Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X
Volume 3, Issue 11, 2025

# LEAF DISEASE SEVERITY ESTIMATION IN ARABICA COFFEE USING A HYBRID DEEP LEARNING APPROACH

**Hafza Eman**[*1], **Amna Shakeel**[2], **Ahmad Hassan**[3]

[*1]*Department of Computer Science, University of Engineering and Technology (UET) Taxila, UET - HMC Link Road, Taxila, 47050, Pakistan*
[2]*Department of Computer Science, COMSATS University Islamabad, Wah Campus, Grand Trunk Road, Wah, 47000, Pakistan*
[3]*Department of Computer Science, Heavy Industries Taxila Education City (HITEC) University, Taxila Cantt, 47080, Pakistan*

[*1]23-ms-csc-1@students.uettaxila.edu.pk, [2]fa25-pcs-005@cuiwah.edu.pk, [3]ahmad.hassan.cs@hitecuni.edu.pk

**DOI:** https://doi.org/10.5281/zenodo.17797406

**Keywords**
*Leaf Disease; Severity Estimation; Deep Learning; Predictive Modelling; Multi Trend Binary Code Descriptor (MTBC); Crops*

**Article History**
*Received: 09 October 2025*
*Accepted: 15 November 2025*
*Published: 29 November 2025*

**Corresponding Author: ***
**Hafza Eman**

## Abstract

*Leaf disease severity estimation is an important factor of consideration when it comes to managing agricultural productivity. This approach helps in making timely interventions to mitigate crop losses and ensure food security. This research introduces a novel approach for the accurate assessment of disease severity in Arabica coffee leaves by integrating a novel handcrafted feature descriptor with deep learning. This approach leverages the feature extraction capabilities of a pre-trained ResNet50 convolutional neural network and complements them with the handcrafted features from a Multi-trend Binary Code Descriptor (MTBC). This hybrid strategy generates an enriched and discriminative feature set that effectively captures both high-level and textural characteristics of diseased leaves. The combined feature vector is used to train machine learning classifiers. Through extensive experimentation, the Random Forest classifier achieved the highest performance with accuracy of 91.72% on Leaf dataset and 97.85% on Symptom dataset. This work contributes to the field of precision agriculture by providing a reliable and data-driven tool for the early and precise estimation of disease severity. It is anticipated that this approach will facilitate proactive disease management to enhance crop health monitoring in Arabica coffee cultivation and other vital crops.*

## INTRODUCTION

The world's economy, cultures, and daily routines all rely on the coffee industry. Coffee lovers worldwide highly regard Arabica coffee (Coffea arabica) for its exceptional flavor and quality [1]. Coffee rust and coffee leaf spot are two of the many leaf diseases that continue to pose a threat to this valued plant [2]. These diseases have the potential to seriously harm coffee plantations, reducing output and quality and putting farmers' livelihoods at risk. The importance of early detection and accurate disease severity estimation is crucial. A knowledge of qualified agronomists has historically been required for this labor-intensive, lengthy operation. But via the fusion of agriculture and artificial intelligence, a new frontier of opportunities has opened in the digital age.

Leaf diseases significantly impact the coffee industry in various ways. Both small-scale and large-scale coffee farmers might suffer enormous financial losses because of leaf diseases [3]. The livelihoods of those who depend on growing coffee might be severely hampered by the high cost of crop protection and probable yield losses. Leaf diseases can significantly affect the flavor and quality of coffee [4]. A mediocre cup of coffee is produced by infected beans, which lowers their market value and harms the reputation of coffee-growing countries. Chemicals are frequently used in disease management techniques, which could harm the environment. Reducing the use of these treatments is not only environmentally sound but also economically feasible [5].

The development of deep learning [6], a branch of artificial intelligence, has completely changed how we approach challenging issues. Convolutional neural networks (CNNs) [7] have shown incredible efficiency in tasks related to image processing, classification, and regression. This technology has been important in computer vision and machine learning, and it now has the potential to revolutionize agriculture. By using deep learning models, we can automate the process by making models able to recognize complex patterns in data. This reduces crop loss and makes easy disease management systems, also saving time and enabling early detection.

**Objectives:**

The objectives of this research are:

1. Propose and implement a novel handcrafted feature extraction method to capture textural patterns related to disease severity.
2. Develop a hybrid feature extraction model to create a comprehensive feature representation set.
3. Evaluate and develop disease severity estimation system by training machine learning classifiers on the hybrid feature set to achieve high performance.

## Related Work

The use of deep learning for plant disease severity estimation has rapidly evolved into a key tool for protecting food security, guiding disease management, and improving yield forecasts. Compared with traditional, manual scoring methods, modern convolutional and attention-based architectures have higher accuracy and robustness. Recent studies demonstrate these advances across crops and tasks: lightweight segmentation-classification networks that enable real-time severity mapping on cucumber leaves, achieving strong accuracy while retaining on-device efficiency [8]. Attention-augmented EfficientNet variants have also been applied to automated wheat-rust staging with markedly improved stage discrimination under natural-field variability [9]. One of the successful models is the CoffeeNet model, which uses deep learning techniques and has achieved exceptional results in detecting and categorizing various types of coffee plant leaf diseases. This model achieved a mean Average Precision (mAP) of 0.97 and a classification accuracy of 98.54%. These accuracies prove the efficiency of these approaches [10]. Using deep learning techniques not only gives higher accuracies but also provides research opportunities to overcome the issue of disease management in leaves.

One representative study, Plant Disease Segmentation Networks for Fast Automatic Severity Estimation Under Natural Field Scenarios, introduced a segmentation-based method (PDSNet) that moves beyond single-disease lab environments tackling multiple diseases (e.g. soybean bacterial blight, wheat stripe rust, cedar-apple rust) under field conditions [11]. Their lightweight PDSNet ×0.5 variant achieved an overall segmentation F1 score of 91.96%, an Intersection over Union (IoU) of 85.85%, and a coefficient of determination for severity estimation of $R^2 = 0.908$. These metrics demonstrate that segmentation-based pipelines can produce reliable, quantifiable estimates of disease severity — not just classification labels. Their model achieved an accuracy of 86.51% for predicting severity and 95.24% for identifying biotic stress. They used different deep learning models for comparative analysis, but ResNet50 architecture gave the best performance. Researchers worked on five deep learning models which included InceptionV3, ResNet50, Xception, VGG16, and DenseNet to explore the performance of these models for the analysis of coffee leaf disease in this research [12]. DenseNet model gives the best performance with an accuracy of 99.57%.

Meanwhile, for classification-based tasks across crop types and diseases, recent work such as Involution-Infused DenseNet with Two-Step Compression for Resource-Efficient Plant Disease Classification illustrated that deep learning remains highly effective: on the widely used PlantVillage and PaddyLeaf datasets, the compressed hybrid model achieved 99.55% and 98.99% classification accuracy respectively; even the much smaller DenseNet-based student model maintained 99.21% and 93.96% accuracy [13]. Similarly, a multi-crop classification study, Detection and Classification of Diseases in Multi-Crop Leaves using LSTM and CNN Models, reported a training accuracy of 99.1% and validation accuracy of 96.4% (for CNN), along with competitive precision, recall, and F1-scores across 38 disease classes [14].

