Udhav Kumar Mishra, 2025, 13:6
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Digital Leaf Image Disease Detection by Content Features and Linear Kernel

Udhav Kumar Mishra, Professor Sujeet Gautam, Professor S Vishwakarma
Department of Computer Science and Engineering
Patel College of Science and Technology, Bhopal, MP, India

Abstract- Agriculture plays a fundamental role in human civilization by ensuring food security and providing essential resources. As plant diseases directly affect crop yield and quality, their early and accurate detection is critically important. The proposed model focuses on extracting Texture and Content features to represent color distribution in leaf images. These features are combined to form a robust representation of healthy and diseased leaf characteristics. A multi-class Support Vector Machine (SVM) with a linear kernel is employed for classification due to its simplicity, efficiency, and suitability for high-dimensional feature spaces. Experimental evaluation on a real tomato plant leaf dataset demonstrates that the proposed approach significantly improves classification accuracy and effectively distinguishes between multiple types of leaf diseases.

Keywords: Image Processing, Plant Leaf, Feature Extraction, Image Classification.

I. INTRODUCTION

Plants play an indispensable role in sustaining the global food system; however, they are constantly exposed to diverse environmental stresses that lead to diseases and considerable yield losses. Conventional disease identification methods, which rely primarily on human observation, are often inaccurate, labor-intensive, and slow, making them ineffective for timely disease diagnosis and control. To overcome these limitations, modern computational approaches such as Machine Learning (ML) and Deep Learning (DL) have gained significant attention due to their ability to detect plant diseases at early stages. Considering the high economic value of agricultural production, continuous crop health monitoring using intelligent technologies is essential to accurately identify crop-specific diseases caused by pathogenic organisms [1].

Plant diseases and pest infestations not only reduce agricultural productivity but also disturb ecological stability. Therefore, early diagnosis and preventive measures are crucial, particularly for large-scale farming systems and orchards. Traditional manual inspection techniques are inefficient, time-consuming, and associated with high operational costs, highlighting the need for automated, accurate, and scalable disease detection systems to enhance crop yield and quality [2].

Recent research has introduced a wide range of ML- and DL-based methods for plant disease identification. However, many existing studies are limited to particular crops or specific diseases, underscoring the need for more generalized and adaptable models. In addition, the scarcity of large, publicly available labeled datasets remains a major challenge. To address these issues, transfer learning, which adapts pre-trained models to new agricultural datasets, has emerged as an effective strategy for improving model performance [3]. Similarly, ensemble learning techniques, which integrate multiple models, have been increasingly adopted to enhance accuracy and reduce dependency on a single classifier [4].

Furthermore, data augmentation techniques, such as image rotation, scaling, and flipping, are widely used to increase dataset diversity and reduce overfitting, thereby minimizing the requirement for extensive labeled data. Collectively, these advancements have the potential to significantly improve plant disease detection and management, contributing to long-term agricultural sustainability and global food security [5].

II. RELATED WORK

Narayanan et al. [6] proposed an innovative hybrid convolutional neural network framework for identifying diseases in banana plants. Their approach focused on preprocessing raw images while preserving original visual information and maintaining uniform image size using a median filtering technique. The model combined Convolutional Neural Networks (CNNs) with Support Vector Machines (SVMs), where an initial SVM stage distinguished healthy leaves from diseased ones, followed by a multiclass SVM during testing to classify specific banana leaf infections.

Jadhav et al. [7] introduced a novel histogram-based transformation technique to improve deep learning performance by generating synthetic samples from low-quality images. Their method enhanced the cassava leaf disease dataset using image degradation techniques such as Gaussian and motion blur, resolution reduction, and overexposure. These augmented samples were integrated with a modified Previous Model model to address data scarcity and improve training effectiveness.

M. Sowmiya et al. [8] developed PLDPNet, a hybrid deep learning architecture for automated potato leaf disease detection. The framework consisted of image acquisition, preprocessing, segmentation, feature extraction, feature fusion, and classification stages. An ensemble of VGG19 and Inception-V3 feature representations was combined with vision transformers to achieve accurate prediction on publicly available potato leaf datasets.

