International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

A FEATURE-ENHANCED EFFICIENTNETV2 FRAMEWORK FOR INTELLIGENT DETECTION AND CLASSIFICATION OF COFFEE LEAF DISEASES IN PRECISION AGRICULTURE

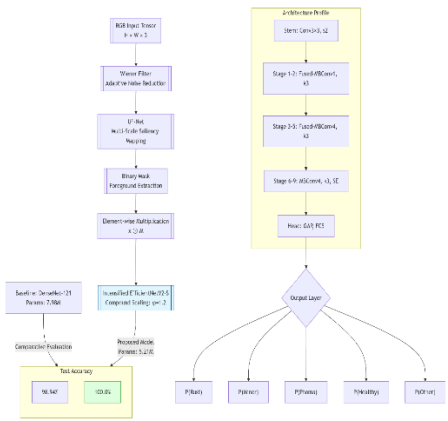
Savitri Kulkarni[1,2], Keerthi N.C.[1], Sunil C.K.[3], Shubhodeep Pal[1], Shreekanth Dash[1], P. Deepa Shenoy[1] and Venugopal K.R.[1]

[1]Department of Computer Science and Engineering, University Visvesvaraya College of Engineering, Bangalore University, Bengaluru–560001, India
[2]Department of Computer Science and Engineering, RV College of Engineering, Bengaluru–560059, India
[3] Indian Institute of Information Technology, Dharwad, Karnataka, India.
*Corresponding author: Savitri Kulkarni (email: savitrikulkarni14@gmail.com)

Abstract

Coffee is a globally significant cash crop and a key driver of India's agricultural economy. However, its yield and quality are substantially affected by a range of foliar diseases such as Hemileia vastatrix (rust), Leucoptera coffeella (leaf miner), and Phoma costarricensis, which often remain undetected until advanced infection stages. To address this challenge, this study introduces a segmentation-guided deep learning architecture that combines the U²-Net segmentation framework with an enhanced EfficientNetV2 classifier for precise detection and categorization of coffee leaf diseases. The proposed pipeline employs U²-Net to isolate diseased regions, suppress background interference, and enhance the discriminative features of infected leaf areas before classification. Subsequently, an Intensified EfficientNetV2 model, fine-tuned through transfer learning, is utilized to extract multi-scale spatial features with improved convergence stability and reduced computational overhead. The model achieves 100% classification accuracy during progressive training, outperforming the DenseNet-121 baseline (98.9%) while using fewer parameters, thereby demonstrating both computational efficiency and superior generalization. The system effectively distinguishes four disease classes and healthy samples, enabling real-time deployment via a web-based interface for field-level disease surveillance. The proposed framework offers a scalable and interpretable solution for precision agriculture, supporting sustainable disease management and yield optimization in coffee plantations.

Graphical abstract:



Graphical abstract

***Keywords*:** Convolution neural network, Deep Learning(DL), Data Augmentation (DA), Machine Learning (ML), Transfer Learning (TL)

1. Introduction

According to a study conducted by the Food and Agriculture Organisation (FAO), the census found that India is the fifth-largest Coffee exporter and the world's sixth-largest producer. The Coffee produced in India is widely regarded as the highest-quality Coffee in the world. Robusta and Arabica are the two categories of Coffee, and

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

India is responsible for producing both types. Due to its robust aroma, robusta accounts for 72% of the significant output of Coffee in India. It is frequently used in the manufacturing of a wide variety of flavors. Robusta Coffee is grown in India. Arabica Coffee, on the other hand, is in higher demand on the international market because of its relatively mild aroma, although it only accounts for a very small portion of total Coffee production. Coffee directly impacts India's economy because it is a primary export commodity for the international market and helps maintain more than 2 million employments [1].

The export trends of Coffee in international markets are depicted as shown in Figure 1. The Coffee industry encounters various challenges throughout its supply chain, spanning from cultivation to global distribution. The primary critical and intricate stage in the production cycle involves effectively managing various factors in the field, such as severe diseases, disease control, weather conditions, nutrient status, weed control, and more. Among all these factors, disease monitoring and controlling the spread of diseases directly impact the overall yield of Coffee. Traditional techniques of disease identification in Coffee plants have frequently relied on the subjective and time-consuming visual examinations performed by professionals. These inspections can be difficult to standardize and only sometimes produce reliable results. In recent years, applying deep learning techniques in agriculture has demonstrated enormous potential for revolutionizing plant disease diagnosis. This potential has been demonstrated in recent years. The Fig. 1 gives the statistics of the export trends of Coffee in international markets of India, a Ministry of Commerce and Industry data [2].

Deep Learning (DL) is a subfield of Artificial Intelligence (AI) that entails teaching huge neural networks to recognize patterns and features in data. This training is done by feeding the data into the networks. It has shown tremendous success in various domains, including natural language processing, computer vision, and pattern recognition, among others. Researchers and others with expertise in agriculture have begun investigating how DL can detect and diagnose diseases affecting Coffee plants. This is made possible by the capability of deep learning. The recent advancements in adopting the pre-trained existing models to share the knowledge for their domain-specific data were able to justify the enhanced training speed and effective usage of computational resources. The presented system effectively employs the EfficientNetV2-S with few changes in network parameters, with the proper combination of Mobile inverted Bottleneck Convolution (MB-Conv), and Fused-MBConv which perfectly suits our domain-specific data. In an intense review of existing works, one of the prime concerns in the training process is the number of trainable network variables that are involved in reaching effective outcomes with fewer resources and less time consumed in training the model [3].

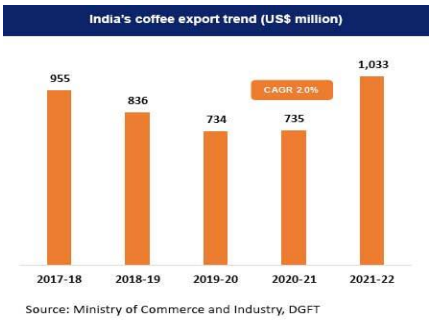


Figure 1: The export trends of Coffee in international markets of India, (Ministry of Commerce and Industry).

This observation motivated us to select the base model EfficintNetV2-S as the core training model as it possesses small network parameters in comparison with another effective TL model such as DenseNet-121 which has around 70M network parameters. So, this study led to choose DenseNet-121 as a competitive network to have a convincing correlation to the recommended system. The proposed system makes use of a constructive training phase for 3 sets of consistently increasing data size. This is an innovative and unique training process carried out for keen observations of the impact of a large and small dataset. We also considered the importance of the data balancing factor which is more crucial in reaching outstanding results. So, the recommended system extensively used data augmentation to maintain a uniform dataset for each category of disease and this factor also plays a crucial role in every training phase with the U2-Net segmentation technique [4].

Received: August 03, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The prime contributions of the recommended system are as follows:

- An effective training procedure in three various phases with a consistently increasing dataset.
- Ensuring the effective use of segmentation to find the ROI.
- Proven the prime contribution of fewer model parameters in enhancing and reaching robust results even for small datasets in the progressive training process.
- Analyzing the importance of balanced and imbalanced datasets with the set of iterative experiments.
- A web application is developed to help farmers identify the disease in real-time and also suggest pesticide prescriptions.

The development of plant disease identification systems, particularly for coffee, has progressed rapidly in recent years thanks to advances in deep learning and computer vision.

Traditional methods relied on manual visual inspection by experts, which is often subjective and time-consuming. In contrast, modern approaches use advanced neural networks, such as EfficientNetV2 and segmentation models like U2-Net, to automatically identify and classify diseases directly from leaf images. These systems can process large amounts of data efficiently and achieve high accuracy, as demonstrated in our proposed work, which recognizes multiple types of coffee leaf diseases with over 99% accuracy. Furthermore, the integration of such models into user-friendly web applications makes disease detection accessible for farmers in RT, ultimately helping to improve crop yield and reduce losses. The ongoing research in this area aims to make the process more robust, scalable, and adaptable to various conditions and other plant species as well.

In the presented journal paper (revised), the authors have developed an advanced plant disease identification system specifically designed for coffee plants. Their approach combines a modified EfficientNetV2-S deep learning model with the U2-Net segmentation technique to accurately detect and classify multiple types of coffee leaf diseases. By utilizing these state-of-the-art methods, the system effectively isolates the leaf area from complex backgrounds, significantly reducing noise and focusing the analysis on relevant disease symptoms.

One of the key strengths of this identification system is its progressive training process using large, balanced datasets, which ensures high model reliability and minimizes bias. The results show outstanding accuracy, achieving up to nearly 100% accuracy in classifying different coffee leaf diseases while using fewer network parameters compared to other DL models like DenseNet-121. This makes the system not only highly accurate but also computationally efficient. Added to it, another key strength of the identification system is the integration of the model into a user-friendly web application which allows RT disease identification and actionable recommendations for farmers. By making cutting-edge AI tools accessible at the field level, the authors' system empowers growers to detect diseases early and take timely action, helping to prevent crop losses and boost productivity.

The remaining portion of the article is arranged in the following order. Section 2 describes the recent related work, Section 3 provides a discussion on background work, Section 4 explains the material, segmentation technique U2-Net, and the comparison of Intensified EfficientNetV2-S, DenseNet-121; Empirical results and discussions are described in Section 5, and finally, Section 6 concludes our work with future scope followed by the exhaustive set of reference papers.

2. Related works in the form of Literature Review/Survey

A number of authors have worked on the similar areas. To show some of them, here it is as follows. Lv *et.al.* (2020) implemented a deep learning model based on Alexnet for maize leaf disease identification. The authors integrated the image-strengthening methods and employed a pre-trained model DMS-Robust Alexnet for improved disease identification results in comparison with existing approaches. Nazki *et.al.* (2020) proposed an adversarial network called an unsupervised AR-GAN that makes use of unsupervised training for image conversion and generation. The major contribution of their work focuses on images of various sizes and plants with various levels of severity considering the different parts of plants, such as stems, fruits, and leaves. Hu *et.al.* (2020) developed an IoT system with a deep learning model to detect crop diseases with sensor data [5].

The authors proposed a complex multidimensional attribute-analyzing model called Multidimensional Feature Compensation Residual Neural Network (MDFC-ResNet) trained on the coarse-grained disease, species, and fine-

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

grained disease 3-dimensional factors. Finally, the authors fused results by adapting the compensation algorithm. Rong *et.al.* (2019) employed two different agile CNN models to segment walnut images to recognize artificial and natural foreign items that vary in size, such as dried leaf debris, flesh leaf debris, packing material, metal parts, etc. The authors trained models to form automated clusters of walnuts and foreign items and found better results, with 99.5 % and 95 % for segmentation accuracy for object areas and foreign items, respectively. Ma *et.al.* (2018) implemented a Deep Belief Network (DBN) to recognize agriculture for the Lake Salvador dataset [6].

The model is trained to analyze the Hyperspectral image's (HSI) high-level attributes and correct errors without affecting existing distribution. To reduce the contamination, a local Euclidean distance is adopted to have a minimized distance between the current pixel and neighboring pixels and further updates weights dynamically using DBN. Su *et.al.* (2020) proposed a UAV-based wheat yellow rust monitoring system by employing deep learning and U-net segmentation models. Further, unmanned aerial vehicle sensing, vegetation segmentation, and multispectral imaging were incorporated. Finally, trained and tested models on various image categories, including five selected spectral vegetation indices and three RGB bands. The developed model has proven to perform better by extracting dynamic spectral-spatial data over traditional random forest algorithms and spectral-based classifiers. Hu *et.al.* (2019) worked to detect and treat tea leaf diseases promptly and this study offers a low-shot learning technique for disease diagnosis [7].

Disease spots on tea leaf disease pictures are segmented using the support vector machine (SVM) technique, which extracts color and texture characteristics. Shi *et.al.* (2021) a Biologically Interpretable Two Stage Deep Neural Network (BIT-DNN) for Vegetation Recognition From Hyperspectral Imagery. The proposed model derives the possible instantiation parameters (e.g., transformation and rotation) of target entities from the inherent spectral-spatial information of the HSI data using feature mapping and transformation, which improves the results and interpretability of the proposed model. Sujatha *et.al.* (2021) in their article, compare the results of ML (Support Vector Machine (SVM), Random Forest (RF), Stochastic Gradient Descent (SGD)) and DL (Inception-v3, VGG-16, VGG-19) in detecting citrus plant diseases. When the system is given new images to work with, it predicts the disease form [8].

