Prediction of Crop Leaf Health by MCCM and Histogram Learning Model Using Leaf Region

Vijay Choudhary[1,*], Archana Thakur[2]

[1]Institute of Engineering and Technology, DAVV/ IPS Academy, Institute of Engineering and Science, Indore, India

[2]School of Computer Science & IT, Devi Ahilya University, Indore, India

Abstract

This study introduces a model called the crop leaf health prediction model (CLHPM) that utilizes a bio-inspired method to accurately identify the leaf region. This approach enhances the process of learning important features and overcomes the challenges posed by the hindrance from the chromatic and structural diversity of each leaf. To train the learning model, a modified co-occurrence matrix (MCCM) in texture analysis is used to overcome the limitations of the leaf region, and a histogram method is also deployed for color analysis. The experiment is conducted on a real dataset of tomato crop leaves. It is observed that the average accuracy has increased by 3.50%. The existing MobileNetV2 model presents an accuracy of 95.73%, and the proposed CLHPM model renders 99.23%. Moreover, an enhancement of 3.72 in the F-measure is also noticed.

1. Introduction

Leaf disease classification is defined as the process of discovering and classifying various plant diseases that damage the leaves and is particularly crucial to diagnose the ensuing impact on the health and productivity of crops [1]. Fungi, bacteria, viruses, and environmental variables are the pathogenic factors of leaf diseases. Given the pathogenesis rendering multiple factors, therefore, identifying the exact cause of a leaf disease is essential for determining the optimal treatment or preventative measure [2]. Recently, the automation of leaf disease detection and classification using computer vision techniques has been widely investigated [2]. These techniques instantly determine the type of disease and furnish insightful information by recognizing a sufficient amount of leaf images.

Moreover, with the abundance of samples photographically examined, the fundament of the disease, its origins, and prognosis can be accurately confirmed and formulated. To assess and categorize the photographs of damaged leaves, multifarious mathematical models can be applied, e.g., a BAT algorithm-based model, named BAT-based crop leaf disease prediction bootstrap model (BCDPBM), convolutional neural network (CNN), and support vector machine (SVM) [3]. The aforementioned models can be trained on a set of labeled images in the context of classifying leaf diseases to differentiate between various leaf diseases [4-5].

Similarly, the MobileNet model, which is known for its specialized architecture crafted for mobile and edge devices, features lightweight CNN optimized for efficient inference on resource-constrained platforms is also promising. Furthermore, an improved MobileNet model (light deep neural network for embedded systems and mobiles) can be used to identify leaf

---

* Corresponding author. E-mail address: vij.choudhary@gmail.com

diseases in classification [6]. The selection of a mathematical model is based on the provision of precisely classifying leaf diseases, including the amount, complexity of the dataset, the degree of precision necessary, and the processing resources available.

In this study, the crop leaf health prediction model (CLHPM), which uses a back propagation neural network (BPNN) for learning, is introduced and gives a promising accuracy of 99.23%. The CLHPM average F-measure value exceeds all sets of testing photos. The average F-measure of the crop leaf class identification model is proximate to 0.99.

2. Literature Review

Ma et al. [7] reviewed the medical image segmentation algorithms proposed from several contexts and divided the algorithms into three groups based on different scenarios, i.e., threshold, pattern recognition methods, and deformable models. Currently, the third algorithm based on deformable models is mostly concerned, and, meanwhile, segmenting tissues and organs in the pelvic cavity area is one of the principal uses of these algorithms. In addition to the contribution previously mentioned, Ma et al. [8] categorized manifold algorithms based on the primary techniques, which are presented as follows: key concepts, application areas, benefits, and drawbacks. Further to the application stated above, the authors studied the female pelvic cavity's tissues and organs using the medical image segmentation algorithms that are divided and categorized, which are intended to highlight the distinctive qualities of each segment. Subsequently, the crucial guidelines for creating pelvic cavity segmentation algorithms are approvingly suggested.

The apple leaf disease dataset (ALDD) was constructed with the aid of picture annotation techniques and data mining algorithms by Jiang et al. [9]. The ALDD comprises intricate photos taken in labs and fields and, given such an attribute, a new deep-CNN-based model for the diagnosis of apple leaf disease was presented, with the GoogleNet inception structure and rainbow concatenation. To identify common apple leaf illnesses, the suggested single-shot multi-box detector (SSD) with inception module and rainbow concatenation (INAR-SSD) model was constructed. The hold-out testing dataset, comprising 26,377 photos of apple leaves with diseases, was used as the samples.

Amin et al. [10] suggested an end-to-end deep learning framework to discriminate between healthy and indisposed leaves of maize plants after meticulous evaluation. To extract various properties from the images of corn plants, the authors have suggested models using two post-cluster CNNs, EfficientNetB0, and DenseNet121. The system effectively merges deep features obtained from individual CNNs using the concatenation method to enhance the comprehension of the dataset to create a more intricate feature set. In this study, image variations were added to the database, which was used to train the model using data augmentation techniques. Therefore, photos boosted qualitatively and quantitatively enabled the model to recognize and resolve more complicated data situations.