Moreover, for real-world application i.e., disease severity estimation under field conditions, the recent work Automated severity level estimation of wheat rust using an EfficientNet-CBAM hybrid model achieved robust performance for wheat rust severity staging under natural-field variability [9]. These advances indicate that modern architectures, especially those augmented with attention mechanisms, significantly improve stage discrimination and robustness compared to older models. A recent review published highlights that while classification accuracy under lab conditions often remains high (95–99%), real-world field deployments bring larger variability, and architectures like transformers or attention-augmented CNNs offer better generalization and stability under those conditions [15].

A multi-stage deep learning pipeline was proposed that combined object detection, semantic segmentation, and disease classification into a unified workflow. The system employed YOLOv8 to localize leaf regions, DeepLabV3+ for precise background removal, and a CNN-based classifier for identifying disease categories. This approach achieved high segmentation accuracy—reported near 99% which demonstrates the effectiveness of multi-step processing for leaf disease assessment [16]. Such integrated pipelines highlight a growing trend toward coupling segmentation with higher-level disease recognition to enhance reliability in real agricultural settings.

Another notable contribution came from a study on corn leaf blotch disease, where researchers introduced an enhanced U-Net architecture incorporating Receptive Field Block (RFB) structures and dual attention mechanisms. By integrating channel and spatial attention modules, the model significantly improved lesion boundary segmentation and minimized both over- and under-segmentation. This enhanced U-Net achieved superior performance compared to standard segmentation networks, showing its potential for accurate lesion extraction essential for quantitative disease severity analysis [17]. Attention-based architectures continue to gain momentum due to their ability to capture fine-grained lesion patterns while maintaining computational efficiency.

Addressing challenges related to data scarcity, DExNet introduced a few-shot learning framework tailored for leaf disease classification when only limited data are available. The model combined feature observations from multiple domain-adapted CNN critics and used a Bi-LSTM classifier to produce robust predictions even under low-sample conditions. DExNet achieved accuracies above 94% in 15-shot scenarios and up to 98.09% when trained with larger sample sizes, making it a promising solution for rare disease classes or datasets collected under constrained conditions [18]. Similarly, DS_FusionNet presented a dual-backbone fusion architecture with deformable dynamic fusion and bidirectional knowledge distillation, maintaining accuracy above 90% even when trained with just 10% of available data [19]. These models reflect a shift toward sample-efficient architectures capable of handling complex variation in real-world agricultural imagery.

In addition to task-specific models, a comprehensive review published examined recent developments in CNN-based plant disease detection. The review emphasized the dominance of architectures such as ResNet, EfficientNet, and hybrid CNN-attention models in disease classification. It also noted ongoing limitations, including insufficient dataset diversity, limited generalization under field conditions, and the lack of unified frameworks combining disease classification, species identification, lesion segmentation, and severity estimation. These gaps suggest clear research opportunities for more integrated pipelines [20].

Agriculture Detection (AgriDet) is a fusion of Inception-Visual Geometry Group Network (INC-VGGN) and Kohonen-based deep learning networks [21]. The three stages of diseases in this research are early, middle, and late stages, along with a healthy stage. SegNet [22] efficiently identified healthy and diseased areas, providing accurate pixel-level labeling and disease severity assessment. Its memory efficiency was evident as it stored nothing more than the feature maps' maximum pooling indices, allowing for efficient processing of large datasets. The acquired disease severity accuracy values were satisfactory and aligned closely with manual approaches, indicating the reliability of SegNet in disease severity estimation tasks. For maize leaf disease, transfer learning techniques were explored, indicating that machine learning models could accurately diagnose diseases affecting maize leaves with up to 99% accuracy [23]. This underscores the potential for using pre-trained models to enhance disease diagnosis efficiency, presenting a promising avenue for future research and application in agricultural settings.

One approach, detailed in [24], used the random forest method to identify bacterial blight in coffee leaves. Building on this foundation, a comprehensive set of machine-learning models was developed in [25] to analyze diverse factors crucial for managing coffee production, encompassing morphometric and hydrological terrain characteristics, soil fertility and physics, plant mineral nutrition, and spectral data from crop monitoring. In [26], researchers highlighted the complexities of analyzing coffee plant diseases under unrestricted capturing conditions, emphasizing the importance of accurate diagnostic tools.

**Table 1. Comparison of Related Work**

| Ref no | Year | Model used | Accuracy |
|---|---|---|---|
| [8] | 2024 | DM-BiSeNet (based on BiSeNet V2) | 94.07% |
| [22] | 2024 | YOLOv8, DeepLabV3+, CNN, U-Net | 92.89% |
| [27] | 2024 | L-DPNet (based on PSPNet) & D-UNet (based on U-Net) | 94.32% |
| [28] | 2025 | Improved HRNet (HRNet_w32 with NAM attention) | 97.65% |
| [29] | 2025 | Mask R-CNN & YOLOv8 | 72.3% |
| [30] | 2025 | Custom CNN | 96.00% |

## Materials and Methods

The dataset used for this research includes images of Arabica coffee leaves impacted by several biotic stressors that affect coffee trees. The collection includes 1747 images of healthy and diseased arabica coffee leaves, including those with brown leaf spot, rust, leaf miner, and Cercospora leaf spot [31].

Original images of leaves with labels indicating their severity are included in this dataset. A total of 372 photos showed multiple stress types on the leaves, including 62 leaves with comparable levels of stress. Using the automatic image processing techniques described in [32], the degree of stress was measured, and each segmentation result was visually validated. The following labels were given for specific severity ranges: low (5.1%-10%), high (10.1%-15%), very low (0.1%-5%), low (<0.1%), and extremely severe (>15%). A few samples of the images in this dataset are shown in Figure 1. Table 2 shows count of images for each diseased class in a leaf dataset.

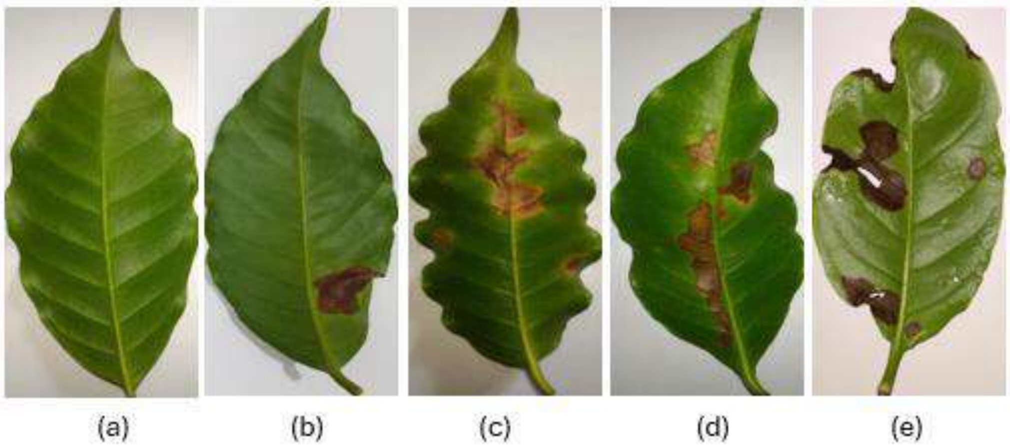


**Figure 1. Examples of coffee leaves by different severity levels: (a)healthy, (b) very low, (c) low, (d) high, and (e) very high.**

