



ARTICLE

## Leveraging Segmentation for Potato Plant Disease Severity Estimation and Classification via CBAM-EfficientNetB0 Transfer Learning

Amit Prakash Singh<sup>1</sup>, Kajal Kaul<sup>1,\*</sup>, Anuradha Chug<sup>1</sup>, Ravinder Kumar<sup>2</sup> and Veerubommu Shanmugam<sup>2</sup>

<sup>1</sup>University School of Information, Communication & Technology, Guru Gobind Singh Indraprastha University, Dwarka, New Delhi, 110078, India

<sup>2</sup>Division of Plant Pathology, Indian Agricultural Research Institute, Indian Council of Agricultural Research, New Delhi, 110012, India

\*Corresponding Author: Kajal Kaul. Email: kajalkaulphd@gmail.com

Received: 23 July 2025; Accepted: 19 September 2025; Published: 06 November 2025

**ABSTRACT:** In agricultural farms in India where the staple diet for most of the households is potato, plant leaf diseases, namely Potato Early Blight (PEB) and Potato Late Blight (PLB), are quite common. The class label Plant Healthy (PH) is also used. If these diseases are not identified early, they can cause massive crop loss and thereby incur huge economic losses to the farmers in the agricultural domain and can impact the gross domestic product of the nation. This paper presents a hybrid approach for potato plant disease severity estimation and classification of diseased and healthy leaves, combining the strengths of classical image processing, computer vision, and deep learning. We propose a pipeline that initially employs OpenCV's cv2 led color-based image segmentation to isolate and highlight diseased brown, yellow-colored lesions or regions and healthy green colored lesion areas associated with various potato leaf diseases. Adaptive Thresholding for illumination and texture feature extraction and U-Net Segmentation for mask refinement for severity estimation. It has a mathematical framework for quantifying the severity based on the spatial area distribution of these regions. This allows for both visual representation of the segmented regions in the form of overlay masks and quantification of distinct disease severity. We detail the implementation of the approach, including color space selection, segmentation strategies, mask creation, area calculation, and a potential mathematical model for severity calculation. Overlay masks generated are then used as input to a CBAM-EfficientNetB0 model, leveraging transfer learning for improved classification accuracy and efficiency. For the Plant Village dataset, the test accuracy achieved is 0.99, whereas the test loss is 0.02, respectively. For the Plant Doc dataset, the test accuracy achieved is 0.97, whereas the test loss is 0.06, respectively. Also, the CBAM attention mechanism model lays emphasis on relevant features within the lesions and overall image context. The results achieved with the Plant Village dataset are slightly better in comparison to the Plant Doc dataset.

**KEYWORDS:** PEB; PLB; PH; UNET; segmentation; overlay mask; CBAM; EfficientNetB0

### 1 Introduction

Potato, also known as *Solanum tuberosum* L., is an everyday staple in most households worldwide, providing a crucial source of carbohydrates and nutrients. However, potato plants are vulnerable to several diseases that significantly decrease productivity and quality. Early and precise detection of diseases is important to implement the control measures and lower economic loss. Conventional diagnosis of diseases relied on visual inspection, which is an error-prone and tedious process. Techniques like DL and Computer



vision are better ways for automated disease detection. DL models like Convolution Neural Network and its variants have achieved humungous success in image classification, including potato plant disease identification. Though large datasets and significant computational resources require transfer learning, comprising a model that is pre-trained and is optimised on a new dataset, provides a more time-efficient and effective approach. Our method leverages cv2 color-based segmentation and adaptive thresholding to isolate and highlight lesion areas associated with different potato diseases for severity estimation. This processed image, i.e., the overlay merged masks, serves as input to the U-Net segmentation model for generating refined ground truth masks for calculating severity. Further, an attention-enabled CBAM-EfficientNetB0 model that has been fine-tuned using transfer learning helped classify the three classes. This particular approach allows the model to emphasise the relevant features of disease symptoms, leading to improved classification accuracy. The datasets used for the above approach were the Plant Village and the Plant Doc datasets. It was observed that the Plant Village dataset performed better in terms of performance metrics. The computational requirements have been discussed in the proposed system requirements. The dataset acquired, pre-processed, image segmentation, overlay mask visualisation, U-Net segmentation, severity estimation and classification for both the datasets have been discussed extensively in the Implementation section. The framework provided a combined classification and severity estimation of potato plant diseases.

## 2 Motivation

The motivation behind this research framework developed in this paper is to address the need to help farmers in the identification of potato leaf disease and its severity. This can be achieved via visualising the potato plant overlays and leveraging them for potato plant disease severity estimation and classification. The aim is to help the agriculturists and farmers in enhancing their yields by letting them know the severity of potato plant leaves and giving them accurate classification results. The research utilises techniques such as image segmentation, utilising colour thresholding for colour-based feature extraction and adaptive thresholding for light feature. The colour range definition, binary mask generation, and colored overlay masks generation using OpenCV's library called cv2. It helped in achieving excellent visualisation of individually colored potato overlay leaf masks. These brown, yellow, and green images were further checked for severity, given the fact that brown and yellow depicted severity and the green part of the leaf is healthy. The lighting feature was also catered to in the adaptive masks generated. Hence, severity estimation was attained by combining colored masks and adaptive thresholding masks. The same set of ground truth images was then classified using a CBAM-based EfficientNetB0 Transfer Learning model. This framework is highly useful for farmers to understand the potato leaf disease classification, and it also helps to measure the severity of various leaves. It is an effort to help farmers enhance their yield, as in India, the potato leaf diseases, such as PEB and PLB, are pretty standard. It is an effective way to protect the farms against these diseases and will help them identify these diseases with efficacy. In accordance with the disease severity, further work can be done by farmers to determine the type of pesticides and chemicals to be used for the particular disease type in the farm fields. This contribution feels noble as it aims to help agricultural farmers in a great way and also helps enhance the GDP of the nation. Hence, both the Plant Village and the Plant Doc datasets were used for this pursuit.

## 3 Literature Review

The Literature Review of various papers has been done in [Table 1](#) as follows:



**Table 1:** Literature review

| Authors            | Dataset                                                                                                                                       | Types of plant diseases                                                                                                                 | Algorithm used                                                                                                                                      | Accuracy                                                                                                                                     |
|--------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------|
| Verma et al. [1]   | Plant village dataset                                                                                                                         | PEB, PLB                                                                                                                                | SqueezeNet;<br>Inception V3;<br>AlexNet; ResNet101;<br>ResNet50; ResNet18                                                                           | AlexNet (96.43%),<br>SqueezeNet (97.32%),<br>ResNet18 (99.11%),<br>ResNet50 (99.55%),<br>ResNet101 (97.32%),<br>and Inception V3<br>(98.66%) |
| Kaul et al. [2]    | Plant village dataset                                                                                                                         | PEB, PLB and PH                                                                                                                         | CNN and<br>GridSearchCV                                                                                                                             | test accuracy (0.988),<br>test loss (0.039)                                                                                                  |
| Li et al. [3]      | Plant village dataset                                                                                                                         | PLB is the disease<br>discussed                                                                                                         | PLB is the disease<br>discussed.                                                                                                                    | –                                                                                                                                            |
| Tonmoy et al. [4]  | Plant village: 54,303<br>leaf images, CCMT,<br>Sugarcane: 6748 leaf<br>images and Coconut:<br>5798 images                                     | Disease categories:<br>PV: 38 CCMT:<br>Cashew, Cassava,<br>Tomato: 5 classes<br>each except for<br>Maize: 7 Sugarcane: 9.<br>Coconut: 5 | MobilePlantViT: A<br>Mobile-friendly<br>Hybrid ViT<br>combining<br>MobileViTv1 and<br>MobileViTv2                                                   | Test accuracies—80%<br>to 99%.                                                                                                               |
| Nyakuri et al. [5] | Plant village dataset                                                                                                                         | Cereal crop diseases                                                                                                                    | TinyLiteNet-<br>MobileNetv2 mode<br>with 6 SE blocks                                                                                                | TinyLiteNet Accuracy<br>(98.6%),                                                                                                             |
| Miao et al. [6]    | Plant village dataset                                                                                                                         | PEB, PLB                                                                                                                                | SqueezeNet;<br>Inception V3;<br>AlexNet; ResNet101;<br>ResNet50; ResNet18                                                                           | AlexNet (96.43%),<br>SqueezeNet (97.32%),<br>ResNet18 (99.11%),<br>ResNet50 (99.55%),<br>ResNet101 (97.32%)<br>and Inception V3<br>(98.66%)  |
| Mahum et al. [7]   | Plant village dataset<br>dataset size: 1700<br>images.                                                                                        | PEB; PLB; Leaf Roll;<br>Potato Verticillium<br>wilt; PH                                                                                 | Pre-trained<br>DenseNet-201 in<br>slightly modified way                                                                                             | DenseNet-201 (97.2%)                                                                                                                         |
| Rashid et al. [8]  | 4062 images; Origin:<br>Central Punjab area,<br>Pakistan.                                                                                     | PEB, PLB                                                                                                                                | Multi-level DL<br>model: CNN; Image<br>Segmentation:<br>YOLOv5                                                                                      | Deep CNN (99.75%)                                                                                                                            |
| Esgario et al. [9] | To capture the images<br>various cellphones<br>were used namely<br>iPhone 6S; ASUS<br>Zenfone 2; Xiaomi<br>Redmi 5A; Xiaomi<br>S2; Galaxy S8. | Healthy, Leaf spot<br>diseases namely Rust,<br>Brown & Cercospora                                                                       | Image Processing &<br>cv2; DL Techniques:<br>MobileNetV2;<br>ResNet50; VGG16;<br>GoogleNet; AlexNet                                                 | Biotic stress<br>classification:<br>ResNet50 (95.24%);<br>Severity estimation<br>(86.51%)                                                    |
| Li et al. [10]     | Origin: Hebei<br>Agricultural<br>University, North<br>China Experimental<br>setup: D7100 camera<br>Camera resolution:<br>6000 × 4000 pixels.  | PEB, PLB, PH                                                                                                                            | Segmentation: Mask<br>R CNN. Semantic<br>segmentation models:<br>PSPNet;<br>DeepLabV3+ UNet.<br>Classification:<br>InceptionV3;<br>ResNet50; VGG16. | Mask R-CNN<br>(95.33%). (MIoU) of<br>the semantic<br>segmentation<br>(89.91%) and MPA<br>(94.24%)                                            |

(Continued)

**Table 1 (continued)**

| Authors                         | Dataset                                                                                                                                   | Types of plant diseases                                                                                                          | Algorithm used                                                                    | Accuracy                                                                                                              |
|---------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------|
| Fu et al. [11]                  | Plant Village Dataset; Images: 486; Image augmentation: 2000 augmented images. Out of which Training images: 1600 validation images: 400. | Anthracnose and PEB, PLB                                                                                                         | In potato plants, leaf segmentation; attention enabled; ResNet50 & UNet model.    | RS-UNet model MioU (79.8%) & Dice (88.86%) which represents improvements of 8.96% and 6.33% compared to standard UNet |
| Johnson et al. [12]             | 1840 images from low end cell phones with image dimensions 3072 × 4096 pixels, 3120 × 4160 pixels, and 2448 × 3264 pixels.                | PEB, PLB, PH                                                                                                                     | Segmentation model: Pre trained Mask R-CNN on dataset—MS COCO                     | –                                                                                                                     |
| Asha Rani and Gowrishankar [13] | Sunflower-467 and 1668 augmented images, Cauliflower and Agri-ImageNet Dataset.                                                           | Gray mold, downy mildew, and leaf scars in sunflower dataset; bacterial spot rot, black rot, downy mildew in Cauliflower dataset | 38 DL models are studied such as MobileNet; VGG; Inception; EfficientNet; ResNet. | –                                                                                                                     |
| Wang et al. [14]                | Plant village dataset having 50,000 images.                                                                                               | Apple leaves; black rot caused by Botryosphaeria obtuse fungus; Healthy leaves                                                   | Transfer Learning on VGG, Inception, and ResNet models were applied.              | VGG16 model (90.4%)                                                                                                   |
| Mutiso and Ndia [15]            | Plant village dataset.                                                                                                                    | Tomato leaves                                                                                                                    | U-Net-Vision Mamba Model (U-net-Vim-HBAM).                                        | U-net-Vim-HBAM (98.63%)                                                                                               |

## 4 Proposed System Requirements

### 4.1 Computational Requirements

The implementation uses the Google Colab platform with a fast NVIDIA v2-8 Tensor Core TPU. Python language utilised OpenCV's cv2 library along with TensorFlow and Keras framework.

#### 4.1.1 System Requirements

GPU-enabled workstation with the device specifications-LAPTOP-P4ORVA2M.

Processor: 11th Gen Intel(R) Core (TM) i5-1135G7 @ 2.40 GHz, 2.42 GHz.

RAM Memory: 16.0 GB (15.8 GB usable) RAM

#### 4.1.2 Operating System Requirements

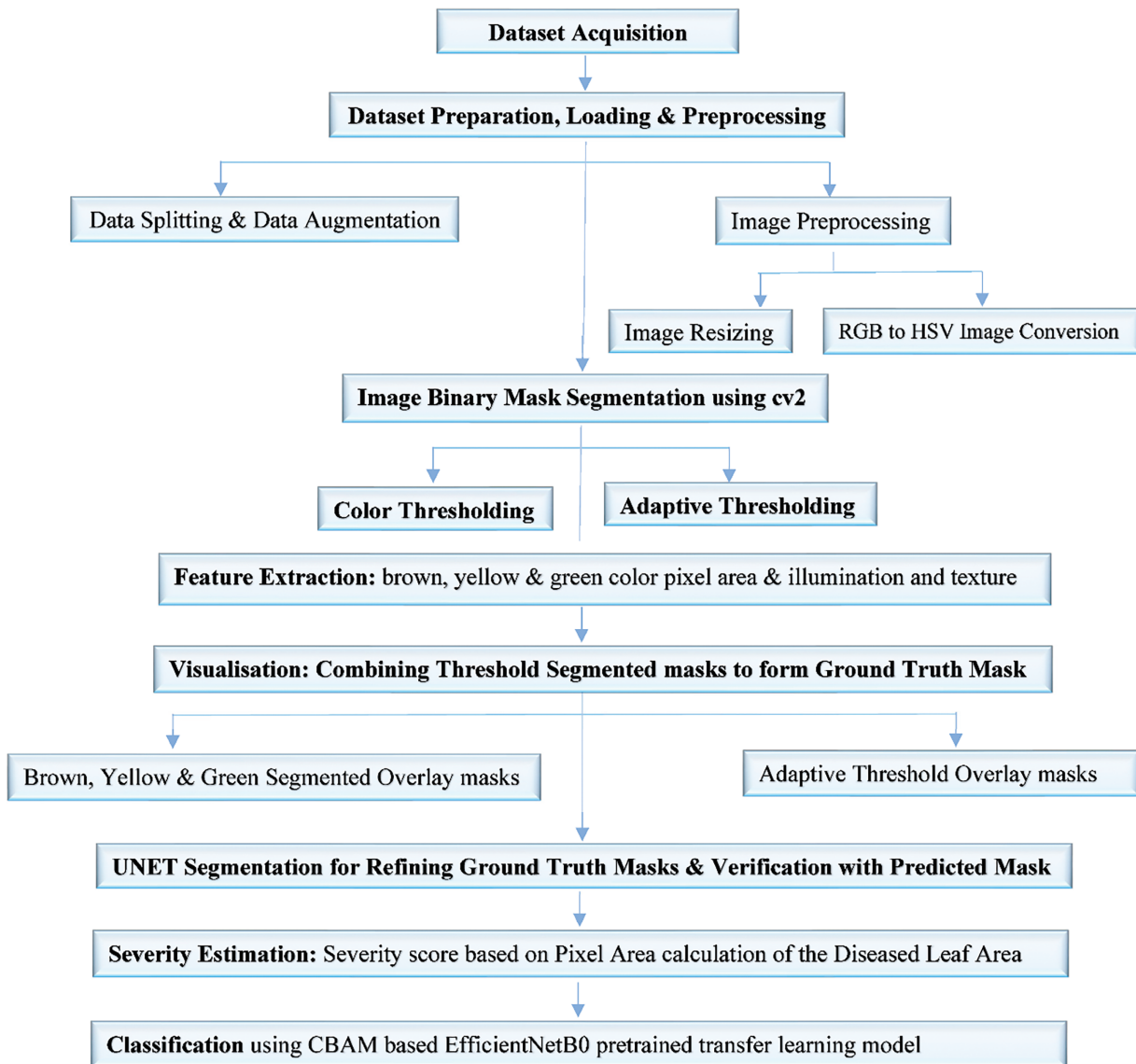
Windows 11, 64-bit OS, ×64-based processor was installed.