Devi and Rajan [11] emphasized using digital image processing for plant leaf detection and disease classification. By utilizing CNN for image segmentation, the study aims to enhance precision agriculture, rendering plant infection analysis more accurate and time efficient. The authors examined the factors causing plant infections and explored various passive detection strategies. The process involved pre-processing, segmentation, feature extraction, and classification.

A successful method for diagnosing rice plant diseases based on the phenotypical appearance and color of lesions in a leaf image was proposed by Upadhyay and Kumar [12]. The proposed model performs picture binarization using Otsu's global threshold technique to eliminate image background noise. The suggested model was trained with 4,000 picture samples of both healthy and diseased rice leaves to identify the three rice diseases. The suggested fully connected CNN method was found to be responsive and efficient, with a 99.7% accuracy rate on the dataset, as interpreted from the results. Such precision outperformed existing techniques for identifying and categorizing plant diseases.

Kiran and Chandrappa [13] described that this study aims to develop a highly effective method for plant leaf disease detection using computer vision techniques. Leaf disease detection comprises histogram equalization, denoising, image color threshold masking, and feature descriptors to extract the salient features of leaf images. The aforementioned features are used to classify the images by training logistic regression, linear discriminant analysis, K-nearest neighbor (KNN), decision tree, random forest, and SVM algorithms using K-fold validation. K-fold validation is used to separate the validation samples from the training samples, where K indicates the frequency of the generalization. The training and validation processes are performed in two ways. The first approach uses default hyperparameters with segmented and non-segmented images. In the second approach, all hyperparameters of the models are optimized to train segmented datasets.

Bouni et al. [14] indicated that a hindrance necessitating constant attention throughout the crop cycle and accounts for a sizeable portion of the total production expenses lies in managing tomato infection. Automation and pre-trained deep neural networks for the strategy to classify tomato illnesses were used. Plant disease monitoring can be done via digital image processing. Deep learning has significantly outperformed the traditional methods in the processing of digital images in recent years. In this study, deep CNNs and transfer learning were used to detect tomato leaf disease. As a research gap, it is found that most of the work was done on the training of the image features for disease classification, whereas feature optimization needs to be improved. Several studies consider the whole image as an input training parameter. However, the image should be processed to remove the background region. The input image leaf identification algorithm should be enhanced, as leaf structure is incredibly fruitful in understanding the texture of the image for feature collection. Learning parameters should be reduced, as scarce studies use large feature sets for training and testing.

This study aims to identify the leaf region in the input image, without prior knowledge. Recognized leaf region is used to identify the feature set for plant leaf disease detection, which removes the limitations of leaf color and structure. Extracted features are used to train a mathematical model for leaf classification. Furthermore, this study contributes to a model that removes the background from the image and identifies leaf regions using the bio-inspired algorithm and Gaussian mixture model (GMM). Meanwhile, the study has identified feature combination modified co-occurrence matrix (MCCM) and histogram improving the learning of neural networks. Hence, the current work proposed a novel approach that identifies leaf regions and feature sets for plant disease prediction.

3. The Proposed Crop Leaf Health Prediction Model (CLHPM)

In this section, a new model of CLHPM is proposed. The visual flow of the model is shown in Fig. 1. The whole work is divided into two modules. The first is pre-processing, as shown in Section 3.1, and the second one is the training of the mathematical model, as depicted in Section 3.2. The block diagram of the proposed CLHPM is delineated below.
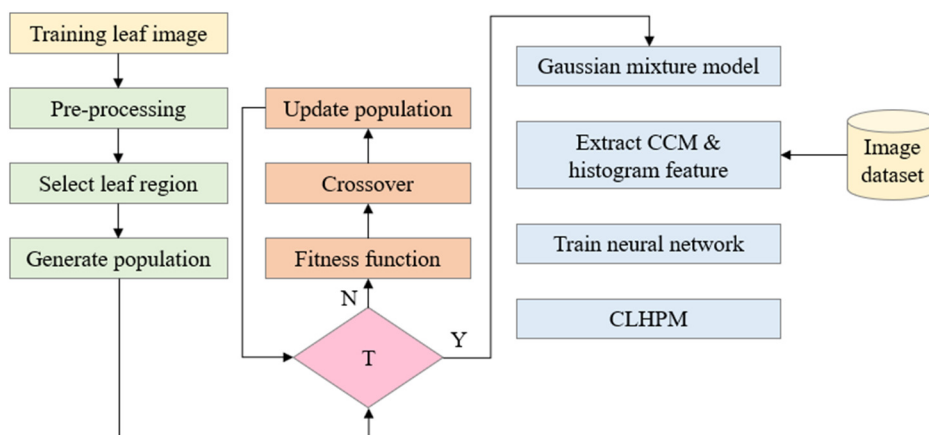

Fig. 1 Architecture of CLHPM model

3.1. Pre-processing

Firstly, the input image dataset is preprocessed to reduce the noise and operational cost of the model. In the pre-processing phase, the image is transformed into a square matrix, which is denoted as processed crop leaf (PCL) [15]. The pre-processing module incorporates multiple steps from selecting leaf region to histogram feature extraction, as shown in Subsections 3.1.1. to 3.1.9.