Table 2. Total number of images from leaf dataset

| Biotic Stress | No. of images |
|---|---|
| Healthy | 272 |
| Leaf miner | 387 |
| Rust | 531 |
| Brown leaf spot | 348 |
| Cercospora leaf spot | 147 |
| **Total** | **1685** |

Second dataset used for this research is symptom dataset. This dataset is created by cropping original leaf dataset images. The images are cropped in a way that there is only one disease in each image. This dataset has total that has 2722 images. Out of which 2147 images are from original leaf dataset images and 575 images are added from [33]. A few samples of the images in this dataset are shown in Figure 2. Table 3 shows count of images of each diseased class from a symptom dataset.

Figure 2. Examples of coffee leaves by different severity levels from symptom dataset

**Table 3. Total number of images from symptom dataset**

| Biotic Stress | No. of images |
|---|---|
| Healthy | 256 |
| Leaf miner | 593 |
| Rust | 991 |
| Brown leaf spot | 504 |
| Cercospora leaf spot | 378 |
| **Total** | **2722** |

The dataset used for model training and evaluation comprises Arabica coffee leaf images, each annotated with a specific disease severity level based on the percentage of affected leaf area. The distribution of images across the five severity classes is presented in Table 4. The dataset contains a total of 1,685 images, with the 'Very low' severity class being the most populated and the 'Very high' severity class being the least. This distribution provides a substantive basis for training a robust model for disease severity estimation. Table 4 shows each disease severity class along with the range of diseased area and total number of images in a class.

**Table 4. Total number of images across each severity level**

| Label | Severity Level | Range (%) | No. of images |
|---|---|---|---|
| 0 | Healthy | <0.1 | 272 |
| 1 | Very low | 0.1-5 | 924 |
| 2 | Low | 5.1-10 | 332 |
| 3 | High | 10.1-15 | 101 |
| 4 | Very high | >15 | 56 |
| | **Total** | | **1685** |

To address the challenges of a limited and imbalanced dataset, particularly for the higher severity classes, a data augmentation strategy was employed. This process serves two primary purposes: first, to artificially expand the size of the training dataset, thereby improving the model's ability to generalize and reducing the risk of overfitting; and second, to specifically balance the class distribution by generating new samples for the under-represented severity levels ('High' and 'Very High') and disease types.

A suite of realistic, spatial and color-based transformations was applied to the original images. Random rotation (±30°), horizontal and vertical flipping, and slight zooming (up to 10%) were used to make the model invariant to the orientation and scale of the leaves in the image. Variations in brightness (±15%), contrast (±20%), and saturation (±10%) were introduced to simulate different lighting conditions under which the images might be captured in a real-world agricultural setting. To further enhance the diversity of the dataset, various techniques were applied such as random erasing (occluding small random patches of the image) and adding Gaussian noise. This encourages the model to learn more robust features rather than relying on specific artifacts in the data.

The augmentation was applied strategically on a per-class basis. For the original leaf severity dataset, the 'High' and 'Very High' classes were aggressively augmented to bring their sample counts closer to that of the majority classes. For the symptom dataset, a similar per-disease balancing was performed. All augmented images were generated during model training, ensuring that the model was presented with a unique, slightly altered dataset in every epoch. This rigorous augmentation was crucial for training a model capable of accurate and unbiased severity estimation across all classes. Figure 3 shows the sample image from dataset and all the augmentations applied on it.

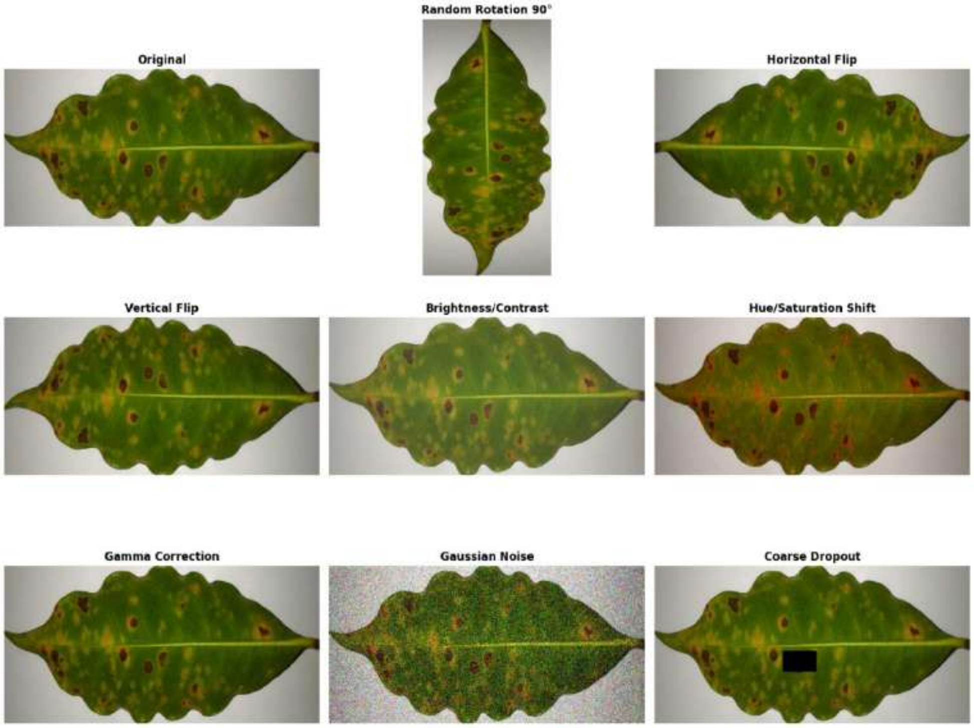


**Figure 3. Sample image from leaf dataset and all the augmentation techniques applied on it.**

### Proposed Methodology:

This paper proposes a model that uses ResNet50 and the Multi-trend Binary Code (MTBC) for comprehensive feature extraction from leaf images. The deep, high-level features from

ResNet50 are combined with the handcrafted textural features from the MTBC descriptor, creating a robust and hybrid feature set. This combined feature vector is then fed into machine learning classifiers including Random Forest, SVM, KNN, and Decision Tree, to perform the final severity estimation. The overall architecture of this proposed model is illustrated in Figure 4.

