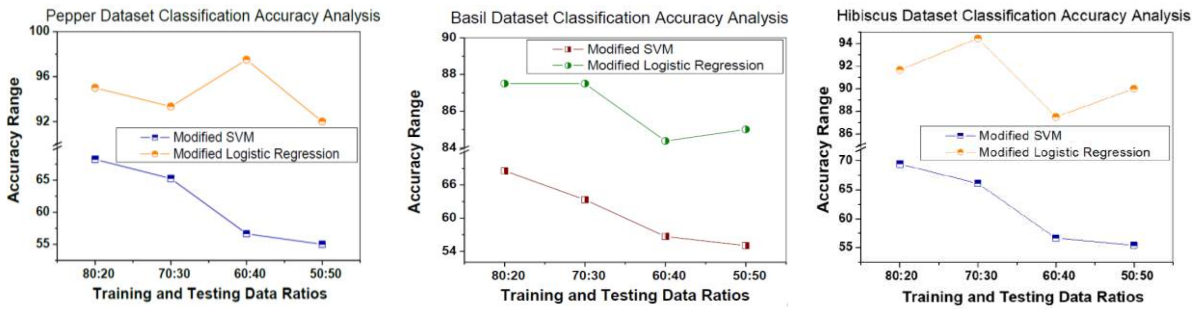
Table.4 Model performance evaluation for Pepper Training dataset (dull)

Training and Testing Ratios	Modified SVM with RBF Kernel					Proposed MLR				
	Accuracy	Precision	Recall	FPR	F-Measure	Accuracy	Precision	Recall	FPR	F-Measure
80:20	68.2351	0.7734	0.6833	0.6822	0.6739	95.0000	0.9545	0.9500	0.0500	0.9523
70:30	65.2523	0.7543	0.6745	0.6599	0.6534	93.3333	0.9333	0.9333	0.0667	0.9333
60:40	56.6672	0.7679	0.5667	0.4333	0.6521	97.5000	0.9762	0.9750	0.0250	0.9756
50:50	55.0000	0.7632	0.5500	0.4500	0.6393	92.0000	0.9310	0.9200	0.0800	0.9255



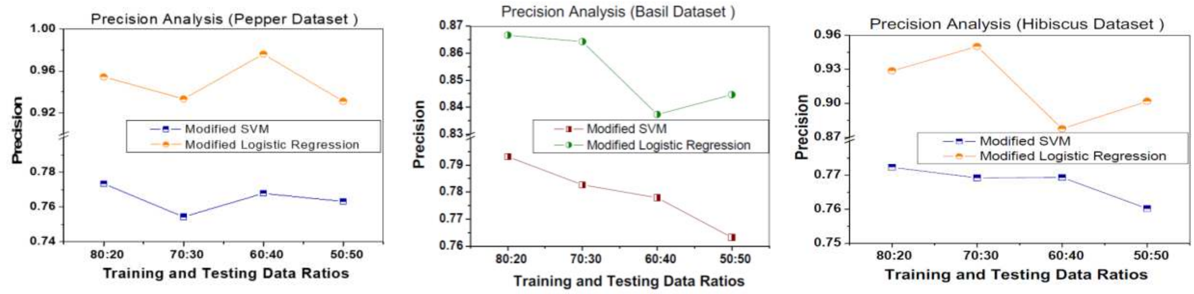
(a) (b) (c)

Fig. 13. Confusion matrix of a) Basil dataset b) Hibiscus dataset c) Pepper dataset



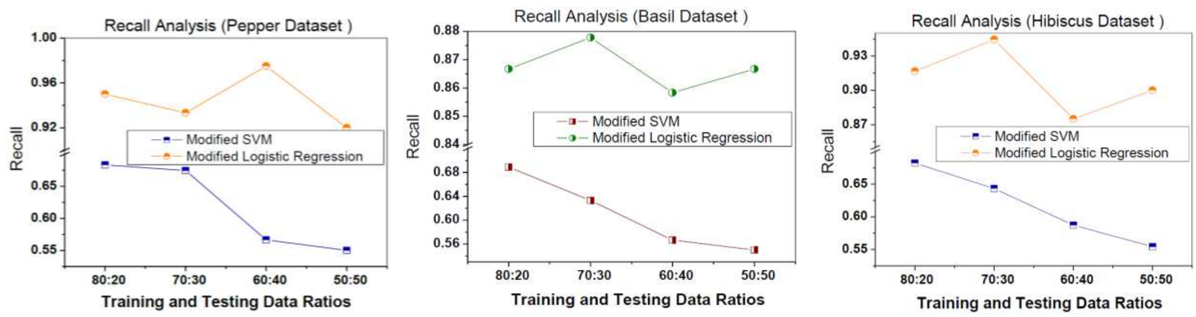
(a) (b) (c)