3.1.1. Selecting leaf region

Secondly, the region of the image containing a leaf apart from the background is found [16]. The selected region image is transformed into Hue Saturation Value (HSV) and gray format, while the input is in RGB format. Subsequently, the leaf region is identified. The input image taken is shown in Fig. 2. After identifying the leaf region, bio-inspired and Gaussian mixture algorithms are used on the resultant image.

Fig. 2 Input leaf image

3.1.2. Bio-inspired algorithm

The algorithm discerns between the foreground and background by working independently on each region. The system creates pixel sets inside each region in turn, assesses their fitness, and modifies the pixel sets according to their fitness ratings. Apropos the algorithm, the ability to thoroughly investigate and improve pixel combinations in both the foreground and background is ensured by this iterative procedure [17]. The dynamic nature of bio-inspired genetic algorithms increases the chances of accurate segmentation without the need for prior training or information. Many existing segmentation models, especially KNN, need prior training for the same conditions. Furthermore, a bio-inspired algorithm is self-adaptable for different sets of complex backgrounds as well.

3.1.3. Population generation

In this process, chromosomes represent the pixel values for each region [18]. Each chromosome, akin to a student S, embodies a potential pixel value set, comprising c cluster center values. Therefore, if the population consists of m chromosomes, the population matrix encompasses $m \times c$ pixels. The selection of pixel values within a chromosome occurs randomly from the available pixel values within region R. The population PCc can be given as follows.

$$PCc \leftarrow GeneratePopulation(m, c, R) \tag{1}$$

3.1.4. Fitness value

Each chromosome is denoted as Cc and undergoes evaluation to determine its fitness value. This evaluation entails computing the summation of the minimum Euclidean distances between each pixel within the region and the pixel set represented by the chromosome. This process resembles clustering evaluation, where a chromosome exhibiting the minimum fitness value emerges as the optimal solution within the population.

3.1.5. Crossover

In this step of the bio-inspired algorithm crossover operation is performed, where according to fitness vector CF, the minimum value position chromosome is acting as best in the current population. Thus, some of the feature values from the chromosome pixel set are copied into other chromosomes of the population. This transfer of feature value from one chromosome to another is termed a crossover operation in a genetic algorithm [19].

3.1.6. Updating population

In this step of the bio-inspired algorithm, the new child fitness value is checked and compared with parent chromosomes. If the fitness value of the parent chromosome is lower than the new child chromosome, the parent will be replaced with the child. In contrast, if the fitness value is higher, the parent will continue, and the new child's chromosome will not be included in the population. After updating the population maximum iteration count is checked and if it has reached its high, the best-obtained solution of the population will act as region pixel representative. Otherwise, the algorithm will be redirected to the fitness value step of the model.

3.1.7. Gaussian mixture model (GMM)

Stauffer and Grimson [16] suggested a versatile customizable GMM to reduce the effect of small recurrent pixels in the foreground of the image. A pixel I located at position x is rendered as a composite of k Gaussian distribution. The present pixel value corresponding to the probability appropriation is provided by the following.

$$P\left(I_x\right) = \sum_{i=1}^{k} W_{x,i} \times \eta\left(I_x, \mu_{x,j}, \sigma_{x,i}^2\right) \tag{2}$$

where η is the density function of the Gaussian distribution.

$$\eta\left(I, \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(1-\mu^2/2\sigma^2\right)} \tag{3}$$

while W, μ, σ^2 represent weight, mean value, and standard deviation of the ith Gaussian in the mixture. To maintain the GMM, the boundaries W must be updated based on the new pixel I_x. Pixels are considered coordinated if I_x exists within standard deviations of a Gaussian. In this situation, 2 is between 1 and 5 for the discovery of modified teaching-learning-based optimization (MTLBO) groups. If one of the k Gaussians is correlated, the correlated Gaussians are updated as follows.

$$\mu_{x,i} = \left(1-\rho\right)\mu_{x,i} + \rho\left(I_x\right) \tag{4}$$

$$\sigma_{x,i}^2 = \left(1-\rho\right)\sigma_{x,i}^2 + \rho\left(I_x, \mu_{x,i}\right)^T \left(I_x, \mu_{x,i}\right) \tag{5}$$

where ρ is given as the rate of learning that rules how quickly μ and σ^2 converge to new observations.

$$\rho = \alpha\eta\left(I_x \big| \mu_{x,i}, \sigma_{x,i}^2\right) \tag{6}$$

The weight of the k Gaussians is accommodated as follows:

$$W_{x,i} = \left(1-\alpha\right)W_{x,i} + \alpha\left(M_i\right) \tag{7}$$

where $M_i = 1$ is set for the identical Gaussian and $M_i = 0$ for all other Gaussian distributions. The rate of learning is utilized to revise the weight; the value ranges from 0 to 1. The allocation with the current pixel value as its mean, a high standard deviation, and a low weight boundary replaces the most unweighted segment if none of the k Gaussian segments match the current pixel value. Subsequently, the loads were rendered uniform.