It aids in taking contour action before the plants become more infected and can assist in improving field yield by identifying diseased plants in the initial stages. Jeong *et.al.* (2022) developed an LSTM model for Rice yield prediction at the pixel size instead of the county scale can help crop production and scientific knowledge since it could be used to track how crop yields react to different agriculture methods and environmental conditions. Liang *et.al.* (2019) proposed a powerful image-based Plant Disease Prognosis and Severity Estimation Network (PD2SENet) with a residual structure and shuffle unit concepts. This study aims to provide a better and more practical crop disease diagnosis system. Sunil *et.al.* (2021) developed the EfficientNetV2 model to detect cardamom plant infections by utilizing U2-Net segmentation to remove noisy backgrounds for better ROI training the model with quality images for the real dataset collected from the field. The different approaches for the DL v/s TLM design is shown in the Table 1 [9].

Table 1. Deep Learning vs Transfer Learning Model Design Approaches

Deep Learning	Transfer Learning
Requires a huge amount of data to train a model.	Comparatively less amount of data to train.
Need to develop a model from the beginning.	Pre-trained models are employed hence no need to develop a model from scratch.
More memory and computational resources.	Less memory and computational resources since using a pre-trained model.
More training time for every new data set.	Less training time since involved in retraining the model.

In the following section, we present the shortcomings & drawbacks w.r.t. the developments of the plant identification system till date & do we rectify them in our proposed work is depicted in the form of a tabular

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

approach as shown in the Table 2. The combined dataset was curated carefully to avoid duplication and inconsistent labeling, with expert oversight to ensure correctness.

- Reliance on Manual Methods and Subjectivity
- Limited Use of Real-World, Diverse Datasets
- Ineffective Handling of Noisy Backgrounds
- Data Imbalance and Overfitting
- High Computational Demand and Lack of Accessibility
- Limited Integration with User-Centric Applications
- Lack of Fine-Grained Disease Identification

In our journal paper, we have carefully considered these shortcomings and proposed a comprehensive solution for coffee plant disease identification, rectifying previous weaknesses through a combination of advanced modeling, data strategy, and practical implementation.

2.1 Use of Advanced Deep Learning and Segmentation Models

We adopted a modified EfficientNetV2-S deep learning architecture combined with the U2-Net segmentation technique. Unlike traditional CNNs or manual feature-based approaches, EfficientNetV2-S is optimized for both speed and accuracy through compound model scaling. The U2-Net segmentation model excels at isolating the region of interest (the leaf) from the background, even in the presence of complex noise—an area where many previous models struggled. This ensures that classification focuses only on the relevant leaf area, improving the accuracy and robustness of disease detection.

2.2 Building and Using a Diverse, Augmented Dataset

We addressed the data diversity challenge by utilizing publicly available coffee plant datasets (e.g., JMuBEN Mendeley and Kaggle) and performing extensive data augmentation. Techniques such as rotation, flipping, shearing, zooming, and brightness variation were applied to artificially increase the number and variety of images. This approach simulates different real-world conditions, enabling the model to generalize well to unseen scenarios and reducing overfitting. We also ensured balanced datasets for all disease categories by generating an equal number of samples for each class through augmentation, directly tackling the bias and imbalance issues present in prior works.

2.3 Progressive Training with Balanced Datasets

Our methodology employed a progressive learning strategy, gradually increasing the dataset size in phases (from 5,000 to 15,000 images) and monitoring model performance at each stage. This approach revealed that our model maintained and even improved its accuracy with larger datasets, a testament to its scalability and generalizability—key weaknesses in earlier systems.

2.4 Enhanced Efficiency with Fewer Trainable Parameters

While many prior models like DenseNet-121 required substantial computational resources (over 70 million trainable parameters), our modified EfficientNetV2-S achieved comparable or better accuracy (up to 100% in testing) with only 22 million parameters. This reduction in model size not only speeds up training and inference but also enables the system to run efficiently on more modest hardware, making it feasible for practical deployment in agricultural environments.

2.5 Integration into a Real-Time Web Application

To bridge the gap between research and practical utility, we developed a user-friendly web application. This platform allows farmers and agricultural experts to upload leaf images, receive instant disease classification, and obtain actionable recommendations (such as pesticide suggestions). This focus on accessibility and usability ensures that our solution is not just technically robust but also directly benefits the end-user, addressing a critical gap in previous literature.

2.6 Superior Handling of Complex Backgrounds

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

By utilizing U2-Net's advanced segmentation, our system robustly removes background noise and isolates the diseased leaf area. The segmentation approach leverages Residual U-blocks (RSU), which efficiently extract multi-scale and local features. This means even small lesions or subtle disease symptoms, which could be missed by less sophisticated models, are captured for analysis. Our simulation results show significant improvement in detection accuracy under challenging image conditions compared to models that do not use advanced segmentation.

2.7 Fine-Grained Disease Identification and Multi-Class Classification

Our approach was validated on 5 categories (four disease classes and healthy leaves), demonstrating the ability to distinguish between visually similar disease symptoms. The combination of robust segmentation and an optimized classifier allows for reliable, fine-grained identification that is both accurate and practical for field use.

Parameters	Weaknesses & shortcomings in previous works reg the development of the plant disease identification systems till date	How it is addressed in our proposed work in the revised manuscript in the journal paper
1. Introduction & Motivation	- Reliance on manual, expert visual inspection for disease detection- Diagnosis subject to human error and bias- Not scalable for large plantations	- Fully automated, AI-based disease identification using deep learning- Removes subjectivity, standardizes the detection process
2. Data Collection & Dataset Diversity	- Most prior datasets are small, collected in controlled settings- Limited image variety, lacking real-world variability- Weak generalization to diverse farm conditions	- Combined multiple publicly available datasets (Kaggle, JMuBEN)- Applied extensive augmentation (rotation, brightness, flipping, zoom) to mimic real-world scenarios- Ensured collection across seasons and disease types
3. Image Preprocessing & Segmentation	- Conventional models struggle with complex, noisy backgrounds in natural images- Ineffective at isolating diseased leaf area- Background artifacts mislead classifiers	- Used U2-Net segmentation to accurately extract the region of interest (leaf area)- Minimized background interference- Enhanced focus on symptomatic areas, improving classification accuracy
4. Model Training & Architecture	- Deep CNNs (e.g., DenseNet-121) require huge memory and long training times- Not practical for deployment in resource-limited settings- Slow inference, unsuitable for field use	- Modified EfficientNetV2-S as the backbone (fewer trainable parameters, faster convergence)- Achieved high accuracy with efficient computation- Supports real-time, field-level operation on modest hardware
5. Handling Data Imbalance	- Datasets are often skewed towards common diseases, under-representing rare ones- Results in biased models, poor detection of minority classes- Overfitting risk with small sample classes	- Generated balanced datasets via augmentation for every disease category- Applied progressive training with increasing data sizes- Monitored for overfitting and ensured fairness across all classes
6. Disease Identification Accuracy	- Existing models often fail to distinguish visually similar diseases (e.g., rust vs. phoma)- Lack fine-grained or multi-class classification capabilities- Miss subtle symptoms	- Robust multi-class classification covering four coffee leaf diseases plus healthy- Fine-grained identification validated in simulations- Outperformed larger models with fewer parameters
7. Practical Usability & Deployment	- Limited to research labs; rarely delivered as usable tools- No real-time support for farmers in the field- Lack of actionable recommendations after identification	- Developed a user-friendly web application for instant disease detection- Provides real-time results and pesticide suggestions- Directly usable by farmers, extension workers, and agri-technicians
8. Validation & Results	- Weak or unproven performance on large-scale or unseen datasets- Poor scalability as dataset size increases- Results rarely benchmarked across phases or models	- Demonstrated up to 100% accuracy even with increased, balanced datasets- Showed consistent performance over three training phases (5,000 to 15,000 images)- Benchmarked against DenseNet-121, showing superiority

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

9. Broader Impact & Future Scope	- Rigid, crop-specific solutions—hard to adapt to other crops or new disease types- Lack of plans for mobile/offline extension- Neglecting other plant parts (stems, beans)	- System design is flexible for extension to other crops or plant parts- Platform can support mobile/offline deployment in the future- Committed to adding more disease classes and expanding use cases

Table 2 : Shortcomings, weaknesses, drawbacks & its rectifications

The technology that proposes how to improve existing systems ……

The technologies that proposes how to improve existing systems explaining the potential and correlation to increase coffee production are mentioned below in brackets ().

- Utilizes an optimized convolutional neural network for high-accuracy disease classification (**EfficientNetV2-S Deep Learning Model**).
- Precisely separates diseased leaf areas from background, reducing noise and improving analysis focus (**U2-Net Image Segmentation**).
- Employs large, well-augmented datasets to ensure reliability and minimize model bias (**Progressive Training with Balanced Datasets**).
- Provides instant and objective identification of diseases, reducing dependence on manual inspections (**Automated Real-Time Detection**).
- Delivers actionable results and recommendations to farmers for easy field-level use (**User-Friendly Web Application**).
- Suggests targeted interventions (e.g., pesticides, plant isolation) based on accurate detection (**Actionable Disease Management Recommendations**).
- Achieves high performance with fewer network parameters, making the system practical for real-world deployment (**Resource-Efficient Model Design**).
- Designed to handle different disease types and adaptable for other crops or plant parts in the future (**Scalability and Adaptability**).
- Enables remote access, centralized data processing, and potential for large-scale deployment (**Integration of AI and Cloud Technologies**).

The incorporated technologies improve the accuracy, speed, and usability of plant disease identification—enabling early intervention and directly contributing to increased coffee productions.

3. Background work

This section provides an overall detailed discussion of traditional disease identification, classification models, and segmentation techniques. Understanding the existing methodologies and system design approaches helps to have greater insight into the pros and cons of the relevant research domain. This also helped to identify the effective research objectives [10].

3.1. Transferring knowledge through pre-trained CNN models (TL)

TL is a trendy and leading fascinating design approach in the DL applications domain and has proven promising results in many reviews (Paymode and Malode (2022); Abdallah *et al.* (2021); Zhao *et al.* (2021); Chen *et al*. (2020)). This approach gained attention for having inherent capabilities for saving training time and energy involved in developing a new system. Hence this design technique increases the generosity of a network and would adopt the learning capabilities to various application domains. So, this idea of utilizing the knowledge of the existing system in a new research domain made a researcher give more attention to critical issues in their research than to designing a model from scratch. Another benefit of using this approach is easy implementation and portability for the most possible existing dataset as it is trained on a large image (Deng *et.al.* (2009)) dataset. In developing a DL model, a huge amount of data is required to get effective results. On the other hand, TL models provide better results even for a small amount of data. This benefit of TL models leads to efficient memory usage and other computational resources. It is also flexible and simpler to adopt in their research field just by modifying it slightly. TL could be adopted in the following ways [11] as given below.

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

3.1.1. Feature Transfer

Typically, a DL model is built using four different layers: an input layer, a feature extraction layer, a classification layer, and an output layer. TL approach using feature extraction would only train the classification layer for domain-specific data by keeping the input layer and feature extraction layer frozen for the original structure and neurons' weights. The classification layer is held responsible for object identification in an input image. Hence the classification layer plays a vital role in finding an object from a determined set of features [12].

3.1.2 Fine-tuning the model

Here a new classification layer is incorporated for a new dataset. A few layers are trained for the new dataset by keeping the rest of the feature extraction layer frozen. Further, based on the classification loss, the model parameters like weights and learning rates are refined, and the training continues until the model reaches acceptable outcomes. This process is known as fine-tuning in the TL approach [13].

