

RESEARCH ARTICLE

Less Is More: Lighter and Faster Deep Neural Architecture for Tomato Leaf Disease Classification

SABBIR AHMED¹, MD. BAKHTIAR HASAN¹, TASNIM AHMED, MD. REDWAN KARIM SONY, AND MD. HASANUL KABIR¹, (Member, IEEE)

Department of Computer Science and Engineering, Islamic University of Technology, Dhaka 1704, Bangladesh

Corresponding author: Md. Bakhtiar Hasan (bakhtiarhasan@iut-dhaka.edu)

ABSTRACT To ensure global food security and the overall profit of stakeholders, the importance of correctly detecting and classifying plant diseases is paramount. In this connection, the emergence of deep learning-based image classification has introduced a substantial number of solutions. However, the applicability of these solutions in low-end devices requires fast, accurate, and computationally inexpensive systems. This work proposes a lightweight transfer learning-based approach for detecting diseases from tomato leaves. It utilizes an effective preprocessing method to enhance the leaf images with illumination correction for improved classification. Our system extracts features using a combined model consisting of a pretrained MobileNetV2 architecture and a classifier network for effective prediction. Traditional augmentation approaches are replaced by runtime augmentation to avoid data leakage and address the class imbalance issue. Evaluation on tomato leaf images from the PlantVillage dataset shows that the proposed architecture achieves 99.30% accuracy with a model size of 9.60MB and 4.87M floating-point operations, making it a suitable choice for low-end devices. Our codes and models are available at <https://github.com/redwankarimsony/project-tomato>.

INDEX TERMS CLAHE, data augmentation, lightweight architecture, MobileNetV2, transfer learning.

I. INTRODUCTION

Tomato, *Solanum lycopersicum*, is one of the most common vegetables grown worldwide. According to recent statistics, around 180.64 million metric tons of tomatoes are grown worldwide which amounts to an export value of 8.81 billion US Dollars [1]. However, the production of tomatoes is on the decline due to the crop being prone to various diseases [2]. Traditional disease detection approaches require manual inspection of diseased leaves through visual cues or chemical analysis of infected areas, which can be susceptible to low detection efficiency and poor reliability due to human error. Adding to the problem, the lack of professional knowledge of the farmers and the unavailability of agricultural experts who can detect the diseases in remote areas also hamper the overall harvest production. Negligence in this

regard poses a significant threat to food security worldwide while causing great losses for the stakeholders involved in tomato production. Early detection and classification of diseases implemented using tools and technologies available to the farmers can go a long way to alleviate all the issues discussed [3].

Several solutions have been proposed using the traditional machine learning approaches for plant disease classification [4]. Moreover, the emergence of deep learning-based methods in the agricultural domain has opened a new door for researchers with outstanding generalization capability removing the dependencies on handcrafted features [5]. Recently, Convolutional Neural Network (CNN) has become a powerful tool for any classification task as it automatically extracts important features from images without human supervision. Moreover, the recent variations of CNN architectures such as AlexNet [6], DenseNets [7], EfficientNets [8], GoogLeNet [9], MobileNets [10], [11], NASNets [12],

The associate editor coordinating the review of this manuscript and approving it for publication was Yongjie Li.

Residual Networks (ResNets) [13], SqueezeNet [14], Visual Geometric Group (VGG) Networks [15] have enabled the machines to understand complex patterns enabling even better performance than humans in many classification problems.

With the introduction of transfer learning where the reuse of a model efficient in solving one problem as the starting point of another problem in a relevant domain has significantly reduced the requirement of vast computational resources [16]. Consequently, the utilization of pretrained AlexNet and GoogLeNet architectures on the publicly available PlantVillage Dataset [17] has been one of the pioneer works of leaf disease classification using transfer learning and paved the way for numerous solutions in the existing literature [18]. These deep neural architectures have been found to be extremely helpful for leaf disease classification tasks for several plants such as, apple [19], cassava [20], corn [21], cucumber [22], grape [23], maize [24], mango [25], rice [26], etc. However, most of these solutions propose deep and complex networks focusing on increasing the accuracy of detection.

Real-life applications, such as agriculture, often require small and low latency models tailored explicitly for devices with small memory and low computational power while also having comparable, if not better, accuracy. Most of the systems focusing on lightweight models had to sacrifice accuracy and/or work with a limited number of diseases/samples. In this regard, this work proposes a lightweight and fast deep neural architecture for tomato leaf disease classification. The system utilizes a pretrained MobileNetV2 as a feature extractor followed by a classifier network. Contrast Limited Adaptive Histogram Equalization (CLAHE) technique has been used to reduce the effect of poor lighting conditions from the leaf images and enhance the disease spots without increasing the noise. We have tackled the dataset imbalance, overfitting, and data leaking issues by applying runtime augmentation in different dataset splits.

The performance of the model has been evaluated on tomato leaf images from the PlantVillage dataset incorporating a healthy and nine disease classes. The baseline accuracy of the MobileNetV2 architecture was 97.27%. The proposed pipeline was robust enough to uplift it to 99.30%. Further comparison with the state-of-the-art tomato leaf disease classification models showed that the proposed approach is competent enough to achieve high accuracy while maintaining a relatively small model size and reduced number of computations. The nearest model producing a similar level of performance required a 2.4 times heavier model with 2.45 times additional FLOPs count requirements. Hence, this approach can pave the way for a suitable solution for designing real-life applications in low-end devices available to the farmers.

The rest of the paper is organized as follows. We discuss the research gaps in the existing literature to justify our motivation in Section II. The details of our proposed methodology are described in Section III. Our experimental results justifying the efficacy of our proposed method is

provided in Section IV. And finally, in Section V, we deliver our concluding remarks, discuss our limitations, and provide directions for future researchers.

II. RELATED WORKS

Current research trends on tomato leaf disease classification tend to focus on developing solutions using Deep Neural Architectures, simplifying networks for faster computation targeting embedded systems, real-time disease detection, etc. The introduction of intelligent systems incorporating these solutions could go a long way to reduce crop yield loss, remove tedious manual monitoring tasks, and minimize human efforts.

Earlier approaches in tomato leaf disease classification involved different image-based hand-crafted feature extraction techniques that were fed into machine learning-based classifiers. These works mainly focused on only a few diseases with extreme feature engineering and were often limited to constrained environments. To extract features, researchers focused on utilizing different image-level feature extraction techniques like Gray-Level Co-occurrence Matrices (GLCM) [27], Geometric and histogram-based features [28], Gabor Wavelet Transformation [29], Moth-Flame Optimization and Rough Set (MFORS) [30], and similar techniques. To segment the diseased portion of the leaves, several works have extracted the Region of Interest (RoI) using k-means clustering [28], Otsu's method [31], etc. To predict the class labels from the extracted features, Support Vector Machine (SVM) [27], [29], Decision Trees [31], and other classifiers were used. Due to their sensitivity to the surroundings of leaf images, machine learning approaches relied on rigorous preprocessing steps like manual cropping of RoI, color space transformation, resizing, background removal, and image filtering for successful feature extraction. This increased complexity due to preprocessing limited the traditional machine learning approaches to classify a handful of diseases from a small dataset, thus failing to generalize on larger ones.

The performances of a significant portion of the prior works were not comparable as they were mostly done on self-curated small datasets. This issue was alleviated to a great extent when the PlantVillage dataset was introduced containing 54,309 images of 14 different crop species and 26 diseases [17]. A subset of this dataset contains nine tomato leaf diseases and one healthy class that has been utilized by most of the recent deep learning-based works on tomato leaf disease classification. Several works on tomato leaf diseases also focused on segmenting leaves from complex backgrounds [32], real-time localization of diseases [33]–[35], detection of leaf disease in early-stage [36], visualizing the learned features of different layers of CNN model [37], [38], combining leaf segmentation and classification [39], and so on. These works mostly targeted removing the restrictions of lighting conditions and uniformity of complex backgrounds.

To alleviate the dependency on hand-crafted features along with achieving better classification accuracy with large datasets, recent transfer learning-based approaches to

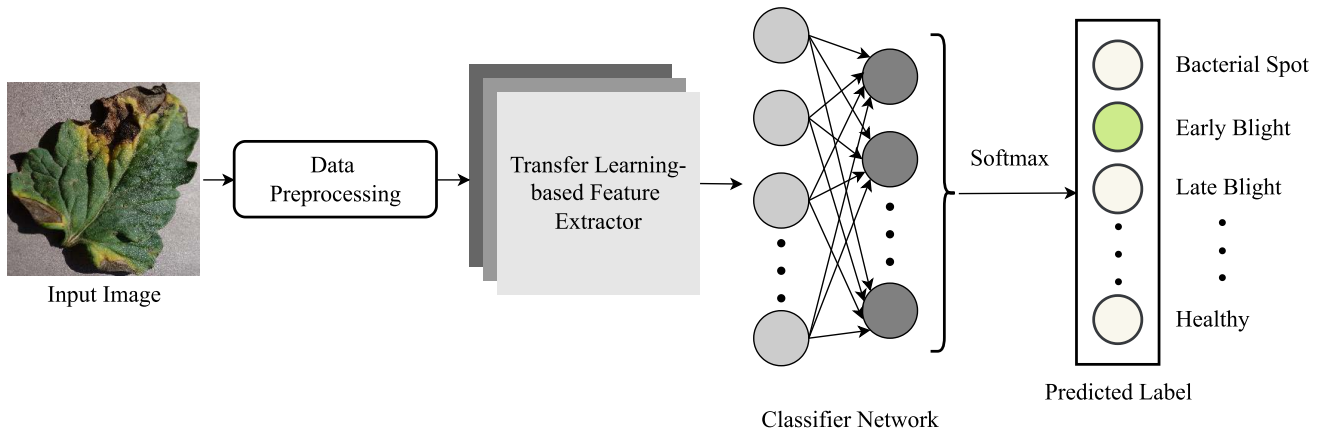


FIGURE 1. Overview of the tomato leaf disease classification architecture.

leaf disease classification have investigated the performance of different pretrained models using various hyperparameters. Based on their results, they recommended the use of GoogleNet [37], [40], [41], AlexNet [42], ResNet [43], DenseNet121 [44] in creating tomato leaf disease detection systems due to their superior performance compared to other models. Some of these works have also investigated the effect of different hyperparameter choices like optimizers, batch sizes, the number of epochs, and fine-tuning the model from different depths to see how they impact its performance [40], [42]. These models were pretrained on massive datasets, making them the perfect choice for extracting relevant features outperforming shallow machine learning-based models. Although these systems achieved high accuracy going up to 99.39% [40], the models were huge and computationally expensive, often making them infeasible for low-end devices.