The k distributions are ordered by W/σ in descending order. This request establishes the most probable foundation with heavy weight and minimal variation at the summit. The principal B Gaussian distributions that exceed a certain edge T are reserved for foundation allocations. If a modest estimate of T is chosen, the foundation model is unimodular, whereas it is multi-modular if a larger estimate of T is opted. If pixel I_x does not match any of the foundation segments, it is designated as the foreground. The foreground image region feature is obtained from the bio-inspired algorithm and is used to identify similar regions in other input images. The image found after the applied method is shown below in Fig. 3.

Fig. 3 Foreground region of the leaf image

3.1.8. Modified co-occurrence matrix (MCCM)

The leaf texture feature is crucial in the identification of image health. Specifically, the image is transformed into RGB and HSV format where red, green, hue, and saturation matrix values are considered [20]. In this paper, the MCCM features are extracted from the image. Instead of using all co-occurrence matrix (CCM) feature sets some of the negative and low values are removed. This updated positive value set is a modified CCM model.

It is observed that with the use of MCCM, learning of the model is boosted. In the present work, four elements are chosen including contrast, energy, inverse difference, and entropy.

$$ID = \sum_{i=1} \sum_{j=1} \frac{1}{1+(i-j)^2} CLR(i,j) \tag{8}$$

where ID is the inverse difference and $CLR(i,j)$ is the intensity value in cell (i,j)

$$Entropy = \sum_{i=1} \sum_{j=1} \frac{1}{1+(i-j)^2} CLR(i,j) \log\left[CLR(i,j)\right] \tag{9}$$

$$Energy = \sum_{i=1} \sum_{j=1} \left[CLR(i,j)\right]^2 \tag{10}$$

$$Contrast = \sum_{i=1} \sum_{j=1} (i-j)^2 CLR(i,j) \tag{11}$$

As per modification, the final feature set C is shown as C = [red energy, red contrast, green energy, green contrast, hue contrast, hue inverse difference, hue entropy, value energy, value contrast].

3.1.9. Histogram

An unhealthy image is visibly evident in the chromatic transformation from green to brown and yellowish. Therefore, color attributes precisely determine the criteria of image classification [21]. Therefore, the histogram feature value set is extracted from the foreground region of the leaf. The image's histogram characteristic is displayed in Fig. 4.
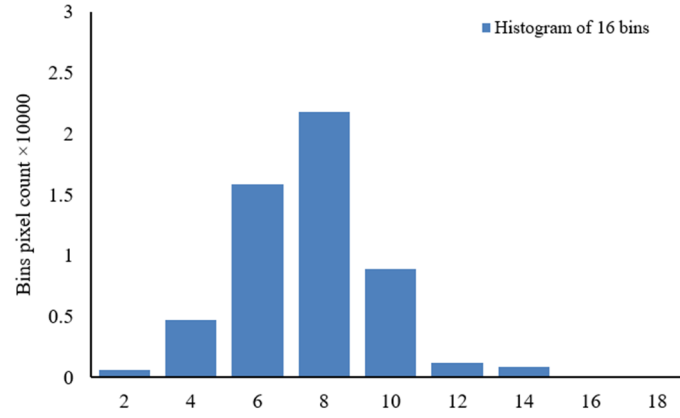
Fig. 4 Histogram feature of the image

3.2. Training of mathematical model

The described module comprises training a neural network with feature vectors and displaying the CLHPM algorithm. After the vector preparation, the feature vectors are sent to a BPNN. The capacity of the neural network for generalization is enhanced, and its ability to classify with the underlying structure of the data is strengthened by incorporating the CLHPM algorithm.

3.2.1. Training feature vector

To train the mathematical model extracted features are combined and normalized for the training. The MCCM proffers 9 feature values as C, and the Histogram provides 16 feature values as H. Hence, the training vector consists of 25 values. MCCM values are normalized by α. The transmissibility function TF is given by the following equation.

$$TF = C \times \alpha \times H \tag{12}$$

3.2.2. BPNN training

A BPNN algorithm is used in the work to learn extracted and optimized feature sets corresponding to the leaf images. The desired output of the model is identified as 1 for the healthy leaf and 0 for the infected leaf. A multilayer neural network comprising two hidden layers is used for learning. BPNN has two hidden layers having 25 neurons in the first hidden layer and 10 neurons in the second layer. The further model turns for approx. 100 epochs for training. The sigmoidal activation function is used for learning the model, the image has non-linear data for training and testing.

