International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 11, No.1 (Mar-2022)

https://dx.doi.org/10.12785/ijcds/1101102

MF-SE-RT:Novel Transfer Learning Method for the Identification of Tomato Disorders in Real-World Using Dilated Multiscale Feature Extraction

Saiqa Khan[1], **Meera Narvekar**[1] **and Makarand Joshi**[2]

[1]*Department of Computer Engineering, DJ Sanghvi College of Engg.,Mumbai, India*
[2]*Department of Plant Pathology, Dr. B.S. Konkan Agril. University, Dapoli, 415 712 Dist. Ratnagiri (M.S.)*

Received 14 Mar. 2021, Revised 8 Dec. 2021, Accepted 9 Jan. 2022, Published 31 Mar. 2022

Abstract: In agriculture domain, plant disorder identification and its classification are one of the emerging problems to study. If a timely and correct diagnosis is not done, it may lead to adverse effects on agricultural productivity and crop yield. The first sign of disease appears on the leaves. Diseases can be detected from the symptoms appearing on leaves. Aiming at tomato, this paper presents a novel disease recognition convolution neural network architecture based on Self-excitation network and ResNet architecture. The main research gap identified was the use of lab controlled standard images, consideration of only biotic disorders and low accuracy on unseen test dataset. The main contribution of this work is to increase generalization. Therefore, to reduce generalization error, augmentation is applied and images are captured in a manner where leaf is surrounded by occlusion areas. To capture minute lesion and spot details, multiscale feature extraction with dilated kernel is applied. Our collected real-world dataset consists of 11 types of biotic and abiotic disorders. Various experiments are carried out to verify proposed method's effectiveness. The proposed method has a recognition accuracy of 81.19% on a real-world validation dataset using 75-10-15(train-validation-test) division ratio on augmented data and average recognition accuracy of 91.76% for the 10-fold cross-validation technique. The comparative analysis with all state-of-the-art techniques exhibited amelioration in the computation time and classification accuracy. The results are used to classify tomato biotic and abiotic diseases in the real-world complex environment and novelty lies in the fact that both biotic and abiotic elements are taken into account.

Keywords: Tomato leaf disease recognition, Image enhancement, Multiscale feature extraction, Residual block

1. Introduction

In recent times, agricultural yield and productivity are largely impacted by plant diseases. The whole scenario becomes more complex when the farmers suffer from the problem of inaccessibility of experts to aid in the identification process. If the access is provided through long-distance travel, the visual inspection of disease symptoms still is a challenge. Apart from this, the process is slow, and it is highly subjective. Over the last decade, there has been tremendous improvement in the field of pattern recognition. This is still a challenging topic for the research fraternity. Countless studies have been published and devoted their attention to the optimization of the work cited earlier. The symptoms that emerge on plant leaves are used to diagnose most diseases and their identification helps in diagnosing disease in time. Researchers across the globe have implemented many algorithms for this automated identification and detection process [1]. The whole identification process can be constituted using 1. Manual identification of features and 2. Automatic identification of features. Manual construction of a feature set can be conducted by considering the texture, colour and shape properties of plant leaves. Literature survey shows that many authors have contributed to this [2], [3], [4]. The main hindrance in the way of detection using this is to find the feature set that is capable to distinguish diseases that are very much similar to each other. Examples are septoria leaf spot, early blight, late blight and other similar diseases. This has given rise to a second method based on automatic identification of features, carried out by deep learning using convolution neural network(CNN). CNN is a recognition and classification network architecture. CNN can be developed from scratch or it can be constructed by fine-tuning existing pioneering models, those who have won the ImageNet classification challenge. A series of models produced every year are VGG (Simonyan and Zisserman, 2014 [5]), AlexNet (Krizhevsky et al. 2012 [6]), ResNet (He et al. 2016 [7]), GoogLeNet model (Szegedy et al. 2015 [8]), and DenseNet (Huang et al.

E-mail address: saiqa.comp@gmail.com, meera.narvekar@djsce.ac.in , msjoshi1234@rediffmail.com

2017 [9]. Several studies have been carried out in the past for automatic disease detection of plants using hand-crafted features and non hand-crafted features by deep learning neural networks. Automatic feature engineering has proven as the main impetus for moving towards automatic feature extraction-based methods where generalization errors would be minimal and this, in turn, gives the advantage to apply the algorithm in the real-world environment. One such effort was conveyed by Lee et al. wherein transfer learning strategies for VGG, InceptionV3 and GoogLeNet along with network from scratch is analysed to prove VGG16 works better than other networks due to limited dataset exposure [10]. In their work, they have tried to use the algorithm on unseen crops to demonstrate its applicability.

Waheed et al. have collected total of 12,332 images from various sources for the corn crop. To increase variation in the dataset, they have applied many augmentation techniques and proposed the optimized denseNet architecture to classify corn diseases into four classes. They have reported accuracy of 98.06% on manually collected images. It has been observed that collected images have a uniform background and this limits the use of the method in practical settings [22]. On the other hand, Picon et al. have addressed this issue by capturing mobile phone images of wheat leaves over two years. They have proven to achieve an average Balanced accuracy(BAC) of 0.98 and were able to remove 71% of the classifier errors [11].

Additionally, Sethy et al. have worked on the rice diseases and proposed ResNet50+SVM based method to achieve an F1 score of 0.9838 [12]. Chen et al. have tried and tested different hybrid methods by proposing combinations such as VGG19+Inception, DenseNet+Inception and image segmentation with CNN from scratch. Rice and maize were the crops considered in their work and they have demonstrated research on around 1200 images after applying various augmentation strategies [13], [14]. Next, in another approach by Argeuso et al. few-shot learning on the Plant Village dataset is used to assess accuracy. Their proposed FSL network was a combination of Inception V3 network and a linear SVM classifier [15].

In another study conducted by Zeng et al., a Self-Attention Convolutional Neural Network (SACNN) is explored. The authors have claimed that a SACNN method has a better anti-interference ability, which demonstrates its strong robust power nature [16]. Additionally, researchers have worked on devising CNN from scratch for the plant disease classification process. One such approach was discussed in the work proposed by Karlekar et al., wherein image segmentation on the PDDB dataset is carried out before implementing CNN from scratch. Results have shown that their system SoyNet was able to detect Soybean diseases with 98.14% for the test data [17]. OPD2SE-Net was put forward by Liang et al. where visualization of hidden layers was used to find and analyse activations of neurons for detection of diseases and severity estimation [18]. Utpal Barman et al. have classified citrus leaf diseases using MobileNet V2 CNN architecture. In their work, they have incorporated the EasyMeasure app to keep the distance constant for capturing leaf images with white plain background [19]. By examining the benefits and drawbacks of recent past works, it has been determined that the majority of the work have either assess their proposed model on 80:20 or with some other ratio or using cross-validation technique. However, it is reported in the literature that systems proposed by Mohanty et al. and Picon et al. are the ones who assessed their proposed architecture on all three train, validation, and test sets. Our main contributions in this research work are summarized as follows:

1) We have collected tomato leaf images across different farms for biotic and abiotic disorders. The key research gap discovered is that many abiotic diseases have symptoms that are similar to biotic disease symptoms and in the past, this is not taken into consideration while designing systems.
2) We have addressed an issue of small disease area recognition by inducting a multiscale feature extraction method, wherein features outputted by different size kernels are fused to generate a robust feature vector.
3) We have introduced the dilated kernel in the multiscale feature extraction process. The proposed model aliased as MF-SE-RT consists of Residual blocks with SeNet to attain maximum feature representation.
4) We conducted extensive experiments on varied datasets and compared performance with many state-of-the-art techniques. Visualization using t-SNE is explored to show the effectiveness of the proposed model.
5) We examined performance using several evaluation measures by considering different background images.