#### 4.1.3 Network Requirements

Network Bandwidth: 5 GHz WIFI Internet connection.

## 5 Implementation

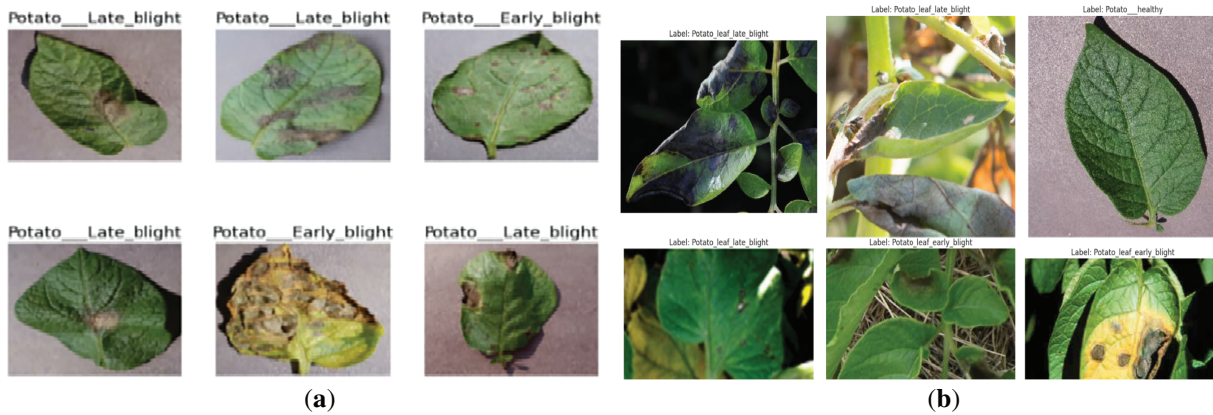
The implementation workflow is represented in Fig. 1 as given below:



**Figure 1:** Flowchart representation of potato plant disease severity estimation & classification

### 5.1 Dataset Acquisition

The dataset was taken from the Kaggle website, having classes as PEB, PLB, and PH leaves. The Plant Village dataset in Fig. 2a consists of 2152 potato plant images. It comprised 1000 PLB, 1000 PEB, and 152 PH plants. The Plant Doc dataset comprises 515 potato plant images, with 208 PLB, 171 PEB, and 136 PH plants. In Fig. 2b, some samples are shown for the Plant Doc dataset used.



**Figure 2:** (a) Representation of all three class labels, PEB, PLB and PH leaves in the Plant Village Dataset. (b) Representation of all three class labels, PEB, PLB and PH leaves in the Plant Doc Dataset

## 5.2 Data Preprocessing

### 5.2.1 Dataset Splitting

The Plant Village dataset, having 2152 images, is trifurcated into testing, validation, and training sets. Data split of 80% data in train\_ds, 10% in val\_ds, and 10% in test\_ds is done. With shuffle: True, train\_split, val\_split, test\_split as the function parameters, the splitting function utilized was `def get_dataset_partitions_tf()`. In the Plant Doc dataset, the total number of images in train\_ds is 309, of which PLB is 125, PEB is 102 and PH is 82. The total number of images in val\_ds is 51, of which PLB is 21, PEB is 17 and PH is 13. The total number of images in test\_ds: 155, of which PH is 41, PEB is 52 and PLB is 62.

### 5.2.2 Data Augmentation

Techniques like cropping, translation, rotation, flipping, scaling/zooming: zooming in or out on a part of the image and shearing are applied.

### 5.2.3 Resizing

The actual image size  $256 \times 256$  is reduced to  $224 \times 224$  for the Sequential layer and the data augmentation layer. The pixels are rescaled to 0–1, and the images are randomly transformed to improve model computational power and generalisation.

### 5.2.4 Color-Based Segmentation via cv2

RGB to HSV conversion is applied. HSV is even more robust to changes in illumination than RGB, making it better suited for color-based segmentation. Defining Color Ranges based on visual observation and experimentation, appropriate HSV ranges are defined for healthy and diseased regions. This stage aims to isolate and highlight the lesion areas associated with different potato diseases using color-based segmentation with the cv2 library. First, images undergo RGB to HSV conversion. This color space is more suitable when variations in lighting and shadow exist compared to RGB. For each disease, specific HSV color ranges are defined based on the characteristic color of the lesions. These ranges are determined empirically through analysis of a representative set of diseased leaves. A binary mask is created by thresholding the HSV image within the defined color ranges. Pixels within the range are allotted value 1, while those outside the range are allotted value 0. Finally, the original image is overlaid with the segmented area using a distinct color

(e.g., red or green) to highlight the lesion. This creates a color-segmented overlay image. Segmentation mask visualisation samples are given.

### 5.2.5 Color Thresholding

The color thresholding is a Global Thresholding technique as it applies a single, fixed range of threshold values across the entire image's color channels to segment the desired pixels. The `cv2.inRange()` function in OpenCV is used to create binary masks that isolate the pixels within the defined HSV ranges. This results in separate masks for healthy and diseased regions.

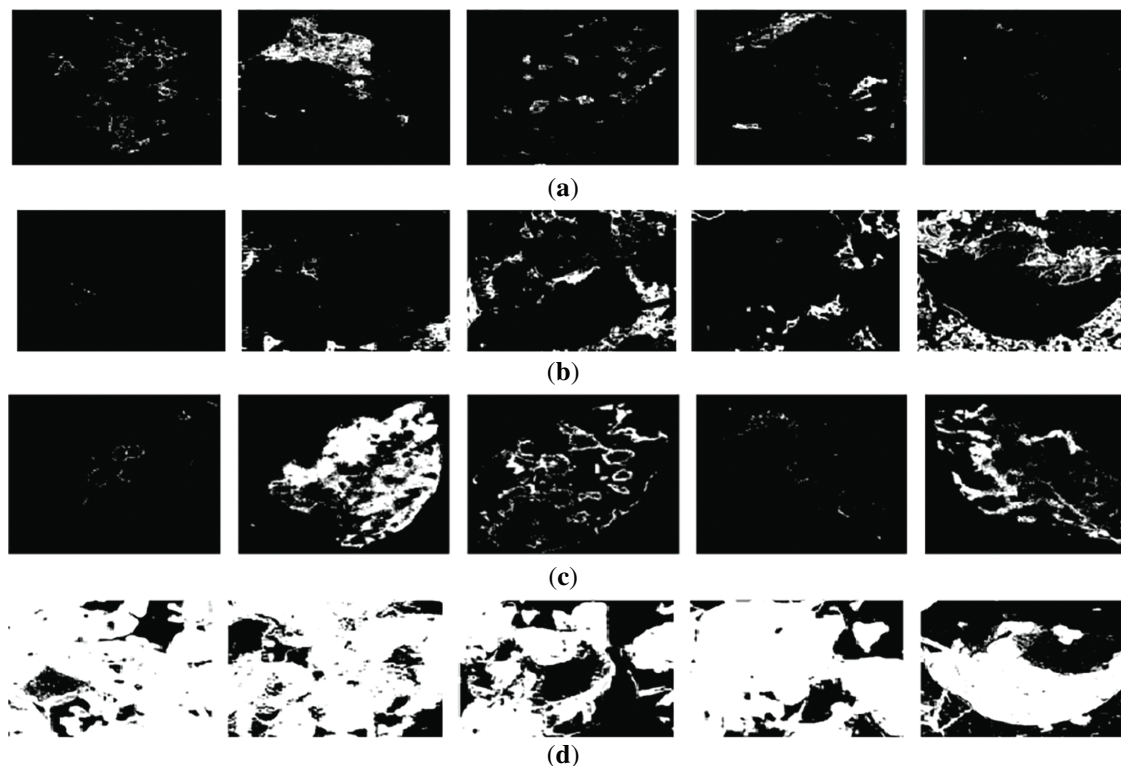
### 5.2.6 Adaptive Thresholding

This technique is a local thresholding technique wherein a separate threshold value for each pixel is calculated. Adaptive thresholding is an image binarization technique that uses a local threshold for each pixel instead of a single global one. It works by dividing an image into small regions. For each pixel, it calculates a threshold value based on the average or weighted sum of the pixel intensities within its neighbourhood. A pixel is then converted to white if its intensity is above this local threshold and black if it's below. This method is effective for images with uneven lighting or shadows because it adapts the threshold to the local conditions.