Fig. 14. Accuracy analysis between modified SVM and Proposed MLR method a) Pepper Dataset b) Basil Dataset c) Hibiscus Dataset.



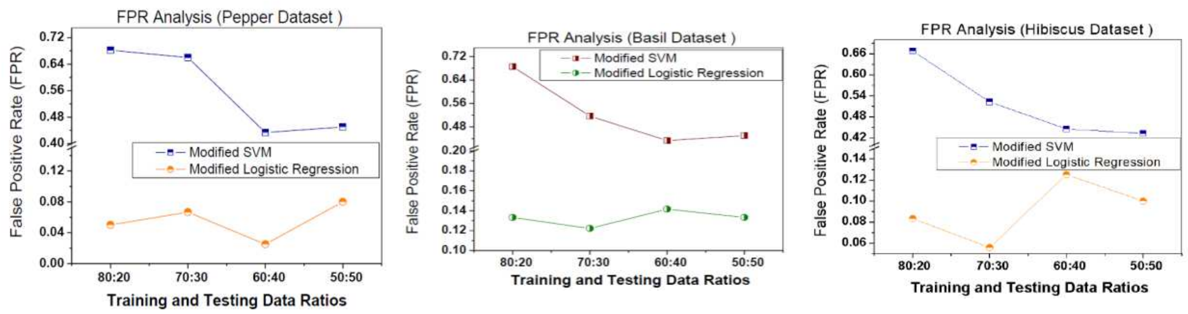
(a) (b) (c)

Fig. 15. Precision analysis between modified SVM and Proposed MLR method a) Pepper Dataset b) Basil Dataset c) Hibiscus Dataset



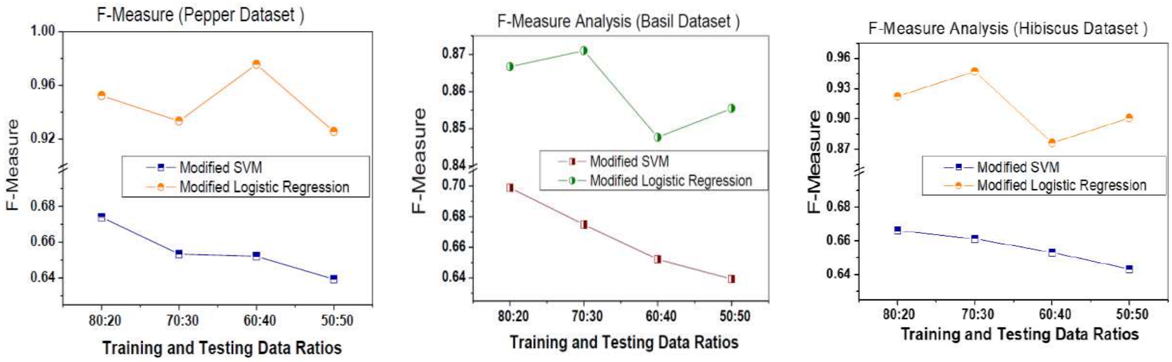
(a) (b) (c)

Fig.16. Recall analysis between modified SVM and Proposed MLR method a) Pepper Dataset b) Basil Dataset c) Hibiscus Dataset