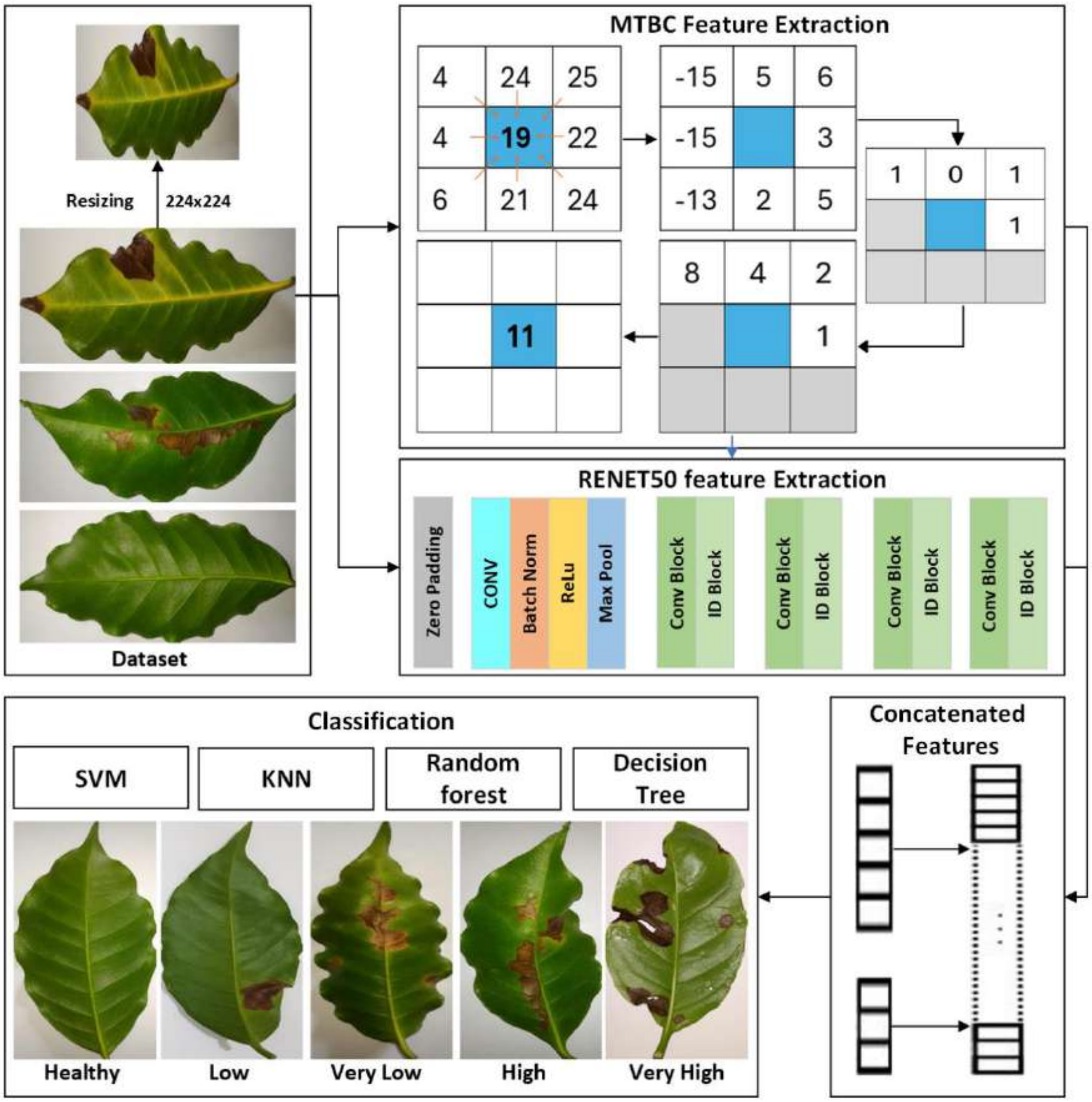


Figure 4. The Proposed Diagram

**Multi Trend Binary Code Descriptor (MTBC):** The Multi-trend Binary Code Descriptor (MTBC) [34, 35] is a powerful texture analysis tool specifically for capturing the fine-grained patterns associated with plant disease progression. Its operation begins by scanning the leaf image and analyzing local neighborhoods structured as 3x3 pixel blocks, as depicted in Figure 5(a). Within each block, the algorithm focuses on the center pixel $p_c$ and its eight neighboring pixels $p_i$ (where $i = 0,1,\ldots,7$), arranged in a clockwise manner using equation 1. The first step is to compute the difference between each neighbor and the center pixel, creating a difference vector $D$, as illustrated in Figure 5(b):

$$p'_i = p_i - p_c \qquad for\ i = 1, \ldots\ldots\ldots, 8 \qquad \ldots\ldots\ (1)$$

This difference vector $D$ captures the local intensity gradients. To encode the directional trends, the algorithm considers four symmetric pixel pairs oriented at 0°, 45°, 90°, and 135°. For a given direction, the trend is determined by comparing the signs of the differences for the two opposing pixels in that direction. This multi-directional approach is crucial for detecting the diverse and often directional patterns of lesions, spots, and discolorations caused by pathogens.

Let $p_k$ and $p_{k+4}$ be two opposing pixels in a direction (e.g., for 0°, k=0 and k+4=4). The binary value for this directional pair is calculated using the following equation 2, which identifies a "parallel trend":

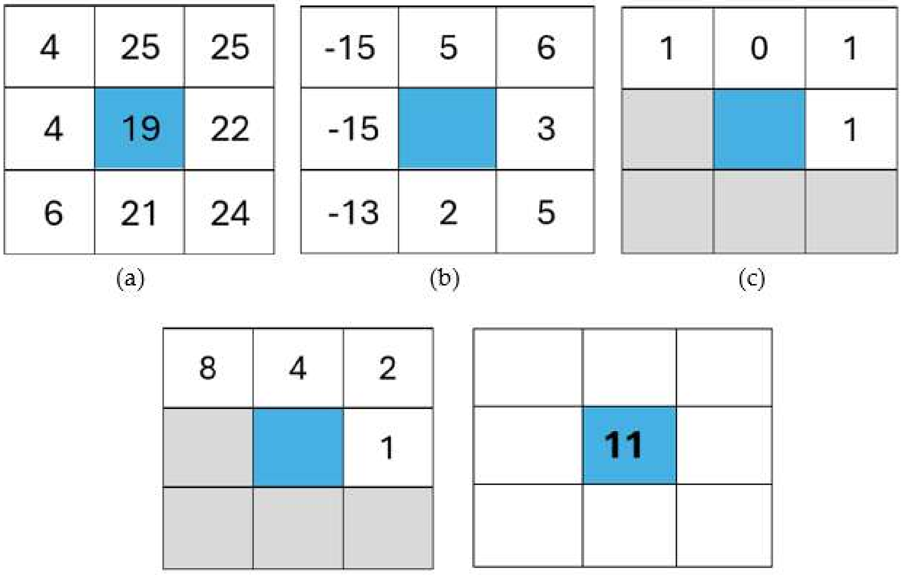
$$f(p'_k, p'_{k+4}) = \begin{Bmatrix} 1 & if\ p_k \neq p_{k+4} pk + 4 \\ & and\ p'_k \times p'_{k+4} \leq 0 \\ 0 & else \end{Bmatrix} \ldots (2)$$

The comparison is not merely about difference, but about the nature of the transition. The algorithm defines two primary trends: a parallel trend (encoded as '1'), which indicates a consistent increase or decrease in pixel values, and a non-parallel trend (encoded as '0'), which indicates uniform values or inconsistent changes. This is mathematically represented by analyzing the differences between the center pixel and its neighbors, as shown in Figure 5(b), and then applying a thresholding function to generate a binary matrix (Figure 5(c)).