## 5.3 Visualisation of Binary Segmented Mask and Overlay Mask

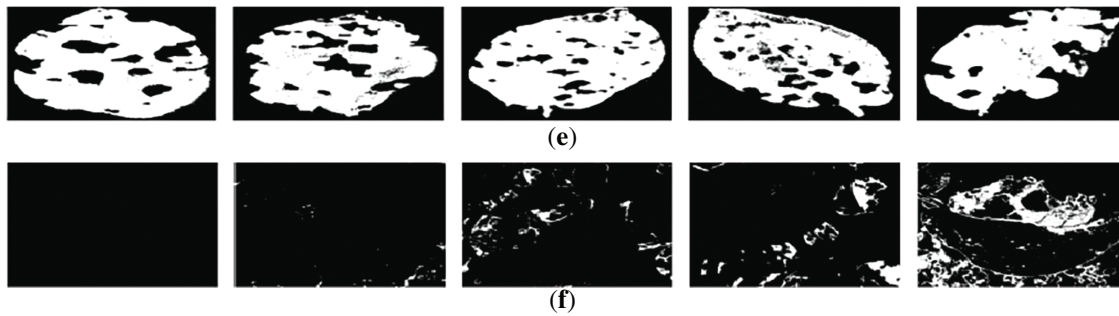
### 5.3.1 Binary Color Segmented Mask

Binary Black and White masks are generated, depicting the colored region under consideration as white, and the rest black, as shown in [Fig. 3a–f](#) below:



**Figure 3:** (Continued)

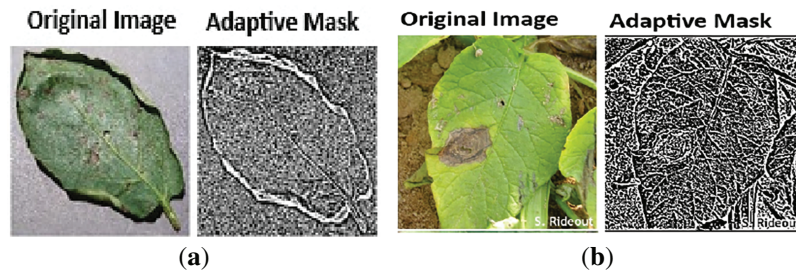




**Figure 3:** (a) Brown segmented mask samples for the Plant Village dataset. (b) Brown segmented mask samples for the Plant Doc dataset. (c) Yellow segmented mask samples for the Plant Village dataset. (d) Yellow segmented mask samples for the Plant Doc dataset. (e) Green segmented mask samples for the Plant Village dataset. (f) Green segmented mask samples for the Plant Doc dataset

### 5.3.2 Adaptive Threshold Binary Segmented Mask Generation

The adaptive threshold segmented masks catering to the lighting feature are crucial to eliminate any light-based distortions. These adaptive masks are given in Fig. 4a,b for the Plant Village and the Plant Doc dataset as follows:



**Figure 4:** (a) Adaptive mask samples for Plant Village dataset. (b) Adaptive Overlay mask samples for Plant Doc dataset

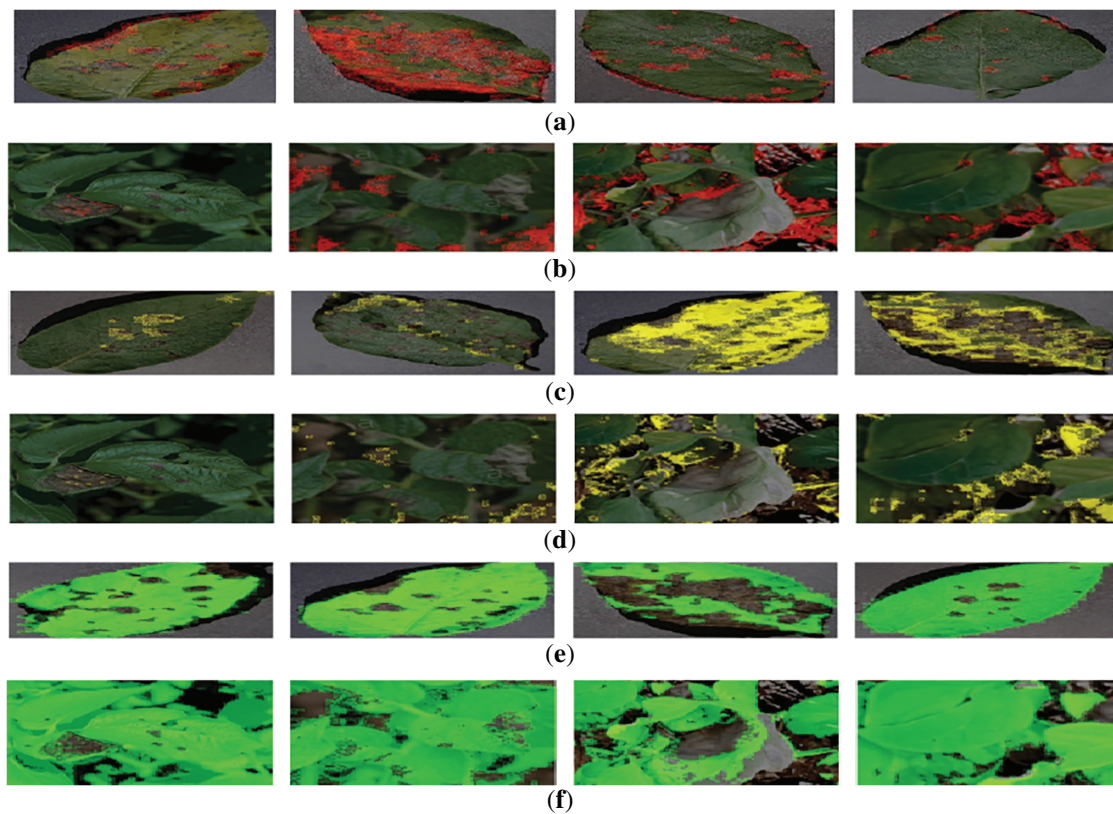
### 5.3.3 Overlay Mask Generation Based on Color Feature

The colored overlay mask highlights the region under consideration. Repositories of Brown, Yellow, and Green highlighted overlays are stored separately in different Google Drive directories. The distinct colored masks are overlaid over original images through the function named `cv2.addWeighted()`. The overlay images of the diseased areas were easily visualised, identified, and assessed. The transparency of the mask can be adjusted to control the visibility of the underlying image. The overlay masks generated are given in Fig. 5a–f as below:

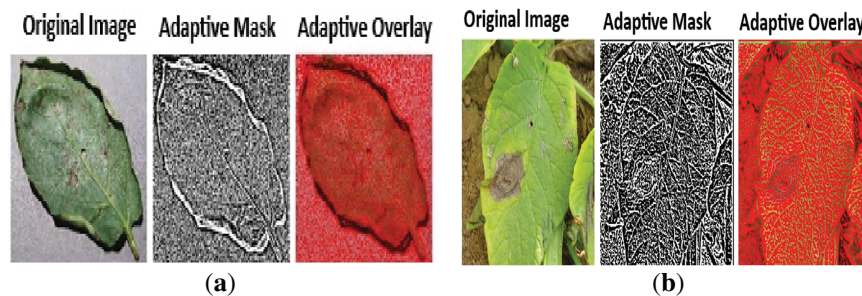
### 5.3.4 Overlay Mask Generation Based on Illumination and Texture Feature