The article is structured as follows. Section 2 presents data acquisition and the proposed method. Section 3 explores the experimental results. Section 4 shows experimental settings and design. Finally, the conclusion is presented.

2. Data Acquisition and Proposed Work

A. Data Acquisition

In this research work, three different data source sets are used for evaluation. The initial dataset constituted real-world images. They are captured at various farms (Tansa, Laasalgaon, Badlapur, Neral) across Maharashtra State, India. Uneven illumination and complex background are the main characteristics of the real-world dataset. 12 MP mobile phone camera, IOS iPhone with 10Xmobile camera and Canon Powershot digital camera with a 12 mm minimum focal length. Another data source utilized for this research work is the Internet Downloaded image dataset. It consists of three disease classes including early blight, late blight, septoria leaf spot and healthy class. To further assess the efficiency, PlantVillage (http://www.plantvillage.org [20]) standard dataset is also used to test recognition accuracy. Compared to real-world images, the PlantVillage dataset has images with a simple background. It has in total 54,305 disease leaf images for 13 plant species. 224 × 224 is the input image size for the model. To achieve this, the input dataset is normalized by the Bi-Cubic interpolation resizing method [21]. The collected plant leaf images are in JPEG format. Figure 1 shows a sample dataset.

B. Data Augmentation

When classes are unbalanced or when there is a paucity of data for training and validation, data augmentation is applied [22]. In the proposed work, using Keras following augmentations are applied 1. Random rotation: Random rotation with the given angle 2. Zoom: Zoom operation for image 3. Horizontal flip & vertical Flip: Toggle between the horizontal and vertical orientations of the image. 4. Fill: Points outside the boundaries of the input are filled according to the given mode. 5. Width shift and height shift: Move the entire picture horizontally or vertically at a certain distance. Both biotic and abiotic disorders are considered in this work for evaluation. Figure 2 shows the Plant Village dataset image after augmentation. The proposed model uses a ResNet residual module as a baseline for building a lighter network that includes multi-scale feature extraction with dilated kernel and ResNet-SeNet combination to improve classification accuracy [23].

Various convolution kernels with feature extraction on different scales are presented based on the features and textures of different tomato illnesses. It can help to extract local features. The large-scale convolution kernel 7 × 7 is used to extract contours. The importance of using varied kernels can be justified due to a fact that disease spots and lesions are relatively small. On the other hand, textures of different diseases are also the same such as early blight, late blight, septoria leaf spot and others. Conclusively, it is decided that to address these issues there is a need to take into consideration both fine-grained features and coarse-grained features. Therefore, dilated kernels of different sizes are utilized with 32 kernels of small size(1 × 1 and 3 × 3) and 16 kernels of large size (5 × 5 and 7 × 7). All the outputted feature maps are combined and pass into the next layer.

Dilated convolution kernel technique was used extensively in image segmentation [24]. After convolution, pooling layers are used in a traditional CNN architecture. They reduce overfitting. They do, however, diminish the spatial information of feature maps. In dilated convolution, a filter is dilated before computing the convolution. For dilation, the convolution filter size is increased, and zeroes are put at all empty position to get the desired width and height of the kernel. In other words, dilated convolutions indicate a type of convolution where holes are inserted between the elements of a kernel to inflate kernel, unlike traditional standard kernel where l(dilation rate) is 1. Technically, 2D dilated convolution kernel can be represented as (Eq .1):

$$(F *_l k)(p) = \sum_{s+lt=p} F(s)\,k(t) \qquad (1)$$

where l is the dilation rate that indicates a degree to which the kernel is widened. In the year 2015 for ImageNet competition, ResNet won first place for an image classification task [21]. It was primarily being designed to solve the vanishing gradient problem. It does so by introducing residual blocks and skip connections as shown in Figure 3. It is considered simpler compared to its previous counterpart such as VGG.

Res-Net uses residual blocks and skip connections. Residual blocks are considered as a special case of networks without the presence of gates. In a neural network, a gate serves as a threshold for determining when the network should employ standard stacked layers vs an identity connection. The output of lower levels is added to the output of subsequent layers in an identity connection. In a nutshell, it allows the network's layers to learn in little steps rather than building transformations from scratch. Gates allow the flow of memory from initial layers to final layers. Gates are missing in the residual block's skip connections; hence, they provide very good performance.

Formally, the underlying mapping is denoted as H(x) and another mapping fit by non-linear layers stacked together is denoted by F(x)=H(x)-x. F(x)+x is a recast of the original mapping. It is stated that optimising residual mapping is more straightforward compared to original mapping. Recasted mapping is called residual mapping. The main intuition is to achieve optimization and it is clearly stated that it is much easier to optimize residual mapping than original mapping. The presence of skip connections makes it easy for identity mapping to be learned. The basic aim of the Squeeze and excitation network(SeNet) is to boost representation quality and this can be carried out by modelling interdependencies in convolution feature channels. The central idea is to apply feature recalibration, which supports the use of global information to concentrate on the most relevant features while eliminating the less important ones.

The structure is represented in Figure 4 where F_{tr} is the transformation that performs mapping of input X to U and $U \in R^{H \times W \times C}$. These produced features are passed through a squeeze network to generate a channel descriptor by the aggregation of feature maps.

t-Distributed Stochastic Neighbor Embedding(t-SNE) is one of the popular methods for visualization [25]. The technique is used to create two-dimensional maps from hundreds of dimensions. This is done by mapping multidimensional data consists of hundreds of dimensions to two dimensions. The algorithm is non-linear, and it transforms underlying data using different operations. Perplexity is a measure of information that is defined as 2 to the power of the Shannon entropy. The perplexity of a fair die with k sides is equal to k. In t-SNE, the perplexity may be viewed as a knob that sets the number of effective nearest neighbours. The original paper on t-sne visualization says, “The performance of SNE is fairly robust to changes in the perplexity, and typical values are between 5 and 50”. t-sne visualization in the optimization process depends on hyperparameters and it will not produce similar outputs on its consecutive runs.

C. Proposed Model architecture

Plant leaf image classification involves various steps to be performed. The flow of steps for classification is illustrated in Figure 5. The model comprises a multi-scale feature extraction module with the expanded kernel, 2 convolution and 2 pooling layers to reduce dimensionality before the application of ResNet and SeNet. 3 ResNet-Senet modules, an average pooling layer, drop out and at the end fully connected layer. The reason for combining residual blocks with squeeze and excitation network is to ensure attention for the region of interest areas such as leaf in a complex background. The output of SeNet is to produce attention weights when combined with convoluted output. They generate feature maps that highlight areas of more importance. This is similar to human brain functioning where we focus on certain areas more compared to other areas. All the convolution layers are followed by batch normalization. The specific structure is shown in Figure 6 and related parameters are shown in Table I.