Several attempts were made to reduce the computational cost and model size. Durmuş *et al.* [45] utilized SqueezeNet to detect tomato leaf diseases. The base SqueezeNet architecture reduces the computational cost by minimizing the number of 3×3 filters, late downsampling, and deep compression. The authors conducted the experiments on an NVIDIA Jetson TX1 device targeting real-time disease detection using robots. Tm *et al.* [46] proposed a variation LeNet, one of the earliest and smallest deep-learning architectures. The authors introduced an additional convolutional and pooling layer to the base architecture and increased the number of filters in different layers to extract complex features. However, the accuracies achieved by these two systems were not on par with the performance of the deeper models. Bir *et al.* [47] utilized pretrained EfficientNet-B0 to achieve a comparable accuracy with the state-of-the-art while keeping the model size and computation low. This architecture applies grid search to find coefficients for width, depth, and resolution scaling to reduce the size of the baseline model with a minimal impact on accuracy. However, when classifying the tomato leaves, the authors had to discard a significant number of samples to gain a comparable accuracy. Reduction of dataset size in this manner, even if balanced with augmentation, might result

in discarding complex samples restricting the generalization capability of the models. All these issues impose the requirement of lightweight models that can achieve state-of-the-art performance with high generalization capability.

III. MATERIALS AND METHODS

Our proposed architecture takes tomato leaf images as input and outputs the class labels. At first, the input image is passed through a preprocessing step where it is enhanced using Adaptive Histogram Equalization. Then, the enhanced image is fed to a transfer learning block, where we utilize a pretrained deep CNN model for efficient feature extraction. To determine a suitable feature extractor, we experimented with nine different pretrained architectures which are DenseNet121, DenseNet201, EfficientNet-B0, MobileNet, MobileNetV2, NASNet-Mobile, ResNet50, ResNet152V2, and VGG19. Based on the results, we have chosen MobileNetV2 due to its smaller size and faster inference while maintaining comparable accuracy. Then the features extracted by the pretrained model are fed through a shallow densely connected classifier network to get the Softmax probabilities for every class using which we predict the final label. The general pipeline of the proposed approach is depicted in Figure 1.

A. DATASET

As of today, the PlantVillage Dataset¹ is the largest open-access repository of expertly curated leaf images for disease diagnosis. The dataset comprises 54,309 images of healthy and infected leaves belonging to 14 crops, labeled by plant pathology experts. Among them, 18,160 images are of tomato leaves, divided into one healthy and nine disease classes. This dataset offers a wide variety of diseases and contains samples of leaves being infected by various diseases to different extents. One sample image from each class can be seen in Figure 2.

¹https://www.tensorflow.org/datasets/catalog/plant_village

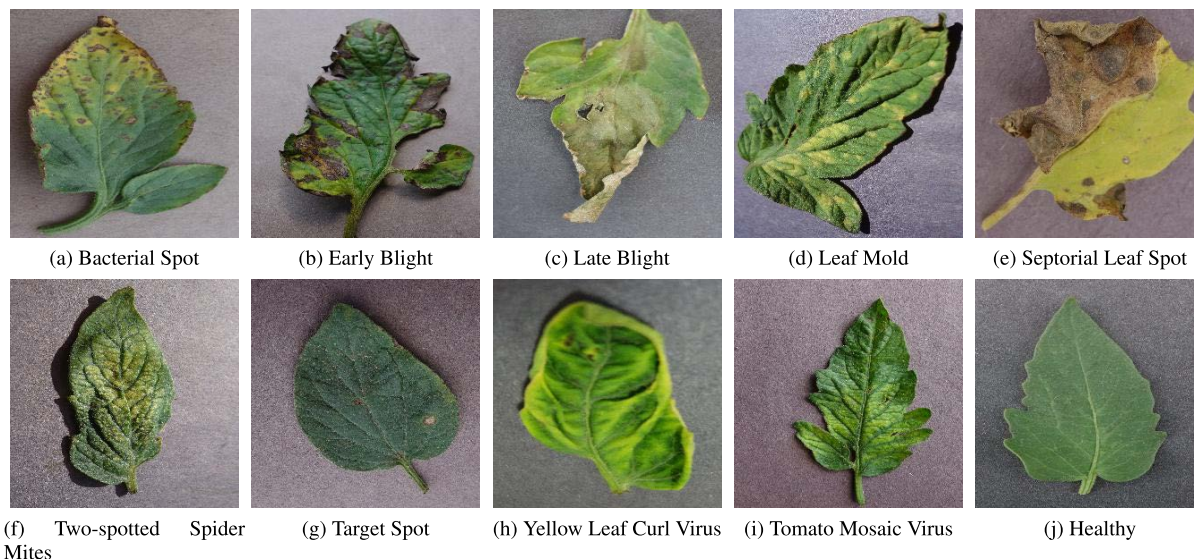


FIGURE 2. Sample tomato leaf images of the 10 classes from the PlantVillage dataset.

TABLE 1. Distribution of samples in the dataset.

Class Label	Sample Count
Bacterial Spot	2127
Early Blight	1000
Late Blight	1909
Leaf Mold	952
Septoria Leaf Spot	1771
Two-spotted Spider Mites	1676
Target Spot	1404
Yellow Leaf Curl Virus	5357
Tomato Mosaic Virus	373
Healthy	1591
Total	18160

From the distribution of the number of samples in different classes shown in Table 1, it is evident that the dataset contains an imbalance as different classes have a significantly varying number of samples. The maximum number of samples is 5357, belonging to Yellow Leaf Curl Virus disease, whereas the number of samples corresponding to Mosaic Virus disease is as low as 373. Few problems arise because of this class imbalance. First, the model does not get a good look at the images of classes with a lower number of samples, leading to less generalization [48]. Moreover, the overall accuracy might still be high even if the model is ignoring these small-sized classes, as they do not contribute much to the overall accuracy [49]. Different techniques involving undersampling and oversampling can be employed to tackle this issue, ensuring that the model is equally capable of identifying all diseases.

B. DATA PREPROCESSING

Disease spots often have close intensity values with the surroundings due to the poor lighting condition of the images

provided in the dataset. Moreover, in real-world applications, images captured by the end-users might not always be adequately illuminated, and this might fail to provide the model with enough details to identify the disease, and hence affect the classification result [50]. Contrast enhancement techniques like histogram equalization can be applied to enhance the details and correct the illumination problem. Generally, histogram-based approaches work globally throughout the image. However, the intensity distribution of the leaf regions can be different from that of the background. So, the same transformation function cannot be applied to the entire image. To tackle the illumination problem by addressing the uneven distribution of intensity, we opted for Contrast Limited Adaptive Histogram Equalization [51].

Furthermore, as mentioned earlier, there exists a class imbalance in the original dataset. This issue has been tackled in various ways in the existing literature. The most common way of dealing with this has been to undersample and/or oversample certain classes [41], [43], [44], [47]. Although it makes the dataset balanced to some extent, it has its own drawbacks. Undersampling may drop some of the challenging images for certain classes that can contain important information for the model to learn, which eventually hinders the generalizing capability of the model. Oversampling utilizes different data augmentation techniques to produce multiple copies of the original images, each having slight variations. But if we perform augmentation before splitting the dataset into train, validation, and test sets, it might inject slight variations of the training set into the test set. As the model learns to classify one variation of the image while training, it is highly likely to correctly classify the other variations in the test set, overestimating the accuracy of the system. This problem is known as data leakage [52]. As each choice has its pros and cons, we decided to perform data augmentation during runtime.

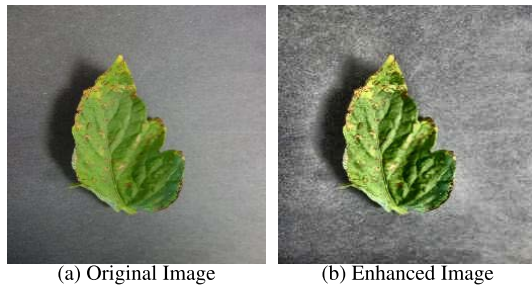


FIGURE 3. Illumination correction using contrast limited adaptive histogram equalization.

1) CONTRAST LIMITED ADAPTIVE HISTOGRAM EQUALIZATION (CLAHE)

CLAHE increases the contrast between diseased spots and the leaf by dividing the image into multiple small regions and applying a transformation function that is proportional to a cumulative distribution function. This function is calculated based on the histogram of the intensity distribution of the pixels inside each region. CLAHE also limits the amplification of the noise, which is prevalent in low light images, near regions with constant intensity by clipping the histogram value beyond a threshold. Figure 3 shows the sample output after applying CLAHE on an original image.

Before applying CLAHE, the leaf image was converted from RGB color space to Hunter Lab color space. The intensity channel of the leaf image was divided into $P \times Q$ regions, where P denotes the number of contextual regions on the x-axis, and Q denotes the number of contextual regions on the y-axis. Our empirical results demonstrated that a value of 7 for both P and Q provided the best results.

A histogram was calculated considering the intensity of the pixels for each contextual region. Then each histogram was clipped based on a threshold β , which was set to be 3 upon experimentation. After that, using the Cumulative Distribution Function, a function was generated to map the input intensities with the desired output intensities. While mapping, the function performed Bilinear Interpolation of the four nearby regions to reduce the blocking effect. Finally, the leaf image was converted from Hunter Lab color space to RGB color space.

To maintain consistency, we have preprocessed all the tomato leaf images of the dataset using CLAHE before feeding them to the model.

2) DATA AUGMENTATION

To reflect real-life scenarios, we have picked height and width shifting, clockwise and counterclockwise rotation, shearing, and horizontal flipping to augment the leaf images during runtime.