The Adaptive Thresholding-based segmentation masks and corresponding adaptive overlay masks are shown in Fig. 6a for the Plant Village dataset and Fig. 6b for the Plant Doc as below:





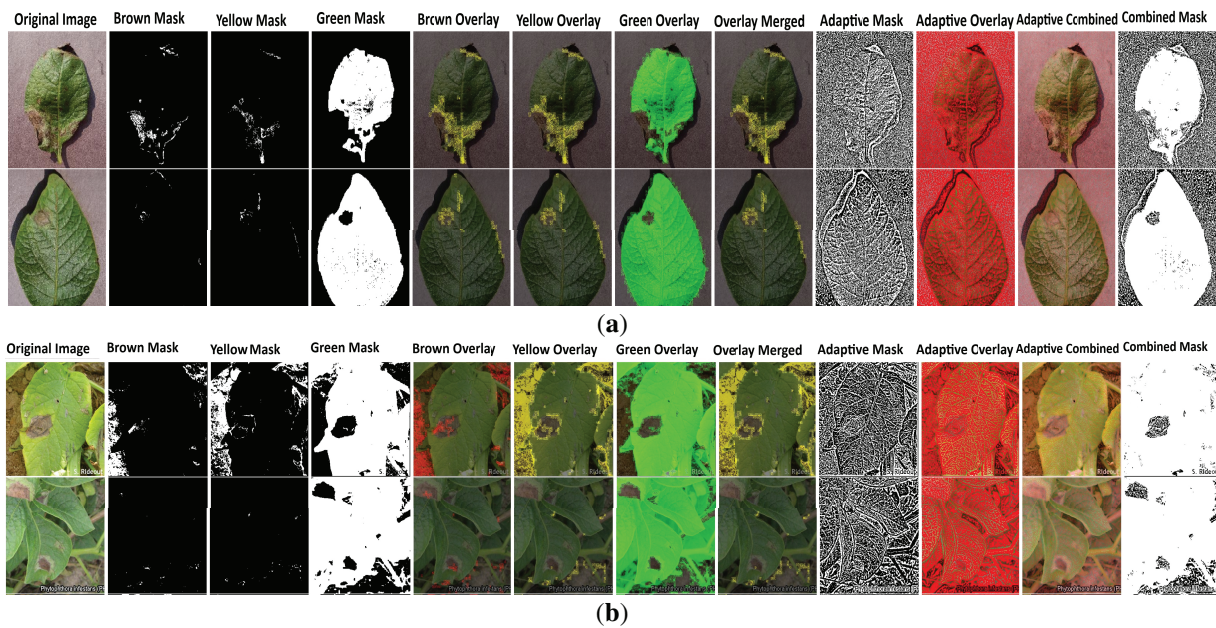
**Figure 5:** (a) Brown segmented overlay mask samples for the Plant Village dataset. (b) Brown segmented overlay mask samples for the Plant Doc dataset. (c) Yellow segmented overlay mask samples for the Plant Village dataset. (d) Yellow segmented overlay mask samples for the Plant Doc dataset. (e) Green segmented overlay mask samples for the Plant Village dataset. (f) Green segmented overlay mask samples for the Plant Doc dataset



**Figure 6:** (a) Adaptive Overlay mask samples for Plant Village dataset. (b) Adaptive overlay mask samples for Plant Doc dataset

### 5.3.5 Overlay Masks Combination Based on Color, Illumination & Texture Features

The Adaptive masks are overlaid over the original images for visualisation. This is based on features like color, illumination and texture of the leaf images as shown in Fig. 7a for the Plant Village dataset and Fig. 7b for the Plant Doc dataset as follows:



**Figure 7:** (a) Combined Final Ground Truth Masks for Plant Village dataset. (b) Combined Final Ground Truth Masks for Plant Doc dataset

#### 5.4 Segmentation and Disease Severity Estimation

The segmentation and disease severity estimation process used in this paper has been systematically mentioned as follows:

##### 5.4.1 Affected Area Calculation

The pixel count in the affected, diseased region mask is counted using the `cv2.countNonZero()` function.

##### 5.4.2 Total Leaf Area Calculation

The total count of pixels in the leaf area is calculated. This is done by isolating the entire leaf area via a segmentation step, which is achieved via the estimation of image dimensions.

##### 5.4.3 Severity Calculation

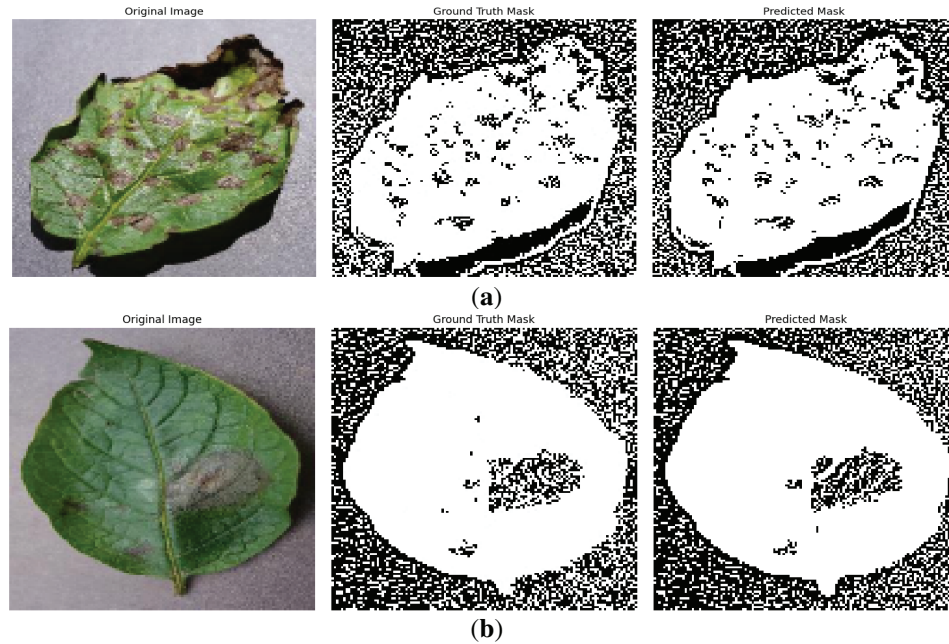
The disease severity is calculated as the ratio of the Diseased\_Area to the Total\_Leaf\_Area, expressed as a percentage using the formula given in Eq. (1) below:

$$\text{Disease Severity (\%)} = \left( \frac{\text{DiseasedArea}}{\text{TotalLeafArea}} \right) * 100 \quad (1)$$

##### 5.4.4 U-Net Segmentation for Mask Refinement

The initial stage focuses on using image segmentation and the DL technique for precise analysis of diseased areas. CV2 and UNET-Refined Masks are created. The process begins by using a U-Net model to create highly accurate segmentation masks of the potato leaves and their diseased spots. The combined overlays serve as the ground truth masks. Ground truth masks are the ideal, correct segmentation masks. In the context of training a model, they serve as the target output that the model tries to learn. Predicted masks

are the masks generated by the trained U-Net model. These are the model's estimations of the segmentation based on the images it has processed. The term “U-Net refined masks” implies that the model's output is clean and precise. In Fig. 8a,b, the Ground Truth Masks and predicted masks are plotted as given below:



**Figure 8:** (a) U-Net Mask Segmentation for Plant Village dataset. (b) U-Net Mask Segmentation for Plant Doc dataset

### 5.5 CBAM-EfficientNetB0 Transfer Learning Based Model Architecture

In this paper, the combination of CBAM and EfficientNetB0 involves two separate, powerful concepts. EfficientNetB0 has a complex architecture with compound scaling, while the CBAM module is an attention mechanism with its own set of mathematical operations. When a pretrained model is used, it means the weights for these mathematical operations have already been learned from a massive dataset, like ImageNet.