3.2.3. Algorithm of CLHPM model

The pseudo-code of the proposed CLHPM model is presented in Algorithm 1 shown below. The algorithm depicts sequentially all the phases of CLHPM including pre-processing of the image followed by the leaf region selection by the bio-inspired algorithm, feature extraction using different feature extraction techniques, and finally the learning phase of the BPNN.

Algorithm 1: CLHPM
Input: ID // image dataset
Output: Trained BPNN // back propagation neural network
Method:
Start
step 1: Loop 1: n // n: number of images in ID
step 2: PCL ← Pre_Process (ICL) // ICL: image of crop leaf
step 3: CLR ← Bio_inspired_Algorithm (PCL) // CLR: crop leaf region

step 4: CLR ← GMM (CLR) // GMM: Gaussian mixture model

step 5: C ← MCCM (CLR) // MCCM features extracted

step 6: H ← Histogram (CLR) // Histogram features extracted

step 7: C ← Normalize (C)

step 8: TF ← [C, H] // Features are combined

step 9: Do ← ID

step 10: End Loop

step 11: BPNN ← Initialize (TF, Do)

step 12: Loop 1: iter // iter: number of iterations

step 13: BPNN ← Train (BPNN, TF, Do) // BPNN is trained

step 14: End Loop

End

4. Experiment Design

The experiment is conducted sequentially using a unique bio-inspired algorithm for preprocessing, MCCM for texture feature extraction, and histogram for color feature extraction. The proposed CLHPM model is trained using a multi-layer BPNN.

4.1. Dataset description and experiments conducted

The experiment is performed on a machine with an intel core i3 processor and 4 gigabytes of RAM hardware configuration. The MATLAB tool is used for developing the proposed model. Experimental work is completed on real datasets of tomato crop leaves. A detailed description of the dataset [22] is depicted in Table 1. Some of the images are taken for the sample from the Kaggle dataset, as described above. Sample images depicting healthy leaves and some of the diseased leaves suffering from various diseases like late blight, late mold, septoria spots, and spider mites, respectively, are shown in Table 2.

Table 1 Image dataset description

Feature	Description
Dimensions	256×256
Type of images	7
Crop	Tomato
Total images	2000

Table 2 Sample images

Type of disease	Sample image	Type of disease	Sample image
Healthy image		Septoria spot	
Late blight		Spider mites	
Late mold			

4.2. Results and discussion

Comparative analysis with different machine learning algorithms must be done for the assessment of the present model [23]. The proposed CLHPM algorithm is compared with the existing algorithm, proposed by Nguyen et al. [6] as an improved MobileNet model. An appreciable work was done by the authors, where the MobileNet model was enhanced. The proposed model is compared with this model and it is found that by using the PlantVillage dataset the accuracy in the case of early blight disease leaf images, and healthy leaf images is 98.68 %, while in the proposed model the accuracy is 98.84%. When the Septoria disease images and healthy leaf images are considered, the improved MobileNet exhibits an accuracy of 98.18%, which is significantly proximal to 100% in the current model. A comparative analysis with the model proposed in Nguyen et al. [6] and similar state-of-the-art models is shown in Table 3.

Table 3 Comparative analysis of the proposed CLHPM with state-of-the-art models

Dataset	Model	Accuracy in %	References
PlantVillage dataset	Fine-tuned MobileNet	97.17	Nguyen et al. [6]
PlantVillage dataset	Support vector machine	91.5	Bhagwat and Dandawate [24]
ImageNet (RGB)	Fine-tuned VGG16	95	Coulibaly et al. [25]
PlantVillage dataset	Fine-tuned MobileNet based with OMNCNN improved	98.7	Ashwinkumar et al. [26]
PlantVillage dataset	Fine-tuned DenseNet	98.17	Kaya and Gürsoy [27]
Kaggle dataset	CLHPM	99.23	Proposed model

A comprehensive comparison with the model, proposed by Ahmed et al. [28] as MobileNetV2 on known hyperparameters, is performed in the next section. Table 4 lists the accuracy of leaf class detection models on diverse types of tomato leaves. As observed from Fig. 5, the use of histogram and MCCM features for the detection of the class has improved the work efficiency. Furthermore, feature extraction from selected regions of the input image has improved the disease detection accuracy. In the previous model MobileNetV2, the use of (CLAHE) alone was done without image leaf region selection, hence accuracy values were compromised.