Height and Width Shifting is performed by translating each pixel of the image respectively in the horizontal and vertical direction by a constant factor. In our case, the constant factor was chosen randomly within the range $[0, 0.2]$. While

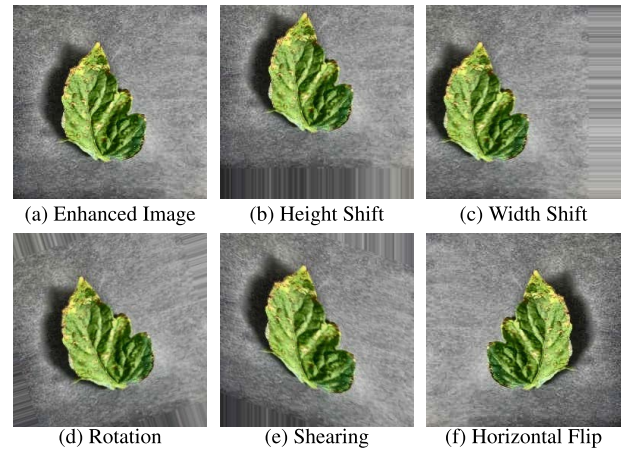


FIGURE 4. Data augmentations. A combination of these augmentations were applied randomly during run-time.

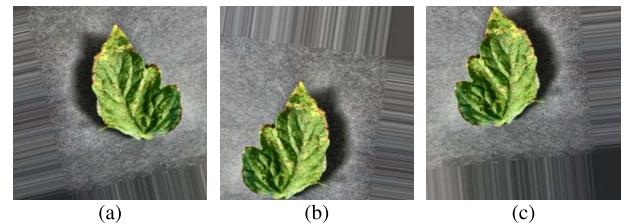


FIGURE 5. Sample augmentations performed on the images during training, validation, and testing phase.

shifting, the pixels going outside the boundary are discarded, and the empty regions are filled with the RGB values of the nearest pixels. Fig. 4b and 4c shows the effect of performing height and width shift, respectively.

Rotation is performed with respect to the center pixel of the image. For our case, the rotation angle was chosen randomly within the range $[-20, 20]$ degrees. Fig. 4d shows the effect of performing rotation. Shearing is performed by moving each pixel towards a fixed direction by an amount proportional to the pixel's distance from the bottom-most pixels of the image based on a shearing factor. We randomly picked the shearing factor within the range $[0, 0.2]$. Fig. 4e shows the effect of performing shearing. Flipping an image horizontally requires mirroring the pixels with respect to the centerline parallel to the x-axis. Fig. 4f shows the effect of performing horizontal flipping.

Multiple random augmentations are applied to the same image to ensure that the model sees a new variation on every epoch and thus learns to recognize a variety of images. Fig. 5 shows the effect of combining different augmentations that are used during the training, validation, and testing phase.

Unlike traditional approaches in the existing literature on tomato leaf disease classification, we decided not to use data augmentations to increase the number of samples before training. Instead, these augmentations were performed randomly on different images during runtime in different splits, ensuring that the model sees different variations of the same

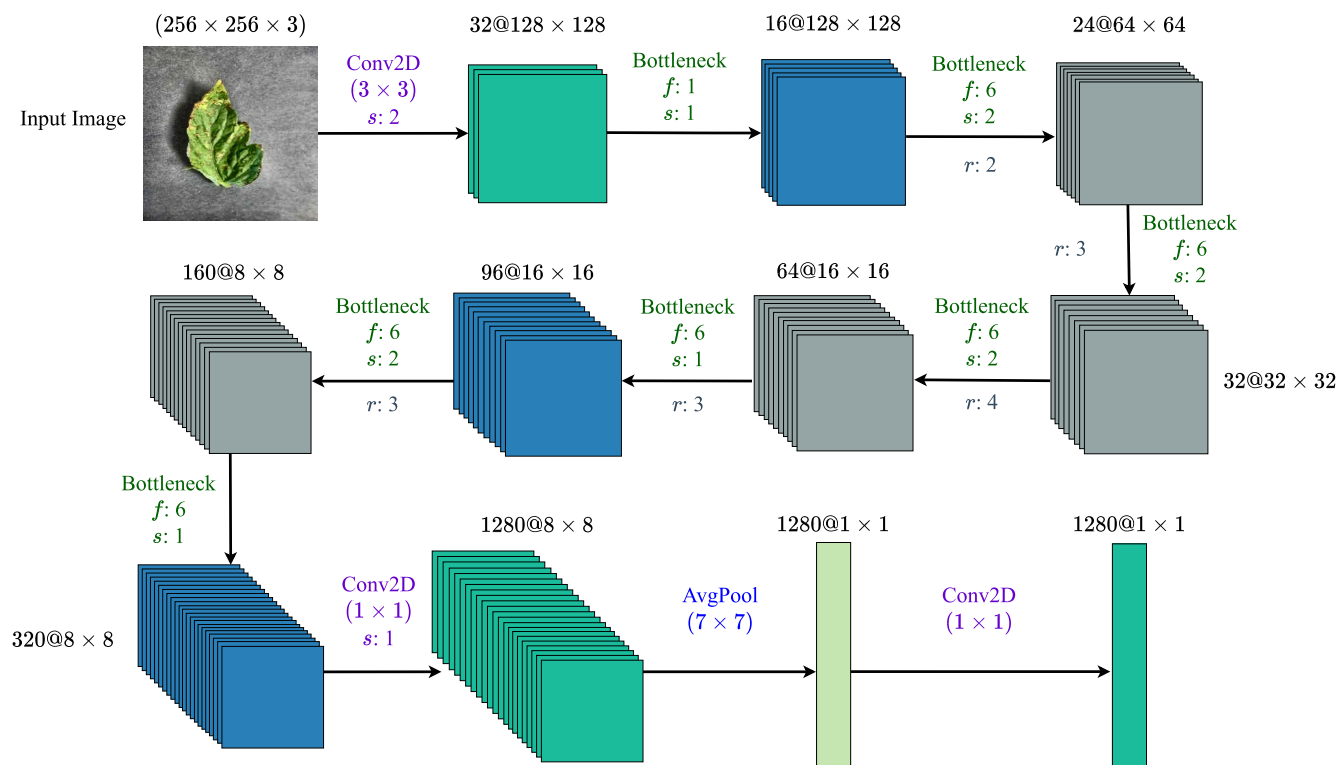


FIGURE 6. MobileNetV2 architecture adopted from [11] and modified for extracting features from $256 \times 256 \times 3$ tomato leaf images. Each box represents the feature maps (not to scale) after going through different layers. Here, f denotes the expansion factor of each Bottleneck Layer. The first layer of each sequence has a stride value of s , and the remaining use stride 1. r denotes the number of times a layer is repeated to produce the next feature map.

image separately in different epochs. This reduces the possibility of overfitting, as it cannot see the same image in every epoch. On the other hand, this ensures that the different variations of the same image do not appear in both the training and the test set, thus eliminating the data leaking problem persistent in the existing literature.

C. TRANSFER LEARNING-BASED FEATURE EXTRACTOR

Earlier machine learning approaches assumed that the training and test data must be in the same feature space. However, recent advances in deep learning approaches have facilitated the use of an architecture trained to extract features on the training data of one domain to be used as a feature extractor for another domain. As the feature extractors in deep learning-based tasks became more and more generalized, this method of knowledge transfer, also known as Transfer Learning, has significantly improved the performance of learning, reducing a considerable amount of computational complexity. In this connection, the MobileNetV2 architecture has enabled real-time applications across multiple tasks and benchmarks using low computational resources. As shown in Figure 6, MobileNetV2 consists of a regular 3×3 convolution with 32 filters, followed by 17 Bottleneck Residual Blocks, a Pointwise convolution layer, a global average pooling layer, and a classification layer. The classification layer usually corresponds to the number of classes of the original dataset. For our system, the classification layer was replaced with a classifier network to classify tomato diseases.

At the heart of the MobileNetV2 architecture resides Bottleneck Residual Block containing three convolutional layers (Fig. 7a). The Expansion Layer increases the number of channels in the input data by performing Pointwise convolution based on an expansion factor. The feature map output by this layer is then fed to a 3×3 Depthwise Convolution layer which works as a filter by applying convolution per channel. The Projection Layer takes these filtered values to generate salient features. Besides, this layer projects the higher dimensional data into a much lower number of dimensions, reducing the number of channels. The Depthwise Convolution layer combined with the Pointwise Convolution performs Depthwise Separable Convolution, reducing the computation by a factor of $O(k^2)$ compared to regular convolutions. Here, k is the size of the Depthwise convolution kernel. Like most modern architectures, each of the three convolution layers is followed by batch normalization to stabilize the learning process. The activation function used by these layers is ReLU6. It bounds the activation within $[0, 6]$, making it more robust than the well-known ReLU function in fixed-point arithmetic. However, the Projection Layer does not contain any activation function due to the low dimensionality of the data produced by this layer. The non-linearity of the ReLU activation function can destroy valuable features. In addition, to reduce the effect of diminishing gradients, inverted residual connections are introduced through the network, which connects the bottleneck blocks with the same number of channels (Fig. 7b).

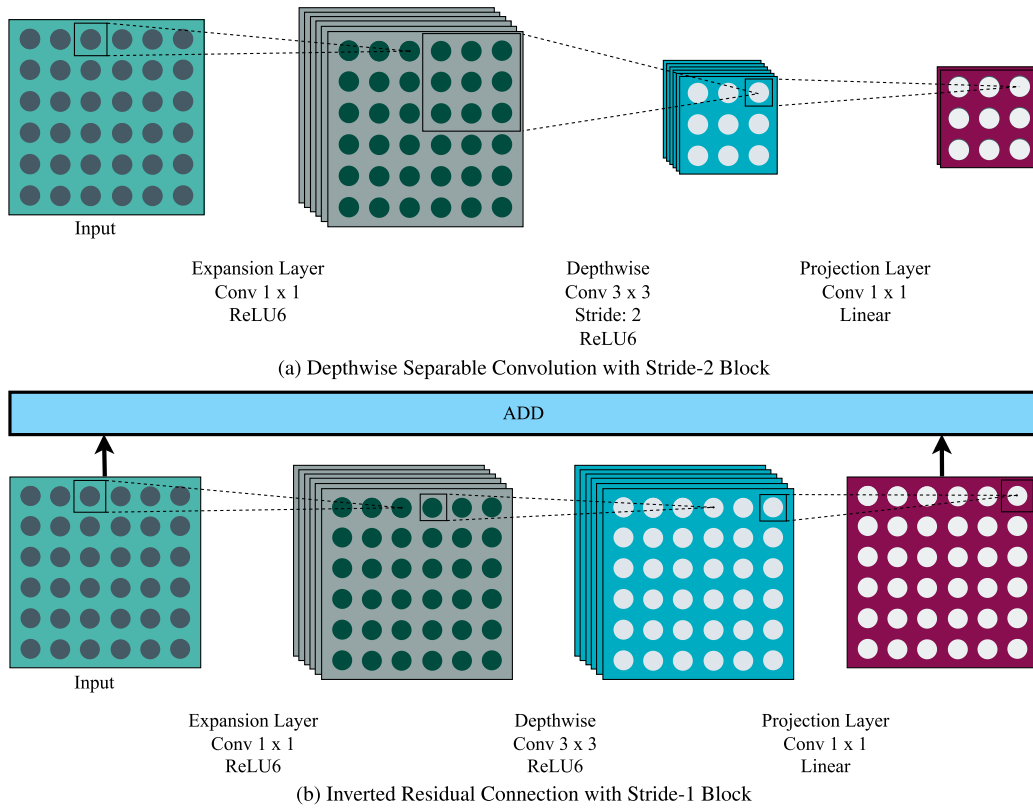


FIGURE 7. Bottleneck residual block. Here, each block represents the feature map output by different layers.