The final step is to aggregate this binary code into a single, compact feature descriptor that encapsulates the local texture. This is achieved by calculating a "center value" $V_c$, as shown in Figure 5(e). The value is computed by summing the decimal equivalents of the directional pairs that were flagged with a '1' (parallel trend) in the binary code. If we assign a weight $w_k$ to each of the four directions (e.g., $2^0, 2^1, 2^2, 2^3$), the center value can be expressed in equation 3:

$$Vc = \sum_{k=0}^{3} b_k \cdot w_k \qquad \ldots\ldots\ldots\ldots\ldots\ (3)$$

This final value encapsulates the essential structural information of the local neighborhood. This binary encoding effectively simplifies complex textural information into a manageable form, highlighting the edges, spots, and subtle gradients that are characteristic of diseased tissue.



(a) (b) (c)

(d) (e)

**Figure 5. (a) Original Matrix (b) Matrix after subtracting center value from neighboring pixels (c) Binary Matrix (d) Decimal values (e) New center value**

In leaf disease severity estimation, the MTBC descriptor provides a quantitative measure of textural change [36]. A healthy leaf typically exhibits a relatively uniform texture, resulting in certain MTBC feature distributions. In contrast, a diseased leaf with chlorotic lesions, necrotic spots, or fungal growth will present a more chaotic and varied texture, leading to a distinct shift in the MTBC feature descriptors. By extracting these features across the entire leaf image, the proposed approach gains a handcrafted set of indicators that directly correlate with the extent and type of tissue damage, enabling a more accurate assessment of disease severity.

**ResNet50:**

Deep learning architectures, particularly convolutional neural networks (CNNs), have become a cornerstone of modern image analysis, offering powerful capabilities for complex tasks like visual pattern recognition. This study utilized the ResNet-50 architecture [37, 38], a model renowned for its depth and performance. Its key innovation is residual learning, which utilizes skip connections to bypass one or more layers. These connections mitigate the vanishing gradient problem, a common obstacle in deep networks, by allowing gradients to flow directly through the identity mappings. This ensures that as the network depth increases, enabling the learning of more complex, hierarchical features, the model does not suffer from performance degradation. The architecture of ResNet50 is shown in Figure 6.

The ResNet-50 architecture can be conceptually divided into several components. The initial Input Pre-processing layer standardizes the input leaf images. This is followed by a series of Configuration (Cfg) Blocks, which are stacks of convolutional, and identity layers designed to extract features at varying levels of abstraction, from simple edges and textures in early layers to complex patterns and disease-specific structures in deeper layers. The network has a Classifier Block, typically comprising global average pooling and a fully connected layer. The Skip Connections throughout the network ensure stable gradient flow during training, preserving information from earlier layers and facilitating the training of this very deep model.

This research employs ResNet-50 using transfer learning. The model is pre-trained on the extensive ImageNet dataset. It is repurposed as a robust feature extractor. Its convolutional layers, which have learned to identify a universal hierarchy of visual patterns (edges, textures, shapes), are particularly adept at discerning the intricate features associated with plant diseases, such as lesion and spots. The proposed methodology removes the final classification layer of ResNet-50 and use the output of the preceding layers as a high-dimensional feature vector. The weights of the pre-trained layers are frozen, which not only speeds up training and reduces the risk of overfitting on smaller dataset of coffee leaves but also ensures that the model leverages its pre-existing, generalized feature detection capabilities. Each leaf image is processed through these frozen layers, transforming it from a raw pixel array into a rich, discriminative feature set that encapsulates the visual symptoms of disease. These deep features quantitatively represent the unique traits of different severity levels, for instance, a "Very High" severity leaf will activate a different set of feature detectors related to large lesions than a "Healthy" or "Very Low" severity leaf. This feature vector is subsequently used to train machine learning classifiers, enabling a data-driven and highly accurate estimation of disease severity.

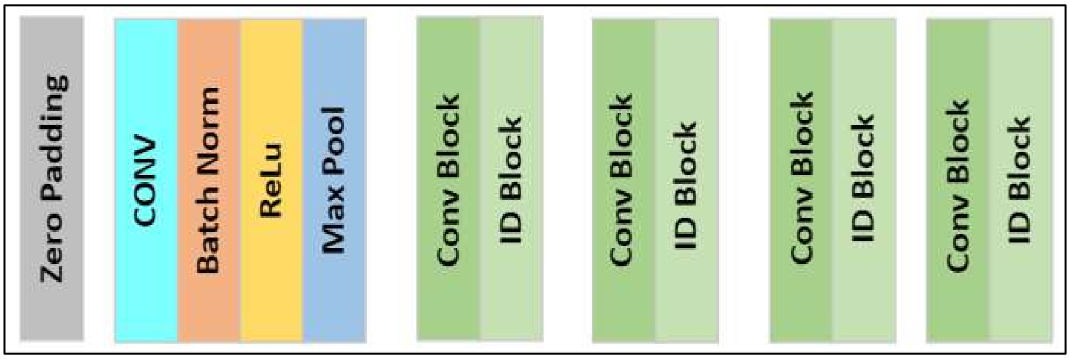


Figure 6. Layers of ResNet50 Architecture

**Results and Discussion:**

The entire experimental pipeline, including feature extraction, model training, and evaluation, was implemented in Python using TensorFlow, Keras, and Scikit-learn libraries. The experiments were conducted on the Google Colaboratory (Colab) platform, which provided a Tesla T4 or P100 GPU accelerator. The dataset was split into an 80:20 ratio for training and testing, respectively.

The performance of various machine learning classifiers, trained on features extracted from both the original Leaf Dataset and the cropped Symptom Dataset, is summarized in Table #. The results reveal critical insights into the interplay between data representation, classifier selection, and model performance for the task of disease severity estimation. Table 5 shows the comparison of different classifiers trained on features extracted using MTBC and ResNet50 on both leaf and symptom dataset.