#### 5.5.1 EfficientNetB0 Architecture

EfficientNetB0 is the smallest model in the EfficientNet family. Its design is based on compound scaling, a technique that balances depth, width, and resolution of the network using a fixed set of scaling coefficients. Instead of scaling a single dimension, EfficientNet scales all three simultaneously to achieve better performance and efficiency. The core of this is the compound coefficient  $\phi$ . The key mathematical relationship for compound scaling is given in Eqs. (2)–(4) as follows:

$$\text{Depth: } d = \alpha\phi \quad (2)$$

$$\text{Width: } w = \beta\phi \quad (3)$$

$$\text{Resolution: } r = \gamma\phi \quad (4)$$

where constants calculated through a small grid search EfficientNetB0 model are represented as  $\alpha$ ,  $\beta$  and  $\gamma$ , subject: **constraint:**  $\alpha.(\beta)^2.(\gamma)^2 \approx 2$ . For EfficientNetB0, these coefficients are taken as  $\alpha = 1.2$ ,  $\beta = 1.1$ , and  $\gamma = 1.15$  with  $\phi = 0$ . The model's architecture is a series of Mobile Inverted Bottleneck Convolution (MBConv) blocks, which use depthwise separable convolutions to reduce computational cost.



### 5.5.2 CBAM (Convolutional Block Attention Module)

The CBAM module is embedded inside the EfficientNetB0 architecture to improve the network's emphasis on essential features. It sequentially applies two Modules, namely Channel Attention and Spatial Attention.

#### (a) Channel Attention Module

This module helps models decide “what” is essential in the image by focusing on the channels. It takes a feature map  $F$  as input, and to accumulate spatial information, it performs average-pooling and max-pooling. Then the pooled features are passed via a common Multi-Layer Perceptron. The outputs are then added element-wise and fed through an activation function—sigmoid to create a channel attention map,  $M_C$ . Mathematical formulation has been given in Eq. (5) below:

$$M_C(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) \quad (5)$$

The modified feature map  $F'$  is attained from the actual feature map  $F$ , which is multiplied by this channel attention map. It is illustrated in Eq. (6) below:

$$F' = M_C(F) \otimes F \quad (6)$$

#### (b) Spatial Attention Module

This module helps the model decide “where” to focus. It feeds input as the feature map  $F'$ , which is channel-refined. It uses two techniques, namely average-pooling and max-pooling, along with the channel axis to form two 2D feature maps. These are combined and convolved to create a spatial attention map. The mathematical formulation is given in Eq. (7) below:

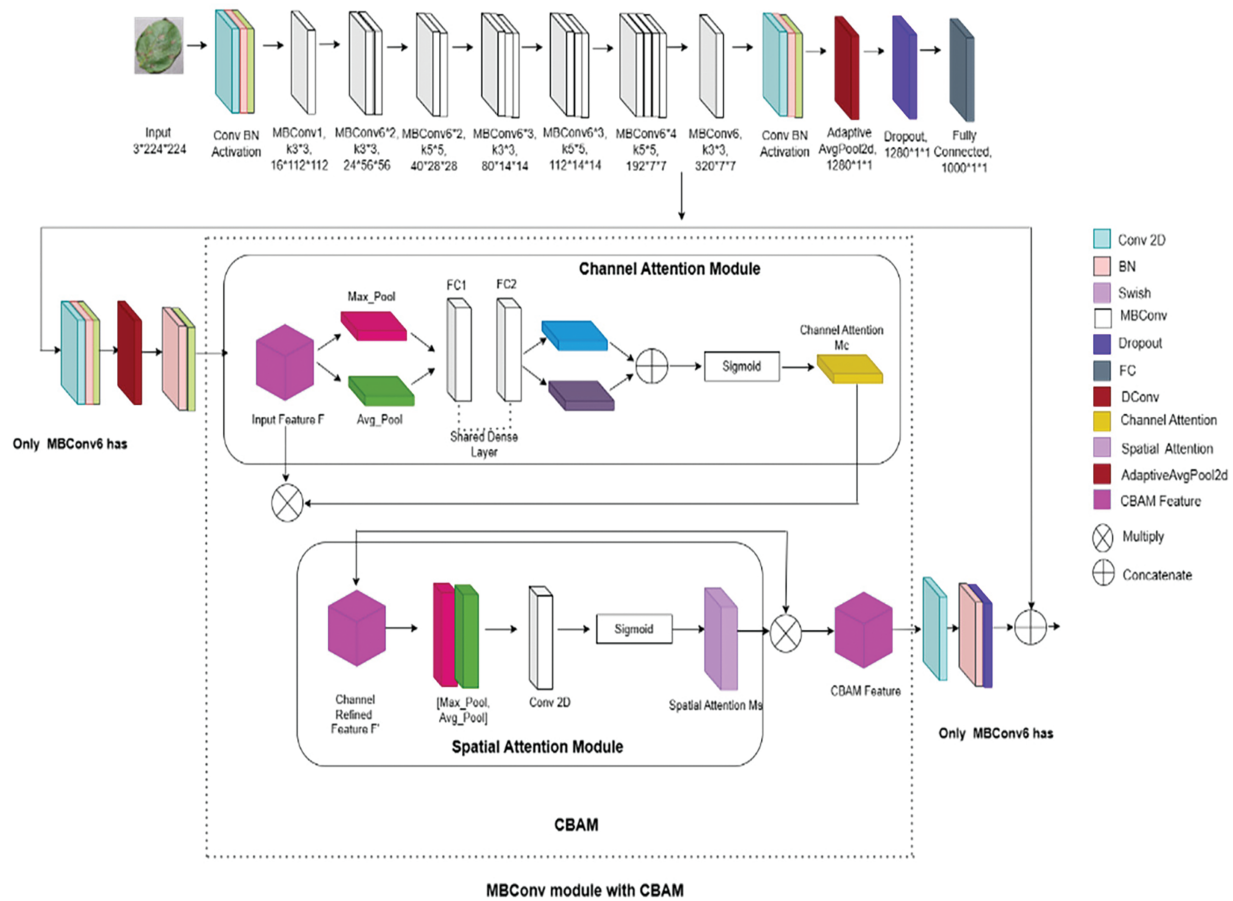
$$M_S(F') = \sigma(f^{7 \times 7}((\text{AvgPool}(F'); (\text{MaxPool}(F')))) \quad (7)$$

The above convolution operation is represented by  $(f)^{7 \times 7}$ , and here the kernel size is  $7 \times 7$ . Finally, the channel-modified feature map  $F'$  is multiplied by the spatial attention map to get  $F''$ , i.e., the final output feature map, which is given in Eq. (8) as below:

$$F'' = M_S(F') \otimes F' \quad (8)$$

### 5.5.3 Pretrained CBAM-EfficientNetB0

For the pretrained CBAM-EfficientNetB0 model, the weights for all the convolutional layers and the MLP in the CBAM module have been learned on a humongous dataset like that of ImageNet. This allows the model to already have a robust understanding of lower-level features such as colours, textures, and higher-level features, i.e., objects. All mathematical operations described above are the core mechanisms by which the pretrained model processes new images and classifies them. Transfer learning is applied to optimise these learned weights on a small dataset like potato plant diseases, allowing the model to quickly adapt to your new task without starting from scratch. The CBAM architecture is shown in Fig. 9 below:



**Figure 9:** CBAM-based EfficientNetB0 architecture

#### 5.5.4 Transfer Learning and Training

The pre-trained CBAM-EfficientNetB0 model on a large dataset, such as ImageNet, learns general image features. The pre-trained model is then refined on the potato plant disease datasets, both Plant Village and Plant Doc. The input to the model consists of two channels: the original image and the segmented overlay image generated in the previous stage. This allows the model to leverage both the overall image context and the highlighted lesion areas. For the Plant Doc dataset, the number of input image channels for classification is one, i.e., it was fed the augmented leaf images due to its small size initially. Transfer Learning makes the training process faster and more efficient. The hyperparameters tuned are given in Table 2 as follows:

**Table 2:** Hyperparameters tuned for plant village dataset & plant doc dataset

| Hyperparameters | Plant village dataset           | Plant doc dataset               |
|-----------------|---------------------------------|---------------------------------|
| Optimiser       | Adam                            | Adam                            |
| Learning rate   | 0.0001                          | 0.00001                         |
| Loss function   | SparseCategoricalCrossentropy() | SparseCategoricalCrossentropy() |
| Epochs          | 50 (17 with early stopping)     | 10                              |



## 6 Results

The proposed method was tested on potato plant leaf overlay merged images with varying degrees of disease severity. The results indicate that both Severity Estimation via cv2 led Image segmentation, overlay mask generation, UNET segmentation, visualisation, and Classification via attention enabled CBAM-based EfficientNetB0 model were performed with utmost efficacy.