In this work, to compare the performance of MobileNetV2, other transfer-learning architectures that are popular in leaf disease detection: DenseNet121, DenseNet201, EfficientNet-B0, MobileNet, NASNet-Mobile, ResNet50, ResNet152V2, and VGG19 were used.

D. CLASSIFIER NETWORK

Instead of directly using the extracted features from pre-trained models for final prediction, we employed a combination of dense, dropout, and batch normalization blocks to fine-tune the extracted traits further. As shown in Figure 8, dense blocks were added before the final output layer. The pretrained MobileNetV2 architecture we used was trained on a large and generalized dataset, making it perfect for feature extraction. The features extracted from the leaf images by MobileNetV2 architecture are then fed into the dense blocks trained from scratch to extract further the relevant features required to classify the diseases. Upon experimentation with different numbers of layers consisting of a varying number of nodes, two dense blocks with 128 and 64 nodes helped us achieve global minima in terms of loss.

A Batch Normalization [53] block was added between the output of the MobileNetV2 and the first densely connected block and one between the second densely connected block and the output layer. A batch normalization block is used to standardize the inputs for the final layer for each mini-batch and stabilize the whole learning process, reducing

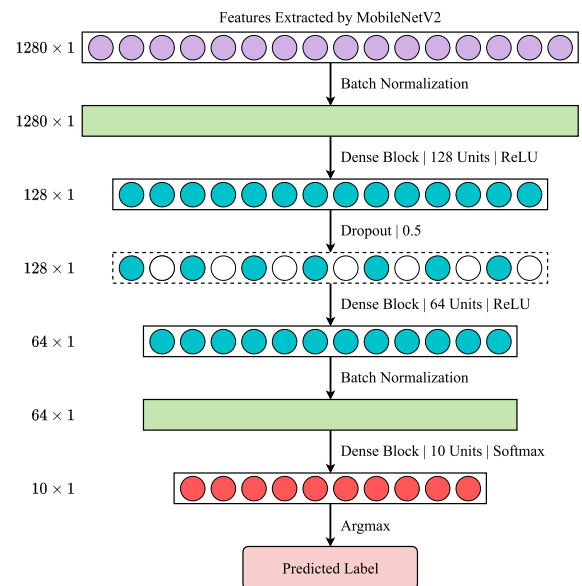


FIGURE 8. Classifier network.

the epochs needed to train the network. Rectified Linear Unit (ReLU) [54] was used as the activation function of the two densely connected blocks. This activation function makes the models easier to optimize and more generalizable. A dropout layer [55] in-between these two dense blocks work as a

regularizer, ensuring that the model does not overfit. The final output layer of the classifier network is also a densely connected block with a Softmax activation function [56]. The output layer of the classifier network contains ten nodes corresponding to each class label. The value of each node represents the probability of the input sample being in that class. Applying Argmax on this layer provides us with the predicted class label.

E. EXPERIMENTAL SETUP

The proposed architecture was trained under a Python environment with TensorFlow, Keras, and other necessary libraries in Google Colab.² All experiments were conducted using an Intel Xeon CPU with a base clock speed of 2.3 GHz and an NVIDIA Tesla T4 GPU with a VRAM of 15 GB. The total usable memory of the machine was 13 GB.

From each class, the sample images were randomly split into 60% for training, 20% for validation, and 20% for the test set. Following the mini-batch gradient descent technique [57], the batch size was selected as 16. Since smaller batch sizes are often noisy, they help create a regularization effect and reduce the generalization error. They also help fit training data into memory.

While working with mini-batches, there is always a possibility of choosing batches that are not representative of the entire dataset, resulting in an inaccurate estimate of the gradient. Thus the training images were shuffled after each epoch throughout the experiments. This increases the probability that the model will converge and not be trapped with too many inferior batches.

For experimenting with mini-batches, it is crucial to shuffle the data after each epoch. The goal of shuffling data is to reduce variance and ensure that models generalize well and do not overfit.

The model was trained for at most 1000 epochs with early stopping. Early stopping helps reduce overfitting and improves the generalization of neural networks. Validation accuracy was selected as the scheme for evaluating the model so that early stopping can be triggered. In our proposed approach, a change in validation accuracy between epochs was considered significant if it was greater than 10^{-4} . Otherwise, it was considered a patient epoch. The training was stopped early if there are ten consecutive patient epochs. Consequently, it was observed that our proposed architecture was able to converge after 70 epochs on average.

To ensure the rapid learning of salient features, we have used Adam optimizer [58] for training our model. Compared to other optimizers, Adam can help multilayer deep learning networks converge faster for computer vision problems. Since this is a multiclass classification task, we have used categorical cross-entropy loss. The initial learning rate was set to 10^{-5} . For every four consecutive patient epochs, the learning rate was decreased by a factor of 0.1 to help the model learn a set

of globally optimal weights that leads to better optimization of the loss function.

The models can be initialized with different weights during the training phase, e.g., 0, random values, or pretrained weight values. In our work, we initialized the feature extractor part of the network with the respective pretrained weights from ImageNet Dataset [59] for the models and the classifier network with random weights. Model Checkpoints were used to save the model with the best validation accuracy so that they can be loaded later to continue the training from the saved state if required.

F. EVALUATION METRICS

1) ACCURACY

Accuracy is the ratio between the total number of predictions that were correct and the total number of predictions. To get a better estimation of the generalization capability of a model, the accuracy is calculated using the samples from the test set, which is unseen to the model during training.

$$\text{Accuracy} = \frac{M}{N} \times 100\% \quad (1)$$

Here, N is the number of samples in the test set and M is the number of samples for which the class labels were correctly predicted by the model.

2) PARAMETER COUNT

Parameters are the model's learnable weights, which are changed during the backward propagation phase based on the chosen optimization algorithm. The number of parameters can not only provide us with an idea regarding the training time of the model but also helps determine the model size and inference time.

$$\text{Parameter Count} = \sum_{i=1}^L p_i \quad (2)$$

Here, p_i is the number of parameters in the i th layer and L is the total number of layers in the model.

3) MODEL SIZE

Trained models can be stored as a Hierarchical Data Format version 5 (HDF5) file. The saved model contains the model's configuration, trained weights, and optimizer state. The model, along with its saved weights, can be loaded again to run inference. The size of the saved model is called the model size. Model size can be measured in MB (Megabyte) or GB (Gigabyte).

4) FLOPs COUNT

FLOPs Count is the theoretical maximum number of floating-point operations that a model requires to perform inference. Since the time taken for inference can vary from device to device, FLOPs Count is a better measurement to compare the relative inference time of deep learning models. It is usually measured in megaFLOPs (MFLOPs), gigaFLOPs

²<https://colab.research.google.com/>

(GFLOPs), or teraFLOPs (TFLOPs). The higher the value, the larger the number of computations required for a model to perform inference.

5) PRECISION

Precision is the ratio of the sum of the number of true positive predictions among all classes and the sum of the number of true positive predictions and false-positive predictions among all classes. In a multiclass problem, for each class, precision is used to evaluate the correctly classified samples of that class among all the samples that were classified as of that class. Precision is also called Positive Predictive Value (PPV).

Precision for each class c can be calculated considering the one-vs-all strategy.

$$\text{Precision}_c = \frac{TP_c}{TP_c + FP_c} \quad (3)$$

Here, TP_c is the number of samples correctly classified as c , and FP_c is the number of samples wrongly classified as c .

For imbalanced classes, macro-average precision is calculated where the precision for each class is calculated separately, and their average is taken. This ensures that the model gets equally penalized for each false positive instance of any class.

For a set of classes C ,

$$\text{Macro Average Precision} = \frac{\sum_{c \in C} \text{Precision}_c}{|C|} \quad (4)$$

Here, Precision_c is the precision value for class c , and $|C|$ is the total number of classes.

6) RECALL

Recall is the ratio of the sum of the number of true positive predictions among all classes and the sum of the number of true positive predictions and false-negative predictions among all classes. In a multiclass problem, recall is used to evaluate how many samples are correctly classified among all the samples that should have been classified as of that class. Recall is also called Sensitivity.

Recall for each class c can be calculated considering the one-vs-all strategy.

$$\text{Recall}_c = \frac{TP_c}{TP_c + FN_c} \quad (5)$$

Here, TP_c is the number of samples correctly classified as c , and FN_c is the number of samples of c that are wrongly classified as other classes.

For imbalanced classes, macro average recall is calculated where the recall for each class is calculated separately and their average is taken. This ensures that the model gets equally penalized for each false-negative instance of any class.

For a set of classes C ,

$$\text{Macro Average Recall} = \frac{\sum_{c \in C} \text{Recall}_c}{|C|} \quad (6)$$

Here, Recall_c is the Recall value for class c , and $N = \text{Total number classes}$.

7) F1-SCORE

F1-Score is the weighted average of precision and recall that considers both the number of false positive predictions and false negative predictions. While working on an imbalanced dataset, having a high F1-Score is crucial to reduce the number of false positive and false negative predictions.

F1-Score for each class c can be calculated using the following formula:

$$\text{F1-Score}_c = \frac{2 \times P_c \times R_c}{P_c + R_c} \quad (7)$$

Here, P_c is the precision value for class c , and R_c is the recall value for class c .