Fizzah Arshad et al. [9] presented IQWO-PCA, a hybrid optimization-based approach for tomato disease classification. The method integrated Improved Quantum Whale Optimization with Principal Component Analysis and employed pretrained models such as AlexNet, VGG16, ResNet50, and DenseNet121. Hyperparameters were systematically optimized to extract discriminative features and enhance classification accuracy.

Sabbir Ahmed et al. [10] proposed a lightweight transfer learning framework for tomato leaf disease detection using a pretrained Previous Model model. Image preprocessing and runtime data augmentation were applied to improve image quality, prevent data leakage, and address class imbalance, resulting in efficient and reliable disease detection.

R. Rashid et al. [11] introduced MMF-Net, a CNN-based model designed for agricultural disease classification. The architecture fused multi-contextual information through RL-blocks and PL-blocks, where RL-blocks extracted local contextual features from coarse images, and PL-blocks captured fine-grained global context along with real-world environmental attributes from multiple data sources.

S. S. Begum [12] presented a structured disease detection framework comprising preprocessing, segmentation, feature extraction, and classification phases. Image enhancement was performed using Improved Contrast Limited Adaptive Histogram Equalization (ICLAHE), while segmentation employed Kernelized Gravity-based Density Clustering (KGDC). For classification, a Gated Self-Attentive Convoluted MobileNetV3 (GSAtt-CMNetV3) model was used, with its parameters optimized through the Osprey Optimization Algorithm (Os-OA) to improve overall performance.

III. PROPOSED METHODOLOGY

This section presents a detailed description of the proposed Digital Leaf Infection Detection (DLID) framework. The method utilizes Co-occurrence Matrix (CCM) features to analyze the texture characteristics of plant leaf images, while histogram-based features are employed to capture color information. The extracted texture and color features are subsequently combined and used to train a multi-class Support Vector Machine (SVM) classifier for effective leaf disease classification.

Plant Leaf Image Preprocessing

Consider an input plant leaf image denoted as PL, having dimensions of $n \times n$, which implies that the image contains a total of n^2 pixels, with n pixels in each row and column. When such an image is read by the system, it is represented in the form of a matrix of the same n × n size, where each element corresponds to the color intensity value of an individual pixel [13]. This matrix representation serves as the basis for subsequent preprocessing and feature extraction operations.

PPL←Pre_Processed_Plant_Leaf(PL)

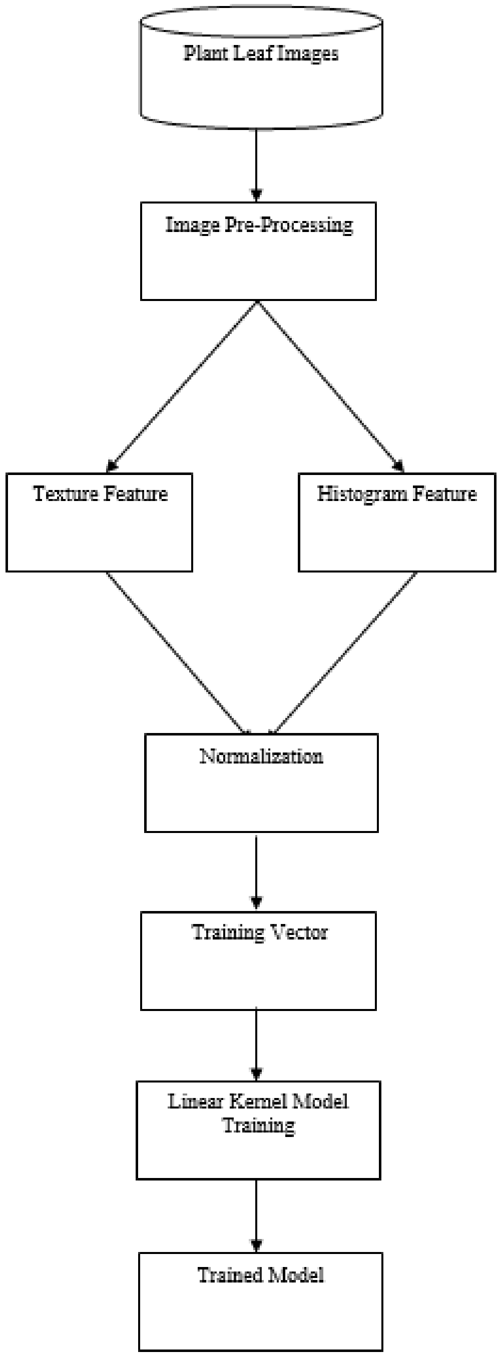

Fig. 1 Block diagram of proposed DLID.