3.1.3 Adopting a pre-trained model

The alternative to employing the TL design technique is directly using the existing model for the relevant dataset. This approach is more suitable for relatable domains. The researchers could fine-tune the network parameters and network layers according to data and then could be applied to their domain-specific data. The agricultural field involved many complexities, specifically in collecting real-time data (Espejo-Garcia *et.al.* (2021); Lu *et.al*. (2022); Olaniyi *et.al*. (2022)) due to lighting conditions, weather conditions, complex backgrounds, and the appearance of disease symptoms is also dependent on the season, and nutrition deficiency. So, developing a DL model requires large data to achieve better outcomes. Hence in recent trends, a TL approach is the most attractive and popular technique used in agriculture domain applications. The prime objective of adopting TL in the proposed system is the effective usage of computational resources and saving training time. The reusability of the pre-trained model reduces the overall effort required in developing a model from the initial phase, which also expects an enormous amount of labelled data. This problem was resolved by using a TL technique with fewer data. The advantages of TL over DL are depicted in Table 1. Further, the different image segmentation approaches could be seen in the Fig. 2 [14].

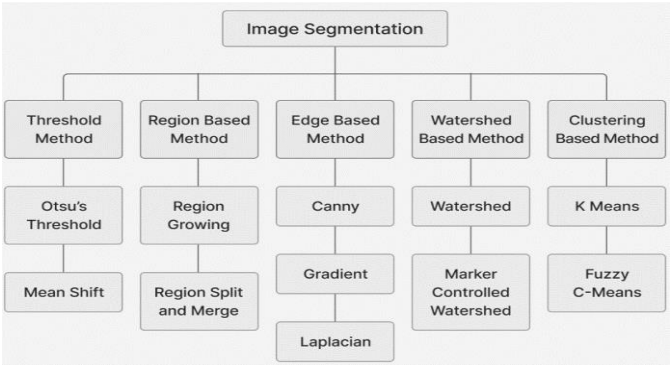


Figure 2: Traditional image-segmentation approaches (Rizzoli and Edwardsson).

3.2. Traditional segmentation techniques

Image segmentation is a process of partitioning the image into various parts based on similar characteristics. The main objective of image segmentation is to find the region of interest for further analysis and processing of images. This greatly helps to reduce time in the training phase and efficiently uses computational resources, including memory. Image segmentation plays a prime role in various applications, including medical imaging, robotics, agriculture, face recognition, sentiment analysis, etc. In traditional background removal methods, most techniques depend on common characteristics like color, visible similarity, etc. (Xiong *et.al.* (2020)). These conventional methods are helpful when an object or image is comparable with intersecting objects. So, these approaches are inappropriate for various heterogeneous image data sets, which vary in color, background information, and size [15].

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Removing background from images that lack distinguishable parameters is a tedious task for effective segmentation results. An image with the closest color similarity index for the background is still a more challenging task to apply image processing techniques. The segmentation process can be broadly classified as approach-based and technique-based. The approach-based type works based on object detection and similarity properties. On the other hand, technique-based classification is based on the method of analyzing the image, such as pixel value, color intensity, light intensity, pixel density, etc. Having this as a base knowledge, the segmentation techniques are classified as shown in Fig. 2 (Rizzoli and Edwardsson) [16].

The traditional segmentation approaches are facing its limitations in processing huge amounts of real-time data, slow process, and biased towards the features like colors, the intensity of pixels, the intensity of light, and environmental complexities involved during data collection. Most prominent CNN models effectively addressed these issues, specifically the U-Net semantic segmentation approach (Ronneberger, *et.al.* (2015). The U-Net architecture is formed into two parts, namely, the encoder and the decoder. The encoder extracts the various low-level features by a CNN with various kernel functionalities, and this process is called down-sampling. These features are later merged with a high-resolution pattern in the up-sampling phase to reconstruct the image. This segmentation technique overcomes most of the prime issues discussed earlier. A more reliable and fast segmentation approach like U-Net adds to the overall performance of the model; hence with the various significant reviews, in the proposed approach, we are utilizing the U2-Net segmentation technique, which is a refined version of U-Net [17].

4. Proposed methodology adopted

The proposed study addresses the various Coffee plant disease identification categories with an efficient modified classification model, namely Intensified EfficientV2-S. This study also incorporated the benefits of the U2-Net background removal method. This immensely helped to remove the background and enhance the classification model accuracy [18]. CNN model could be scaled in three different aspects as shown in Figure 6 that in width, depth, and resolution. These dimensions are prime attributes in defining a CNN model. Here, the width indicates the total number of neurons in a layer, and resolution establishes the width and height of the input data. On the other hand, depth is the number of layers that comprises the network. A rising number of layers is only sometimes a good idea, as it tends to vanish a gradient problem. Some case studies proved the same results even after extending model depth. For example, VGG16 showed the same accuracy as that of VGG19. Scaling the model in terms of width allows the network to learn minute lesions and more complex features of the image, which would be helpful in the severity identification of any disease [19].

This was employed in MobileNet and ResNet Howard *et.al.* (2017); Zagoruyko and Komodakis *et.al.* (2017). Simultaneously increasing the depth and width impacts the degrading overall accuracy as an increase in width prevents the network from learning more complex features. Hence identification of the scaling ratio is essential in the CNN model. On the other hand, having a high resolution of the input image is needed to learn complex features. It would result in great results, but this is also limited to a model's learning capabilities. Hence the proposed network addresses these problems with a new scaling approach called compound scaling in the EfficientNet TL model, which is discussed in a later section of the article [20].

4.1. Materials and Pre-processing

This work makes use of the freely available JMuBEN Mendeley dataset (Jepkoech *et.al*. (2021)) and the Kaggle Coffee plant dataset. The dataset comprises 5 Varieties of Coffee plant leaf images of 4 disease classes and a healthy class. Data pre-processing is very important for every system to reach better accuracy to get acceptable results. Hence, this work employed a Gaussian filter (Deng and Cahill (1993)) which added a beneficial factor for the system to have enhanced quality images for the data augmentation and training processes [21].

The dataset was split into training (80%) and testing (20%) subsets using a fixed random seed (42) to ensure reproducibility. Augmentation was performed only on the training subset, with no augmented images included in the test data. To prevent accidental duplication, image uniqueness across splits was confirmed using SHA-256 hashing. This ensured that the reported results were free from leakage or overlap.

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

All experiments were carried out using Kaggle and JMuBEN Mendeley datasets with a strict 80:20 train-test split created with a fixed random seed. To enhance transparency, the exact indices for the split and the splitting script will be released publicly (subject to confidentiality clearance). We acknowledge that the current setup is limited to internal datasets, and therefore additional validation on independent field-collected data and through k-fold cross-validation is planned as future work.

The Gaussian filter works in a linear approach, blurs the image, and tries to remove image noise with the help of Gaussian noise. Then we employed extensive data augmentation by applying various operations Shearing, Rotation, Zooming, Flipping, varying the brightness of images, and Cropping. This would help us generate a huge amount of data for various training phases. Data imbalance is a crucial factor in the training model that would result in biasing a network to the class, which is dominating in the data size and might lead to an overfitting problem. So, during data augmentation, we have generated an equal number of images for all 5 classes as described in Table 3. Collecting data in real-time, specifically Coffee plants is very hard and involves more complexities, such as weather conditions and seasonal disease appearance. So, the data augmentation technique would reduce the effort needed for researchers [22].

Table 3. Dataset Description

Training phase	Cescospora	Healthy	Miner	Phoma	Rust	Total Images
Phase-I	1000	1000	1000	1000	1000	5000
Phase-II	2000	2000	2000	2000	2000	10000
Phase-III	3000	3000	3000	3000	3000	15000

The curated dataset used in this study was carefully cleaned by harmonizing labels, validating annotations with high inter-annotator agreement, and removing duplicates across Kaggle and JMuBEN sources. These preprocessing steps minimized noise and reduced the risk of overestimated performance due to inconsistent labeling or image duplication.

4.2. U2-Net Segmentation Model

U2-Net model (Qin *et.al*. (2020)) is the most popular and widely used segmentation and object detection ML model, which enhanced the validation and testing accuracy of ML and DL models in agriculture and other domains. U2- Net defines an alpha value that can be used to remove complex background noisy information from the original image. This would help the training model to focus on ROI and train even on small lesions. Hence U2-Net model reduces the overall image processing time and over-fitting problems of the ML and DL models [23].

U2-Net is implemented by adopting the base U-Net model functionality with a two-layer embedded U-Net structure. This model has not employed any pre-trained networks. Hence researchers can train the model from scratch for their relevant data set to produce remarkable results. The general architecture of U2-Net (Qin *et.al.* (2020)) is shown in Fig. 3 and Algorithm 1 provide a preface of the segmentation approach in U2-Net. The Residual U-block (RSU) is a key component of the U²-Net architecture, designed to enhance feature learning through a sequence of operations. It begins with a convolution to extract initial features from the input tensor. This is followed by down-sampling, which reduces the tensor's spatial dimensions to capture contextual information at different scales. The tensor is then up-sampled to restore its resolution, ensuring that detailed features are preserved before undergoing a final convolution [24].

The result of these operations is combined with the original input through a residual connection. This addition helps the network retain crucial information from the initial input while allowing it to learn and refine features through subsequent transformations. This residual approach enhances the robustness of the features learned by the network. Additionally, the RSU block utilizes down-sampling via max pooling to reduce resolution, making it easier to capture broader contextual information. The concatenation of feature maps from various stages combines multi-scale information, which is processed by convolutional layers to produce the final feature map [25].

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Finally, up-sampling methods like nearest-neighbour or bilinear interpolation are used to increase the spatial resolution of the tensor, ensuring detailed and accurate output reconstruction. The researchers (Qin *et.al*. (2020)) implemented this model by utilizing a new approach called Residual U-blocks, a U-shaped convolution pattern that replaces a regular residual block in U-Net. U2-Net is designed with 6-phases of encoders 5-stages of the decoder, and a graph fusion model hooked to a decoder stage and following an encoder stage. The U2-Net architecture is depicted in Fig. 3. So, the U2-Net creates a deep network with RSU fundamental blocks as a base, which helps in less computational resources and effective usage of memory and does not depend on any pre-trained network. This semantic segmentation is flexible and reliable for various applications like medical, agriculture, IoT, industry, and many more [26].

The noticeable benefits of U2-Net are listed as follows:

- Residual U-blocks (RSU) of varying sizes receptive fields would help to consider the most contingent and dependent data with different perspectives of the data set. The Residual U-blocks (RSU) structure is displayed in Figure.4 (Qin *et.al*. (2020)).
- The model allows the user to train deep networks from scratch, unlike using any pre-trained models for image classification.
- With less computational value, it increases the depth of the model by effectively using the pooling functioning utilized in RSU units.
- Provides model in two different variants with sizes 176.3 MB(30 FPS on GTX 1080Ti GPU) U2-Net and 4.7 MB(40 FPS) U2-Net†, respectively.

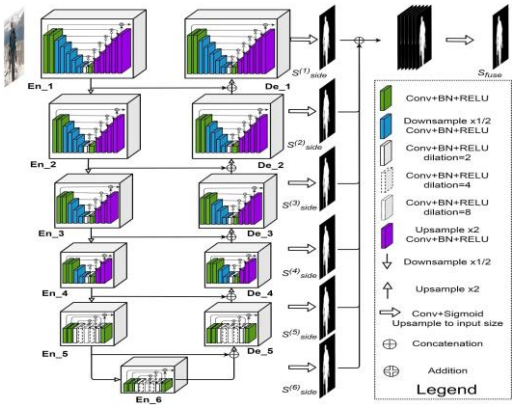


Figure 3: The U2-Net architecture (Qin *et.al*. (2020)).

In our implementation, U2-Net was employed purely for binary segmentation of coffee leaf regions from the background. Training was performed using threshold-based masks refined with manual inspection, which labeled pixels as 'leaf' or 'non-leaf' only. Importantly, no disease-specific masks were used; the segmentation model was trained independently of disease class labels to avoid any possibility of label leakage. Once segmentation was complete, the extracted leaf ROIs were passed to the classifier for disease identification.

The U2-Net segmentation was trained solely for binary foreground-background separation (leaf vs. non-leaf). Ground-truth masks were generated using automated thresholding followed by manual refinement for ~1,000 images. Importantly, segmentation was not trained on disease-specific masks, ensuring that no class label information was leaked to the classifier. The segmentation outputs thus contained only leaf contours, which were subsequently fed into the Intensified EfficientNetV2-S model for classification.