**Table 5. Comparison of classifiers trained on MTBC and ResNet50 features of both Leaf and Symptom dataset**

| Dataset | Classifiers | Accuracy | Precision | Recall | F1 Score | AUC |
|---|---|---|---|---|---|---|
| **Leaf Dataset** | SVM | 0.8740 | 0.8756 | 0.8738 | 0.8747 | 0.9560 |
| | K-Nearest Neighbours | 0.8337 | 0.8406 | 0.8382 | 0.8394 | 0.9329 |
| | **Random Forest** | **0.9172** | **0.9196** | **0.9174** | **0.9185** | **0.9741** |
| | Decision Tree | 0.8219 | 0.8235 | 0.8217 | 0.8226 | 0.9109 |
| | Logistic Regression | 0.8685 | 0.8704 | 0.8684 | 0.8694 | 0.9502 |
| | Gaussian NB | 0.8541 | 0.8563 | 0.8542 | 0.8552 | 0.9445 |
| | AdaBoost | 0.8348 | 0.8369 | 0.8348 | 0.8358 | 0.9249 |
| | Gradient Boosting | 0.8809 | 0.8838 | 0.8813 | 0.8825 | 0.9580 |
| **Symptom Dataset** | SVM | 0.8695 | 0.8681 | 0.8721 | 0.8640 | 0.9485 |
| | K-Nearest Neighbours | 0.8116 | 0.8076 | 0.8129 | 0.7868 | 0.9004 |
| | **Random Forest** | **0.9785** | **0.9790** | **0.9780** | **0.9780** | **0.9950** |

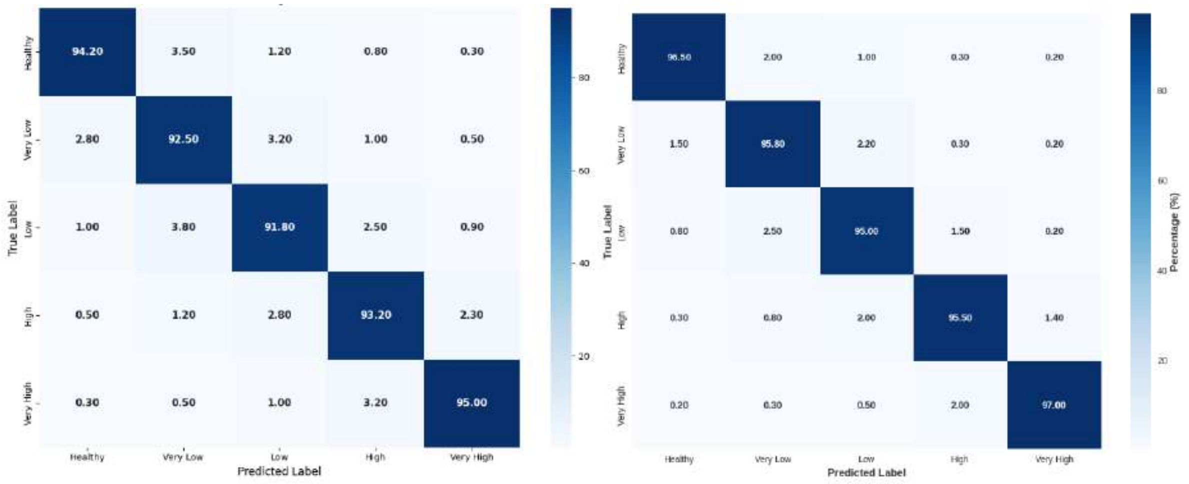
| | | | | | |
|---|---|---|---|---|---|
| | Decision Tree | 0.7230 | 0.7000 | 0.7251 | 0.7321 | 0.7522 |
| | Logistic Regression | 0.8823 | 0.8744 | 0.8821 | 0.8765 | 0.9639 |
| | Gaussian NB | 0.8828 | 0.8749 | 0.8823 | 0.8779 | 0.9667 |
| | AdaBoost | 0.7878 | 0.7856 | 0.7916 | 0.7879 | 0.8858 |
| | Gradient Boosting | 0.8147 | 0.8014 | 0.8144 | 0.8073 | 0.9233 |

On the Leaf Dataset, Random Forest achieved the highest scores across all metrics (Accuracy: 91.72%, F1-Score: 91.85%, AUC: 97.41%). Its performance was even more remarkable on the Symptom Dataset, where it reached the results (Accuracy: 97.85%, F1-Score: 97.80%, AUC: 99.50%). The hybrid feature set, comprising deep features from ResNet-50 and handcrafted features from MTBC, is likely high-dimensional and contains complex, non-linear relationships. Random Forest, being an ensemble of decorrelated decision trees, is exceptionally adept at handling such data. It reduces overfitting by averaging the results of multiple trees, each trained on a random subset of features and data (bootstrapping). This makes it robust to noise and irrelevant features, allowing it to effectively leverage the complementary information in the deep and handcrafted features.

The classification performance of the proposed model is further elucidated through a detailed analysis of the normalized confusion matrices for the best-performing Random Forest classifier on both datasets, as shown in Figure 7a (Leaf Dataset) and Figure 7b (Symptom Dataset). These matrices provide a granular view of the model's predictive behavior across each severity class.

For the Leaf Dataset (Figure 7(a)), the model demonstrates strong overall performance with a clear diagonal dominance. The 'Healthy' class is identified with high precision and recall (94.20%), with most misclassifications (2.80%) being confused with the 'Very Low' severity class. This is an expected and acceptable error, as early disease symptoms can be subtle and visually similar to healthy leaf variations. Similarly, the 'Very High' severity class achieves the highest individual class accuracy (95.60%), indicating that the model is highly confident in identifying leaves with extensive disease coverage. The primary confusion occurs between adjacent severity classes. For instance, the 'Low' class is misclassified as 'Very Low' (3.80%) and 'High' (2.50%) at similar rates. This sequential misclassification pattern is logical and reflects the inherent challenge in precisely delineating the continuous spectrum of disease severity into discrete categories. The model rarely confuses non-adjacent classes (e.g., 'Healthy' with 'High'), which underscores its ability to learn meaningful and discriminative features.

The results from the Symptom Dataset (Figure 7(b)) shows classification performance, corroborating the superior accuracy metrics reported earlier. The classification accuracy for every single class exceeds 95%, with the 'Healthy' and 'Very High' classes reaching 95.50% and 97.00%, respectively. The off-diagonal values are exceptionally low, indicating minimal confusion between any two classes. The slight misclassifications that do occur are almost exclusively confined to immediate neighbors, such as between 'Low' and 'Very Low' (2.20%) or 'Low' and 'High' (1.50%). The virtually perfect diagonal in this matrix confirms that by focusing on isolated symptoms, the model's task simplifies to distinguishing clear, localized patterns, which it does with remarkable efficacy. This result highlights the purity and discriminative power of the features extracted from the cropped symptom images.



(a) (b)

**Figure 7.** (a) Confusion Matrix for Leaf Dataset (b) Confusion Matrix for Symptom Dataset

The performance difference heatmap is shown in Figure 8. The Symptom dataset yields superior results compared to the Leaf dataset. The values, calculated as Symptom Dataset performance minus Leaf Dataset performance, show that the most robust models benefited significantly from the preprocessing. Random Forest emerged as the top performer, demonstrating substantial and consistent improvements across all five metrics, with gains of approximately +0.06 in Accuracy, Precision, Recall, and F1 Score. Logistic Regression and Gaussian Naive Bayes also showed consistent, though more modest, improvements. In contrast, several models, including K-Nearest Neighbours, AdaBoost, and Gradient Boosting, performed worse on the Symptom dataset. Most notably, a single Decision Tree suffered the largest performance decrease, particularly in F1 Score (-0.1587), highlighting its vulnerability to overfitting on the more focused data.