Image Block In this study, the image was divided into smaller blocks measuring bxb pixels each. Since the entire analysis was conducted on grayscale images, the pixel values fall within the range of 0 to 255, representing different shades of gray. Consequently, each block in the analysis comprises a collection of 16 values, as detailed in the literature.

B←Plant_Leaf_Blocked(PL)

Feature Extraction After preprocessing, the plant leaf image is divided into smaller blocks of fixed n × n pixel size to facilitate detailed analysis. Feature extraction is then performed on these image blocks to obtain meaningful information that can effectively distinguish between healthy and diseased leaf regions. In the proposed model, histogram-based features are extracted to represent the distribution of pixel intensity values, particularly capturing color variations within the leaf image. The inclusion of histogram features enhances computational efficiency while preserving essential visual characteristics relevant to disease identification.

In addition to color information, texture-related features are extracted using the Co-occurrence Matrix (CCM) approach. The CCM employed in this work generates a total of 16 feature values, providing a compact yet informative representation of texture patterns present in the leaf image. These combined features contribute significantly to improving the overall classification performance.

Co-occurrence Matrix (CCM)

To effectively characterize the surface and texture properties of plant leaf images, the co-occurrence matrix serves as a powerful analytical tool. The CCM captures the spatial relationship between neighboring pixels by evaluating how frequently specific pixel intensity pairs occur at a defined distance and orientation. This statistical representation quantitatively describes the texture structure of the image.

In this study, four key texture descriptors are derived from the co-occurrence matrix: contrast, which measures local intensity variation; energy, which reflects textural uniformity; inverse difference, indicating homogeneity; and entropy, which quantifies the randomness or complexity of texture patterns [14]. Together, these features enable a robust assessment of leaf surface characteristics and play a crucial role in accurately identifying plant diseases.

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

$$Entropy = -\sum_{i=1} \sum_{j=1} I(i,j) \log[I(i,j)]$$

$$Energy = \sum_{i=1} \sum_{j=1} (I(i,j))^2$$

$$Contrast = \sum_{i=1} \sum_{j=1} (i-j)^2 * I(i,j)$$

Histogram Feature In the proposed approach, image color characteristics are represented using a histogram composed of B discrete bins. The histogram feature is computed by counting the number of pixels whose intensity values fall within predefined ranges. These ranges are defined as (1–B),(B+1–2B),…,(PB–M), where M denotes the maximum possible pixel intensity value, and P is calculated as (M/B−1) [15]. For example, in the case of an image containing 256 gray-level intensity values, the histogram bins are formed as (0–15),(16–31),(32–47),…,(250–255). This representation effectively summarizes the distribution of pixel intensities and serves as an informative descriptor of the image's color content.

PH←Plant_Leaf_Histogram(B)

Linear Kernel Learning Model

The Multiclass Support Vector Machine (MSVM) is a supervised linear kernel learning technique widely used for both classification and regression tasks [16]. When provided with labeled training data, the MSVM learning algorithm constructs a model that assigns each new input sample to one of multiple predefined classes. MSVM is particularly effective for handling non-linear classification problems by mapping input features into a higher-dimensional space using the kernel trick. In this feature space, the algorithm identifies an optimal separating hyperplane, or a set of hyperplanes, that maximizes class separation.

T_Vector←[PH, E Et, I, C]
D_Class←PLD // D_Class: Desired Class
MSVM←Train_MSVM(T_Vector, D_Class)

A key concept in SVM learning is the functional margin, which refers to the distance between the separating hyperplane and the nearest data point from any class. A larger margin generally results in better generalization performance and reduced classification error, as it ensures a clearer separation between classes.