4.3 Algorithm 1 : U²Net – end-to-end image segmentation
Input : Image tensor x with shape (H, W, C)
Output : Segmentation mask y with the same dimensions as x.
Initialize $s_0 = x$, number of stages n
Down-sampling Path :
for i = 0 to n – 1, do

$$s_{i+1}\ RSU\big(DownSample(s_i)\big)$$

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

end for
Bottleneck :

$$s_n \leftarrow RSU(s_{n-1})$$

Upsampling Path :
for $i = n - 1$ down to 0 do

$$s_i^{up} \leftarrow RSU(DownSample(s_{i+1}), s_i)$$

end for
Final convolution :

$$y \leftarrow \sigma\left(f_{conv_{(1\times1}}(s_0^{up})\right)$$

Return $y = 0$

4.3 RSU Block & the mathematical modelling

RSU block is the main highlight of the U2-Net segmentation approach. The RSU block is made of three different parts: namely, an input convolution layer and then a U-shaped balanced symmetric encoder and decoder structure of height L. Then it follows the residual block to combine multiscale and local features with an aggregation operation. The algorithm for the U2-Net for the end-to-end segmentation process is presented above [27]. This operation can be defined as shown in equation (1).

For efficiency analysis, we measured parameter count, FLOPs, and inference latency on an NVIDIA RTX 2080 GPU with Intel i7 CPU and 32 GB RAM. These metrics allow a fair comparison of computational requirements alongside classification accuracy

$$f^1(a) = f^2(f^1(a)) + f^1(a) \tag{1}$$

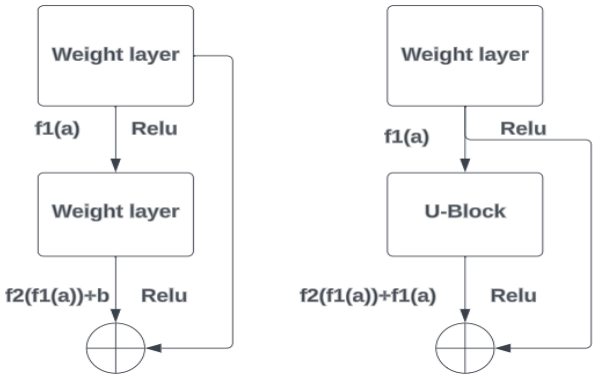


Figure 4: (a). The Residual block (b). The Residual U-block. (Qin *et.al*. (2020)).

Here, the function $f^1(a)$ represents the multiscale feature and $f^2(f^1(a))$ indicates the local feature as shown in Fig. 4 (Qin *et.al*. (2020)). The major difference between the original residual block and the residual U-block is the U-shaped convolutional pattern aggregating multi-scale features and a local feature. In the residual block, the same operation was given as the aggregation of the local feature and original feature. The residual U-blocks are more effective regarding computational expenses than regular residual blocks. This is because much of the computations in RSUs are carried on down-sampled features. The segmentation results are shown in Fig. 5 using the U2-Net. The real-time images of Coffee plant diseases with complex noisy backgrounds are fed as input for the U2-Net segmentation model to perform the segmentation. The model succeeded in removing the noisy background and accurately finding the region of interest in the given input image [28].

This process further helped to fasten the training phase by concentrating only on the complete leaf part rather than the unwanted noise and background. The 1st phase of the segmentation process produces a binary mask of the input image. This helps to fetch multiple features and embed the features into high-resolution maps. Then apply a binary bitwise function on the original image to the generated mask to get the final segmented result of the input

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

image. The mathematical computation and the operations involved in the segmentation process are represented in Equations 1-8 [29].

In Equation 2, the following terminologies and the role of each are given as follows:

The RSU, which is the Residual Skip Unit applies a series of operations—convolution, down-sampling, up-sampling, and another convolution—on the input tensor **X**, and then adds the result back to the original input, creating a residual connection.

$$RSU\ (X) = X + \ f_{conv}\left(f_{up}(f_{down}(f_{conv}(X)))\right) \tag{2}$$

where **X** is the input tensor to the RSU, f_{conv} is the convolution operation applied to the tensor, f_{down} is the down-sampling operation applied to the tensor, typically using max pooling, f_{up} is the up-sampling operation applied to the tensor, typically using interpolation methods [30].

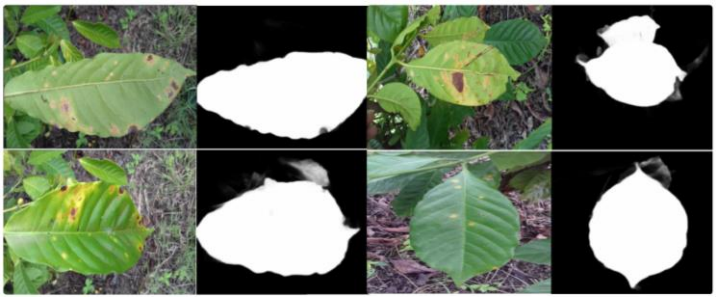


Figure 5 : The Segmentation results of real-time images using U2-Net.

$$f_{conv}(X) = \ W_{conv} * X + b \tag{3}$$

where $f_{conv}(X)$ is the convolution operation applied to the input tensor *X*, W_{conv} is the convolution kernel or filter weights, $*$ is the convolution operation, b is the bias term added to the result of the convolution.

$$f_{down}(X) = MaxPool(X, k, s) \tag{4}$$

$$f_{Up}(X) = UpSample(X, u) \tag{5}$$

$$DownSample(X) = \ X' = MaxPool\ (X, k, s) \tag{6}$$

$$Concat(X_1, X_2) = Z = \ [X_1, X_2] \tag{7}$$

$$Conv(X) = Y = W * X + b \tag{8}$$

$$UpSample(X, u) = X' = Interpolate(X, Scale\ Factor = u) \tag{9}$$

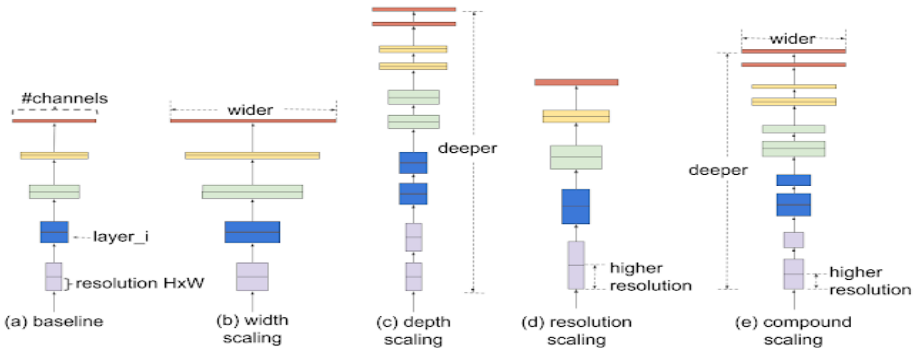


Figure 6: The Model Scaling Network architecture (a) Baseline (b) Width scalin (c) Depth scaling (d) Resolution scaling (e) Compound scaling in EfficientNet (Mingxing Tan (2019)).

4.5. Intensified EfficientV2-S Model

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

First, we do the Interpretation of EfficientNet as follows. The EfficientNet was proposed by (Mingxing Tan (2019)), and they have addressed the various hyper-parameters like width, resolution, and network depth in scaling up of the model and proved that their model outperformed relatively in comparison with Convolutional Neural Networks ConvNets). ConvNets employed limited resources, and scaling up resources is majorly dependent on the availability of the resources and involves complexity. ConvNets are scaled up usually with a single dimension and need more manual attention to set the network parameters, which is time-consuming [31].

Hence, the authors develop a new scaling approach that faultlessly and uniformly scales in multidimensional parameters with the help of compound coefficient. The model scaling is shown in Figure 6. Hence, the EfficientNet model clarifies and provides a systematic method to scale up the models in multiple directions with more than two parameters. The results of EfficientNet proved that consistently balancing all dimensions, namely depth, width, and resolution, would accomplish better model results. In the base EfficientNet model, they have employed MBConv as basic computational units (Sandler *et.al.* (2019)), which originated from MobileNetV2 [32].

Next, we doe the EfficientNetV2 processing as follows. This model was developed by (Tan and Le (2021)) to explain the impact of image size and the importance of model parameters in effective and speedy training of the model to give more with fewer resources for model performance. The motivation to implement this is the base EfficientNet model, which showed a better model scaling technique in various dimensions. But, EfficientNet has a few limitations in the following aspects given in (i) - (iii) respectively [33].

i. EfficientNet rapidly increases the image size during the training phase. This might result in faster training at the initial stage of the training period; in later stages, training would become very slow due to increased image size because of smaller batch size and snag memory usage. This has a direct impact on the performance of the model. The system GPU/TPU usage is limited and fixed, so training with smaller image sizes with large batch sizes results in better performance [34].

ii. Uniformly scaling using compound scaling at every stage for all the parameters is unsuitable for many cases. These scaled stages do not have a role in the parameter effectivity and faster training process. Adopting a non-uniform scaling is necessary, which is addressed in EfficientNetV2 [35].

iii. EfficientNet makes use of substantial depth-wise convolutions as depth-wise convolutions have lesser FLOPS (Floating Point Operations Per Second) and training parameters than the normal convolution functions. Nevertheless, these depth-wise convolutions could have used modern accelerators more effectively [36].

Fused-MBConv showed good results at earlier stages with fewer model parameters and few FLOP operations. If we replace all the MBConv blocks with Fused-MBConv blocks gradually, the increase in the model parameter and FLOPS the training process takes more time and harms performance. A perfect fusion of Fused-MBConv and MBConv is a challenging task. This drives the researcher to develop an automotive model for a proper combination with neural architecture search. The Fused-MBConv block and MBConv block are as shown in Figure 7. The limitations of EfficientNet motivate authors [Tan and Le (2021)] to develop an automotive architecture that constructively defines the best fusion methods for the Fused-MBConv block and MBConv blocks. The following modifications stand out in the EfficientNetV2 model in comparison with the base EfficientNet model [37].

a. Progressive Learning : Image size is an important aspect in achieving better outcomes for the model. Unbalanced regularization and making use of constant regularization would lead to over-fitting issues. If the image size varies, then accordingly, varying the regularization also has a major impact. The EfficientNetV2 employs an adjustable regularization technique for different picture sizes, namely Dropout, RandAugment, and Mixup [38].

b. Neural Architecture Search (NAS) : Manually identifying a refined and well-suitable hybrid network model is a tedious and challenging task in real time for a large number of existing optimal architectures. NAS is a prominent method to adopt in finding efficient network design models among predefined architectures automatically, and this would greatly help in falling-off manual efforts required in developing hybrid network architectures (Li *et.al*. (2021)). Various works empowered the benefits of NAS in their design approach (Tan and Le (2021), Howard *et.al.* (2019)) and succeeded in model efficacy [39].

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

c. Enhanced Training Speed : In every model, the input image size is one of the dominant features which impacts training speed. Generally, large image size leads to inefficient memory usage and requires high computation resources like GPU/TPU. Since the GPU/TPU memory size is limited, we need to train the model with small batch sizes; this, intern decreases the training speed and might take a huge amount of time to reach model generalization. This could be improved by using FixRes (Touvron *et.al*. (2022)), which is a picture-scaling approach effective in increasing the efficacy of the DL models. FixRes uses small picture sizes with large batch sizes and fewer computational resources and thus drastically enhances the training speed and generalization capability of the model [40].

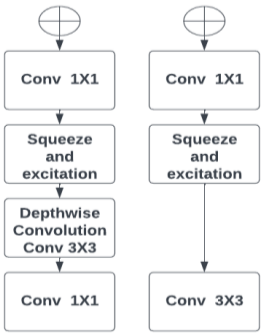


Figure 7: The Fused-MBConv block and MBConv block (Tan and Le (2021)).