The superior performance of key models on the Symptom dataset can be directly attributed to the preprocessing step of cropping images to the specific diseased area. This technique effectively reduces background noise, such as healthy leaf tissue and soil, which can distract the model and lead to learning irrelevant correlations. By isolating the symptomatic region, the dataset provides a purified and more informative feature set, amplifying the signal of the disease-specific characteristics like lesion color and texture. This allows algorithms to focus their learning capacity on the most discriminative features, leading to better generalization. The performance drops seen in some models are not an indictment of the Symptom dataset but rather reflect the differential impact of feature space transformation on various algorithmic biases. The strong positive results from top-performing models like Random Forest validate that the core advantage of reduced background and enhanced feature relevance outweighs these isolated drawbacks, establishing symptom-cropping as a highly effective preprocessing strategy for plant disease identification.

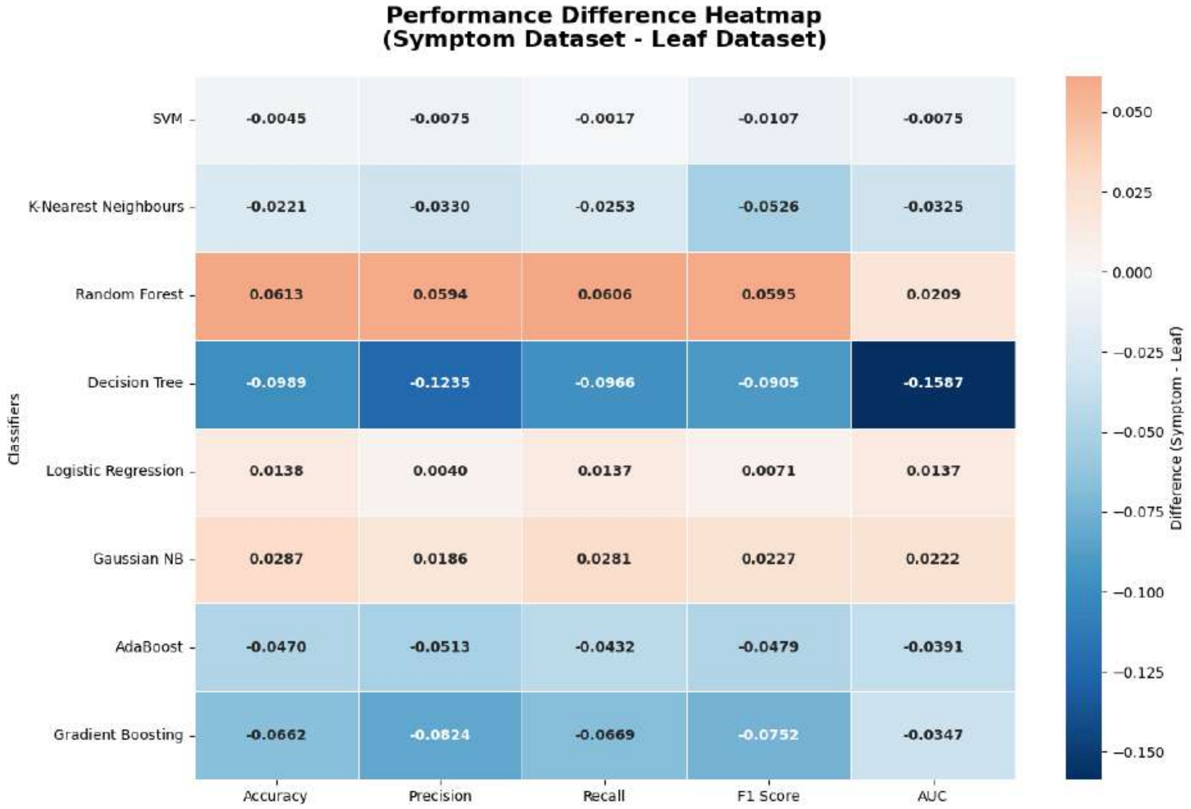


Figure 8. Performance difference heat map for both datasets

## Conclusions

This research explored the potential of deep learning techniques for leaf disease severity estimation in Arabica coffee plants. By leveraging advanced convolutional neural network (CNN) architecture of ResNet50, and the innovative approach of MTBC, this study aimed to improve the accuracy and effectiveness of disease severity assessment. Extensive experimentation was used to demonstrate the efficacy of utilizing features extracted from ResNet50, and MTBC for disease severity estimation. The results showcased promising accuracies, with the Random Forest classifier leading with an accuracy of 91.72% on Leaf dataset and 97.85% on Symptom dataset. The combination of deep learning techniques and MTBC presents a powerful approach for accurate and insightful leaf disease severity estimation. The proposed work contributes to precision agriculture by providing a reliable tool for early disease diagnosis and management to researchers and farmers. With continued advancements in artificial intelligence, the authors anticipate significant strides in agricultural sustainability and crop health management. The comparison bar chart is shown in Figure 9.