3. Experimental Results and Evaluations

A. Image Classification Experiments

In all our experiments; pre-processing, data augmentation and model implementation was implemented using Python 3.6, Keras and Tensorflow back end. Model training was implemented by Google Colab. The equipment configuration is shown in Table II. Mini-batches are used in training and validation. The batch size is kept as 30. The maximum number of epochs is set to 100. The weight initialization is set to glorot Xavier and bias to zero. The glorot Xavier initializes each weight with a small Gaussian value with mean = 0.0 and variance based on the fan-in and fan-out of the weight. The optimization parameter is Adam and the final classification activation function is softmax. The momentum is 0.9 and the 0.1 is selected as the initial learning rate. Reduce learning rate is adapted in which validation accuracy is monitored and if there has been no improvement for three epochs, then the learning rate is reduced by half. Nevertheless, early stopping is used, which means if loss of validation does not decrease after six iterations, training is considered complete. For training and validation, the input image size is set to 224 × 224. Table III shows all the hyperparameters. Aside from that, the Cross-Entropy Loss function is used. Cross entropy loss also called log loss is used to assess the performance of a model whose output can be characterized as probabilities. The loss is increased when the projected probability differs from the actual probability. So, the ideal case is to have zero loss. Cross-entropy loss for binary classification when the number of classes is two can be calculated as(Eq .2):

$$-(y\log(q) + (1-y)\log(1-q)) \tag{2}$$

where q is the probability. When $M > 2$, loss is calculated per class label per instance and summing the result gives loss for a multiclass classification problem as (Eq 3):

$$-\sum_{c=1}^{M} y_{o,c}\log(q_{o,c}) \tag{3}$$

where M indicates number of classes for which loss is calculated, log is the natural logarithm, y is the indicator (binary) for some instance o when class label denoted as c is the accurate answer and q indicates final predicted probability outcome.

Model evaluation parameters are accuracy, precision, recall and F1 Score. They are calculated as:

$$Accuracy = (TP+TN)/(TP+FP+FN+TN) \tag{4}$$

$$Precision = TP/(TP+FP) \tag{5}$$

$$Recall = TP/(TP+FN) \tag{6}$$

$$F1\,Score = 2*(Recall*Precision) \quad /(Recall+Precision) \tag{7}$$

where TP stands for True Positive, FP for False Positive, FN for False Negative and TN for True Negative [26] (Eq 4-7).

For the first set of experiments, the PlantVillage dataset is selected. The PlantVillage dataset is a standard plant disease dataset consisting of 24 classes of crops for different diseases. Many augmentation experiments are conducted on tomato leaf images by considering variations. The augmented PlantVillage dataset’s distribution is shown in Figure 7.

Table IV shows the performance of the different models for the PlantVillage dataset. It is observed that proposed model achieves 97.27% accuracy on validation data. For the second set of experiments, proposed model is tested on tomato leaf images captured in a real-world environment

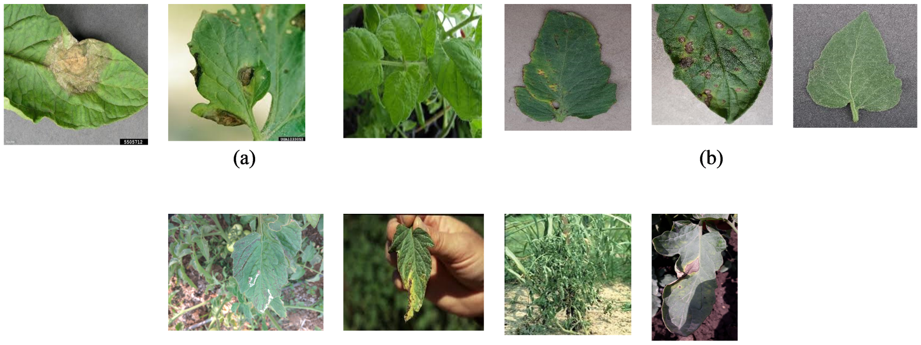


Figure 1. Sample dataset images of tomato leaves (a)Internet Dataset, (b)Plant Village Dataset images and (c)Real-world Dataset

TABLE I. Model parameters

Model Layer	Layer Type	Used Kernel size	Stride	Neuron size	Feature Maps
Multiscale	Multiscale Feature extraction	–	–	224*224	96
max_pooling2d	MaxPooling2D	3*3	2	112*112	96
conv2d_4	Conv2D	3*3	1	112*112	192
batch_normalization_4	Batch normalization	–	–	112*112	192
max_pooling2d_1	MaxPooling2D	3*3	2	56*56	192
conv2d_5	Conv2D	3*3	1	56*56	16
batch_normalization_5	Batch normalization	–	–	56*56	16
Resnet1	Residual Module	–	–	56*56	16
SeNet1	Self Excitation	–	–	56*56	16
Add	Add	–	–	56*56	16
Activation_2	Activation	–	–	56*56	16
conv2d_10	Conv2D	3*3	2	28*28	32
batch_normalization_8	Batch Normalization	–	–	28*28	32
ResNet2	Residual Module	–	–	28*28	32
SeNet2	Self excitation	–	–	28*28	32
Add_1	Add	–	–	28*28	32
Activation_5	Activation	–	–	28*28	32
conv2d_15	Convolution	3*3	2	14*14	64
batch_normalization_11	Batch normalization	–	–	14*14	64
ResNet3	Residual Module	–	–	14*14	64
SeNet3	Self excitation	–	–	14*14	64
Add_2	Add	–	–	14*14	64
average_pooling2d	Average Pooling	7*7	7	2*2	64
Dropout	Dropout	–	–	2*2	64
Flatten	Flatten	–	–		256
Dense	Softmax regression	Classifier	–	1*1*4	no_of_classes

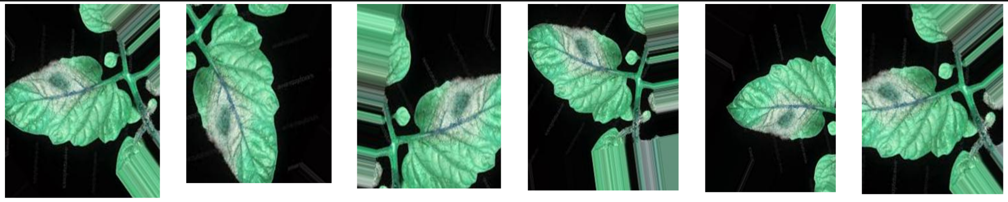


Figure 2. Sample augmented images

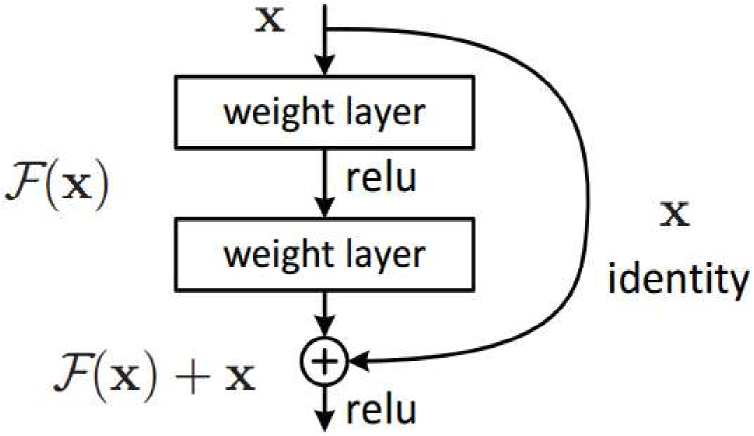

x
weight layer
$\mathcal{F}(\mathbf{x})$
relu
weight layer
x
identity
$\mathcal{F}(\mathbf{x}) + \mathbf{x}$
relu