For imbalanced classes, the macro average F1-score is calculated where the F1-score for each class is calculated separately and their average is taken. This ensures that each class gets equal priority in classification.

For a set of classes C ,

$$\text{Macro Average F1-Score} = \frac{\sum_{c \in C} F_c}{|C|} \quad (8)$$

Here, F_c is the F1-Score for class c , and $|C|$ is the total number of classes.

8) AUC-ROC SCORE

Precision, recall, and the majority of the commonly utilized metrics have their individual set of restrictions. Precision is a measurement that determines how accurate a classification task is and it is only based on true positive and false positive predictions; a score of 1.0 for precision can be achieved with just one true positive prediction. On the other hand, recall is all about completeness and is solely based on true positive and false negative responses. As a result, predicting all the samples as positive will result in a recall of 1.0, but the precision will be very low. The Receiver Operating Characteristic (ROC) curve and the area under the ROC curve (AUC-ROC) are utilized as evaluation methods in this regard by combining the True Positive Rate (TPR) and False Positive Rate (FPR).

These methods allow models to be evaluated according to how well they separate classes from one another. The FPR and TPR for a series of predictions generated by the model at various thresholds are calculated to summarize the behavior of the model and can also be used to examine its ability to differentiate classes. Each probability threshold is represented by a point, linked to form a curve in the ROC graph. A model with no discriminating power is depicted by a diagonal line from FPR 0 and TPR 0 (coordinates: 0, 0) to FPR 1 and TPR 1. (co-ordinates: 1, 1). A perfect model is represented as a point in the upper-left corner of the plot.

TABLE 2. Comparison of the performance and characteristics among the baseline architectures on the original dataset.

Architecture	Accuracy (%)	Parameters Count (Millions)	Model Size (MB)	FLOPs Count (MFLOPs)
DenseNet121	97.96	7.1	27.58	14.1
DenseNet201	99.36	18.35	71.11	36.69
EfficientNet-B0	96.94	4.1	15.89	8.1
MobileNet	96.53	3.2	12.51	6.5
MobileNetV2	97.27	2.28	8.98	4.54
NASNet-Mobile	97.21	4.3	17.53	8.6
ResNet50	98.70	23.62	98.29	51.11
ResNet152V2	98.62	58.36	223.52	116.61
VGG19	99.48	20.02	76.48	40.05

IV. RESULT AND DISCUSSION

For our experiments, we first investigated the performance of different baseline Deep CNN architectures to choose the best fit for our requirements. After that, an ablation study was conducted to justify how the different considerations in our proposed pipeline and modifications over the baseline contributed to improving our model's performance. Next, we inspected per-class precision, recall, and F1-score to evaluate how the proposed architecture addresses the class imbalance issue. Then, we compared the performance of our model with the existing state-of-the-artwork of tomato leaf disease classifications to establish its superiority. Finally, an error analysis was conducted to figure out where to invest the future improvement efforts.

A. PERFORMANCE OF DIFFERENT BASELINE ARCHITECTURES

To choose the baseline model, several state-of-the-art Deep CNN architectures were implemented to perform tomato leaf disease classification. A comparison of their performance is shown in Table 2.

The models were initialized with their pretrained weights on the ImageNet dataset and fine-tuned using the original tomato leaf samples from the PlantVillage dataset. The benefit of this initialization was that the models were already capable of learning complex patterns leading to faster convergence. Since our goal was to pick the best-suited baseline for the proposed system, we only changed the final softmax layer with the number of classes of our dataset and trained without any enhancement or augmentations.

While choosing the appropriate architecture, we have considered the accuracy, number of trainable parameters, estimation of the number of floating-point operations (FLOPs), and the model's size. The VGG19 and DenseNet201 architectures achieved an accuracy higher than 99% percent, and the performance of the ResNets also came close. These models are superior in terms of accuracy but have a significant disadvantage considering the other metrics. For example, the VGG19 model has achieved 99.48% accuracy, which is 2.2% higher

TABLE 3. Ablation study of different components of the proposed pipeline.

CLAHE	Augmentation	Classifier Network	Accuracy
×	×	×	97.27
×	×	✓	98.29
×	✓	×	98.46
×	✓	✓	99.03
✓	×	×	97.71
✓	×	✓	98.60
✓	✓	×	98.84
✓	✓	✓	99.30

than the accuracy of MobileNetV2 architecture. However, this improvement is costly in terms of memory and inference time. The model consumed 8.5 times the storage space and 8.8 times higher FLOPs count than MobileNetV2. Similar can be said for DenseNet201 as well. On the other hand, the relatively lighter models such as EfficientNet-B0, MobileNet, and NASNet-Mobile had lower accuracy than MobileNetV2 despite having higher values in terms of other metrics.

The MobileNetV2 architecture has the smallest model size and the lowest FLOPs count, making it ideal for real-time disease detection in devices with memory constraints. In addition to that, the fewer parameters of MobileNetV2 architecture result in faster training and inference. For these reasons, we chose MobileNetV2 as our base transfer learning architecture. We further aimed to improve the baseline performance by utilizing preprocessing techniques and an additional classifier network.

B. ABLATION STUDY

An ablation study was conducted to understand the contribution of different components of the proposed pipeline to the overall performance. We considered several combinations of the design choices like the preprocessing steps, such as CLAHE, data augmentation, and the introduction of a classifier network to analyze their effects. A summary of the result in different settings can be found in Table 3.

A positive impact can be seen on the results when the images are preprocessed using CLAHE. This can be attributed to CLAHE for enhancing the leaf images' disease spots, making them more prominent and easier to identify for the models. For example, the baseline performance of 97.27% was improved to 97.71% after we introduced CLAHE. We noticed a further improvement in the results when data augmentation was introduced. The runtime augmentations allow the model to learn from different representations of the images in every epoch, allowing the model to focus on the features highlighted by CLAHE.

Experiments were performed to find out how data augmentation in different splits affects the overall performance. We found that performing data augmentation on all three splits resulted in the best accuracy. As a result of the augmentation, the model learns to recognize different variations of the original image during the training phase. However, without augmenting the test set, no such variations are to be found. This violates the key assumption in dataset splitting for general classification tasks, that the distribution of images found in the training and validation set should be similar to the distribution in the test set.

One key factor here is that, as all our samples are being augmented with random probability during run-time, the model never sees the same version of an image twice. Augmentation in training and validation splits ensures that the model hardly gets any chance to overfit and learns generic feature representations. In addition, augmentations performed on the test set ensure that those samples represent real-life scenarios, making the classification task even more challenging. However, this begets a problem. Since the augmentation is performed randomly, the model sees different images in each test run. As a result, the accuracy for each run might not be the same; instead, it gives us a value within a range. To resolve this issue, we tested the trained model 100 times whenever we used augmentation and reported the average accuracy. The benefits of doing this are two-fold. First, as the test set is randomly augmented, the average accuracy is a better descriptor of the model's performance, preventing any chance of getting lucky. Moreover, these trials are testing the model with a variety of samples, more than what could be done using a static test set. So a model being able to do well in this setup will be robust and can be expected to achieve similar accuracy in real-life scenarios. It is worth mentioning that the maximum accuracy achieved by our best model was 99.53%.

Our hypothesis of introducing the classifier network was that the model would be able to consider further combinations of the extracted features from the MobileNetV2 network, leading to improved overall performance. Since this network was trained from scratch on the provided information from the feature extractor network, it extracted even more meaningful features for leaf disease classification. Thus we found an improvement in the overall accuracy every time the classifier network was introduced in different setups.

Initially, the performance of the baseline MobileNetV2 model was only 97.27%. The combination of the

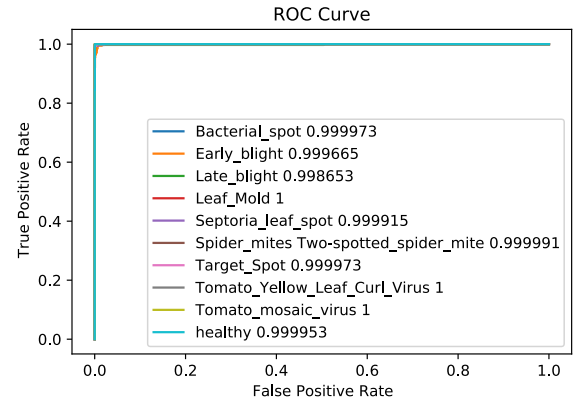


FIGURE 9. ROC curve for the tomato leaf diseases.

preprocessing techniques increased it up to 98.84% showing how these choices improve the generalizing capability of the model. Finally, the model's competence was further enhanced with the classifier network leading to a mean accuracy of 99.30% (Standard Deviation: 0.00095) over 100 runs.

C. ADDRESSING THE CLASS IMBALANCE

As mentioned earlier, there exists a class imbalance in the PlantVillage dataset. Thus drawing a conclusion on a model's performance solely based on the accuracy metric might be unwise as the accuracy might still be in the 90th percentile even if the model is incapable of classifying half of the samples of the least populated classes. To tackle this issue, macro-averaged precision, recall, and F1-score values were taken under consideration, which gives equal importance to all the classes regardless of the number of samples. Our proposed model achieves 99.18 precision, 99.07 recall, and 99.12 F1-score. The high values of precision and recall signify that our model does a great job identifying the True Positives. At the same time, it penalizes the accidental False Positive and False Negative cases. Taking a harmonic mean of these two metrics, the 99.12 value of the F1-score proves the robustness of the proposed architecture even in imbalanced datasets.

Furthermore, Table 4 shows the precision, recall, and F1-score for each class. From the table, it is evident that our data augmentation technique solved the class imbalance problem as these values are high even for the classes with a low number of samples.

D. CLASS SEPARABILITY

The AUC-ROC curve can assess the performance of a predictive model by describing the trade-off between the True Positive Rate and False Positive Rate by employing multiple probability thresholds. A perfect classifier will have a ROC where the graph reaches 100% true positives and zero false positives. We generally measure the number of positive classifications with an increment in the rate of false positives.

As shown in Figure 9, the ROC curves overall each other at the top-left corner. That means, our proposed architecture

TABLE 4. Per class precision, recall, F1-Score for the test set.