### 6.1 Severity Estimation

The estimated average severity values correlated well with visual assessments performed by experts. The estimated Disease Severity average calculation for all the leaf images of the respective class labels is shown in Table 3 for the Plant Village dataset and the Plant Doc dataset as below:

**Table 3:** Average disease severity estimation

| Class labels        | Plant village dataset | Plant doc dataset |
|---------------------|-----------------------|-------------------|
| Potato early blight | 0.0431                | 0.0449            |
| Potato late blight  | 0.0567                | 0.508             |
| Potato healthy      | 0.0427                | 0.0428            |

### 6.2 U-Net Segmentation

The U-Net-based segmentation has several evaluation metrics as given in Table 4 below:

**Table 4:** U-Net segmentation evaluation metrics

| Evaluation metrics     | Plant village dataset | Plant doc dataset |
|------------------------|-----------------------|-------------------|
| IOU                    | 0.735                 | 0.810             |
| Dice coefficient       | 0.847                 | 0.894             |
| Testing precision      | 0.947                 | 0.959             |
| Testing recall         | 0.895                 | 0.902             |
| Testing F1 score       | 0.920                 | 0.929             |
| Testing specificity    | 0.881                 | 0.870             |
| Testing pixel accuracy | 0.981                 | 0.894             |

### 6.3 Classification

For the classification of our dataset into three respective classes, we utilised the CBAM-enabled EfficientNetB0 model. The following evaluation metrics for performance are represented in Eqs. (9)–(12) as follows:

**Accuracy:** Correctly classified images percentage;

$$Accuracy = (TP + TN) / (TP + FP + TN + FN) \quad (9)$$

**Precision:** The ratio of true positives over the overall count of predicted positives;

$$Precision(p) = \frac{TP}{TP + FP} \quad (10)$$

**Recall:** The ratio of true positives over the overall count of actual positives.

$$Recall(r) = TP / (TP + FN) \quad (11)$$

**F1-score:** The mean(harmonic) of precision and recall.

$$F1\ Score = 2 * p * r / (p + r) \quad (12)$$

where:

TP = True Positives (precisely identified affected(diseased) samples)

TN = True Negatives (precisely identified healthy samples)

FP = False Positives (healthy samples inaccurately identified as diseased)

FN = False Negatives (diseased samples inaccurately identified as healthy).

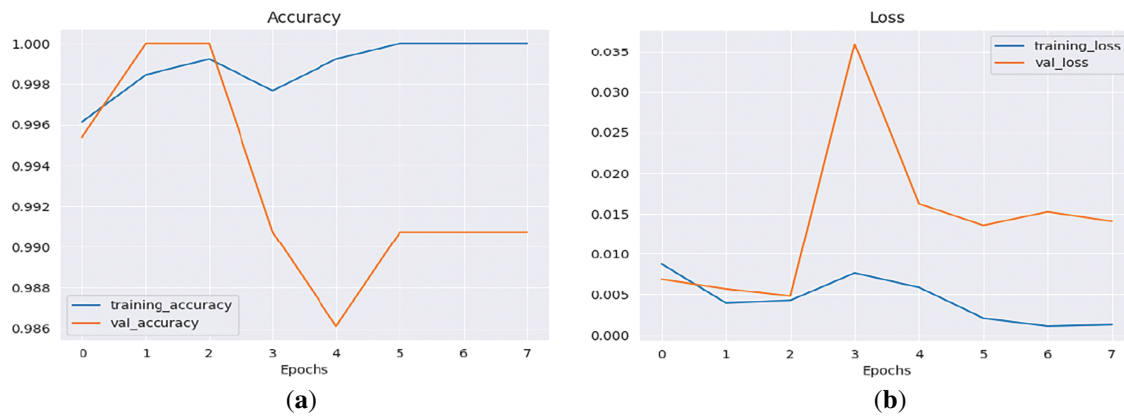
The performance measures used in the research are Accuracy score, Precision, Recall, F1 Score, Classification Report, Confusion matrix and Loss. All the performance metrics are given in [Table 5](#) for the Plant Village dataset and the Plant Doc dataset as follows:

**Table 5:** Classification performance metrics for plant village and plant doc dataset

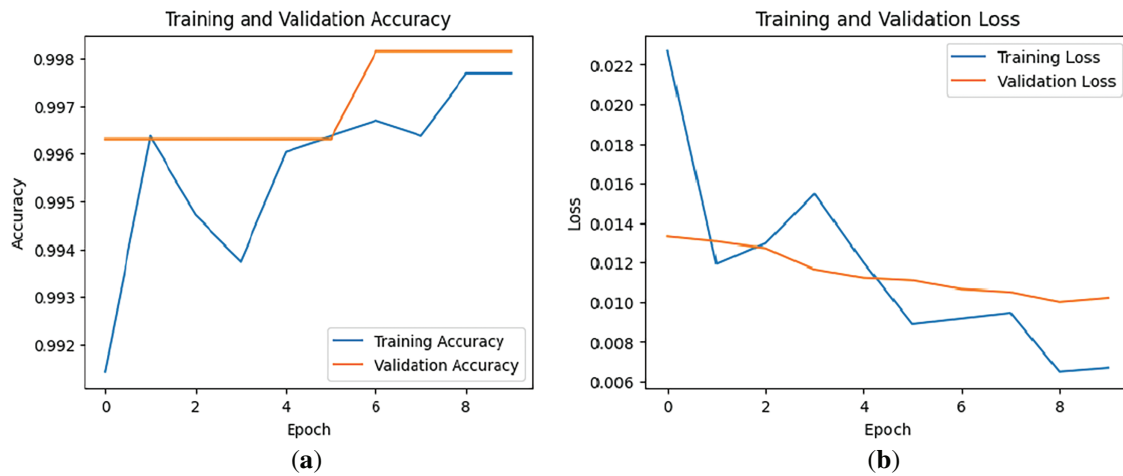
| Dataset       | Accuracy score | Precision         | Recall           | F1 score                          | Classification_Report |        |          |         | Confusion_Matrix                     | Loss                           |
|---------------|----------------|-------------------|------------------|-----------------------------------|-----------------------|--------|----------|---------|--------------------------------------|--------------------------------|
| Plant Village | Test: 0.9907   | Macro average     | Macro average    | Macro average F1 Score: 0.98      | Precision             | Recall | F1-score | Support | [[300 0 0]<br>[0 294 0]<br>[0 0 46]] | Test: 0.0203                   |
|               |                | Precision: 0.96   | Recall: 0.99     | Score: 0.98                       | EB                    | 1.00   | 1.00     | 1.00    | 300                                  |                                |
|               |                | Weighted average  | Weighted average | Weighted average F1 Score: 0.99   | LB                    | 1.00   | 0.98     | 0.99    | 300                                  |                                |
|               |                | Precision: 0.99   | Recall: 0.99     | Score: 0.99                       | Healthy               | 0.88   | 1.00     | 0.94    | 46                                   |                                |
|               |                |                   |                  |                                   | Accuracy              |        |          | 0.99    | 646                                  |                                |
|               |                |                   |                  |                                   | Macro avg             | 0.96   | 0.99     | 0.98    | 646                                  |                                |
|               |                |                   |                  |                                   | Wt. avg               | 0.99   | 0.99     | 0.99    | 646                                  |                                |
| Plant Doc     | Test: 0.9737   | Macro average     | Macro average    | Macro average F1 Score: 0.9737    | Precision             | Recall | F1-score | Support | [[230 5 0]<br>[15 229 0]]            | Test: 0.06366<br><br>[0 0 224] |
|               |                | Precision: 0.9741 | Recall: 0.9737   | Score: 0.9737                     | EB                    | 0.94   | 0.98     | 0.96    | 235                                  |                                |
|               |                | Weighted average  | Weighted average | Weighted average F1 Score: 0.9737 | LB                    | 0.98   | 0.94     | 0.96    | 244                                  |                                |
|               |                | Precision: 0.9741 | Recall: 0.9737   | Score: 0.9737                     | Healthy               | 1.00   | 1.00     | 1.00    | 244                                  |                                |
|               |                |                   |                  |                                   | Accuracy              |        |          | 0.97    | 723                                  |                                |
|               |                |                   |                  |                                   | Macro avg             | 0.97   | 0.97     | 0.97    | 723                                  |                                |
|               |                |                   |                  |                                   | Wt. avg               | 0.97   | 0.97     | 0.97    | 723                                  |                                |