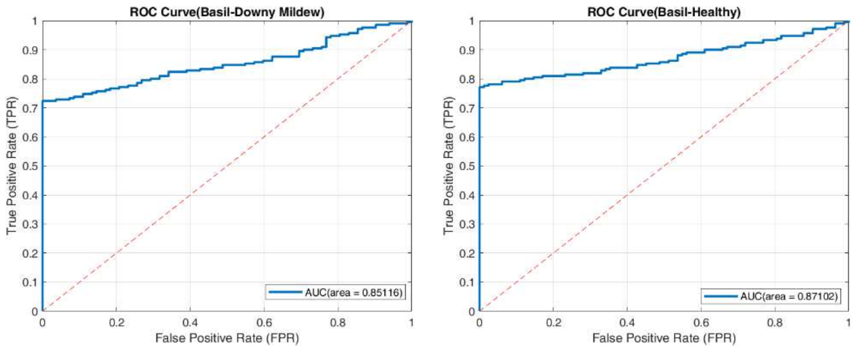
(a) (b) (c)

Fig. 17. FPR analysis between modified SVM and Proposed modified LR method a) Pepper Dataset b) Basil Dataset c) Hibiscus Dataset



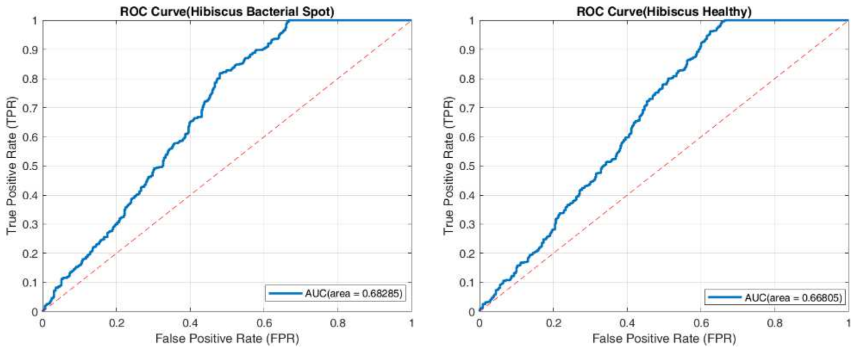
(a) (b) (c)

Fig. 18. F-Measure analysis between modified SVM and Proposed modified LR method a) Pepper Dataset b) Basil Dataset c) Hibiscus Dataset.



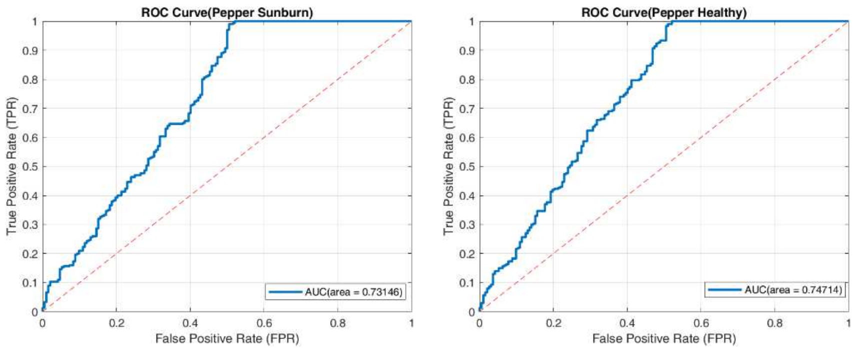
(a) (b)

Fig. 19. ROC curve of a) Basil downy mildew b) Basil Healthy



(a) (b)

Fig. 20. ROC curve of a) Hibiscus Bacterial spot b) Hibiscus Healthy



(a) (b)

Fig. 21. ROC curve of a) Pepper Sunburn b) Pepper Healthy

Fig. 14 presents graphs showing the accuracy of the modified SVM model for both training and testing across various data splits: The ratios are 80:20, 70:30, 60:40 and 50:50. It is obtained that, the proposed method having 39.3%, 43%, 72% and 67% of improvement for Pepper leafs, 27.6%, 38.2%, 48.89% and 54.54% improvement for Basil leafs, 31.98%, 42.85%, 54.41% and 62.3% improvement for Hibiscus leaves with respect existing modified logistic regression.

The Fig.15 plots evaluation of metric precision for different ratios 80:20, 70:30, 60:40, and 50:50. Further it is observed that 23.41%, 23,73%, 27.12% and 21.98% of growth in pepper leaves, 9.28%, 10.42%, 7.63% and 10.66% of growth in Basil leaves,20.23%, 23.5%, 14% and 18.62% of growth in Hibiscus leaves.