Class Label	Sample Count	Precision	Recall	F1-Score
Bacterial Spot	425	0.9976	0.9953	0.9965
Early Blight	200	0.9745	0.9550	0.9646
Late Blight	381	0.9794	0.9974	0.9883
Leaf Mold	190	1.000	0.9895	0.9947
Septoria Leaf Spot	354	0.9915	0.9915	0.9915
Two-spotted Spider Mite	335	0.9852	0.9940	0.9896
Target Spot	281	0.9928	0.9858	0.9893
Yellow Leaf Curl Virus	1071	1.0000	0.9981	0.9991
Tomato Mosaic Virus	74	1.0000	1.0000	1.0000
Healthy	318	0.9969	1.0000	0.9984

achieves a commendable performance for all 10 classes. Among the classes, our model achieved an AUC score of 1 for Leaf Mold, Yellow Leaf Curl Virus, and Mosaic Virus. Scores for the other classes are also fairly high, indicating a satisfactory class separability.

E. COMPARISON WITH STATE-OF-THE-ART METHODS

Table 5 presents a comparison of our proposed architecture with the state-of-the-art models of tomato leaf disease classification. Our model achieves a commendable accuracy of 99.30% while keeping the model size and the number of operations low. Comparing it to the state-of-the-art models, we can notice that only [40] achieved a mere 0.09% increase in accuracy, having 2.4 times the model size and 59.27% increase in FLOPs count. Our model's smaller size and low computational cost without sacrificing performance make it suitable for low-end devices.

Some of the works mentioned in the table did not utilize all the samples from the subset of the PlantVillage dataset, which leaves a possibility of accidentally missing some critical samples [41], [43], [44], [47]. Additionally, some of the models did not consider all the classes, which might lead to the misclassification of unseen samples. For example, [37] achieved an accuracy of 99.18%, but the experiment did not include any healthy samples of tomato leaves. This results in labeling a healthy leaf sample to any of the disease classes.

Further analysis shows that the space requirement of our proposed architecture is only 9.6MB. In contrast, different works in the existing literature required at least twice of this storage space, if not more, to produce similar accuracy (Fig. 10a). Although [45] has a smaller model size than that of ours, the accuracy is far less. Fig. 10b shows that our model significantly reduced the FLOPs requirement without compromising the accuracy. Hence it removes the requirement for high-performance hardware along with reducing the inference time of the model. It can be observed that despite using deeper models, some works could not achieve comparable performance to the state-of-the-art. This further

justifies the usefulness of the different components of our proposed architecture.

F. QUALITATIVE ANALYSIS

1) FOCUSING ON THE DISEASED PORTION

To get an intuitive understanding of whether our model is learning to correctly predict considering the relevant features or not, we examine the Gradient-weighted Class Activation Mapping (GradCAM) output for the correctly classified samples [60] (Figure 11). This visualization can show us how our model classifies a diseased or healthy sample by highlighting the region in which our model focuses while making the class label decision.

As shown in Fig. 11(a) and 11(b), our model focused on the diseased portion of the leaf image to provide the class decision. Conversely, in the case of a healthy leaf (Fig. 11(c)), the model focuses on the entire leaf image finding no diseased portion and then classifies the image as healthy. This goes to show the capability of our model in understanding what to look for in a leaf image to make the correct class decision.

2) ERROR ANALYSIS

According to the confusion matrix of our best performing model (Figure 12), for half of the classes, our model was able to predict all the unseen test samples correctly. For the rest, the accuracy is comparable to other state-of-the-art methods. However, the most misclassified samples were from the 'Early Blight' class. A few of the misclassified samples from this class were predicted as 'Late Blight'. Upon reviewing the misclassified samples, we identified visually similar leaves from both classes. For example, in the original dataset, the class label for Fig. 13(a) is 'Early Blight', which was misclassified to 'Late Blight' class during inference. The GradCAM output for the sample shows that our model correctly focuses on the diseased region of the leaf image (Fig. 13(a)).

However, in the training set of the Late Blight class, there are several images (e.g. Fig. 13(b)) that are similar to Fig. 13(a). Since during training, the model learns to classify

TABLE 5. Performance comparison with the State-of-the-Art models for tomato leaf disease classification.

Reference	Class Count	Image Count	Accuracy (%)	Model Size (MB)	FLOPs Count (MFLOPS)
Durmuş et al. [45]	10	N/A	94.30	2.94	1.44
Brahimi et al. [37]	9	14828	99.18	23.06	11.95
Tm et al. [46]	10	18160	94.85	156.78	82.18
Rangarajan et al. [42]	7	13262	97.49	350.25	183.53
Zhang et al. [43]	9	5550	97.28	98.29	51.1
Bir et al. [47]	10	15000	98.60	15.59	8.11
Maeda-Gutierrez et al. [40]	10	18160	99.39	23.06	11.95
Wu et al. [41]	5	5300	94.33	23.06	11.95
Abbas et al. [44]	10	16012	97.11	27.58	28.09
Proposed Method	10	18160	99.30	9.60	4.87

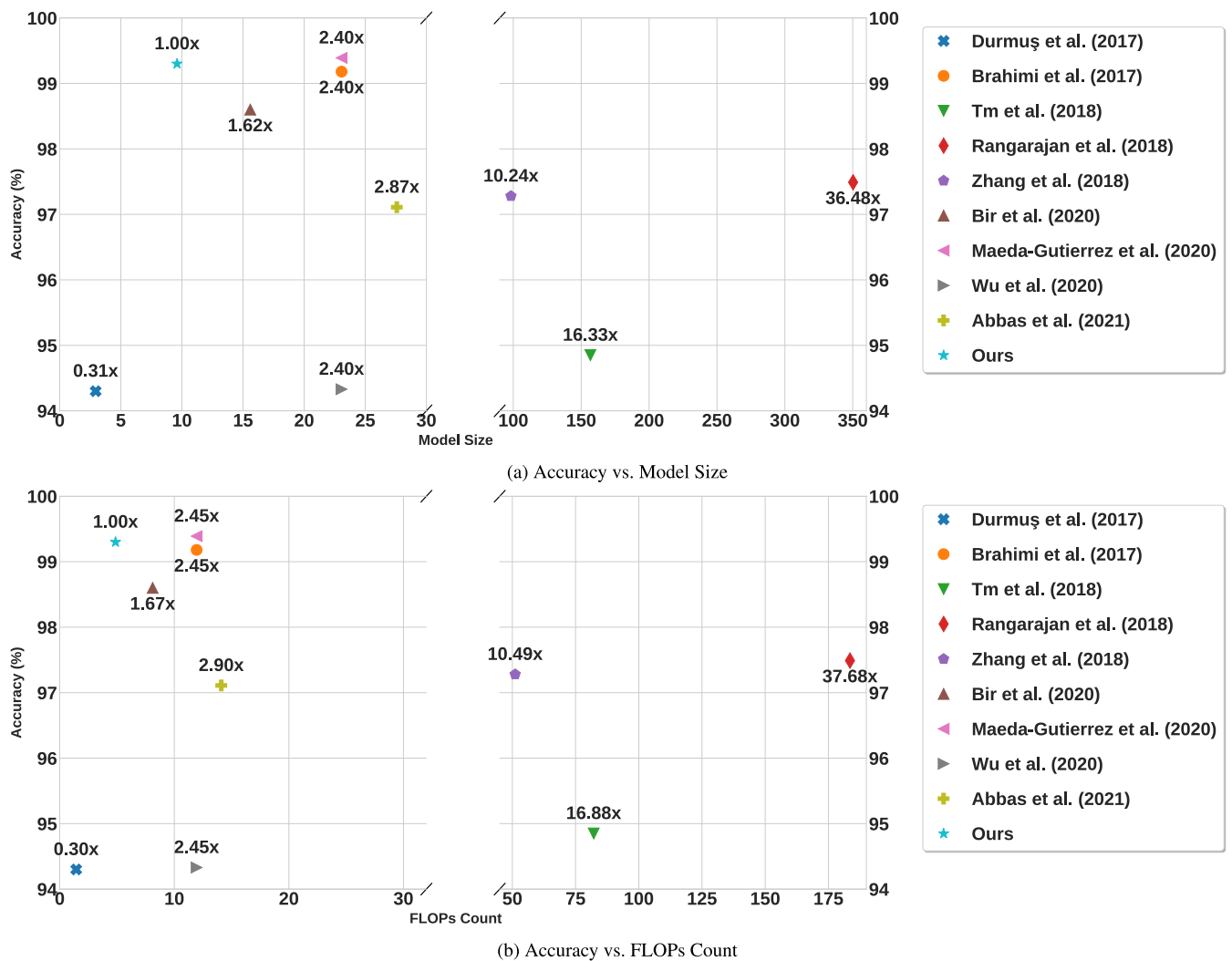


FIGURE 10. Performance comparison with state-of-the-art tomato leaf disease classification architectures based on model size and FLOPs count.

these images as of the class ‘Late Blight’, it is expected that similar images from the test set will also be classified in the same class.

To conclude, after analyzing the misclassified samples, we found some inter-class similarities in the infected regions among some of the diseases. A few of the leaves were

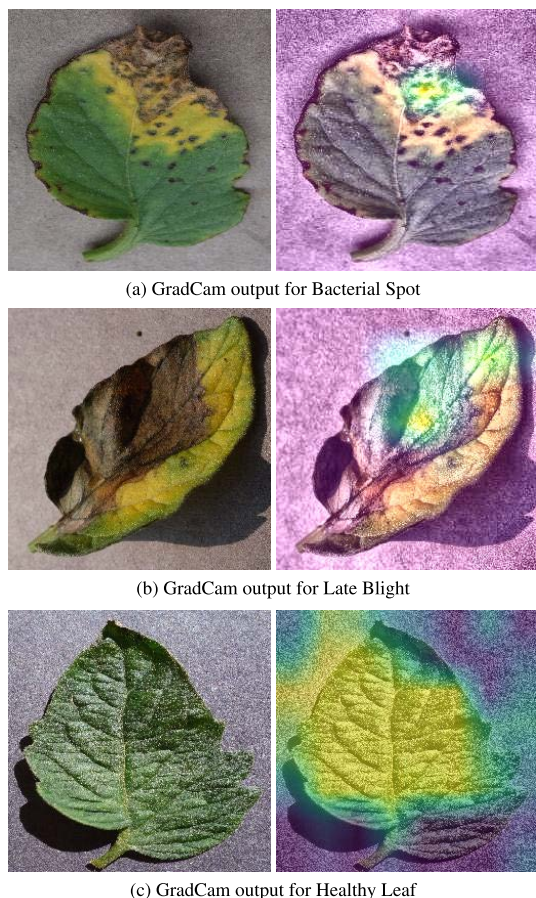


FIGURE 11. Qualitative results showing GradCam output for correctly classified samples.