Proposed DLID Algorithm
Input: PLD – Plant Leaf Dataset
Output: Trained MSVM model

1. For each image d in PLD
2. obtain the preprocessed plant leaf image:
3. PPL←Pre_Processed_Plant_Leaf(PL[d])
4. B←Plant_Leaf_Blocked(PPL)
5. For each block b
6. Extract CCM texture features:
7. [E,Et,I,C]←PCCM(B)
8. PH←Plant_Leaf_Histogram(B)
9. End loop
10. T_Vector←[PH,E,Et,I,C]
11. D_Class←PLDD
12. End loop
13. Train the MSVM classifier using extracted features:
14. MSVM←Train_MSVM(T_Vector,D_Class)

The extracted histogram and CCM-based texture features are organized according to their respective class labels and used to train the MSVM model. Once trained, the model is applied to classify and identify diseases in previously unseen plant leaf images.

IV. EXPERIMENT AND RESULTS

This section describes the experimental setup along with details of the implementation environment used in the study. The performance of the proposed model is systematically evaluated and compared with existing approaches reported in [10]. To ensure a fair and comprehensive evaluation, multiple performance metrics are employed for model comparison. The experimental results indicate that the proposed approach achieves superior performance across several evaluation parameters when compared with the reference models. All experiments, including model development, training, and testing, were implemented and executed using the MATLAB software environment.

Dataset

Experiment was performed on different image dataset size. Plant leaf image set have 256x256 dimension. Total image dataset is of 1500 images. Plant disease leaf detection dataset was taken from [17], some of sample images were shown in table 2.

Results

Table 1 illustrates that the proposed DLID approach achieves higher precision in plant leaf disease identification when compared with the Previous Model model. The improvement is mainly attributed to the effective use of histogram-based color features and CCM texture descriptors for training the multiclass SVM (MSVM) classifier. The utilization of an optimal hyperplane-based learning mechanism enables better separation of image classes, resulting in enhanced precision performance.

Table 1 Plant leaf precision value based disease detection model.

Testing Image Set	Previous Model	DLID
40	0.9052	0.9321
80	0.9184	0.9417
160	0.9369	0.9583
320	0.9642	0.9495

Table 2 Plant leaf recall value based disease detection model.

Testing Image Set	Previous Model	DLID
40	0.5421	0.9153
80	0.4986	0.9084
160	0.4712	0.8926
320	0.4567	0.8369

Table 2 demonstrates that the recall rate of the proposed DLID model is significantly improved due to the use of transformed and discriminative features during training. The experimental analysis reveals that the DLID model enhances recall performance by approximately 28.4% when compared with the conventional Previous Model -based approach.

Table 3 Plant leaf f-measure value based disease detection model.

Testing Image Set	Previous Model	DLID
40	0.6843	0.9236
80	0.6379	0.9324
160	0.6187	0.9412
320	0.6095	0.8874

Table 3 indicates that the proposed DLID framework substantially improves the F-measure, reflecting a balanced enhancement in both precision and recall. The integration of histogram and CCM features for MSVM training contributes to improved class discrimination and overall robustness of the detection model.

Table 4 Plant leaf accuracy value based disease detection model.

Testing Image Set	Previous Model (%)	DLID (%)
40	68.25	94.87
80	64.18	95.42
160	63.05	96.13
320	62.74	93.68

Table 4 presents the classification accuracy of plant leaf disease detection models. The results confirm that the balanced combination of 16-bin histogram features and four CCM texture parameters significantly enhances the learning capability of the MSVM classifier. Feature transformation also reduces computational complexity while improving prediction reliability.

V. CONCLUSIONS

Plant leaves play a vital role in food production and directly influence agricultural sustainability. Under adverse environmental conditions, plants are vulnerable to diseases, which can severely reduce crop yield and quality. This study proposed an effective approach for identifying unhealthy plants through leaf image analysis. The model employs color histogram features and CCM-based texture descriptors to train a multiclass support vector machine, enabling accurate classification of multiple leaf disease types.

Experiments conducted on a real-world plant leaf image dataset demonstrate that the proposed DLID framework improves detection accuracy by approximately 30.8% compared to existing methods. Additionally, precision performance shows an improvement of nearly 0.47%. Future work may focus on further enhancing performance by eliminating background noise and incorporating advanced image segmentation techniques to improve feature quality.