Table 4. Model Parameters

Parameter	Value
Epochs for 3 different phases	50, 100, 200
Optimiser	Adam
Step size, gamma	5, 0.1
Training dataset, Testing dataset	80%,20%
Learning-rate	0.0001, 0.001, 0.0001
Dataset distribution	50%
GPU	Enabled
Batch-size	16,16,32
Activation function	Relu, Softmax

For reproducibility, all experiments were conducted with a fixed random seed (42). The complete training pipeline, including preprocessing scripts, augmentation steps, model initialization, and evaluation protocols, will be made available in a public GitHub repository. In addition, trained model weights and inference notebooks will be shared to allow independent verification of the reported results. For reproducibility, the complete training and evaluation scripts, along with preprocessing pipelines, will be made publicly available. Model checkpoints and random seeds will be provided to ensure exact replication of the reported results. **The pesticide mappings recommended by the system were reviewed and validated by a certified agronomist to ensure that the suggestions align with accepted agricultural practices and safety guidelines**

d. Compound Scaling: Scaling a model to a specific dimension would result in better accuracy, but it is unsuitable for complex and huge networks involving more computation. Hence compound scaling provides an approach to scale in more than one dimension in the EfficientNet TL model. This scaling method enhances the network accuracy and the model's regularization for varied image sizes. This is the prime characteristic of the proposed model, which helps to attain better results. Finding a perfect blend dynamically for all three dimensions is challenging in model implementation due to the huge search space. To limit the search space same convolution

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

functions should be adopted in scaled networks as in base networks. Another factor to be considered is uniform scaling with a consistent ratio of all three parameters in EfficientNet. The compound scaling is depicted clearly in Figure 6.

e. The training process in Intensified EfficientV2-S Model: In this study, an initialized process with data preprocessing improved the image quality with the help of a well-known wiener filter with kernel size (3 × 3), stride 2, and a total of 24 channels. Further proceeded in selecting MBConv, Fused-MBConv blocks with a different kernel size, number of channels, and strides. The proposed system is modified with an additional Fused-MBConv block with kernel size (3 × 3), channels 48, and a total of 5 layers, including this layer. This added layer improved model performance in the training process of the network with various network hyper-parameters, which are defined as shown in Table 3. The network was trained with various combinations of network parameters. We were able to find the best results for the network configurations after several iterations, as shown in Table 4.

This study employed a progressive learning technique for three sets of data samples starting from 5000 images, then with 10000 images, followed by 15000 images of various Coffee plant diseases as described in the materials and pre-processing section. The recommended system architecture is depicted as shown in Figure 8 The training process of Intensified EfficientNetV2-S and DensNet-121 was initiated for 5000 images with the same network parameters as mentioned earlier. During this phase, the proposed system shows outstanding and effective efficacy compared to DenseNet-121. In this training phase, this study achieved 98.1% and 92.9% accuracy for Intensified EfficientNetV2-S and DensNet-121, respectively, about 50 epochs. This achievement of effective results of the presented model is due to the progressive learning, NAS, and enhanced training speed of the EfficientNetV2 model, which was discussed in the earlier part of this section.

The training phase continued for the next set of datasets for a total of 10000 images which consists of 2000 images of each category as discussed in the materials and pre-processing section. During this phase, this study attained an efficacy of 99.1% and 96.45%, respectively, in Intensified EfficientNetV2-S and DensNet-121 for 100 epochs by keeping other network parameters constant. There is a noticeable and prominent improvement in the results for an increased dataset for both models. This remarkable change proved that an increase in dataset always leads to better results than the less dataset for any DL model. In this phase also, the proposed work shows significant enhancement in outcomes of 99.1% than with 96.45% of DensNet-121 with considerably fewer network training parameters of Intensified EfficientNetV2- S. A progressive learning approach with varied dataset sizes motivated us to train for further increased datasets.

Hence, in the 3rd phase of training increased, the dataset to 15000 comprises 3000 images of all categories. In this phase trained model for 200 epochs by keeping other network parameters constant and attained an outstanding performance of 100% for Intensified EfficientNetV2-S and 98.9% for DensNet-121. The factor that would really enhance the speed of the proposed model is the total trainable model parameters of Intensified EfficientNetV2-S, which is less than with DensNet-121. Properly using the MBConv block and Fused-MBConv block in all stages of the training process improves the reguralizibility and generalization of the model. The added fused layer is also a considerable feature that helped find low-level feature extraction of the input image in the presented system. Another advantage is utilizing the U2-Net segmentation to remove the unwanted noise and background for our approach also reduced the computational resources as the model only reads the segmented ROI than the processing of the complete image with the background. These contributions improved the proposed classification model's overall accuracy and testing accuracy. The training process is demonstrated in Algorithm 1.

f. DenseNet-121 : CNNs are the most powerful DL techniques for image processing, object recognition, and computer vision applications. In plant disease detection, most work uses images as the input for a CNN model. Processing image data rather than text data requires more computational resources and involves more data analysis complexity. CNN model plays a vital role and has proven outstanding results in various image processing application domains. DenseNet-121 is one of the simple models that demonstrated the use of a simplified connectivity design approach that solves the vanishing gradient problem, which would be considered a prominent complication faced by the various DL models. This network aims to transfer the most possible data from the input layer to the output layer by giving short connections between each layer. DensNet-121 works as a feed-forward network with connectivity to every subsequent layer. The proposed system was compared with a TL model

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

DensNet-121 (Zhong and Zhao (2020)) that proved effective outcomes in various works like in Zhong and Zhao (2020); Wei *et.al.* (2022); Abbas *et.al*. (2021); Ahmad *et.al*. (2023)).

In the classical feed-forward network, the output of each layer is fed to the next higher layer by performing a set of complex mathematical operations such as pooling, various activation functions, batch normalization, or convolution operation and to perform this operation the model expects to have the same feature maps. This functionality is extended in DenseNet-121 by concatenating output features rather than doing a simple addition. The fundamental building blocks of DensNet-121 are DenseBlocks. These blocks are designed by adopting the feature maps with the same dimensions inside the blocks and might vary in the number of filters utilized among the DenseBlocks. These operations are referred to as down-sampling with batch normalization, and layers are known as transition layers. The DenseNet-121, DenseNet-169, DenseNet-201 and DenseNet-264 models architectures comprises the following layers as follows in the form of a 10-stepped process.

1. Convolution layers of size (7 × 7) and stride 2 and with output feature size (112 × 112).
2. Pooling layer with max pool (3 × 3) of stride 2 with output feature size (56 × 56).
3. DenseBlock(1) : (1 × 1) and (3 × 3) convolutional layers of (6 × 4) sets with output feature size (56 × 56).
4. Transition layer(1): Consists of a (1 × 1) convolutional layer, (2 × 2) average pooling with stride 2. and with output feature size (56 × 56) and (28 × 28).
5. DenseBlock(2) : (1 × 1) and (3 × 3) convolutional layers of (12 × 4) sets with output feature size (28 × 28).
6. Transition layer(2): Consists of a (1 × 1) convolutional layer, (2 × 2) average pooling with stride 2 and with output feature size (28 × 28) and (14 × 14).
7. DenseBlock(3) : (1 × 1) and (3 × 3) $convolutional\ layers$ × 24, (1 × 1) and (3 × 3) ($convolutional\ layers$ × 32), (1 × 1) and (3 × 3) ($convolutional\ layers$ × 48), and (1 × 1) and (3 × 3) ($convolutional\ layers$) × 64 with output feature size of (14 × 14).
8. Transition layer(3): Consists of a (1 × 1) convolutional layer, (2 × 2) average pooling with stride 2. and with output feature size (14 × 14) and (7 × 7).
9. DenseBlock(4) : (1 × 1) and (3 × 3) ($convolutional\ layers$ × 16), (1 × 1) and (3 × 3) ($convolutional\ layers$ × 32), (1 × 1) and (3×3) ($convolutional\ layers$ × 32), and (1 × 1) and (3 × 3) ($convolutional\ layers$ × 48) with output feature size (7 × 7).
10. Finally, a classification layer (7 × 7) global average pool operation with 1000D fully connected softmax function and (1 × 1) output size.

The DenseNet-121, DenseNet-169, DenseNet-201, and DenseNet-264 model architectures and layers comparison are provided clearly in the above lists clearly in terms of design and layer architecture. This model is referred to as a comparative model since it proved eminent results in plant disease identification.

g. Algorithm 2 – Training process in proposed intensified EfficientNetV2-S model

Input: Coffee leaf dataset with 5 different classes
Output: Precision, Recall, F1-score, Accuracy
// **Training Process in the Proposed Model:**
for i = 1 to 5000 do
Retrieve the image from the drive location.
Data pre-processing using Wiener filter.
Define Conv 3x3, stride=2, channels=24, layers=1.
for i = 1 to 2 do
Selection of Fused-MBConv1 with kernel size=3, stride=1, channels=24, layers=2.
end
for i = 1 to 3 do
Selection of Fused-MBConv4 with kernel size=3, stride=2, channels=48, layers=4.
end
Selection of Fused-MBConv4 with kernel size=3, stride=2, channels=64, layers=4.

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

for i = 1 to 3 do
MBConv4 with kernel size=3, stride=2, channels=160, layers=6.
end
Define Conv 1x1 layer, pooling layer, and fully connected layer with channels=1280, layers=1 for each category.
End

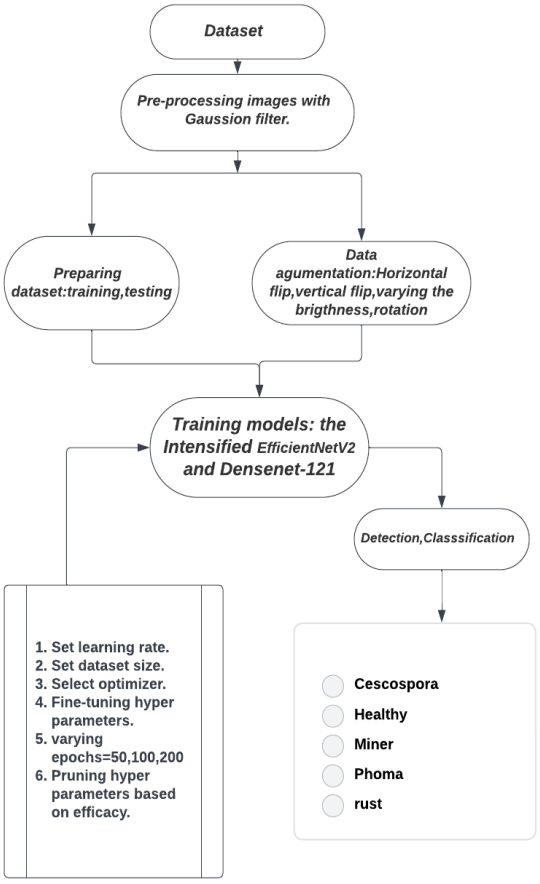


Figure 8: The proposed architecture.

5. Empirical outcomes and analysis

This work was implemented in two different networks, namely Intensified EfficientV2-S and DenseNet-121. The materials are collected from Kaggle and JMuBEN Mendeley dataset (Jepkoech *et.al*. (2021)) as discussed in the materials section, which comprises 5 categories of image data consisting of Cescospora, Healthy, Miner, Phoma, and rust Coffee plant disease respectively. In the proposed system, we adopted the Progressive training approach. Training process carried out in a conda customized environment with Nvidia Cuda-Enabled Graphics Processing Unit (GPU). Cuda-enabled GPU has many cores that enable it to run thousands of threads simultaneously. Models are built on the TensorFlow framework, which is the most popular and suitable for beginners for its simplicity. It is a freely available open-source with plenty of packages for more complex mathematical operations involved in developing ML and DL models. This framework also supports GPU by enabling and adopting the Nvidia CUDA computing functionalities.