Figure 3. Residual block[21]

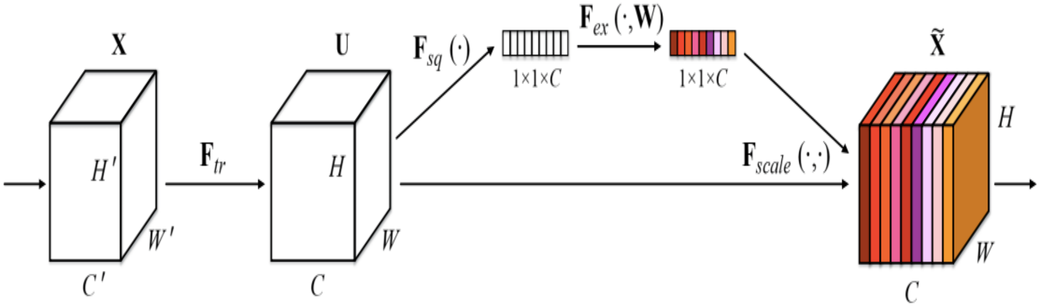


Figure 4. Squeeze and excitation block

with uneven illumination and cluttered background conditions. To reduce the overfitting issue with a small input dataset, augmentation techniques are applied to enrich the dataset (Figure 8).

Following is a description of the process adopted in the proposed system:

1) First step is to change the size of all input images to 224 × 224.
2) All the images are categorized to train, validation and test dataset and some images are kept there for unseen data testing purpose.
3) Secondly, augmentation techniques are applied to increase diversity in the dataset and to make it balanced.
4) Next, images are pre-processed to make them normalized in the range 0 to 255. As discussed in

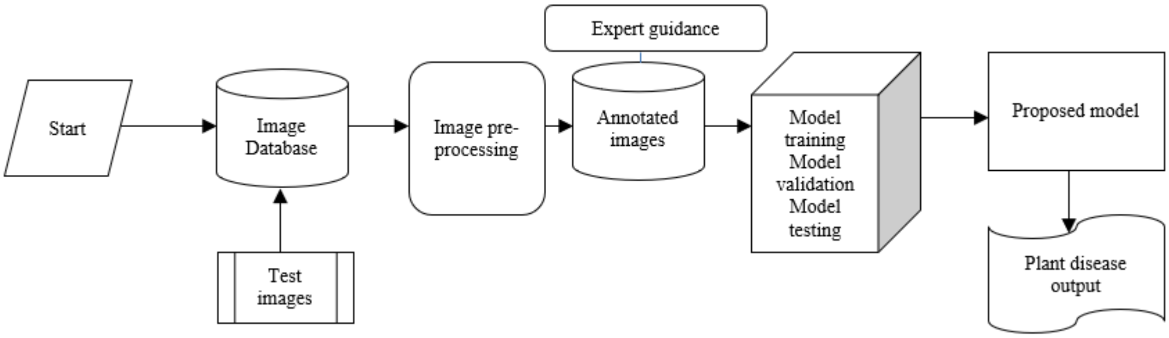


Figure 5. Framework for proposed work

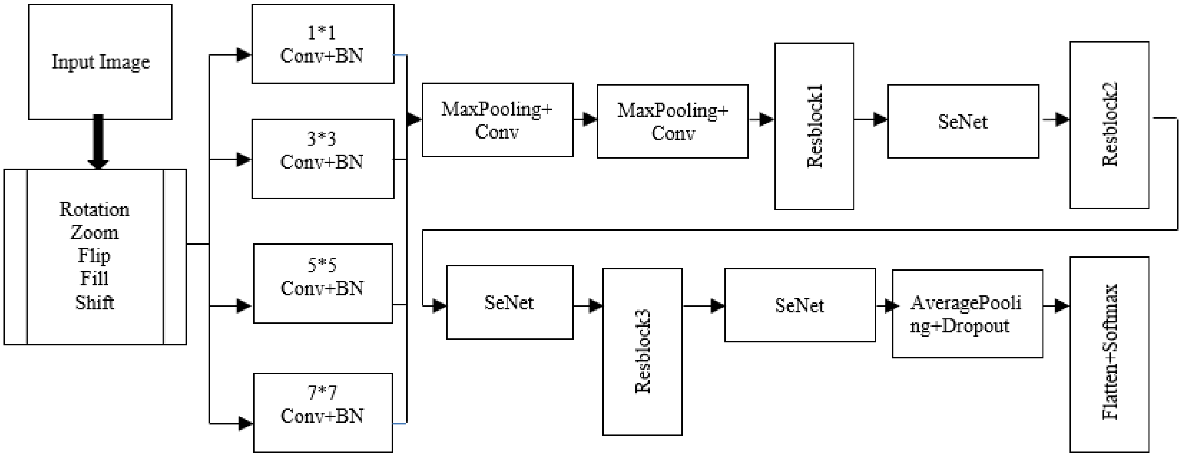


Figure 6. Proposed model architecture

TABLE II. Equipment Configuration

Name	Parameter
Memory(RAM)	12.0 GB
Processor	Intel Core i5 @2.6 ghz
Graphics card	Nvidia Graphics card GeForce 920MX
Language	Python

section 2,the training set is used to train the model. Numerous experiments are carried out.

5) Validation data is used to assess the performance of model and finally, testing is carried out for unseen test data. The actual results are compared with the predicted categories and various performance evaluators will be assessed to check its effectiveness.

Figure 9 shows validation loss, validation accuracy, training loss and training accuracy. The a,b,c and d part of Figure 8 respectively represent the curves of recognition accuracy and loss rate obtained by proposed and other models on training and validation dataset for real-world 11 disorders of the tomato plant. It can be noticed that in the proposed model there is not much difference between training accuracy and validation accuracy and as a result, there is no overfitting. When it comes to the training and validation data sets, the suggested model, as shown in Figure 6, works well. DenseNet is a model used for disease recognition by the advent of a transfer learning strategy. Chen et al. have developed their model by combining the VGG model with Inception blocks. The accuracy of training and validation differs dramatically, as seen in Figure 8(b). As a result, the system is overfitting. There is an improvement in training accuracy after the tenth epoch, but validation accuracy appears to be steady. There is no consistent improvement for validation loss, as seen in Figure 8(c). As a result, validation accuracy behaves similarly. Figure 8(d) depicts the results

TABLE III. Hyperparameters

Hyperparameter	Value
Solver type	Adam
Learning rate	Initial as 0.1
Batch size	30 for training and validating