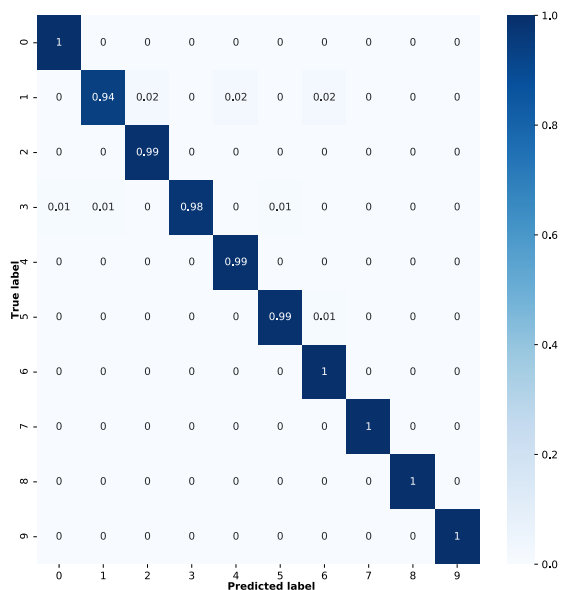


FIGURE 12. Confusion matrix. Here, 0 = Bacterial spot, 1 = Early blight, 2 = Late Blight, 3 = Leaf Mold, 4 = Septoria Leaf Spot, 5 = Two-spotted Spider Mite, 6 = Target Spot, 7 = Yellow Leaf Curl Virus, 8 = Tomato Mosaic Virus, 9 = Healthy.

severely damaged by the virus, which eventually restricted the model from extracting meaningful features leading to misclassification.



FIGURE 13. Misclassified sample with visually similar samples of the predicted class.

V. CONCLUSION

Fast and accurate recognition of leaf diseases can go a long way to meeting the ever-increasing demand in food production. In this regard, we have proposed a lightweight deep neural network by combining a fine-tuned pretrained model and a classifier network. The utilization of the adaptive contrast enhancement technique has eliminated the illumination problem persistent in the dataset. Runtime data augmentation techniques have been applied to address the class imbalance issue while avoiding data leakage. All these components of the pipeline enabled the model to focus on the disease spots and extract high-level features leading to an accuracy of 99.30%. We achieved this performance with a significantly smaller model size and FLOPs count compared to the state-of-the-art models. This makes the proposed pipeline a suitable choice for building applications for low-end devices.

However, one of the limitations of the PlantVillage dataset is that the samples are taken in laboratory conditions. Further experiments can be performed with tomato leaf images with varying backgrounds taken from the field. Such images might contain occlusion and background clutter. Advanced segmentation techniques can be taken into account to locate the infected regions before classification. Moreover, only a single disease can be found in each of the samples used in our experiment. Identifying multiple diseases within a single leaf will be another challenging task to solve. The classification goal can also include detection of the severity of infection on leaves, which intelligent systems can utilize to decide the amount of pesticide to be used. Finally, this work can be extended to classify diseases from a broader range of crops.

ACKNOWLEDGMENT

The authors would like to thank A.B.M. Ashikur Rahman, Mohammad Ishrak Abedin, and Shahriar Ivan, Department of Computer Science and Engineering, Islamic University of Technology, and Nura Jahan, Department of Entomology, Bangabandhu Sheikh Mujibur Rahman Agricultural University for their valuable time and support in making this study possible.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

- [1] Tridge Co., Ltd. (2020). *Global Production of Tomato*. Accessed: Aug. 12, 2021. [Online]. Available: <https://www.tridge.com/intelligences/tomato/production>
- [2] S. Panno, S. Davino, A. G. Caruso, S. Bertacca, A. Crnogorac, A. Mandić, E. Noris, and S. Matic, "A review of the most common and economically important diseases that undermine the cultivation of tomato crop in the Mediterranean basin," *Agronomy*, vol. 11, no. 11, p. 2188, Oct. 2021. [Online]. Available: <https://www.mdpi.com/2073-4395/11/11/2188>
- [3] L. Li, S. Zhang, and B. Wang, "Plant disease detection and classification by deep learning—A review," *IEEE Access*, vol. 9, pp. 56683–56698, 2021. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9399342>
- [4] K. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine learning in agriculture: A review," *Sensors*, vol. 18, no. 8, p. 2674, 2018. [Online]. Available: <https://www.mdpi.com/1424-8220/18/8/2674>
- [5] A. Kamilaris and F. X. Boldú, "Deep learning in agriculture: A survey," *Comput. Electron. Agricult.*, vol. 147, pp. 70–90, Apr. 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0168169917308803>
- [6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, vol. 25, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Red Hook, NY, USA: Curran Associates, 2012. [Online]. Available: <https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html>
- [7] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269. [Online]. Available: <https://ieeexplore.ieee.org/document/8099726>
- [8] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Mach. Learn.*, in Proceedings of Machine Learning Research, vol. 97, K. Chaudhuri and R. Salakhutdinov, Eds., Jun. 2019, pp. 6105–6114. [Online]. Available: <http://proceedings.mlr.press/v97/tan19a.html>
- [9] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9. [Online]. Available: <https://ieeexplore.ieee.org/document/7298594/>
- [10] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.
- [11] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 4510–4520. [Online]. Available: <https://ieeexplore.ieee.org/document/8578572>
- [12] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8697–8710. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2018/html/Zoph_Learning_Transferable_Architectures_CVPR_2018_paper.html
- [13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778. [Online]. Available: <https://ieeexplore.ieee.org/document/7780459>
- [14] F. N. Iandola, M. W. Moskewicz, K. Ashraf, S. Han, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <1MB model size," 2016, *arXiv:1602.07360*.
- [15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, Y. Bengio and Y. LeCun, Eds., 2015.
- [16] L. Torrey and J. Shavlik, "Transfer learning," in *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*. Hershey, PA, USA: IGI Global, 2010, pp. 242–264. [Online]. Available: <https://www.igi-global.com/chapter/transfer-learning/36988>
- [17] D. P. Hughes and M. Salathé, "An open access repository of images on plant health to enable the development of mobile disease diagnostics," 2015, *arXiv:1511.08060*.
- [18] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers Plant Sci.*, vol. 7, p. 1419, Sep. 2016. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fpls.2016.01419/full>
- [19] Z. Rehman, M. A. Khan, F. Ahmed, R. Damasevicius, S. R. Naqvi, M. W. Nisar, and K. Javed, "Recognizing apple leaf diseases using a novel parallel real-time processing framework based on MASK RCNN and transfer learning: An application for smart agriculture," *IET Image Process.*, vol. 15, no. 10, pp. 2157–2168, 2021. [Online]. Available: <https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.12183>
- [20] O. O. Abayomi-Alli, R. Damaševičius, S. Misra, and R. Maskeliūnas, "Cassava disease recognition from low-quality images using enhanced data augmentation model and deep learning," *Expert Syst.*, vol. 38, no. 7, p. e12746, 2021. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/exsy.12746>
- [21] A. Waheed, M. Goyal, D. Gupta, A. Khanna, A. E. Hassani, and H. M. Pandey, "An optimized dense convolutional neural network model for disease recognition and classification in corn leaf," *Comput. Electron. Agricult.*, vol. 175, Aug. 2020, Art. no. 105456. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0168169920302180>
- [22] S. Zhang, S. Zhang, C. Zhang, X. Wang, and Y. Shi, "Cucumber leaf disease identification with global pooling dilated convolutional neural network," *Comput. Electron. Agricult.*, vol. 162, pp. 422–430, Jul. 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0168169918317976>
- [23] B. Liu, C. Tan, S. Li, J. He, and H. Wang, "A data augmentation method based on generative adversarial networks for grape leaf disease identification," *IEEE Access*, vol. 8, pp. 102188–102198, 2020. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9104723/>
- [24] X. Zhang, Y. Qiao, F. Meng, C. Fan, and M. Zhang, "Identification of maize leaf diseases using improved deep convolutional neural networks," *IEEE Access*, vol. 6, pp. 30370–30377, 2018. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8374024>
- [25] U. P. Singh, S. S. Chouhan, S. Jain, and S. Jain, "Multilayer convolution neural network for the classification of mango leaves infected by anthracnose disease," *IEEE Access*, vol. 7, pp. 43721–43729, 2019. [Online]. Available: <https://ieeexplore.ieee.org/document/8675730>
- [26] Y. Lu, S. Yi, N. Zeng, Y. Liu, and Y. Zhang, "Identification of rice diseases using deep convolutional neural networks," *Neurocomputing*, vol. 267, pp. 378–384, Dec. 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0925231217311384>
- [27] U. Mokhtar, N. El-Bendary, A. E. Hassenian, E. Emary, M. A. Mahmoud, H. A. Hefny, and M. F. Tolba, "SVM-based detection of tomato leaves diseases," in *Intelligent Systems'2014: Proceedings of the 7th IEEE International Conference Intelligent Systems IS'2014, September 24-26, 2014, Warsaw, Poland, Volume 2: Tools, Architectures, Systems, Applications* (Advances in Intelligent Systems and Computing), vol. 323, D. P. Filev, J. Jablkowski, J. Kacprzyk, M. Krawczak, I. Popchev, L. Rutkowski, V. S. Sgurev, E. Sotirova, P. Szykarczyk, and S. Zadrozny, Eds. Cham, Switzerland: Springer, 2015, pp. 641–652. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-11310-4_55
- [28] U. Mokhtar, M. A. Ali, A. E. Hassenian, and H. Hefny, "Identifying two of tomatoes leaf viruses using support vector machine," in *Information Systems Design and Intelligent Applications* (Advances in Intelligent Systems and Computing), J. Mandal, S. Satapathy, S. Kumar, P. Sarkar, and A. Mukhopadhyay, Eds. New Delhi, India: Springer, vol. 339, 2015, pp. 771–782. [Online]. Available: https://link.springer.com/chapter/10.1007/978-81-322-2250-7_77