The Recall parameter analysis is presented in the above graph shown in Fig.16 for different ratios, 80:20, 70:30, 60:40 and 50:50. In pepper leaves 39.03%, 38.36%, 72.04% and 67.27% growth is observed. In Basil leaves, 25.8%, 38.62%, 51.45%, and 57.58% growth is obtained. Further for Hibiscus leaves 34.33%, 46.82%, 48.98% and 62.33% of improvement is obtained.

From the graphs shown Fig.17 it can be inferred that FRP for Pepper leaves -92.67%, -89.89%, -94.23% and -82.22% low rate is observed, in case of Basil -80.56%, -76.34%, -67.29% and -70.37 rate is obtained. Further for Hibiscus leaves -87.52%, -89.35%, -71.92% and -76.86% of lower rate is achieved with respect to modified RBF kernel. In case of FRP lower value indicates fewer false positive precisions, which generally necessary.

The F-Measure parameter analysis is presented in the above graph shown in Fig.18 for different ratios, 80:20, 70:30, 60:40 and 50:50. In pepper leaves 41.43%, 42.83%, 49.60% and 44.76% growth is observed. In Basil leaves, 24.01%, 29.07%, 29.99%, and 33.81% growth is obtained. Further for Hibiscus leaves 38.48%, 43.23%, 34.13% and 40.04% of improvement is obtained.

The Receiver Operating Characteristic (ROC) curve is an AUC and is one of the more common benchmark measurements that can be implemented for learning algorithms. ROC is the quite popular graph depicting the true positive rate and the false positive rate of the classification results. The ROC curves for Basil, Hibiscus and Pepper are shown in Fig 19 to 21.

4.3 Baseline model comparison

The Table.5 represented above indicating performance metrics evaluated with different models. The features for the model obtained by using DarkNET-19 [31], in case of base line models and proposed method features are obtained by using GLCM. The peak accuracy achieved by the proposed method was 94.45% (average of all ratios). The second and third highest accuracies obtained with the Linear SVM and Linear SVM+PCA.

Table.5 Pepper leaf disease detection for different models

Model	**Accuracy**	**Precision**	**Recall or Sensitivity**	**F1-Score**
DarkNET-19 Features				
Gaussian NB[31]	85.3%	0.88	0.82	0.82
Gaussian NB+PCA[31]	78.4%	0.80	0.75	0.72
Linear SVM[31]	93.5%	0.94	0.94	0.92
Linear SVM+PCA[31]	93.1%	0.92	0.94	0.91
K-NN.[31]	90.5%	0.87	0.97	0.87
K-NN. +PCA[31]	73.3%	0.69	1.00	0.50
DT[31]	82.5%	0.84	0.80	0.78
DT+PCA[31]	81.3%	0.82	0.81	0.75
GLCM Features				
Logistic Regression [9]	79.59%	0.79	0.79	0.78

SVM Classifier[30]	62.59%	0.76	0.63	0.64
Modified SVM classifier with RBF kernel[30]	68.23%	0.77	0.68	0.67
The proposed method (MLR)	**95.0%**	**0.95**	**0.94**	**0.95**

As a result potential performance obtained with GLCM features when compared with the DarkNET-19 features. Furthermore, the Modified Logistic Regression having the one half constant (proposed method) yielded the highest value in accuracy, precision, sensitivity, and F measure.

Table.6 Hibiscus sunburn leaf disease detection for different models

Model	**Accuracy**
CNN Model[32]	85%
Logistic Regression[9]	52.07%
SVM Classifier[30]	63.2231%
Modified SVM classifier with RBF kernel[30]	69.4533%
The proposed method (MLR)	**90.90%**

The Table 6 indicating the accuracy evaluation of Hibiscus dataset with different models. The proposed model reaches the highest accuracy of 90.90%(average of all ratios). The second and third highest accuracies obtained with CNN model [32] and Modified SVM classifier with RBF kernel.

Table.7 Basil leaf disease detection for different models

Model	**Accuracy**
CNN Model[33]	75.0%
Logistic Regression[9]	66.66%
SVM Classifier[30]	61.11%
Modified SVM classifier with RBF kernel[30]	68.56%
The proposed method (MLR)	**86.09%**

The Table.7 indicating the accuracy evaluation of Basil dataset with different models. The proposed model reaches the highest accuracy of 86.09% (average of all ratios). The second and third highest accuracies obtained with CNN model [33] and Modified SVM classifier with RBF kernel.