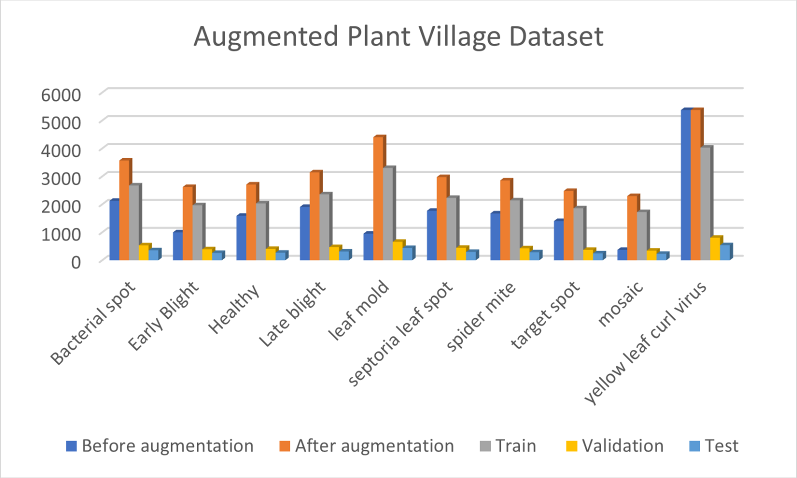


Figure 7. Augmented PlantVillage dataset

TABLE IV. Comparative results on PlantVillage dataset

Pre-trained model	Training accuracy(%)	Training Loss	Validation accuracy(%)	Validation Loss	Early stopping epoch number
Dense-Net [9]	100	0.00033	94.44	0.2682	311
Karlekar Model(without background removal) [17]	99.70	0.0126	96.35	0.1497	127
Chen et al. [27]	100	0.00033	94.16	0.2687	34
VGG-GAP	67.69	0.9564	63.45	0.9861	238
Proposed Model	99.20	0.0268	97.27	0.0835	340

when the system converges more quickly. When its accuracy is tested on unseen test data, it degrades dramatically. Table V shows the performance comparison with other models. Karlekar et al. in their work designed the model from scratch and applied pre-processing for the background removal using Hue, Saturation and Value (HSV) colour space. However, when applied the same background removal technique on real-complex images it doesn't produce effective results due to fixed thresholds. They tested their findings on the PDDB database, which has a simple background. PDDB can be accessed in the link https://www.digipathos-rep.cnptia.embrapa.br/. Other notable work discussed here for the comparison purpose is Visual Geometry Group (VGG) with Global Average Pooling (GAP) where VGG pre-trained model on ImageNet weights are combined with global average pooling.

Table VI shows how proposed model delivers for real-world images with 50 epochs for 10 folds. The average recognition accuracy achieved is 91.76 with a loss of 0.26126. For fold 1 to fold 9, accuracy is higher than 90%. For the third set of experiments, all the images are downloaded from the Internet, acquired using disease name in tomato plants. Images are categorized into three disease classes and one healthy class. All these images have a cluttered complex background that resembles real-world characteristics. The next step was to enrich the dataset with augmented images. The main objective of network designing that it should be able to distinguish various classes. Before training, images are resized for processing

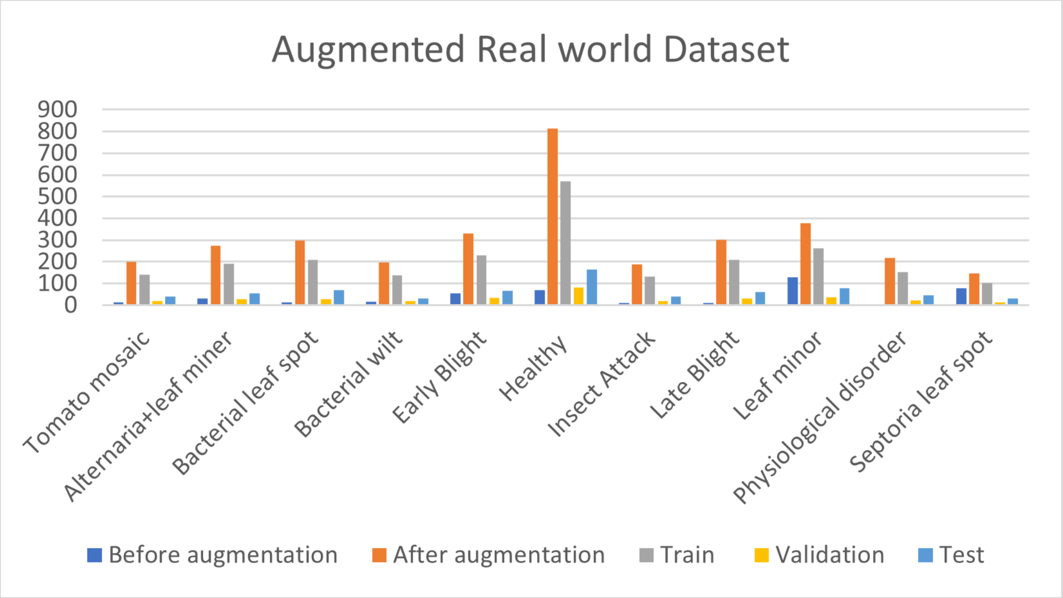


Figure 8. Augmented real world dataset

TABLE V. Comparative results on real-world dataset

Pre-trained model	Training accuracy(%)	Training Loss	Validation accuracy(%)	Validation Loss	Early stopping epoch number
Dense-Net [9]	92.87	0.2817	80.19	0.4676	33
Karlekar Model(without background removal) [17]	24.46	2.2690	24.08	2.2690	11
VGG-Inception [27]	99.48	0.0580	79.70	0.6146	24
Proposed Model	86.82	0.3524	81.19	0.4857	39

and 0-255 range is selected for normalization for training by written script in Python using OpenCV framework.

t-sne visualization technique(Figure 10) is adapted to show the feature representation of validation data for the MF-SE-RT model.