The distinct accuracy graph and loss graph for the classification model are given in [Fig. 10a,b](#) for Plant Village dataset and [Fig. 11a,b](#) for Plant Doc dataset as depicted below:

The results demonstrated the efficacy of our novel method for disease classification and severity estimation in the potato plant. The CBAM-EfficientNetB0 model with color and adaptive threshold segmented overlays achieved slightly higher accuracy for the Plant Village dataset compared to the Plant Doc dataset. Specifically, we observed the following:



**Figure 10:** (a) Accuracy graph for the plant village dataset. (b) Loss graph for the plant village dataset



**Figure 11:** (a) Accuracy graph for plant doc dataset. (b) Loss graph for the plant village dataset

### 6.3.1 Improved Accuracy

The cv2 color segmentation, illumination, lighting feature, and texture features using adaptive threshold segmented masks enhanced our UNET model to refine these masks, allowing the model to focus on the areas of interest. This resulted in improved classification accuracy, particularly for diseases with subtle visual differences.

### 6.3.2 Enhanced Generalisation

The transfer learning approach enabled the model to generalise well to unseen images, even with little training data.

### 6.3.3 Effective Attention Mechanism

The attention mechanism-based CBAM module focused on informative features of the dimensions, both channel and spatial. This qualifies the model's capability to discriminate between different diseases.

#### 6.3.4 Reasonable Severity Estimation

The method provides a reasonable estimation of disease severity, aligning with visual assessments.

### 7 Conclusion

This paper has a novel take on the integrated approach served by severity estimation via cv2 colour segmentation overlays fed to the CBAM-EfficientNetB0 transfer learning model. The results presented show that this novel workflow and systematic structure of the methodology achieve superior results compared to conventional DL approaches, highlighting the efficacy of integrating classical image processing and computer vision for enhanced disease detection and severity estimation as well. The primary feature extracted is the colour of the leaf, based on which severity is estimated, since it's one of the most quintessential features for detecting severity in leaves. The color segmentation step aids in the same.

**Acknowledgement:** Really grateful to the Department of Biotechnology (DBT) for kind support in undertaking this research project.

**Funding Statement:** Funding has been received for this study. This research was done under Department of Biotechnology (DBT) project titled “Application of Machine Learning for Hyperspectral Imaging and Remote Sensing aimed at Early Detection of Fungal Foliar Diseases and Bacterial Wilt Diseases in Potato Crop”, DBT/Reference. No. BT/PR45388/133/58/2022.

**Author Contributions:** The research work was done by Kajal Kaul, Amit Prakash Singh, Anuradha Chug, Ravinder Kumar and Veerubommu Shanmugam. Kajal Kaul contributed in the following areas: introduction, motivation, literature review, proposed system requirements, implementation, results and conclusion. Amit Prakash Singh and Kajal Kaul helped in proofreading and formatting, and Anuradha Chug, Ravinder Kumar, and Veerubommu Shanmugam reviewed the results. All authors reviewed the results and approved the final version of the manuscript.

**Availability of Data and Materials:** The dataset used for this paper will be made available only upon seeking consent from the authors.

**Ethics Approval:** Not applicable.

**Conflicts of Interest:** The authors declare no conflicts of interest to report regarding the present study.

### Abbreviations

|      |                                    |
|------|------------------------------------|
| PEB  | Potato Early Blight                |
| PLB  | Potato Late Blight                 |
| PH   | Potato Healthy                     |
| CBAM | Convolution Block Attention Module |
| GDP  | Gross Domestic Product             |
| DL   | Deep Learning                      |
| CNN  | Convolution Neural Networks        |

### References

1. Verma S, Chug A, Singh AP, Singh D. Plant disease detection and severity assessment using image processing and deep learning techniques. *SN Comput Sci.* 2023;5(1):83. doi:10.1007/s42979-023-02417-5.
2. Kaul K, Chug A, Singh AP, Kumar R, Shanmugam V. Hyperparameter tuning of CNN based potato plant disease detection model integrated with GridSearchCV. *J Inf Optim Sci.* 2025;46(1):1–14. doi:10.47974/jios-1846.
3. Li J, Wu J, Liu R, Shu G, Liu X, Zhu K, et al. Potato late blight leaf detection in complex environments. *Sci Rep.* 2024;14(1):31046. doi:10.1038/s41598-024-82272-3.

4. Tonmoy MR, Hossain MM, Dey N, Mridha MF. MobilePlantViT: a mobile-friendly hybrid ViT for generalized plant disease image classification. *arXiv:2503.16628*. 2025. doi:10.48550/arxiv.2503.16628.
5. Nyakuri JP, Nkundineza C, Gatera O, Nkurikiyeyezu K, Mwitende G. AI and IoT-powered edge device optimized for crop pest and disease detection. *Sci Rep*. 2025;15(1):22905. doi:10.1038/s41598-025-06452-5.
6. Miao Y, Meng W, Zhou X. SerpensGate-YOLOv8: an enhanced YOLOv8 model for accurate plant disease detection. *Front Plant Sci*. 2025;15:1514832. doi:10.3389/fpls.2024.1514832.
7. Mahum R, Munir H, Mughal ZUN, Awais M, Sher Khan F, Saqlain M, et al. A novel framework for potato leaf disease detection using an efficient deep learning model. *Hum Ecol Risk Assess*. 2023;29(2):303–26. doi:10.1080/10807039.2022.2064814.
8. Rashid J, Khan I, Ali G, Almotiri SH, AlGhamdi MA, Masood K. Multi-level deep learning model for potato leaf disease recognition. *Electronics*. 2021;10(17):2064. doi:10.3390/electronics10172064.
9. Esgario JGM, Krohling RA, Ventura JA. Deep learning for classification and severity estimation of coffee leaf biotic stress. *Comput Electron Agric*. 2020;169:105162. doi:10.1016/j.compag.2019.105162.
10. Li X, Zhou Y, Liu J, Wang L, Zhang J, Fan X. The detection method of potato foliage diseases in complex background based on instance segmentation and semantic segmentation. *Front Plant Sci*. 2022;13:899754. doi:10.3389/fpls.2022.899754.
11. Fu J, Zhao Y, Wu G. Potato leaf disease segmentation method based on improved UNet. *Appl Sci*. 2023;13(20):11179. doi:10.3390/app132011179.
12. Johnson J, Sharma G, Srinivasan S, Masakapalli SK, Sharma S, Sharma J, et al. Enhanced field-based detection of potato blight in complex backgrounds using deep learning. *Plant Phenomics*. 2021;2021(1):9835724. doi:10.34133/2021/9835724.
13. Asha Rani KP, Gowrishankar S. Pathogen-based classification of plant diseases: a deep transfer learning approach for intelligent support systems. *IEEE Access*. 2023;11(20):64476–93. doi:10.1109/access.2023.3284680.
14. Wang G, Sun Y, Wang J. Automatic image-based plant disease severity estimation using deep learning. *Comput Intell Neurosci*. 2017;2017:2917536–8. doi:10.1155/2017/2917536.
15. Mutiso G, Ndia J. A unified U-net-vision mamba model with hierarchical bottleneck attention for detection of tomato leaf diseases. *J Artif Intell*. 2025;7(1):275–88. doi:10.32604/jai.2025.069768.