REFERENCES

1. Fu, L.; Gao, F.; Wu, J.; Li, R.; Karkee, M.; Zhang, Q. Application of consumer RGB-D cameras for fruit detection and localization in field: A critical review. Comput. Electron. Agric. 2020, 177, 105687.
2. SepuLveda, D.; Fernández, R.; Navas, E.; Armada, M.; González-De-Santos, P. Robotic aubergine harvesting using dual-arm manipulation. IEEE Access 2020, 8, 121889–121904.
3. Agarwal Mohit, Gupta Suneet K. r., and Biswas K. K.. 2019. Grape disease identification using convolution neural network. In Proceedings of the 2019 23rd International Computer Science and Engineering Conference. 224–229.
4. Hussain, M.; Bird, J.J.; Faria, D.R. A Study on CNN Transfer Learning for Image Classification. In Proceedings of the UK Workshop on Computational Intelligence, Nottingham, UK, 5–7 September 2018; Volume 840, pp. 191–202.
5. Agarwal Mohit, Singh Abhishek, Arjaria Siddhartha, Sinha Amit, and Gupta Suneet. 2020. ToLeD: Tomato leaf disease detection using convolution neural network. Procedia Computer Science 167, 2 (2020), 293–301.
6. Narayanan, K.L.; Krishnan, R.S.; Robinson, Y.H.; Julie, E.G.; Vimal, S.; Saravanan, V.; Kaliappan, M. Banana Plant Disease Classification Using Hybrid Convolutional Neural Network. Comput. Intell. Neurosci. 2022, 2022, 9153699.
7. Abayomi-Alli, O.O.; Damaševičius, R.; Misra, S.; Maskeliūnas, R. Cassava disease recognition from low-quality images using enhanced data augmentation model and deep learning. Expert Syst. 2021, 38, e12746.
8. M. Sowmiya, S. Krishnaveni. "IoT enabled prediction of agriculture's plant disease using improvedϖ quantum whale optimization DRDNN approach". Measurement: Sensors, Volume 27, 2023.
9. Fizzah Arshad, Muhammad Mateen, Shaukat Hayat, Maryam Wardah, Zaid Al-Huda, Yeong Hyeon Gu, Mugahed A. Al-antari. "PLDPNet: End-to-end hybrid deep learning framework for potato leaf disease prediction". Alexandria Engineering Journal, Volume 78, 2023.
10. Sabbir Ahmed, Md. Bakhtiar Hasan , Tasnim Ahmed, Md. Redwan Karim Sony, And Md. Hasanul Kabir. "Less Is More: Lighter and Faster Deep Neural Architecture for Tomato Leaf Disease Classification". IEEE Access, 23 June 2022.
11. R. Rashid, W. Aslam, R. Aziz and G. Aldehim, "An Early and Smart Detection of Corn Plant Leaf Diseases Using IoT and Deep Learning Multi-Models," in IEEE Access, vol. 12, pp. 23149-23162, 2024, doi: 10.1109/ACCESS.2024.
12. S. S. A. Begum and H. Syed, "GSAtt-CMNetV3: Pepper Leaf Disease Classification Using Osprey Optimization," in IEEE Access, vol. 12, pp. 32493-32506, 2024.
13. Shalinee Jain Asst.Prof. Ravi Gedam. "A Survey on Image Retrieval Approaches with Features Utilization". International Journal of Scientific Research & Engineering Trends Volume 4, Issue 4, July-Aug-2018.
14. Dr. Pushpendra Anuragi, Dr. Pratima Gautam. "KNN Based Medical Image Diagnosis by Content Features". International Journal of Science, Engineering and Technology, 11:4, 2023.
15. Mehboob, R., Javed, A., Dawood, H. et al. Histogram of Low-Level Visual Features for Salient Feature Extraction. Arab J Sci Eng 47, 10589–10604 (2022).

16. Jianqun Zhang, Jun Zhang, Min Zhong, Jinde Zheng, Ligang Yao. "A GOA-MSVM based strategy to achieve high fault identification accuracy for rotating machinery under different load conditions", Measurement, Volume 163, 2020.
17. https://www.kaggle.com/datasets/vipoooool/new-plant-diseases-dataset