Table 4 Accuracy value of tomato healthy leaf prediction

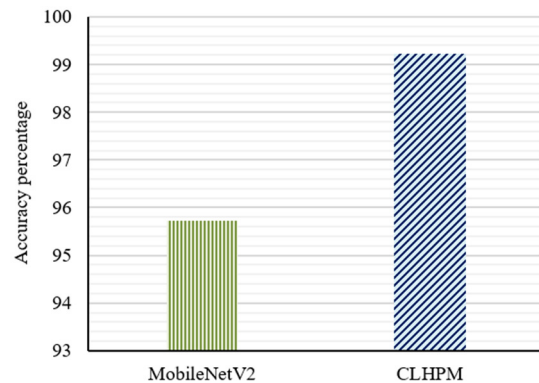| Leaf types	MobileNetV2	CLHPM
Early blight + healthy	94.19	98.84
Bacterial spot + healthy	96.51	97.83
Late blight + healthy	93.02	100
Late mold + healthy	95.35	98.75
Septoria spot + healthy	96.51	100
Spider mites + healthy	98.84	100

Fig. 5 Comparison of average accuracy values of leaf class detection models

Recall values of leaf class detection are shown in Table 5. It is observed that the present CLHPM model yields a high recall value, and an increase in the average recall value of 7.11% is noted, as compared to the MobileNetV2. The use of a bio-inspired algorithm for leaf region detection has escalated the learning model performance.

Table 5 Recall the value of tomato healthy leaf prediction

Leaf types	MobileNetV2	CLHPM
Early blight + healthy	0.8837	0.9767
Bacterial spot + healthy	0.9302	0.9592
Late blight + healthy	0.8605	1
Late mold + healthy	0.907	0.973
Septoria spot + healthy	0.9302	1
Spider mites + healthy	0.9767	1

The F-measure value in Table 6 indicates that the feature combination of MCCM, and histogram with pre-processing by the bio-inspired algorithm has improved the leaf class detection. Fig. 6 shows that the CLHPM average F-measure value is high among all sets of testing images. Learning BPNN has enhanced work performance. The average F-measure of the crop leaf class detection model is graphed in Fig. 6.

Table 6 F-measure value for different leaf classes of tomato healthy leaf prediction

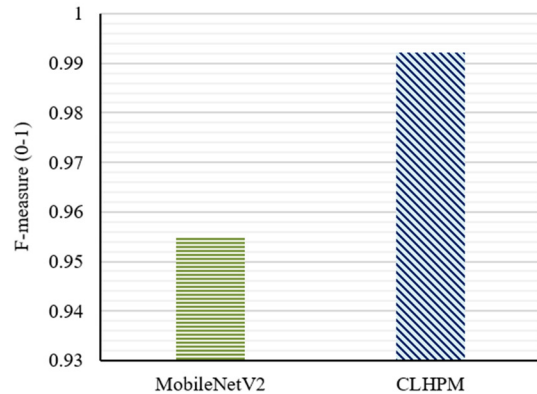| Leaf types	MobileNetV2	CLHPM
Early blight + healthy	0.9383	0.9882
Bacterial spot + healthy	0.9639	0.9792
Late blight + healthy	0.925	1
Late mold + healthy	0.9512	0.9863
Septoria spot + healthy	0.9639	1
Spider mites + healthy	0.9882	1

Fig. 6 Comparison of average F-measure values of leaf class detection models

True positive rate (TPR) is the probability of getting a positive prediction on the actual positive value [29]. Table 7. shows the TPR value at different input testing datasets. In the case of the given leaf sets, the best TPR of 0.9605 is achieved by the proposed model, while the best TPR value shown by the existing model is 0.92. The TPR value for distinct size datasets can be observed graphically in Fig. 7.

Table 7 TPR value for distinct size leaf class of tomato healthy leaf prediction

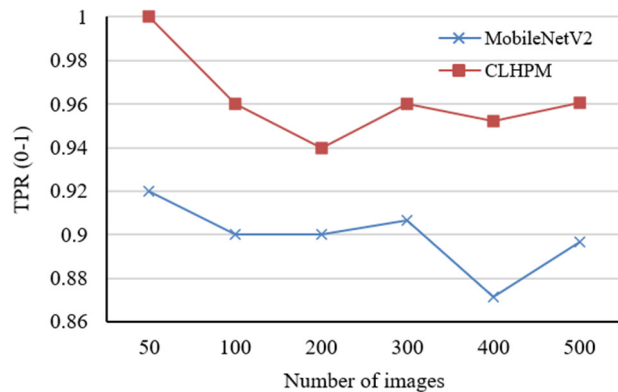| Early blight + healthy leaf sets	MobileNetV2	CLHPM
50	0.92	1
100	0.9	0.96
200	0.9	0.94
300	0.9067	0.9603
400	0.8713	0.9523
500	0.8968	0.9605

Fig. 7 TPR values for distinct size datasets

Experiments are performed on distinct sized tomato leaf datasets and get different readings, corresponding to the accuracy of the existing undertaken model MobileNetV2 and proposed CLHPM model. Table 8 illustrates the results. Fig. 8 exhibits accuracy values with different sized input testing datasets graphically.