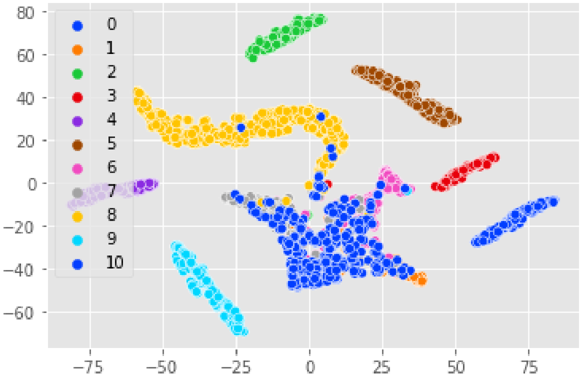


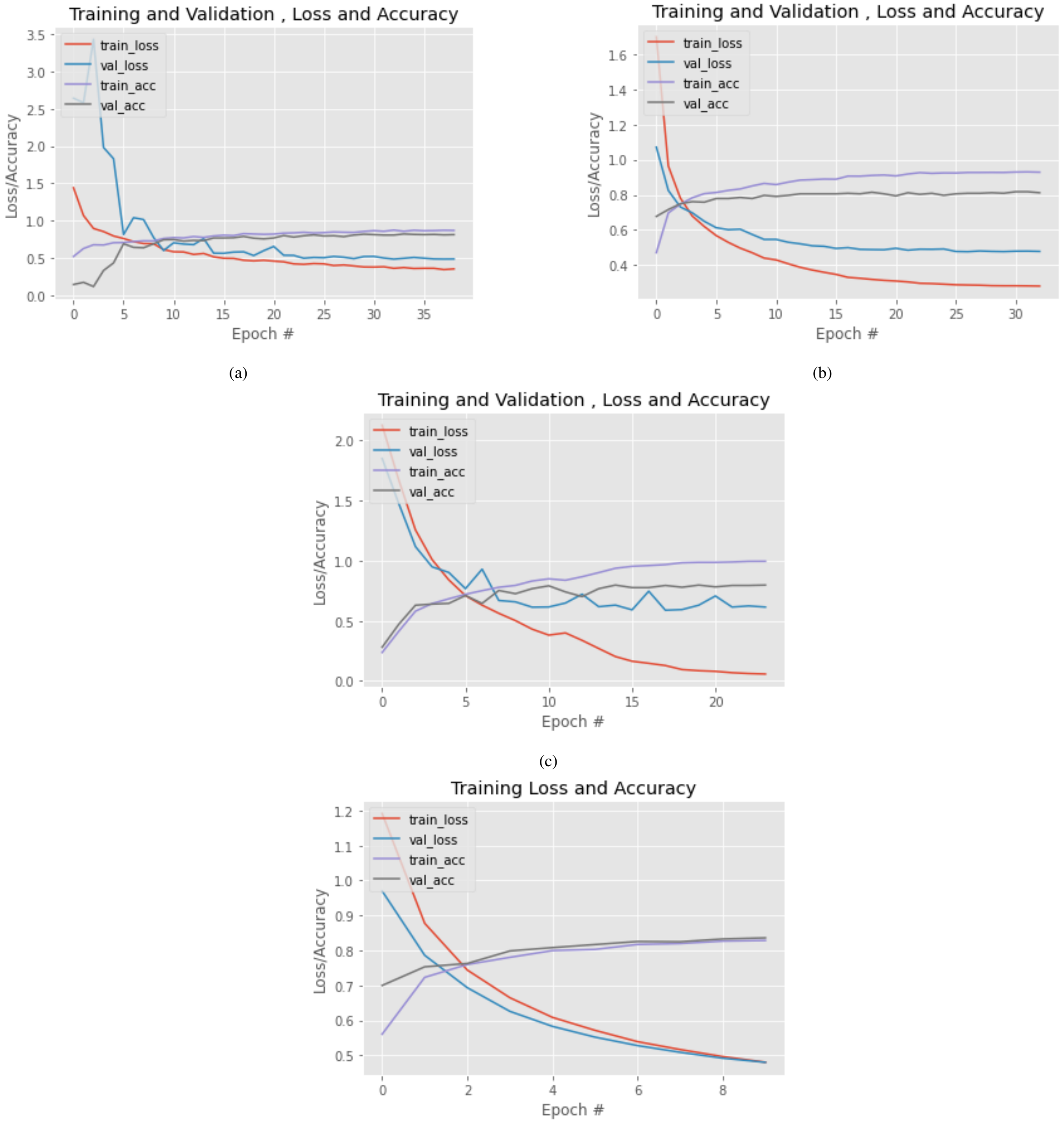
Figure 10. t-sne visualization of the validation

Figure 11(a) shows the confusion matrix for test data for 11 biotic and abiotic disorder classes. From Figure 11(b) it can be observed that class 6(Insect attack) and class 7(late blight) have low F1 score compared to other classes on test data. Trainable parameters in a model are related to time and space complexity. Therefore, apart from accuracy, it is important to ensure time and space requirement for the proposed model.

Table VII shows comparative results, and it is observable by experimental results that the proposed method performs better than other techniques on the Internet dataset. The primary intuition behind this is to have distinguishable characteristics extracted for different classes. Even though DenseNet and VGG+GAP were trained using ImageNet weights, the proposed model employs a random initialization strategy. Optimal results were not achieved by these models.

All metrics for assessment are used for comparison purpose as shown in Figure 12.

Furthermore, a confusion matrix is created for the analysis of validation data and it is depicted in Figure 13.



(a)

(b)

(c)

(d)

Figure 9. Validation loss, validation accuracy, training loss and training accuracy (a) Proposed model, (b)DenseNet, (c)Chen et al.(d)VGG+GAP

We have compared a number of trainable parameters in a model to other state-of-the-art models and the same is depicted in Table VIII.

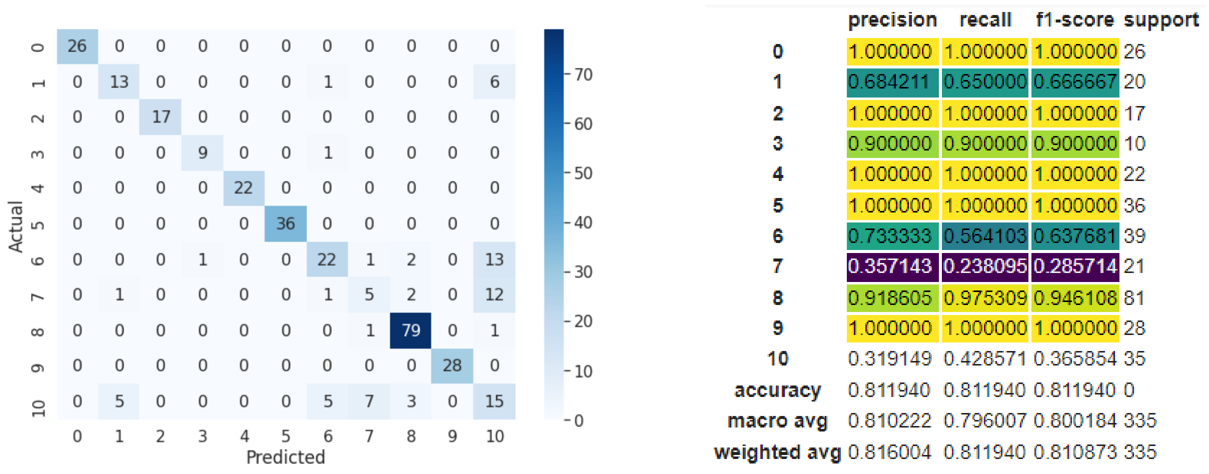
4. Conclusions and Future Work

This work has presented architecture based on Residual with SeNet for the classification of tomato leaf diseases. For the verification of the architecture robustness, a new dataset of real-world tomato leaf images is produced. Many existing results are shown for comparative analysis. It is verified that the combination of Residual blocks with SeNet and a multiscale feature extraction gives good performance benefits.