In the implement of proposed work following limitations need to consider.

- GLCM might not fully capture all related information about leaf diseases. It is not easy to characterize the texture changes with GLCM.
- Sometimes visible cues are might overlook in texture-based features by GLCM.
- GLCM model having limited generalization capabilities.
- It is need to consider bias performance of proposed classification model.

5 Conclusion

Intelligent systems are crucial for handling real-world challenges, and patterns are essential for extracting features and generating unique classes. Large-scale information gathering is crucial for classification, but it can be complicated by redundant data. Curse dimensionality can help reduce complexity. Pre-processing is essential to remove barriers during data collection and improve classification performance. Agriculture faces challenges due to environmental changes, and diseases affect crop growth and production. Early detection of diseases is crucial to prevent pesticide use and contamination. Medicinal leaves have been used for treating human diseases, but there is a gap in information and identification. The proposed work aims to improve disease detection in medicinal plants and increase crop yields. The proposed method uses CLAHE for disease detection, K-Mean algorithm, GLCM, and Logistic Regression classifier with integral kernel sigmoid function for segmentation, feature extraction, and classification. The proposed system is better than existing disease detection systems, generating more accurate results and improving farmers' lives. It can be a boon to the agricultural sector, advancing crop production and management, and facilitating the growth of per capita income.

Future scope The detection of illnesses in plant leaves is a significant research area, with IoT augmentation focusing on automating the spraying of specific disinfectants. The system can be implemented through web-interface enhancements and mobile apps. Deep learning methods, such as CNNs, can predict more accurately. However, challenges include the availability of large datasets and consistent performance assessment. Despite these challenges, deep learning has the potential to transform disease detection and treatment, leading to more sustainable and effective agriculture.

Author contributions: All authors contributed significantly to this research. S. Kiran and N. Subramanyan conceptualized the study and designed the methodology. M. Bhavsingh and K. Samunnisa were responsible for data collection, preprocessing, and analysis. B. Swathika and Lavanya A. conducted model implementation, experimental evaluation, and result interpretation. Jaime Lloret provided critical revisions, guidance, and final approval of the manuscript. All authors reviewed, refined, and approved the final manuscript for publication.

Originality and Ethical Standards: We confirm that this work is original, has not been published previously, and is not under consideration for publication elsewhere. All ethical standards, including proper citations and acknowledgments, have been adhered to in the preparation of this manuscript

Data Availability The basil leaf images utilized during the implementation of proposed leaf disease detection method are collected from Railway Koduru region, Annammaiah District, Hibiscus leaf images are collected from Korrapadu and Gopavarm villages belongs to Proddatur city, YSR Kadapa district and the Pepper leaf images utilized are available in “Pepper disease classification” repository:

https://www.kaggle.com/code/longqua69/pepper-disease-classification/output.
www.kaggle.com/code/longqua69/pepper-disease-classification/output.

Conflict of Interest: There is no conflict of Interest.

Funding: The research received no external funding.

Similarity checked: Yes.

References

[1] T. N. Quoc and V. T. Hoang, “VNPlant-200 – A Public and Large-Scale of Vietnamese Medicinal Plant Images Dataset,” 2021, pp. 406–411. doi: 10.1007/978-3-030-49264-9_37.

[2] M. Fitzgerald, M. Heinrich, and A. Booker, “Medicinal Plant Analysis: A Historical and Regional Discussion of Emergent Complex Techniques,” *Front. Pharmacol.*, vol. 10, Jan. 2020, doi: 10.3389/fphar.2019.01480.

[3] A. Jain, S. Sarsaiya, Q. Wu, Y. Lu, and J. Shi, “A review of plant leaf fungal diseases and its environment speciation,” *Bioengineered*, vol. 10, no. 1, pp. 409–424, Jan. 2019, doi: 10.1080/21655979.2019.1649520.