Table 8 Accuracy value for distinct size leaf class of tomato healthy leaf prediction

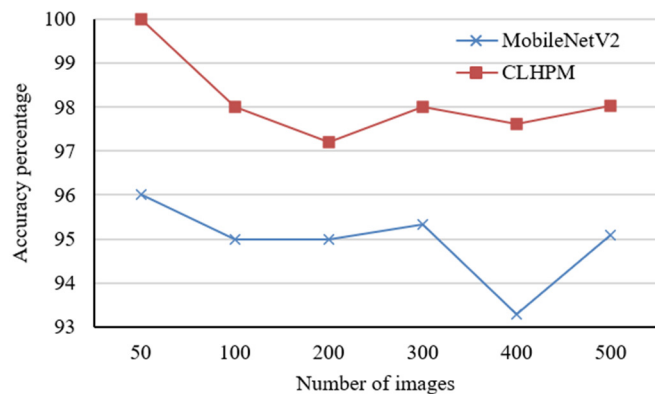| Early blight + healthy leaf sets	MobileNetV2	CLHPM
50	96	100
100	95	98
200	95	97.2
300	95.33	98.01
400	93.3	97.62
500	95.08	98.02

Fig. 8 Accuracy values for distinct size datasets

As observed from Fig. 7 and Fig. 8, the present CLHPM model yields a high detection accuracy in a discrete size set of testing datasets, and the accuracy reaches above 95%, corresponding to all the sets undertaken. Table 9 shows the execution time of the models and delineates that the proposed model takes less time as compared to the previous one in a set of testing datasets. The use of the CNN model for testing is the time taking process in MobileNetV2, and, as a result, the deployment of BPNN is suggested instead of CNN.

Table 9 Execution time (in seconds) for distinct size leaf
classes of tomato healthy leaf prediction

Early blight + healthy leaf sets	MobileNetV2	CLHPM
50	178.0349	8.1844
100	200.8028	12.8731
200	322.5644	24.2822
300	490.9811	31.9987
400	678.2183	37.449
500	982.7412	44.7441

5. Conclusions and Future Scope

Vegetables are useful sources for ingesting multifarious nutrients. Hence, monitoring such vegetable plants is a significant task. To be proximate to infallible monitoring, researchers have proposed manifold models. In this study, a model for leaf health identification and infection prediction is presented. Pre-processing of an image has been done by a bio-inspired algorithm that finds the leaf region in the image for feature extraction. The proposed model uses histogram and MCCM features for training the neural network. Experiments have been done on a real dataset of tomato images having 7 different classes. Results show that the proposed model exhibits an accuracy of 99.23%, compared to the existing model's accuracy of 95.73%. Similarly proposed CLHPM model has enhanced the TPR by 6.55% at different testing sets. It has been observed that the application of genetic algorithms and GMM for region identification has improved the feature extraction process. Moreover, optimized neural networks have improved learning and detection accuracy. Given the attributes stated above, the model introduced herein contributes to extensive application.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] G. Saranya and A. Pravin, "A Comprehensive Study on Disease Risk Predictions in Machine Learning," International Journal of Electrical and Computer Engineering, vol. 10, no. 4, pp. 4217-4225, August 2020.
[2] S. Zhang, Z. You, and X. Wu, "Plant Disease Leaf Image Segmentation Based on Superpixel Clustering and EM Algorithm," Neural Computing and Applications, vol. 31, no. 2 supplement, pp. 1225-1232, February 2019.
[3] V. Choudhary and A. Thakur, "BAT Algorithm-Based Multi-Class Crop Leaf Disease Prediction Bootstrap Model," Proceedings of Engineering and Technology Innovation, vol. 26, pp. 72-82, February 2024.
[4] G. K. V. L. Udayananda, C. Shyalika, and P. P. N. V. Kumara, "Rice Plant Disease Diagnosing Using Machine Learning Techniques: A Comprehensive Review," Discover Applied Sciences, vol. 4, no. 11, article no. 311, November 2022.
[5] K. Kumar K and K. Eswariah, "An Efficient Deep Neural Network for Disease Detection in Rice Plant Using XGBOOST Ensemble Learning Framework," International Journal of Intelligent Systems and Applications in Engineering, vol. 10, no. 3, pp. 116-128, September 2022.
[6] H. T. Nguyen, H. H. Luong, L. B. Huynh, B. Q. H. Le, N. H. Doan, and D. T. D. Le, "An Improved MobileNet for Disease Detection on Tomato Leaves," Advances in Technology Innovation, vol. 8, no. 3, pp. 192-209, July 2023.

[7]   Z. Ma, J. M. R. S. Tavares, and R. M. N. Jorge, "A Review on the Current Segmentation Algorithms for Medical Images," Proceedings of the First International Conference on Computer Imaging Theory and Applications, vol. 1, pp. 135-140, February 2009.

[8]   Z. Ma, J. M. R. S. Tavares, R. N. Jorge, and T. Mascarenhas, "A Review of Algorithms for Medical Image Segmentation and Their Applications to the Female Pelvic Cavity," Computer Methods in Biomechanics and Biomedical Engineering, vol. 13, no. 2, pp. 235-246, 2010.