One limitation of this work is the limited real-world dataset. However, architecture has proven efficient and consistent in performance due to testing on varied datasets with train-test split and cross-validation strategies. In the future, we intend to expand dataset by adding more crops and their respective diseases. We plan to add the age factor at the time of capturing images. So, multimodal analysis can

TABLE VI. 10-Fold cross validation results on real-world dataset

Fold Number	Loss	Accuracy
Fold1	0.161441	94.03%
Fold2	0.188695	93.43%
Fold3	0.10275	96.72%
Fold4	0.114066	95.52%
Fold5	0.25182	91.34%
Fold6	0.461352	85.37%
Fold7	0.177619	94.03%
Fold8	0.211399	93.13%
Fold9	0.09485	95.81%
Fold10	0.748863	95.81%
Average scores for all folds:		
Accuracy: 91.7666150463952 (+- 5.075302627477456)		
Loss: 0.2612682721681065		



	0	1	2	3	4	5	6	7	8	9	10
0	26	0	0	0	0	0	0	0	0	0	0
1	0	13	0	0	0	0	1	0	0	0	6
2	0	0	17	0	0	0	0	0	0	0	0
3	0	0	0	9	0	0	1	0	0	0	0
4	0	0	0	0	22	0	0	0	0	0	0
5	0	0	0	0	0	36	0	0	0	0	0
6	0	0	0	1	0	0	22	1	2	0	13
7	0	1	0	0	0	0	1	5	2	0	12
8	0	0	0	0	0	0	0	1	79	0	1
9	0	0	0	0	0	0	0	0	0	28	0
10	0	5	0	0	0	0	5	7	3	0	15

(a)

	precision	recall	f1-score	support
0	1.000000	1.000000	1.000000	26
1	0.684211	0.650000	0.666667	20
2	1.000000	1.000000	1.000000	17
3	0.900000	0.900000	0.900000	10
4	1.000000	1.000000	1.000000	22
5	1.000000	1.000000	1.000000	36
6	0.733333	0.564103	0.637681	39
7	0.357143	0.238095	0.285714	21
8	0.918605	0.975309	0.946108	81
9	1.000000	1.000000	1.000000	28
10	0.319149	0.428571	0.365854	35
accuracy	0.811940	0.811940	0.811940	0
macro avg	0.810222	0.796007	0.800184	335
weighted avg	0.816004	0.811940	0.810873	335

(b)

Figure 11. Classification report for unseen test data (a)Confusion matrix, (b) Classification summary

TABLE VII. Comparative results on Internet dataset

Pre-trained model	Training accuracy(%)	Training Loss	Validation accuracy(%)	Validation Loss	Early stopping epoch number
Dense-Net [9]	92.99	0.2696	84.62	0.3896	356
Karlekar Model(without background removal) [17]	29.19	1.3814	26.56	1.3868	112
Chen et al. [27]	96.86	0.1214	87.91	0.2473	11
VGG+GAP [28]	68.79	0.8584	64.47	0.8961	237
Proposed Model	98.92	0.0456	95.97	0.1179	340

be carried out by considering text and image data together.

Severity estimation by considering age will bolster this whole setup and will be effective for the farmers for efficient

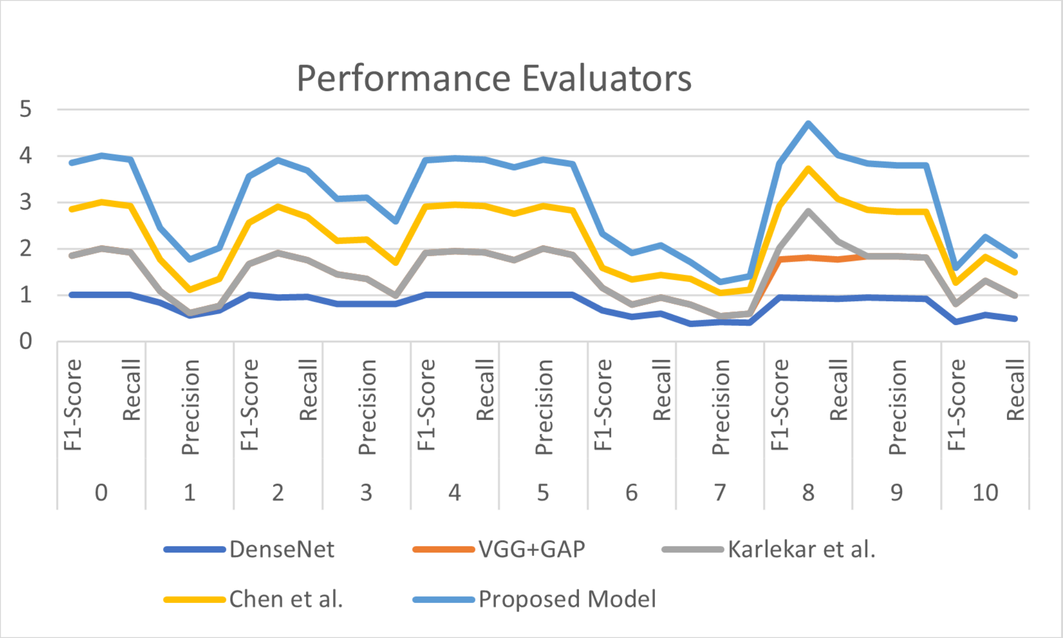


Figure 12. Evaluation parameters for comparative study

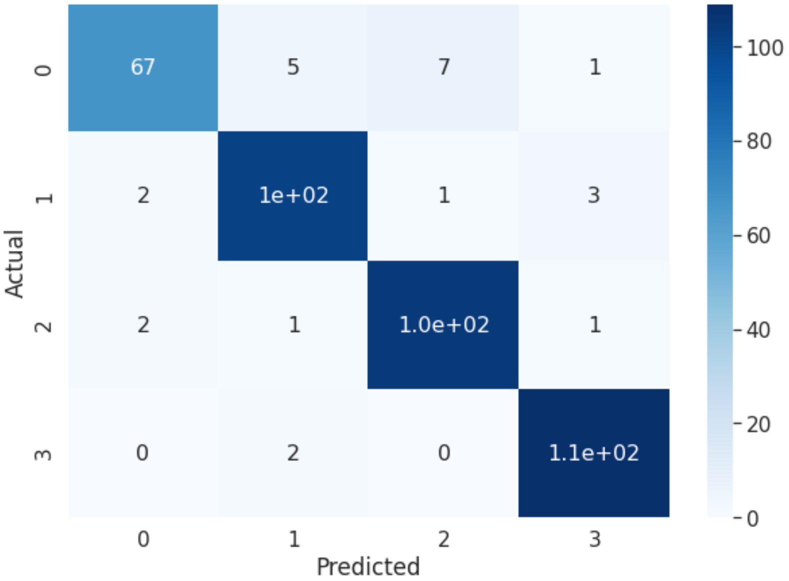


Figure 13. Confusion matrix for validation data of Internet dataset

decision making in the real-world environment. Visualization strategies including Class Activation Maps(CAM), Gradient Class Activation Maps(GradCAM) etc. can be utilized to find interclass and intraclass performance characteristics as well as the role of different filters used in architecture.

Trainable parameters Table shows that there is significant different between models in terms of number of trainable parameters. This eventually affects model performance when it is to be operated in real-world environment where speed and accuracy have to be balanced to achieve performance.

There could be a further improvement by guiding the farmer for automating adjustment in the orientation of a camera angle that could prevent shadows at the time of clicking images. Besides, image capturing practices there could be an improvement in the segmentation method, reliability of their usage in the real world.