- [29] U. Mokhtar, M. A. S. Ali, A. E. Hassenian, and H. Hefny, "Tomato leaves diseases detection approach based on support vector machines," in *Proc. 11th Int. Comput. Eng. Conf. (ICENCO)*, Dec. 2015, pp. 246–250. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/7416356>
- [30] A. E. Hassaniien, T. Gaber, U. Mokhtar, and H. Hefny, "An improved moth flame optimization algorithm based on rough sets for tomato diseases detection," *Comput. Electron. Agricult.*, vol. 136, pp. 86–96, Apr. 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0168169916308225>
- [31] H. Sabrol and K. Satish, "Tomato plant disease classification in digital images using classification tree," in *Proc. Int. Conf. Commun. Signal Process. (ICCCSP)*, Apr. 2016, pp. 1242–1246. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/7754351>
- [32] L. C. Ngugi, M. Abdelwahab, and M. Abo-Zahhad, "Tomato leaf segmentation algorithms for mobile phone applications using deep learning," *Comput. Electron. Agricult.*, vol. 178, Nov. 2020, Art. no. 105788. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0168169920306529>
- [33] J. Liu and X. Wang, "Tomato diseases and pests detection based on improved YOLO v3 convolutional neural network," *Frontiers Plant Sci.*, vol. 11, p. 898, Jun. 2020. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fpls.2020.00898>
- [34] Y. Zhang, C. Song, and D. Zhang, "Deep learning-based object detection improvement for tomato disease," *IEEE Access*, vol. 8, pp. 56607–56614, 2020. [Online]. Available: <https://ieeexplore.ieee.org/document/9044330>
- [35] F. Alvaro, Y. Sook, K. Sang, and P. Dong, "A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition," *Sensors*, vol. 17, no. 9, p. 2022, 2017. [Online]. Available: <https://www.mdpi.com/1424-8220/17/9/2022>
- [36] J. Liu and X. Wang, "Early recognition of tomato gray leaf spot disease based on MobileNetV2-YOLOv3 model," *Plant Methods*, vol. 16, no. 1, pp. 1–16, Dec. 2020. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/32523613>
- [37] M. Brahimi, K. Boukhalifa, and A. Moussaoui, "Deep learning for tomato diseases: Classification and symptoms visualization," *Appl. Artif. Intell.*, vol. 31, no. 4, pp. 299–315. [Online]. Available: <https://www.tandfonline.com/doi/abs/10.1080/08839514.2017.1315516>
- [38] A. Fuentes, D. H. Im, S. Yoon, and D. S. Park, "Spectral analysis of CNN for tomato disease identification," in *Artificial Intelligence and Soft Computing (Lecture Notes in Computer Science)*, vol. 10245, L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L. A. Zadeh, and J. M. Zurada, Eds., Cham, Switzerland: Springer, 2017, pp. 40–51. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-59063-9_4
- [39] G. Yang, G. Chen, Y. He, Z. Yan, Y. Guo, and J. Ding, "Self-supervised collaborative multi-network for fine-grained visual categorization of tomato diseases," *IEEE Access*, vol. 8, pp. 211912–211923, 2020. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9264241/>
- [40] V. Maeda-Gutiérrez, C. E. Galván-Tejada, L. A. Zanella-Calzada, J. M. Celaya-Padilla, J. I. Galván-Tejada, H. Gamboa-Rosales, H. Luna-García, R. Magallanes-Quintanar, C. A. G. Méndez, and C. A. Olvera-Olvera, "Comparison of convolutional neural network architectures for classification of tomato plant diseases," *Appl. Sci.*, vol. 10, no. 4, p. 1245, Feb. 2020. [Online]. Available: <https://www.mdpi.com/2076-3417/10/4/1245>
- [41] Q. Wu, Y. Chen, and J. Meng, "DCGAN-based data augmentation for tomato leaf disease identification," *IEEE Access*, vol. 8, pp. 98716–98728, 2020. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9099295>
- [42] A. K. Rangarajan, R. Purushothaman, and A. Ramesh, "Tomato crop disease classification using pre-trained deep learning algorithm," *Proc. Comput. Sci.*, vol. 133, pp. 1040–1047, Jan. 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1877050918310159>
- [43] K. Zhang, Q. Wu, A. Liu, and X. Meng, "Can deep learning identify tomato leaf disease?" *Adv. Multimedia*, vol. 2018, pp. 1–10, Sep. 2018. [Online]. Available: <https://www.hindawi.com/journals/am/2018/6710865/>
- [44] A. Abbas, S. Jain, M. Gour, and S. Vankudothu, "Tomato plant disease detection using transfer learning with C-GAN synthetic images," *Comput. Electron. Agricult.*, vol. 187, Aug. 2021, Art. no. 106279. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0168169921002969>
- [45] H. Durmuş, E. O. Güneş, and M. Kırıcı, "Disease detection on the leaves of the tomato plants by using deep learning," in *Proc. 6th Int. Conf. Agro-Geoinform.*, Aug. 2017, pp. 1–5. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8047016>
- [46] P. Tm, A. Pranathi, K. SaiAshritha, N. B. Chittaragi, and S. G. Koolagudi, "Tomato leaf disease detection using convolutional neural networks," in *Proc. 11th Int. Conf. Contemp. Comput. (IC3)*, Aug. 2018, pp. 1–5. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8530532/>
- [47] P. Bir, R. Kumar, and G. Singh, "Transfer learning based tomato leaf disease detection for mobile applications," in *Proc. IEEE Int. Conf. Comput., Power Commun. Technol. (GUCON)*, Oct. 2020, pp. 34–39. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9231174>
- [48] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. 1, pp. 321–357, 2002. [Online]. Available: <https://www.jair.org/index.php/jair/article/view/10302>
- [49] J. L. Leevy, T. M. Khoshgoftaar, R. A. Bauder, and N. Seliya, "A survey on addressing high-class imbalance in big data," *J. Big Data*, vol. 5, no. 1, pp. 1–30, Dec. 2018. [Online]. Available: <https://link.springer.com/article/10.1186/s40537-018-0151-6>
- [50] Y. Li, H. Lu, J. Li, X. Li, Y. Li, and S. Serikawa, "Underwater image de-scattering and classification by deep neural network," *Comput. Elect. Eng.*, vol. 54, pp. 68–77, Aug. 2016. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0045790616302075>
- [51] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. T. Haar Romeny, J. B. Zimmerman, and K. Zuiderveld, "Adaptive histogram equalization and its variations," *Comput. Vis., Graph., Image Process.*, vol. 39, no. 3, pp. 355–368, 1987. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0734189X8780186X>
- [52] S. Kaufman, S. Rosset, C. Perlich, and O. Stitelman, "Leakage in data mining: Formulation, detection, and avoidance," *ACM Trans. Knowl. Discovery From Data*, vol. 6, no. 4, pp. 1–21, 2012, doi: 10.1145/2020408.2020496
- [53] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn. (ICML)*, in Proceedings of Machine Learning Research, vol. 37, F. Bach and D. Blei, Eds., Lille, France, Jul. 2015, pp. 448–456. [Online]. Available: <http://proceedings.mlr.press/v37/loffe15.html>
- [54] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, in Proceedings of Machine Learning Research, vol. 15, G. Gordon, D. Dunson, and M. Dudík, Eds., Fort Lauderdale, FL, USA, Apr. 2011, pp. 315–323. [Online]. Available: <http://proceedings.mlr.press/v15/glorot11a.html>
- [55] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2670313>
- [56] I. J. Goodfellow, Y. Bengio, and A. C. Courville, *Deep Learning (Adaptive Computation and Machine Learning)*. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: <http://www.deeplearningbook.org>
- [57] S. Khirirat, H. R. Feyzmahdavian, and M. Johansson, "Mini-batch gradient descent: Faster convergence under data sparsity," in *Proc. IEEE 56th Annu. Conf. Decis. Control (CDC)*, Dec. 2017, pp. 2880–2887. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8264077>
- [58] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, Y. Bengio and Y. LeCun, Eds., 2015.
- [59] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015. [Online]. Available: <https://link.springer.com/article/10.1007/s11263-015-0816-y>
- [60] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626. [Online]. Available: <https://ieeexplore.ieee.org/document/8237336>



SABBIR AHMED was born in Dhaka, Bangladesh, in 1996. He received the B.Sc. Engg. degree (Hons.) in computer science (CS) from the Islamic University of Technology (IUT), Gazipur, Bangladesh, in 2017, where he is pursuing the M.Sc. degree in CS.

Since 2018, he has been working as a Lecturer with the Department of Computer Science and Engineering, IUT. His research interests include pattern recognition, deep learning in computer vision, and intelligent agriculture. He received the IUT Gold Medal for his B.Sc. Engg. Degree.



MD. BAKHTIAR HASAN received the B.Sc. Engg. and M.Sc. Engg. degrees in computer science and engineering (CSE) from the Islamic University of Technology (IUT), in 2018 and 2022, respectively.

Since 2019, he has been working as a Lecturer with the Department of Computer Science and Engineering, IUT. His research interests include the use of deep learning and computer vision techniques in human biometrics and smart agriculture.

Mr. Hasan received Huawei Seeds for the Future Scholarship, in 2018. He was awarded IUT Gold Medal in recognition of his outstanding performance in the pursuit of the B.Sc. Engg. degree, in 2018.



TASNIM AHMED was born in Kushtia, Bangladesh, in 1997. He received the B.Sc. degree in computer science and engineering from the Islamic University of Technology, Gazipur, Bangladesh, where he is currently pursuing the M.Sc. degree.

Since 2020, he has been working as a full-time Lecturer with the Computer Science and Engineering Department, Islamic University of Technology. His research interests include computer vision, natural language processing, bioinformatics, and software engineering.



MD. REDWAN KARIM SONY received the B.Sc. degree in computer science and engineering from the Islamic University of Technology (IUT), in 2016. From 2017 to 2021, he was a Lecturer at the Department of Computer Science and Engineering, IUT. He is now on study leave and currently pursuing the Ph.D. degree with Michigan State University, USA. His research interests include facial biometrics, model visualization, and the explainability of deep learning models.



MD. HASANUL KABIR (Member, IEEE) received the B.Sc. degree in computer science and information technology from the Islamic University of Technology, Bangladesh, and the Ph.D. degree in computer engineering from Kyung Hee University, South Korea.

He is currently a Professor with the Department of Computer Science and Engineering, Islamic University of Technology. His research interests include feature extraction, visual question answering, and sign language interpretation by combining image processing, machine learning, and computer vision.

...