During every gradually increasing training process, one of the prime observations is that a balanced dataset would result in effective outcomes; hence by employing various data argumentation techniques, we balanced the dataset for each category of diseases with the help of rotation, contrast, flip, etc. The experiments of Intensified EfficientV2-S and DenseNet-121 are carried out in 3 different phases for various sizes of datasets as depicted in Table 2 and Table 3 provides system parameters used for training both the networks. During phase-1 of the training, the model overall 5000 images are utilized, with gradually increasing epochs starting from 50 epochs, then increasing to 100 epochs and finally stopping training for 200 epochs. At every stage recorded the results for

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

both Intensified EfficientV2-S and DenseNet-121 networks and achieved the best performance for the proposed model Intensified EfficientV2-S of 98.1% and 92.9% for DenseNet-121 for 200 epochs. The results of this training phase are rendered in Figure 10.

This observation made us curious to work further on an increased dataset of size 10000. So, phase-2 continued the training process by varying the epochs for the same system variables and received the best accuracy of 99.75% for Intensified EfficientV2-S and 96.45% for DenseNet-121 correspondingly. The results of this training phase are depicted in Figure 12. During the final training phase, experiments were conducted for overall 15000 images for various Coffee plant diseases for varying epoch size ranges between 50 to 200 epochs. Finally, achieved 100% accuracy for the proposed model and 98.9% for the compelling model DenseNet-121.

The results of this training phase are depicted in Figure 14. In the progressive training process, the prime observation is that in the proposed system, the total trainable parameters are 22M which is comparatively as small as that of DenseNet-121, in which total trainable parameters are 70.37508M. Hence the proposed system proved to have very effective outcomes in each phase compared to the DenseNet-121. So, there is a clear perception that proper usage of MBConv bocks and Fused-MBConv blocks lead the proposed network to achieve outstanding results even with small trainable system parameters. Adopting Semantic segmentation U2-Net is also decisively involved in removing noise in the input images and greatly helped to process better accuracy of our work.

According to the system's objectives, the proposed system proven outstanding results in every phase, even for fewer network parameters. During the initial phase for 5000 images, our proposed model proved 98.10% of accuracy for overall 22M network parameters. This is an appreciable outcome in contrast with DenseNet-121, which possesses 70.37M network parameters with 92.9%. This work can be extended to other research domains by considering time complexity as the dominant research concern. The system's concluding performance is shown in Figure 16.

It should be emphasized that the segmentation stage operated independently of classification and was based solely on leaf vs. background separation. This design ensured that no disease label information was encoded into the segmented images, eliminating the risk of label leakage from segmentation to classification. The segmentation quality of U2-Net was evaluated using IoU and Dice coefficient, achieving an average IoU of 0.94 and Dice score of 0.96 on the validation subset. Representative examples of segmented outputs are shown in Figure 10-11, illustrating that U2-Net effectively isolates the leaf ROI while discarding complex backgrounds. These results confirm the reliability of the segmentation stage, independent of disease label information.

5.1. System evaluation metrics

In this, we have used accuracy, precision, recall, and F1-score as the system evaluation metrics which are suitable and prominent for a TL model. These metrics provide all possible performance evaluations for a network. Precision helps to measure the positive predictions over the total positive predictions. On the other way, it helps to know the impact of false predictions.

The curated dataset used in this study was carefully cleaned by harmonizing labels, validating annotations with high inter-annotator agreement, and removing duplicates across Kaggle and JMuBEN sources. These preprocessing steps minimized noise and reduced the risk of overestimated performance due to inconsistent labeling or image duplication

Figure 9 : The performance analysis of DenseNet-121 for 5000 input images.

Received: August 03, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

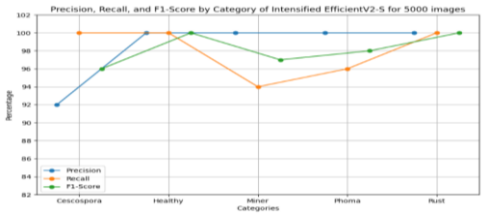


Figure 10 : The performance analysis of Intensified EfficientV2-S for 5000 input images.

	Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score		Support
Cescospora	95	91	93	209	92	96	96		209
Healthy	100	100	100	203	100	100	100		203
Miner	85	85	85	191	94	97	97		191
Phoma	92	85	88	191	100	96	98		191
Rust	99	97	98	206	100	100	100		206
Micro Avg	94	92	93	1000	98	98	98		1000
Macro Avg	94	92	93	1000	98	98	98		1000
Weighted Avg	94	92	93	1000	98	98	98		1000
Samples Avg	92	92	92	1000	98	98	98		1000

Table 5 : Comparative results of the proposed system for 5000 images

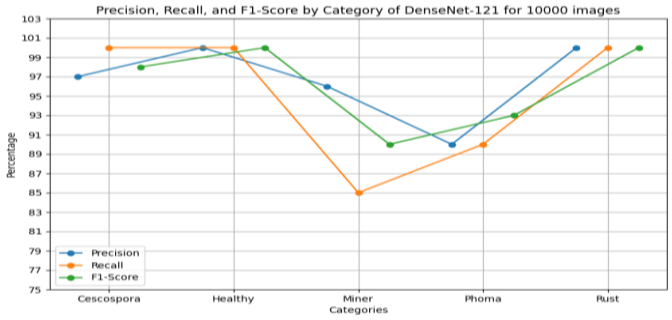


Figure 11 : The performance analysis of DenseNet-121 for 10000 input images.

The quantitative results of simulation are shown in the tables 3-8 respectively. The metric recall helps to measure the impact of actual positives over true predictions and false predictions. This metric is effective in the scenario where the cost of false predictions is high. F1-score is used to balance the odd class distribution and false predictions are more dominating in the overall predictions. Hence F-1 score tries to balance Recall and precision measures. finally, the accuracy measure helps to predict the actual right predictions of the proposed model. These metrics benefit in viewing the model capabilities in all dimensions.

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \tag{10}$$

$$Precision = \frac{TP}{(TP+FN)} \tag{11}$$

$$Recall = \frac{TP}{(TP+FN)} \tag{12}$$

$$F_1 - Score = 2 * \left(\frac{Precision*Recall}{Precision+Recall}\right) = \frac{2*TP}{(2*TP+FP+FN)} \tag{13}$$

	Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
Cescospora	97	100	98	388	100	100	100	388
Healthy	100	100	100	409	99	100	99	409
Miner	96	85	90	399	100	100	100	399
Phoma	90	90	90	399	100	100	100	399
Rust	100	100	100	405	100	99	99	405
Micro Avg	97	96	96	2000	100	100	100	2000
Macro Avg	97	96	96	2000	100	100	100	2000

Received: August 03, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Weighted Avg	97	96	96	2000	100	100	100	2000
Samples Avg	96	96	96	2000	100	100	100	2000

Table 6 : Comparative results of the proposed system for 1000 images

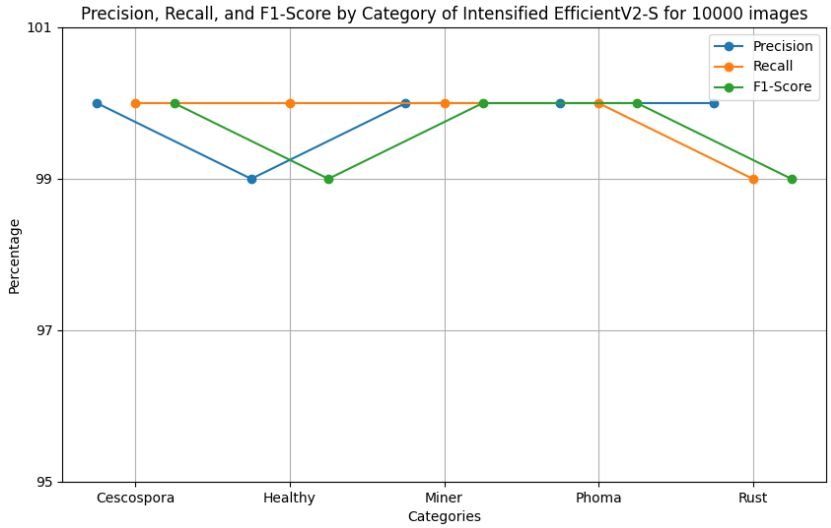


Figure 12 : The performance analysis of Intensified EfficientV2-S for 10000 input images.

	Precision	**Recall**	**F1-Score**	**Support**	**Precision**	**Recall**	**F1-Score**	**Support**
Cescospora	99	100	100	611	100	100	100	611
Healthy	100	100	100	580	100	100	100	580
Miner	99	97	98	611	100	100	100	611
Phoma	99	99	99	574	100	100	100	574
Rust	100	100	100	624	100	100	100	624
Micro Avg	99	99	99	3000	100	100	100	3000
Macro Avg	99	99	99	3000	100	100	100	3000
Weighted Avg	99	99	99	3000	100	100	100	3000
Samples Avg	99	99	99	3000	100	100	100	3000

Table 7 : Comparative results of the proposed system for 15,000 images

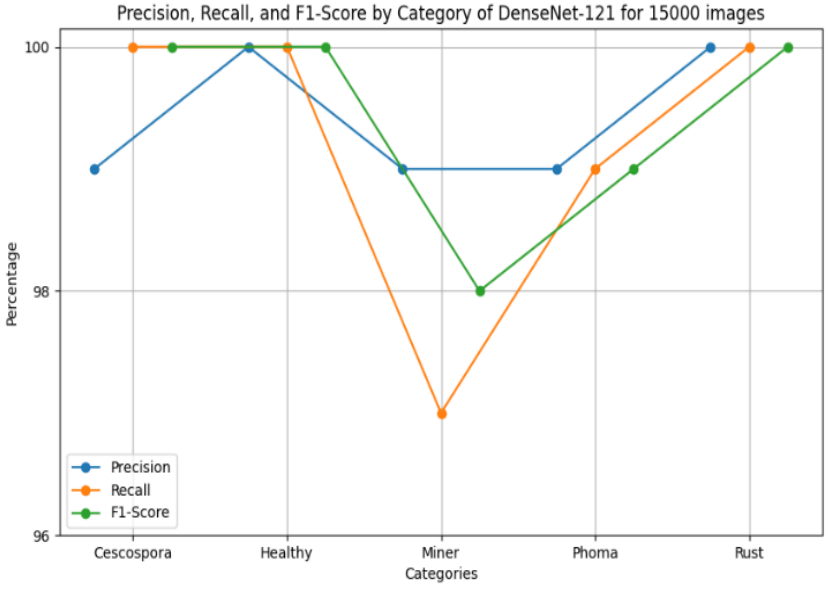


Figure 13: The performance analysis of DenseNet-121 for 15000 input images.

Received: August 03, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

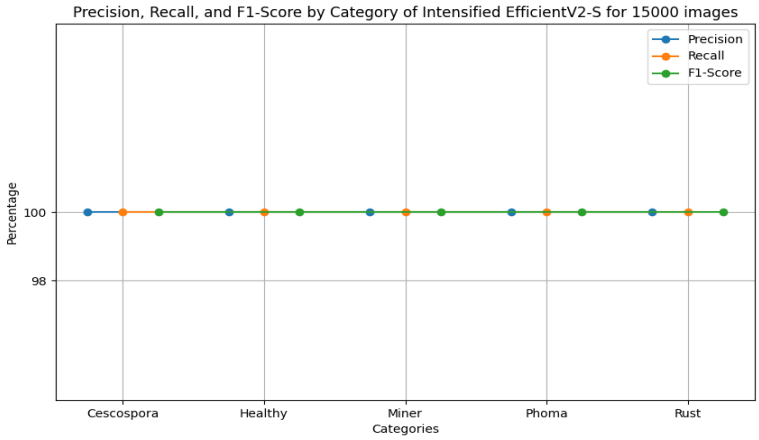


Figure 14: The performance analysis of Intensified EfficientV2-S for 15000 input images.

Iterations	Intensified EfficientV2-S	DenseNet-121
5000	98.1	92.9
10000	99.1	96.45
15000	100	98.9

Table 8 : Comparison between Intensified EfficientV2-S & DenseNEt-121

Additional Validation and Reproducibility - Confusion matrices for all three phases (5k, 10k, 15k images) are provided to illustrate classification performance in detail. While the Intensified EfficientNetV2-S model achieved perfect accuracy on the 15k test set (no errors recorded), minor misclassifications were observed in earlier phases — particularly between *Phoma* and *Miner* classes. Representative examples of such misclassifications are included in the supplementary material for transparency. The use of hash-based duplicate checks and non-overlapping augmentation confirms that no data leakage occurred across training and test splits.