TABLE VIII. Layer parameters

Sr. No	Model	Total parameters	Trainable Parameters
1	TCCNN [29]	2,492,940	2,492,940
2	VGG+GAP	20,030,027	5,643
3	Unified Model(Li et al.,2019)	49,588,899	49,501,347
4	GPDCNN [30]	2,55,571	2,54,467
5	DenseNet-Inception [27]	20,242,984	20,013,928
6	DenseNet [9]	7,037,504	7,037,504
7	MF-SE-RT Proposed	2,57,370	256,266

References

[1] J. G. A. Barbedo, "Plant disease identification from individual lesions and spots using deep learning," *Biosystems Engineering*, vol. 180, pp. 96–107, 2019.

[2] J. Basavaiah and A. A. Anthony, "Tomato leaf disease classification using multiple feature extraction techniques," *Wireless Personal Communications*, vol. 115, no. 1, pp. 633–651, 2020.

[3] X. E. Pantazi, D. Moshou, and A. A. Tamouridou, "Automated leaf disease detection in different crop species through image features analysis and one class classifiers," *Computers and electronics in agriculture*, vol. 156, pp. 96–104, 2019.

[4] S. S. Chouhan, U. P. Singh, U. Sharma, and S. Jain, "Leaf disease segmentation and classification of jatropha curcas l. and pongamia pinnata l. biofuel plants using computer vision based approaches," *Measurement*, vol. 171, p. 108796, 2021.

[5] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[8] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[9] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[10] S. H. Lee, H. Goëau, P. Bonnet, and A. Joly, "New perspectives on plant disease characterization based on deep learning," *Computers and Electronics in Agriculture*, vol. 170, p. 105220, 2020.

[11] A. Picon, M. Seitz, A. Alvarez-Gila, P. Mohnke, A. Ortiz-Barredo, and J. Echazarra, "Crop conditional convolutional neural networks for massive multi-crop plant disease classification over cell phone acquired images taken on real field conditions," *Computers and Electronics in Agriculture*, vol. 167, p. 105093, 2019.

[12] P. K. Sethy, N. K. Barpanda, A. K. Rath, and S. K. Behera, "Deep feature based rice leaf disease identification using support vector machine," *Computers and Electronics in Agriculture*, vol. 175, p. 105527, 2020.

[13] J. Chen, D. Zhang, and Y. A. Nanehkaran, "Identifying plant diseases using deep transfer learning and enhanced lightweight network," *Multimedia Tools and Applications*, vol. 79, no. 41, pp. 31 497–31 515, 2020.

[14] J. Chen, J. Chen, D. Zhang, Y. Sun, and Y. A. Nanehkaran, "Using deep transfer learning for image-based plant disease identification," *Computers and Electronics in Agriculture*, vol. 173, p. 105393, 2020.

[15] D. Argüeso, A. Picon, U. Irusta, A. Medela, M. G. San-Emeterio, A. Bereciartua, and A. Alvarez-Gila, "Few-shot learning approach for plant disease classification using images taken in the field," *Computers and Electronics in Agriculture*, vol. 175, p. 105542, 2020.

[16] W. Zeng and M. Li, "Crop leaf disease recognition based on self-attention convolutional neural network," *Computers and Electronics in Agriculture*, vol. 172, p. 105341, 2020.

[17] A. Karlekar and A. Seal, "Soynet: Soybean leaf diseases classification," *Computers and Electronics in Agriculture*, vol. 172, p. 105342, 2020.

[18] Q. Liang, S. Xiang, Y. Hu, G. Coppola, D. Zhang, and W. Sun, "Pd2se-net: Computer-assisted plant disease diagnosis and severity estimation network," *Computers and electronics in agriculture*, vol. 157, pp. 518–529, 2019.

[19] U. Barman, R. D. Choudhury, D. Sahu, and G. G. Barman, "Comparison of convolution neural networks for smartphone image based real time classification of citrus leaf disease," *Computers and Electronics in Agriculture*, vol. 177, p. 105661, 2020.

[20] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers in plant science*, vol. 7, p. 1419, 2016.

[21] R. Keys, "Cubic convolution interpolation for digital image processing," *IEEE transactions on acoustics, speech, and signal processing*, vol. 29, no. 6, pp. 1153–1160, 1981.

[22] M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," *Neural Networks*, vol. 106, pp. 249–259, 2018.

[23] L. M. De Carvalho, F. W. Acerbi, J. G. Clevers, L. M. Fonseca, and S. M. De Jong, "Multiscale feature extraction from images using wavelets," in *Remote sensing image analysis: Including the spatial domain*. Springer, 2004, pp. 237–270.

[24] R. Hamaguchi, A. Fujita, K. Nemoto, T. Imaizumi, and S. Hikosaka, "Effective use of dilated convolutions for segmenting small object instances in remote sensing imagery," in *2018 IEEE winter conference on applications of computer vision (WACV)*. IEEE, 2018, pp. 1442–1450.

[25] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne." *Journal of machine learning research*, vol. 9, no. 11, 2008.

[26] U. Shafi, R. Mumtaz, H. Anwar, A. M. Qamar, and H. Khurshid, "Surface water pollution detection using internet of things," in *2018 15th International Conference on Smart Cities: Improving Quality of Life Using ICT & IoT (HONET-ICT)*. IEEE, 2018, pp. 92–96.

[27] J. Chen, D. Zhang, Y. A. Nanehkaran, and D. Li, "Detection of rice plant diseases based on deep transfer learning," *Journal of the Science of Food and Agriculture*, vol. 100, no. 7, pp. 3246–3256, 2020.

[28] Q. Yan, B. Yang, W. Wang, B. Wang, P. Chen, and J. Zhang, "Apple leaf diseases recognition based on an improved convolutional neural network," *Sensors*, vol. 20, no. 12, p. 3535, 2020.

[29] J. Ma, K. Du, F. Zheng, L. Zhang, Z. Gong, and Z. Sun, "A recognition method for cucumber diseases using leaf symptom images based on deep convolutional neural network," *Computers and electronics in agriculture*, vol. 154, pp. 18–24, 2018.

[30] J. Ma, K. Du, F. Zheng, L. Zhang, and Z. Sun, "A segmentation method for processing greenhouse vegetable foliar disease symptom images," *Information Processing in Agriculture*, vol. 6, no. 2, pp. 216–223, 2019.

Saiqa Khan Saiqa Khan is a PhD candidate in the Department of Computer Engineering at DJ Sanghvi College of Engineering. She has completed her masters in Computer Engineering from Thadomal Shahani Engineering college , Mumbai. She has authored more than 50 publications. Her research interest areas are computer vision, machine learning and deep learning.

Meera Narvekar Dr.Meera Narvekar received the Ph.D. degree in Computer Science and Technology from SNDT University, Mumbai. She is currently professor and Head of department of computer engineering, DJSCE Mumbai. She is a member of board of studies in Mumbai University. Her research interest are in mobile computing, data science and machine learning.

M.S. Joshi Dr. M.S.Joshi is currently heading plant pathology department of Dr. B. S. Konkan Krishi Vidyapeeth, Dapoli, Dist. Ratnagiri, India. His area of interest are Mycology and Plant Pathology. His current research work focusses Rice Pathology. He has experience over 22 years in agriculture. He is a recipient of 'Baliraja' award for writing book in Marathi on Mango diseases. He is a editor/co-editor for various journals.