[9]   P. Jiang, Y. Chen, B. Liu, D. He, and C. Liang, "Real-Time Detection of Apple Leaf Diseases Using Deep Learning Approach Based on Improved Convolutional Neural Networks," IEEE Access, vol. 7, pp. 59069-59080, 2019.

[10] H. Amin, A. Darwish, A. E. Hassanien, and M. Soliman, "End-to-End Deep Learning Model for Corn Leaf Disease Classification," IEEE Access, vol. 10, pp. 31103-31115, 2022.

[11] P. S. Devi and A. S. Rajan, "An Inquiry of Image Processing in Agriculture to Perceive the Infirmity of Plants Using Machine Learning," Multimedia Tools and Applications, in press. https://doi.org/10.1007/s11042-023-18052-4

[12] S. K. Upadhyay and A. Kumar, "A Novel Approach for Rice Plant Diseases Classification With Deep Convolutional Neural Network," International Journal of Information Technology, vol. 14, no. 1, pp. 185-199, February 2022.

[13] S. M. Kiran and D. N. Chandrappa, "Plant Leaf Disease Detection Using Efficient Image Processing and Machine Learning Algorithms," Journal of Robotics and Control, vol. 4, no. 6, pp. 840-848, 2023.

[14] M. Bouni, B. Hssina, K. Douzi, and S. Douzi, "Impact of Pretrained Deep Neural Networks for Tomato Leaf Disease Prediction," Journal of Electrical and Computer Engineering, vol. 2023, no. 1, article no. 5051005, January 2023.

[15] H. S. Nagamani and H. Sarojadevi, "Tomato Leaf Disease Detection Using Deep Learning Techniques," International Journal of Advanced Computer Science and Applications, vol. 13, no. 1, pp. 305-311, 2022.

[16] C. Stauffer and W. E. L. Grimson, "Adaptive Background Mixture Models for Real-Time Tracking," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 246-252, June 1999.

[17] Y. Lu and H. Guo, "Background Removal in Image Indexing and Retrieval," Proceedings 10th International Conference on Image Analysis and Processing, pp. 933-938, September 1999.

[18] S. R. Patel and P. Verma, "An Unsupervised TLBO Based Drought Prediction by Utilizing Various Features," International Journal of Scientific Research & Engineering Trends, vol. 3, no. 6, pp. 239-243, November-December 2017.

[19] N. K. Sah and A. Bhargava, "Genetic Algorithm Based Load Data Analysis for Demand Side Management," International Journal of Science, Engineering and Technology, vol. 7, no. 6, article no. 157, 2019.

[20] S. S. Jain and S. Malviya, "Digital Image Retrieval Using Annotation, CCM and Histogram Features," International Journal of Scientific Research & Engineering Trends, vol. 4, no. 5, pp. 859-864, September-October 2018.

[21] C. Kumar, A. K. Singh, and P. Kumar, "A Recent Survey on Image Watermarking Techniques and Its Application in E-Governance," Multimedia Tools and Applications, vol. 77, no. 3, pp. 3597-3622, February 2018.

[22] S. Bhattarai, "New Plant Diseases Dataset," http://www.kaggle.com/datasets/vipoooool/new-plant-diseases-dataset, October 20, 2023

[23] V. Choudhary and A. Thakur, "Comparative Analysis of Machine Learning Techniques for Disease Prediction in Crops," IEEE 11th International Conference on Communication Systems and Network Technologies, pp. 190-195, April 2022.

[24] R. Bhagwat and Y. Dandawate, "A Review on Advances in Automated Plant Disease Detection," International Journal of Engineering and Technology Innovation, vol. 11, no. 4, pp. 251-264, September 2021.

[25] S. Coulibaly, B. Kamsu-Foguem, D. Kamissoko, and D. Traore, "Deep Neural Networks With Transfer Learning in Millet Crop Images," Computers in Industry, vol. 108, pp. 115-120, June 2019.

[26] S. Ashwinkumar, S. Rajagopal, V. Manimaran, and B. Jegajothi, "Automated Plant Leaf Disease Detection and Classification Using Optimal MobileNet Based Convolutional Neural Networks," Materials Today: Proceedings, vol. 51, part 1, pp. 480-487, 2022.

[27] Y. Kaya and E. Gürsoy, "A Novel Multi-Head CNN Design to Identify Plant Diseases Using the Fusion of RGB Images," Ecological Informatics, vol. 75, article no. 101998, July 2023.

[28] S. Ahmed, M. B. Hasan, T. Ahmed, M. R. K. Sony, and M. H. Kabir, "Less is More: Lighter and Faster Deep Neural Architecture for Tomato Leaf Disease Classification," IEEE Access, vol. 10, pp. 68868-68884, 2022.

[29] K. Al Jallad, M. Aljnidi, and M. S. Desouki, "Anomaly Detection Optimization Using Big Data and Deep Learning to Reduce False-Positive," Journal of Big Data, vol. 7, article no. 68, 2020.