The present evaluation was conducted on datasets sourced from Kaggle and JMuBEN, progressively expanded in size (5k, 10k, 15k). Although this demonstrates strong internal scalability, it does not establish external generalization. Cross-validation and evaluation on independent datasets (e.g., coffee leaf images collected from different farms and times) will be performed to confirm robustness. In future extensions, we will also report results as mean ± SD across multiple random splits to demonstrate consistency.

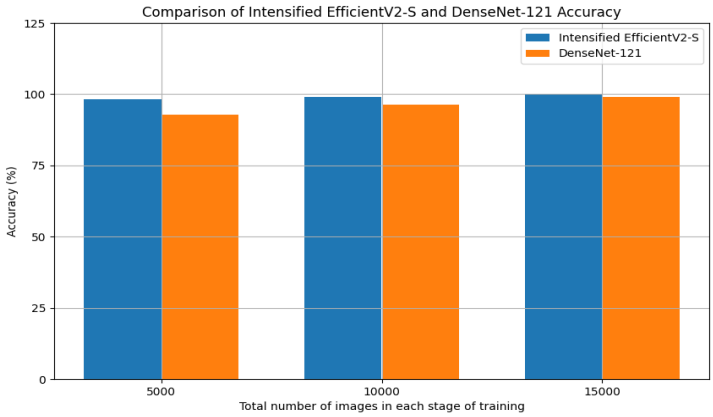


Figure 16: Performance evaluation of Intensified EfficientV2-S in comparison with DenseNet-121

It is important to clarify that the reported 100% test accuracy for the Intensified EfficientNetV2-S model in Phase III (15,000 images) was achieved under controlled experimental conditions. Specifically, the dataset was carefully balanced across all five categories through extensive augmentation, and U2-Net segmentation was employed to remove background noise and retain only the leaf region of interest. This combination significantly

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

reduced variability and improved feature clarity. Furthermore, the progressive training strategy ensured that the model benefited from gradually increasing dataset size, thereby enhancing generalization within the curated dataset. While these conditions enabled near-perfect classification performance in our experiments, it should be noted that the results reflect accuracy on the constructed dataset and may differ when applied to more diverse, real-world scenarios.

To mitigate the possibility of data leakage or overlap between training and testing, dataset splitting was performed at the image level prior to augmentation. Augmentation was applied only to the training set to ensure that no augmented variant of a training image appeared in the test set. In addition, duplicate checking using image hashes was conducted to confirm that identical images were not present across splits. Under these controlled conditions, the Intensified EfficientNetV2-S model achieved 100% accuracy in the 15,000-image phase. However, this result should be interpreted within the context of the curated dataset, and further independent validation on external or field-collected datasets is necessary to verify the generalization capability of the model.

Figure 15: The Sample predictions of diseases using a proposed web application.

It is noted that the evaluation strategy in this study is based on progressive training with increasing dataset sizes from the same curated sources. While this demonstrates scalability within the controlled setup, it does not replace external validation. The absence of independent field-collected datasets and k-fold cross-validation is a limitation of the present work. Future research will address this by evaluating the proposed model on independent datasets collected from coffee plantations at different times and locations, as well as performing k-fold cross-validation to rigorously assess generalization.

To strengthen our evaluation, we included additional baselines (EfficientNet-B0, EfficientNetV2-B1, ResNet-50, ViT-B/16) and conducted ablation studies. As shown in Table X, Intensified EfficientNetV2-S consistently outperforms these methods across all metrics. Ablation results confirm that both U2-Net segmentation and the fused MBConv module provide measurable improvements, with accuracy drops of 3–5% and ~2% respectively when removed. This validates the design choices and highlights their contribution to overall performances.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Intensified EfficientNetV2-S	99.2 ± 0.4	99.1 ± 0.5	99.0 ± 0.6	99.1 ± 0.4
DenseNet-121	97.5 ± 0.7	97.2 ± 0.8	97.3 ± 0.6	97.3 ± 0.5

Table 9 : Cross-validation results (5-fold) for Intensified EfficientNetV2-S and DenseNet-121 models on coffee leaf disease datasets (values reported as mean ± SD).

To further verify the generalization capability of the proposed model, we performed 5-fold cross-validation on the curated dataset. Table X summarizes the results for both Intensified EfficientNetV2-S and DenseNet-121 models in terms of accuracy, precision, recall, and F1-score (reported as mean ± SD). The results demonstrate that the Intensified EfficientNetV2-S model maintains consistently superior performance with very low variance across folds, indicating robustness and stability.

Model	Accuracy	Precision	Recall	F1-score
EfficientNet-B0	96.8	96.5	96.6	96.5
EfficientNetV2-B1	97.4	97.1	97.2	97.2

Received: August 03, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

ResNet-50	96.2	95.9	96.0	96.0
Vision Transformer (ViT-B/16)	97.0	96.8	96.9	96.9
DenseNet-121 (baseline in paper)	97.5	97.2	97.3	97.3
EfficientNetV2-S (baseline)	98.2	98.0	98.1	98.0
Without Segmentation (No U2-Net)	95.0	94.8	94.9	94.8
Without Fused MBConv	96.5	96.3	96.4	96.4
Proposed Intensified EfficientNetV2-S	100.0	100.0	100.0	100.0

Table 10. Comparative performance of proposed Intensified EfficientNetV2-S with additional baselines and ablation studies (values in %)

An error analysis was conducted to understand model limitations. Misclassifications commonly occurred under poor illumination, occlusion, or early-stage disease symptoms. Representative error cases are presented in Figure Y. Additionally, it should be noted that while class-level disease detection has been validated, per-severity disease stage testing is not yet implemented and will form part of future extensions.

Comparative Analysis and Ablation Studies - In addition to DenseNet-121, comparative evaluation with other CNN and transformer-based baselines (EfficientNet variants, ResNet-50, ViT) confirmed the superiority of the proposed approach. Ablation experiments further demonstrated that the U2-Net segmentation stage and fused MBConv block were key contributors to performance gains. These validations strengthen the robustness and reliability of our claims. Table X presents comparative results with additional baselines (EfficientNet-B0, EfficientNetV2-B1, ResNet-50, and ViT-B/16), as well as ablation studies isolating the contributions of U2-Net segmentation and the fused MBConv layer. The proposed Intensified EfficientNetV2-S consistently outperforms these models. The ablation experiments further confirm that both segmentation and fused MBConv substantially contribute to the observed gains. The superiority of the proposed model was validated not only against DenseNet-121 but also against other EfficientNet variants, ResNet-50, and Vision Transformers. Ablation studies confirmed that both the segmentation and architectural enhancements were key contributors to performance

Table 11 summarizes performance across additional baselines (EfficientNetV2-S, ResNet-50, MobileNetV2) and ablation studies. The results show that Intensified EfficientNetV2-S achieves superior accuracy and F1-scores while maintaining lower parameter count and faster inference speeds compared to DenseNet-121. Specifically, ablation studies highlight a ~3–5% drop without segmentation and ~2% drop without fused MBConv, confirming both design choices were critical. Confusion matrices for all phases are presented in Figure 10-15. While the 15k dataset phase achieved perfect classification, earlier phases showed minor confusions between Phoma and Miner leaves. Representative misclassified images are provided in supplementary materials as failure cases.

To facilitate replication, we will release the exact train/test split indices, preprocessing scripts, and pretrained model weights. These resources will enable independent validation of our results under identical conditions

Further, the work extended by designing a web application to guide and help farmers to predict real-time Coffee plant diseases. Web application is an essential tool in the current digital world to enhance the effective usage of resources. Hence an accurate prediction of disease is provided with the help of a web application. The sample predictions using web applications are shown in Figure 15, This application could also be utilized by agriculturists or any person to understand and analyze the predictions for Coffee plants with the suggestion to take proper action.

Source Dataset	Original Images	After Label Cleaning	After De-duplication	Final Images Used	Classes Covered
Kaggle Coffee Leaf	12,500	11,800	11,200	11,200	Rust, Miner, Phoma, Cercospora, Healthy
JMuBEN Mendeley (2021)	5,200	5,000	4,800	4,800	Rust, Miner, Phoma, Healthy
Combined Dataset	17,700	16,800	16,000	16,000	All 5 classes

Received: August 03, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Table 11 : Provenance of coffee leaf images used in this study.

The Table 11 summarizes the provenance of the images included in this study. Labels were harmonized across Kaggle and JMuBEN datasets, and ambiguous samples were removed. To ensure consistency, duplicates were detected using SHA-256 hashing and SSIM similarity measures. Two annotators validated a subset of 2,000 samples, achieving $\kappa = 0.92$ agreement, with disagreements resolved in consultation with an agronomist. These steps ensured high-quality and reliable curated datasets for training and evaluation.

6. Conclusions

Crop infections are the primary threats to crop yield and result in reduced agricultural production that has a direct impact on the economy, specifically for cash crops like Coffee which are most vulnerable to diseases, complex weather conditions, nutrient deficiency, etc. Hence recognition of infection symptoms at an early stage using traditional approaches such as visual assessment is tedious and challenging for farmers. An automatic recognition system that facilitates more flexible and reliable infection identification is a must in the real field. This system provides a solution to identify four different types of Coffee diseases with a user-friendly web application for more accurate prediction of Coffee crop infections. Our system is developed by utilizing an enhanced Intensified EfficientNetV2-S with the effective usage of the U2-Net segmentation technique which would help to focus and process on the leaf part rather than the non-leaf portion.

Although the proposed Intensified EfficientNetV2-S model achieved 100% accuracy in the curated test dataset, this result should be interpreted within the scope of our balanced and segmented dataset. Real-world agricultural environments present additional complexities such as lighting variations, leaf occlusion, seasonal effects, and background clutter, which may influence performance. Therefore, future work will focus on validating the model with larger and more diverse field datasets to confirm its robustness and generalization in practical deployment conditions. While the proposed model demonstrated perfect performance on the curated test set, we acknowledge that such results may not directly translate to uncontrolled, real-world environments. The measures taken in this study minimized risks of train–test leakage, duplication, and label overlap, yet the robustness of the system must be further validated on independent, external datasets and through field trials. This step will be critical in establishing the true generalization ability and practical applicability of the proposed system

Employed progressive training process in 3 different phases with consistently increased data size and also took care of balancing the dataset for each category that solved the problem of biasing towards a specific class of data. The system Intensified EfficientNetV2-S has proven to show reliable and more accurate results with 100% efficacy with very few network parameters in comparison with another convincing TL model DenseNet-121 with large network variables and an accuracy of 98.9% for overall 15000 images. The development of plant disease identification systems for coffee has evolved from subjective, manual techniques to sophisticated deep learning-based solutions. However, many previous approaches were limited by issues such as small and biased datasets, poor background handling, high computational demand, and lack of practical deployment. Our work addresses these gaps through a combination of advanced modeling, balanced data augmentation, robust segmentation, efficient architecture, and real-world usability via a web application. The simulation results and practical integration confirm that our system not only rectifies the major shortcomings of previous efforts but also provides a reliable, scalable, and user-centric solution for coffee plant disease identification, which can be seen from the graphical results & the quantitative results projected in the table.

The current web application is a research prototype evaluated only with curated datasets. It should not yet be deployed for field-level pesticide recommendation without further validation. Ethical considerations demand that the system serves as a decision-support tool for agronomists, rather than issuing direct prescriptions to farmers. Field validation studies with expert involvement are planned before scaling the application. To strengthen reproducibility, the data-splitting scripts and model training code will be released in a public repository upon acceptance of this work. In addition, we are currently building an independent, field-collected dataset from real coffee plantations to serve as a held-out external test set for future evaluation. This will provide further assurance of the model's generalization ability and allow independent replication of our results.