[4] C. G. Dhaware and K. H. Wanjale, “A modern approach for plant leaf disease classification which depends on leaf image processing,” in *2017 International Conference on Computer Communication and Informatics (ICCCI)*, Jan. 2017, pp. 1–4. doi: 10.1109/ICCCI.2017.8117733.

[5] M. Ali Khan, “Detection and Classification of Plant Diseases Using Image Processing and Multiclass Support Vector Machine,” *Int. J. Comput. Trends Technol.*, vol. 68, no. 4, pp. 5–11, Apr. 2020, doi: 10.14445/22312803/IJCTT-V68I4P102.

[6] Namita M. Butale, & Dattatraya.V.Kodavade. (2018). Survey Paper on Detection of Unhealthy Region of Plant Leaves Using Image Processing and Soft Computing Techniques. *International Journal of Computer Engineering in Research Trends*, *5*(12), 232–235.

[7] V. Pooja, R. Das, and V. Kanchana, “Identification of plant leaf diseases using image processing techniques,” in *2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR)*, Apr. 2017, pp. 130–133. doi: 10.1109/TIAR.2017.8273700.

[8] M. P. Mathew and T. Y. Mahesh, “Leaf-based disease detection in bell pepper plant using YOLO v5,” *Signal, Image Video Process.*, vol. 16, no. 3, pp. 841–847, Apr. 2022, doi: 10.1007/s11760-021-02024-y.

[9] T. Meenakshi, “Automatic Detection of Diseases in Leaves of Medicinal Plants Using Modified Logistic Regression Algorithm,” *Wirel. Pers. Commun.*, vol. 131, no. 4, pp. 2573–2597, Aug. 2023, doi: 10.1007/s11277-023-10555-5.

[10] S. Shreya, P. Likitha, G. Saicharan, and S. Bhargava Choubey, “Plant Disease Detection Using Deep Learning,” 2023. [Online]. Available: www.ijcrt.org

[11] V. Sathiya, M. S. Josephine, and V. Jeyabalaraja, “Plant Disease Classification of Basil and Mint Leaves using Convolutional Neural Networks,” *Int. J. Intell. Syst. Appl. Eng.*, vol. 11, no. 2, pp. 153–163, 2023.

[12] X. Gong and S. Zhang, “An Analysis of Plant Diseases Identification Based on Deep Learning Methods,” *Plant Pathol. J.*, vol. 39, no. 4, pp. 319–334, Aug. 2023, doi: 10.5423/PPJ.OA.02.2023.0034.

[13] K. Suresh, K. Thapan, K. Vamshi Reddy, K. Polaiah, & T. Abhinav Surya. (2024). A Hybrid Framework for Detecting Automated Spammers on Twitter: Integrating Machine Learning and Heuristic Approaches. *International Journal of Computer Engineering in Research Trends*, *11*(1s), 53–60.

[14] M. Dai *et al.*, “Pepper leaf disease recognition based on enhanced lightweight convolutional neural networks,” *Front. Plant Sci.*, vol. 14, Aug. 2023, doi: 10.3389/fpls.2023.1230886.

[15] R. G. Malapela, G. Thupayagale-Tshweneagae, and W. M. Baratedi, “Use of home remedies for the treatment and prevention of coronavirus disease: An integrative review,” *Heal. Sci. Reports*, vol. 6, no. 1, Jan. 2023, doi: 10.1002/hsr2.900.

[16] R. A. Manju, G. Koshy, and P. Simon, “Improved Method for Enhancing Dark Images based on CLAHE and Morphological Reconstruction,” *Procedia Comput. Sci.*, vol. 165, pp. 391–398, 2019, doi: 10.1016/j.procs.2020.01.033.

[17] P. Musa, F. Al Rafi, and M. Lamsani, “A Review: Contrast-Limited Adaptive Histogram Equalization (CLAHE) methods to help the application of face recognition,” in *2018 Third International Conference on Informatics and Computing (ICIC)*, Oct. 2018, pp. 1–6. doi: 10.1109/IAC.2018.8780492.

[18] S. M. Pizer *et al.*, “Adaptive histogram equalization and its variations,” *Comput. Vision, Graph. Image Process.*, vol. 39, no. 3, pp. 355–368, Sep. 1987, doi: 10.1016/S0734-189X(87)80186-X.