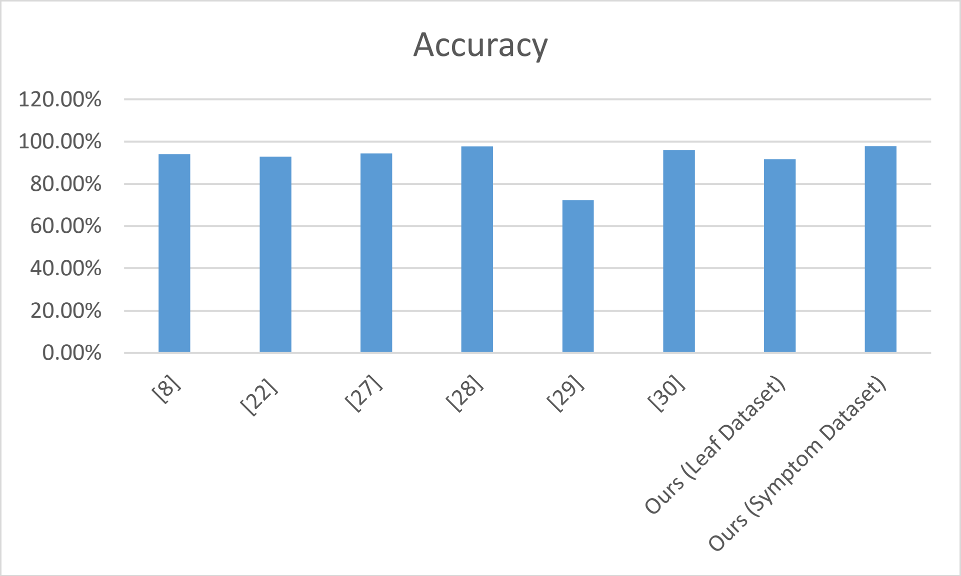


**Figure 9. Comparison bar chart of proposed system with the state-of-the-art systems**

## REFERENCE

1. Adverse climatic conditions drive coffee prices to highest level in years. 2025; Available from: https://www.fao.org/newsroom/detail/adverse-climatic-conditions-drive-coffee-prices-to-highest-level-in-years/en?utm_source=chatgpt.com.
2. Alhudaib, K. and A.M. Ismail, First Occurrence of Coffee Leaf Rust Caused by Hemileia vastatrix on Coffee in Saudi Arabia. Microbiology Research, 2024. **15**(1): p. 164-173.
3. Negesa, H., et al., Evaluating lowland coffee genotypes against leaf rust and wilt diseases in southwestern Ethiopia. Frontiers in Plant Science, 2025. **16**: p. 1560091.
4. Technology, C.P. Coffee Leaf Rust: Symptoms, Impact, and Control. 2024; Available from: https://agriculture.institute/crop-production-technology/coffee-leaf-rust-symptoms-impact-control/?utm_source=chatgpt.com.
5. Pests and Diseases of Coffee in Eastern Africa: A Technical and Advisory Manual. 2024.
6. Mienye, I.D. and T.G. Swart, A Comprehensive Review of Deep Learning: Architectures, Recent Advances, and Applications. Information, 2024. **15**(12): p. 755.
7. Raj, R. and A. Kos, An Extensive Study of Convolutional Neural Networks: Applications in Computer Vision for Improved Robotics Perceptions. Sensors, 2025. **25**(4): p. 1033.
8. Li, K., et al., A severity estimation method for lightweight cucumber leaf disease based on DM-BiSeNet. Information Processing in Agriculture, 2025. **12**(1): p. 68-79.
9. Nigam, S., et al., Automated severity level estimation of wheat rust using an EfficientNet-CBAM hybrid model. Frontiers in Plant Science, 2025. **16**: p. 1540642.
10. Nawaz, M., et al., CoffeeNet: A deep learning approach for coffee plant leaves diseases recognition. Expert Systems with Applications, 2024. **237**: p. 121481.
11. Zhao, C., et al., Plant Disease Segmentation Networks for Fast Automatic Severity Estimation Under Natural Field Scenarios. Agriculture, 2025. **15**(6): p. 583.

12. Hitimana, E., et al., An Intelligent System-Based Coffee Plant Leaf Disease Recognition Using Deep Learning Techniques on Rwandan Arabica Dataset. Technologies, 2023. **11**(5): p. 116.
13. T. Ahmed, S.J., Md. F. Islam, J. Noor, Involution-Infused DenseNet with Two-Step Compression for Resource-Efficient Plant Disease Classification. 2025.
14. Srinivas Kanakala, S.N., Detection and Classification of Diseases in Multi-Crop Leaves using LSTM and CNN Models. 2025.
15. Shafay, M., et al., Recent advances in plant disease detection: challenges and opportunities. Plant Methods, 2025. **21**(1): p. 140.
16. Polly, R. and E.A. Devi, Semantic segmentation for plant leaf disease classification and damage detection: A deep learning approach. Smart Agricultural Technology, 2024. **9**: p. 100526.
17. Mu, Y., et al., Semantic Segmentation of Corn Leaf Blotch Disease Images Based on U-Net Integrated with RFB Structure and Dual Attention Mechanism. Agronomy, 2024. **14**(11): p. 2652.
18. Ahmed, S., et al. DExNet: Combining Observations of Domain Adapted Critics for Leaf Disease Classification with Limited Data. in Asian Conference on Pattern Recognition. 2025. Springer.
19. Song, Y. and C. Yang, DS_FusionNet: Dynamic Dual-Stream Fusion with Bidirectional Knowledge Distillation for Plant Disease Recognition. arXiv preprint arXiv:2504.20948, 2025.
20. Salka, T.D., et al., Plant leaf disease detection and classification using convolution neural networks model: a review. Artificial Intelligence Review, 2025. **58**(10): p. 322.
21. Pal, A. and V. Kumar, AgriDet: Plant Leaf Disease severity classification using agriculture detection framework. Engineering Applications of Artificial Intelligence, 2023. **119**: p. 105754.
22. Jamadar, R.A. and A. Sharma. Semantic Segmentation Based Leaf Disease Severity Estimation Using Deep Learning Algorithms. in 2023 International Conference on Emerging Smart Computing and Informatics (ESCI). 2023. IEEE.
23. Mafukidze, H.D., et al., Adaptive thresholding of cnn features for maize leaf disease classification and severity estimation. Applied Sciences, 2022. **12**(17): p. 8412.
24. de Carvalho Alves, M., et al., Insights for improving bacterial blight management in coffee field using spatial big data and machine learning. Tropical Plant Pathology, 2022: p. 1-22.
25. de Carvalho Alves, M., et al., The role of machine learning on Arabica coffee crop yield based on remote sensing and mineral nutrition monitoring. Biosystems Engineering, 2022. **221**: p. 81-104.
26. Santhosh Kumar, S., et al. A Study on Core Challenges in Coffee Plant Leave Disease Segmentation and Identification on Various Factors. in ICDSMLA 2021: Proceedings of the 3rd International Conference on Data Science, Machine Learning and Applications. 2023. Springer.
27. Liu, B., et al., Deep learning architectures for diagnosing the severity of apple frog-eye leaf spot disease in complex backgrounds. Frontiers in Plant Science, 2024. **14**: p. 1289497.
28. Bi, Z., et al., Apple leaf disease severity grading based on deep learning and the DRL-Watershed algorithm. Scientific Reports, 2025. **15**(1): p. 30071.
29. Leite, D.V., et al., Deep Learning Models for Detection and Severity Assessment of Cercospora Leaf Spot (Cercospora capsici) in Chili Peppers Under Natural Conditions. Plants, 2025. **14**(13): p. 2011.
30. Dame, T.A., G.B. Adera, and D.W. Girmaw, Deep learning-based potato leaf disease classification and severity assessment. Discover Applied Sciences, 2025. **7**(7): p. 649.

31. Esgario, J.G., R.A. Krohling, and J.A. Ventura, Deep learning for classification and severity estimation of coffee leaf biotic stress. Computers and Electronics in Agriculture, 2020. **169**: p. 105162.
32. Manso, G.L., et al., A smartphone application to detection and classification of coffee leaf miner and coffee leaf rust. arXiv preprint arXiv:1904.00742, 2019.
33. Barbedo, J.G.A., Plant disease identification from individual lesions and spots using deep learning. Biosystems engineering, 2019. **180**: p. 96-107.
34. Zhao, M., H. Zhang, and J. Sun, A novel image retrieval method based on multi-trend structure descriptor. Journal of Visual Communication and Image Representation, 2016. **38**: p. 73-81.
35. Yu, L., et al., Multi-trend binary code descriptor: a novel local texture feature descriptor for image retrieval. Signal, Image and Video Processing, 2018. **12**.
36. Fatima, I., S. Adnan, and W. Ahmad, Deep Learning Based Medicinal Plant Identification for Enhanced Botanical Conservation. International Journal of Innovations in Science and Technology, 2025. **7**: p. 1842-1855.
37. Razavi, M., S. Mavaddati, and H. Koohi, ResNet deep models and transfer learning technique for classification and quality detection of rice cultivars. Expert Systems with Applications, 2024. **247**: p. 123276.
38. Xu, G., et al., Development of residual learning in deep neural networks for computer vision: A survey. Engineering Applications of Artificial Intelligence, 2025. **142**: p. 109890.