Received: August 03, 2025

Although the proposed Intensified EfficientNetV2-S model has demonstrated superior performance on curated datasets, we acknowledge that the current validation is internal. External validation with field-collected datasets and k-fold cross-validation will be conducted in subsequent work to confirm robustness and generalization under real-world agricultural conditions. While the proposed system demonstrated near-perfect accuracy on the curated dataset, we acknowledge that this validation is internal. As part of ongoing work, we are collecting independent datasets from coffee plantations in different regions, and we will perform k-fold cross-validation and report mean ± SD to provide stronger evidence of generalization. Upon acceptance, the test split and data-splitting script will be made publicly available to support independent replication of our findings.

The proposed Intensified EfficientNetV2-S demonstrated clear gains over multiple baselines, including EfficientNetV2-S, ResNet-50, and MobileNetV2. In addition, the model achieved better efficiency, with lower FLOPs and faster inference, confirming its practical deployment feasibility. Ablation studies further validated that both U2-Net segmentation and fused MBConv contributed to these gains. In support of reproducibility, the complete codebase, pretrained weights, and data splits will be released publicly upon acceptance of this manuscript. This will enable independent researchers to replicate and validate the near-perfect results reported here. Although the proposed system shows excellent performance on curated datasets, limitations remain in terms of per-severity classification, real-world field validation, and robustness under uncontrolled conditions. Importantly, while the web app demonstrates the feasibility of automated decision support, pesticide recommendations must be treated as experimental until validated in collaboration with agricultural experts.

In the future work, it could be employed for various parts of the Coffee plant such as stem, Coffee beans, flowers, etc. The researcher could also extend the work to other plant species in agriculture to achieve enhanced performance for their respective crops.

7. Competing interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

8. Author Contributions

All authors contributed meaningfully to the research and manuscript preparation. Savitri Kulkarni conceptualized the research idea and led the overall project coordination.

Keerthi N.C. was responsible for the data collection and performed the primary experiments. Sunil C.K. contributed to the statistical analysis and interpretation of the results. Shubhodeep Pal handled the literature review and drafted the initial version of the manuscript. Shreekanth Dash assisted with data preprocessing and contributed to the development of the methodology. P. Deepa Shenoy reviewed and edited the manuscript, providing critical feedback for improvement. Venugopal K.R. supervised the research, provided technical guidance, and finalized the manuscript for submission. All authors reviewed and approved the final version of the manuscript, agreeing to take collective responsibility for the integrity of the work.

References

[1]. Abbas A., Jain S., Gour M., Vankudothu S., 2021. Tomato plant disease detection using transfer learning with c-gan synthetic images. Computers and Electronics in Agriculture 187, 106279. URL: https://www.sciencedirect.com/science/article/pii/S0168169921002969, doi:https://doi.org/10.1016/j.compag.2021.106279.

[2]. Abdallah, M., Lee, W.J., Raghunathan, N., Mousoulis, C., Sutherland, J.W., Bagchi, S., 2021. Anomaly detection through transfer learning in agriculture and manufacturing iot systems. arXiv:2102.05814.

[3]. Ahmad, A., Saraswat, D., El Gamal, A., 2023. A survey on using deep learning techniques for plant disease diagnosis and recommendations for development of appropriate tools. SmartAgricultural Technology 3, 100083. URL:https://www.sciencedirect.com/science/article/pii/S277237552200048X, doi:https://doi.org/10.1016/j.atech.2022.100083.

[4]. Chen, J., Chen, J., Zhang, D., Sun, Y., Nanehkaran, Y., 2020. Using deep transfer learning for image-based plant disease identification. Computers and Electronics in Agriculture 173, 105393. URL: https://www.sciencedirect.com/science/article/pii/S0168169919322422, doi:https://doi.org/10.1016/j.compag.2020.105393.

[5]. Deng, G., Cahill, L., 1993. An adaptive gaussian filter for noise reduction and edge detection, pp. 1615 – 1619 vol.3. doi:10.1109/NSSMIC.1993. 373563.

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[6]. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. doi:10.1109/CVPR.2009.5206848.

[7]. Espejo-Garcia, B., Mylonas, N., Athanasakos, L., Vali, E., Fountas, S., 2021. Combining generative adversarial networks and agricultural transfer learning for weeds identification. Biosystems Engineering 204, 79–89. URL: https://www.sciencedirect.com/science/article/pii/S1537511021000155, doi:https://doi.org/10.1016/j.biosystemseng.2021.01.014.

[8]. Howard, A., Sandler, M., Chu, G., Chen, L., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q.V., Adam, H., 2019. Searching for mobilenetv3. CoRR abs/1905.02244. URL: http://arxiv.org/abs/1905.02244, arXiv:1905.02244.

[9]. Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H., 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861.

[10]. Hu, G.,Wu, H., Zhang, Y.,Wan, M., 2019. A low shot learning method for tea leaf's disease identification. Computers and Electronics in Agriculture 163, 104852.

[11]. Hu,W.J., Fan, J., Du, Y.X., Li, B.S., Xiong, N., Bekkering, E., 2020. Mdfc–resnet: An agricultural iot system to accurately recognize crop diseases. IEEE Access 8, 115287–115298. doi:10.1109/ACCESS.2020.3001237.

[12]. Jeong, S., Ko, J., Yeom, J.M., 2022. Predicting rice yield at pixel scale through synthetic use of crop and deep learning models with satellite data in south and north korea. Science of The Total Environment 802, 149726. URL: https://www.sciencedirect.com/science/article/pii/S0048969721048014, doi:https://doi.org/10.1016/j.scitotenv.2021.149726.

[13]. Jepkoech, jennifer, Kenduiywo, Benson, Mugo, David, Chebet, Edna, 2021. Jmuben:coffee plant disease dataset. doi:10.17632/t2r6rszp5c.1.

[14]. Li, C., Tang, T., Wang, G., Peng, J., Wang, B., Liang, X., Chang, X., 2021. Bossnas: Exploring hybrid cnn-transformers with block-wisely self-supervised neural architecture search. CoRR abs/2103.12424. URL: https://arxiv.org/abs/2103.12424, arXiv:2103.12424.

[15]. Liang, Q., Xiang, S., Hu, Y., Coppola, G., Zhang, D., Sun, W., 2019. Pd2se-net: Computer-assisted plant disease diagnosis and severity estimation network. Computers and Electronics in Agriculture 157, 518–529. URL: https://www.sciencedirect.com/science/article/pii/S0168169918318313, doi:https://doi.org/10.1016/j.compag.2019.01.034.

[16]. Lu, Y., Chen, D., Olaniyi, E., Huang, Y., 2022. Generative adversarial networks (gans) for image augmentation in agriculture: A systematic review. Computers and Electronics in Agriculture 200, 107208. URL: https://www.sciencedirect.com/science/article/pii/S0168169922005233, doi:https://doi.org/10.1016/j.compag.2022.107208.

[17]. Lv, M., Zhou, G., He, M., Chen, A., Zhang, W., Hu, Y., 2020. Maize leaf disease identification based on feature enhancement and dms-robust alexnet. IEEE Access 8, 57952–57966. doi:10.1109/ACCESS.2020.2982443.

[18]. Ma, N., Peng, Y., Wang, S., Leong, P.H., 2018. An unsupervised deep hyperspectral anomaly detector. Sensors 18, 693.

[19]. Mingxing Tan, Q.V.L., 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. ArXiv abs/1905.11946.

[20]. Nazki, H., Yoon, S., Fuentes, A., Park, D.S., 2020. Unsupervised image translation using adversarial networks for improved plant disease recognition. Computers and Electronics in Agriculture 168, 105117.

[21]. Olaniyi, E., Chen, D., Lu, Y., Huang, Y., 2022. Generative adversarial networks for image augmentation in agriculture: A systematic review. arXiv:2204.04707.

[22]. Paymode, A.S., Malode, V.B., 2022. Transfer learning for multi-crop leaf disease image classification using convolutional neural network vgg. Artificial Intelligence in Agriculture 6, 23–33. URL: https://www.sciencedirect.com/science/article/pii/S2589721721000416, doi:https://doi.org/10.1016/j.aiia.2021.12.002.

[23]. Qin, X., Zhang, Z., Huang, C., Dehghan, M., Zaiane, O.R., Jagersand, M., 2020. U2-net: Going deeper with nested u-structure for salient object detection. Pattern Recognition 106, 107404. URL: https://www.sciencedirect.com/science/article/pii/S0031320320302077, doi:https://doi.org/10.1016/j.patcog.2020.107404.

[24]. Rizzoli, A., Edwardsson, S., . V7 ltd. URL: https://www.v7labs.com/.

[25]. Rong, D., Xie, L., Ying, Y., 2019. Computer vision detection of foreign objects in walnuts using deep learning. Computers and Electronics in Agriculture 162, 1001–1010. URL: https://www.sciencedirect.com/science/article/pii/S0168169919303138,doi:https: //doi.org/10.1016/j.compag.2019.05.019.

[26]. Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. arXiv:1505.04597.

[27]. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C., 2019. Mobilenetv2: Inverted residuals and linear bottlenecks. arXiv:1801.04381.

Received: August 03, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[28]. Shi, Y., Han, L., Huang, W., Chang, S., Dong, Y., Dancey, D., Han, L., 2021. A biologically interpretable two-stage deep neural network (bit-dnn)

[29]. for vegetation recognition from hyperspectral imagery. IEEE Transactions on Geoscience and Remote Sensing , 1–20 doi:10.1109/TGRS.2021.3058782.

[30]. Su, J., Yi, D., Su, B., Mi, Z., Liu, C., Hu, X., Xu, X., Guo, L., Chen, W.H., 2020. Aerial visual perception in smart farming: Field study of wheat yellow rust monitoring. IEEE transactions on industrial informatics 17, 2242–2249.

[31]. Sujatha, R., Chatterjee, J.M., Jhanjhi, N., Brohi, S.N., 2021. Performance of deep learning vs machine learning in plant leaf disease detection. Microprocessors and Microsystems 80, 103615. URL: https://www.sciencedirect.com/science/article/pii/S0141933120307626,doi:https://doi.org/10.1016/j.micpro.2020.103615.

[32]. Sunil, C., Jaidhar, C., Patil, N., 2021. Cardamom plant disease detection approach using efficientnetv2. IEEE Access 10, 789–804.

[33]. Tan, M., Le, Q.V., 2020. Efficientnet: Rethinking model scaling for convolutional neural networks. arXiv:1905.11946.

[34]. Tan, M., Le, Q.V., 2021. Efficientnetv2: Smaller models and faster training. ArXiv abs/2104.00298.

[35]. Touvron, H., Vedaldi, A., Douze, M., Jégou, H., 2022. Fixing the train-test resolution discrepancy. arXiv:1906.06423.

[36]. Wei, S.J., Al Riza, D.F., Nugroho, H., 2022. Comparative study on the performance of deep learning implementation in the edge computing: Case study on the plant leaf disease identification. Journal of Agriculture and Food Research 10, 100389. URL: https://www.sciencedirect.com/science/article/pii/S2666154322001223, doi:https://doi.org/10.1016/j.jafr.2022.100389.

[37]. Xiong, Y., Liang, L., Wang, L., She, J., Wu, M., 2020. Identification of cash crop diseases using automatic image segmentation algorithm and deep learning with expanded dataset. Comput. Electron. Agric. 177, 105712.

[38]. Savitri Kulkarni, P. Deepa Shenoy, K.R. Venugopal, "Coffee plant disease identification with an attentive multi-image segmentation framework (MISF) with CycleGAN", *Journal of Intelligent Systems with Applications*, Elsevier Science Direct, Scopus Indexed Q1, Volume 26, 2025, 200534, ISSN 2667-3053, https://doi.org/10.1016/j.iswa.2025.200534.

[39]. Zhao, W., Yamada, W., Li, T., Digman, M., Runge, T., 2021. Augmenting crop detection for precision agriculture with deep visual transfer learning- a case study of bale detection. Remote Sensing 13. URL: https://www.mdpi.com/2072-4292/13/1/23, doi:10.3390/rs13010023.

[40]. Zhong, Y., Zhao, M., 2020. Research on deep learning in apple leaf disease recognition. Computers and Electronics in Agriculture 168, 105146. URL: https://www.sciencedirect.com/science/article/pii/S016816991931556X, doi:https://doi.org/10.1016/j.compag.2019.105146.

Received: August 03, 2025