[19] P. Shan, “Image segmentation method based on K-mean algorithm,” *EURASIP J. Image Video Process.*, vol. 2018, no. 1, p. 81, Dec. 2018, doi: 10.1186/s13640-018-0322-6.

[20] S. M.AqilBurney and H. Tariq, “K-Means Cluster Analysis for Image Segmentation,” *Int. J. Comput. Appl.*, vol. 96, no. 4, pp. 1–8, 2014, doi: 10.5120/16779-6360.

[21] A. Dash, P. Kanungo, and B. P. Mohanty, “A modified gray level co-occurrence matrix based thresholding for object background classification,” in *Procedia Engineering*, 2012, vol. 30, pp. 85–91. doi: 10.1016/j.proeng.2012.01.837.

[22] R. M. Haralick, I. Dinstein, and K. Shanmugam, “Textural Features for Image Classification,” *IEEE Trans. Syst. Man Cybern.*, vol. SMC-3, no. 6, pp. 610–621, 1973, doi: 10.1109/TSMC.1973.4309314.

[23] R. A. Anggraini, F. F. Wati, M. J. Shidiq, A. Suryadi, H. Fatah, and D. N. Kholifah, “IDENTIFICATION OF HERBAL PLANT BASED ON LEAF IMAGE USING GLCM FEATURE AND K-MEANS,” *J. Techno Nusa Mandiri*, vol. 17, no. 1, pp. 71–78, Mar. 2020, doi: 10.33480/techno.v17i1.1087.

[24] K. Pankaja and V. Suma, “Leaf Recognition and Classification Using GLCM and Hierarchical Centroid Based Technique,” in *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*, Jul. 2018, pp. 1190–1194. doi: 10.1109/ICIRCA.2018.8597184.

[25] M. T. Meenakshi and P. Rani, “An efficient enhanced medicinal leaf image production using Adaptive Gamma Correction Weighting Distribution (AGCWD) algorithm with guided filter,” *Int. J. Health Sci. (Qassim).*, pp. 1650–1677, Aug. 2022, doi: 10.53730/ijhs.v6nS7.11434.

[26] Walter Ocimati, Sivalingam Elayabalan, & Nancy Safari. (2024). Leveraging Deep Learning for Early and Accurate Pre-diction of Banana Crop Diseases: A Classification and Risk Assessment Framework. *International Journal of Computer Engineering in Research Trends*, *11*(4), 46–57.

[27] C.-Y. J. Peng, K. L. Lee, and G. M. Ingersoll, “An Introduction to Logistic Regression Analysis and Reporting,” *J. Educ. Res.*, vol. 96, no. 1, pp. 3–14, Sep. 2002, doi: 10.1080/00220670209598786.

[28] A. M. El-Habil, “An Application on Multinomial Logistic Regression Model,” *Pakistan J. Stat. Oper. Res.*, vol. 8, no. 2, p. 271, Mar. 2012, doi: 10.18187/pjsor.v8i2.234.

[29] H. Anh, “Pepper disease classification.” https://www.kaggle.com/code/longqua69/pepper-disease-classification/output (accessed May 06, 2024).

[30] T. Meenakshi and P. Rani, “Dull Intensity Medicinal Plant Leaf Disease Detection using Machine Learning Algorithms,” *NeuroQuantology*, vol. 20, no. 8, pp. 7228–7242, 2022, doi: 10.14704/nq.2022.20.8.NQ44745.

[31] A. ÖZCAN and E. DÖNMEZ, “Bacterial Disease Detection for Pepper Plant by Utilizing Deep Features Acquired from DarkNet-19 CNN Model,” *DÜMF Mühendislik Derg.*, pp. 573–579, Sep. 2021, doi: 10.24012/dumf.1001901.

[32] C. Shewale, A. Sawadkar, K. Kadale, P. Sangle, and S. Saswadkar, “Plant Disease Identification Using Convolutional,” vol. 9, no. January, pp. 2320–2882, 2023, [Online]. Available: www.ijcrt.org

[33] S. S. Patil *et al.*, “Tulsi Leaf Disease Detection using CNN,” in *2022 IEEE Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation, IATMSI 2022*, 2022. doi: 10.1109/IATMSI56455.2022.10119405.