



vCrop: an automated plant disease prediction using deep ensemble framework using real field images

KAVITHA LAKSHMI RAMANADHAM* and NICKOLAS SAVARIMUTHU

Department of Computer Applications, NIT Trichy, Trichy, India
e-mail: kavitha.rachabattuni@gmail.com; nickolas@nitt.edu

MS received 31 July 2022; revised 23 September 2022; accepted 19 October 2022

Abstract. Plant disease monitoring and management are essential for ensuring reliable and lucrative crop production in all kinds of plantations and guaranteeing sustainable agriculture production. Most traditional approaches depend significantly on human effort, which is liable to time delay and is expensive. Moreover, plant pathogens are nearly identical to other non-harmful species in many circumstances. Recently, computer vision based deep learning algorithms have not been deceived by these similar diseases causing false warnings. This paper proposes a novel deep ensemble neural network (D-ENN) framework for automated plant disease detection. The dataset collected in real cultivated fields contains healthy and diseased images with specific class labels. Since there are limited images of a few specific crops, a conditional generative adversarial network (CGAN) is leveraged to generate the additional synthetic images. Then, the total dataset is split into the training set, validation set, and test set, in the ratio of 70:10:20 used to avoid overfitting problems. The proposed model is trained using real and synthetic images utilizing the transfer learning mechanism. Finally, the experimental outcomes are assessed using standard performance measures evaluating the performance of the proposed method. The proposed vCrop framework attained Precision, Recall, and F1-Measure, and Accuracy of 95.71%, 95.32%, 95.51%, and 96.02% respectively, in classifying the plant diseases in comparison with the other state-of-the-art approaches. The proposed D-ENN model can be a potentially helpful tool for farmers and agronomists in diagnosing and quantifying cotton diseases.

Keywords. Deep Learning; ensemble network; conditional GAN; transfer learning; precision agriculture.

1. Introduction

Plant diseases have a catastrophic impact on agriculture production and pose a severe threat to food security globally. Therefore, diagnosing and control of plant diseases is vital in assuring the maximum yield [1]. The conventional approach of identifying plant infections involves a great deal of expertise and knowledge of professionals. With the rapid development of deep learning technologies, farmers also browse the crop diseased images from the database or interact with skilled agronomists and plant pathologists to examine crop diseases remotely. The downside is that the evaluation is simple but inaccurate with poor efficiency.

Many studies have investigated automated plant diseases detection based on image processing and machine learning methods to increase the accuracy and speed of diagnostic findings [2–4]. In addition, advanced techniques such as digital image processing methods [5], Principle component analysis approach [6], and fuzzy inference system [7] are used to diagnose the various crop diseases. Deep learning-based strategies have been used successfully in multiple

applications during the last decade [8, 9]. Recently, deep convolutional neural networks (CNNs) have elevated in the field of agriculture and, and it has become a standard technique for plant disease detection applications.

Coulibaly *et al* [10] used a transfer learning-based VGG-16 convolutional neural network for mildew disease in pearl millet crop and achieved an accuracy of 95%. In Selvaraj *et al* [11] work, a total of 18,000 field images of banana were collected. They developed a custom deep CNN through transfer learning and reported a mean Average Precision of 98%. With large open-source datasets, various crop diagnostic experiments such as paddy [13], apple [14], and grape [15] reached an average accuracy of more than 85%. In this line, Mohanty *et al* [16] used a large open-source plant leaf dataset of 26 different classes of healthy and diseased leaves images. They evaluated two different CNN models for plant disease detection, and outcomes were quite promising with an accuracy of up to 99.35%. However, a key shortcoming was the complete images are taken in laboratory conditions rather than in real field cultivated images.

Atila *et al* [17] designed a diagnostic system for plant disease identification using recent EfficientNet CNN

*For correspondence

models. They trained the models with and without data augmentation images. Their accuracy rates are between 99.91% and 99.97%, depending on original and augmented image datasets. Abbas *et al* [12], adopted DenseNet121 architecture for tomato plant leaves diagnosis. They considered both original and synthetic tomato diseased leaves of ten prevalent classes. Original images are collected from the plant village repository, and Conditional-GAN generates synthetic images. After training, they attained an overall accuracy of 99.51% (5 classes), 98.65% (7 classes), and 97.11% (10 classes), respectively. Similar experiments on tomato crops [21, 22] have also been undertaken.

Numerous plant disease frameworks have also been presented for real field condition images. In xu *et al* [18], an Attentional Pyramid framework (APN) for custom herbal image recognition was developed with satisfiable accuracy. Mukhopadhyay *et al* [19] proposed a genetic algorithm based on an image clustering approach for tea plant leaves diagnosis, with a success rate of 83%. Pantazi *et al* [20], reported a recognition accuracy 95% on vine leaves images. Lu *et al* [15] developed an automated CNN diagnostic system using a cross-validation approach to recognize the ten rice disease classes, achieving 95.48% overall accuracy.

Despite the effectiveness, numerous critical issues are still there in the solutions indicated above. Initially, the efficacy of the deep learning CNN models significantly relies on the massive amount of datasets for training. Currently, the datasets available for plant disease diagnosis tasks are limited. These datasets contain laboratory condition standard single leaf images with coherent backgrounds. Recent investigations have demonstrated that the detection performance of the diagnostic system is drastically diminished when tested with real field images [16].

Consequently, their detection approaches cannot be implemented in practice. On the other hand, the limited trials for disease recognition result in low accuracy and lead to the temporal analysis of disease conditions. To overcome the abnormalities mentioned barriers, we collect the real field images captured at different lighting conditions and viewpoint locations.

This study includes the following significant contributions:

1. An end-to-end deep ensemble neural network (D-ENN) for automated prevention and control of vegetable crop diseases and sensible use of fertilizers is proposed.
2. The loss functions of the proposed framework makes the system extremely resilient against test data using real cultivated field images.
3. The empirical results of the D-ENN model surpass existing benchmark classification methods.

The rest of the paper is organised as follows: Section 2 discusses the proposed architecture and implementation details. Section 3 describes the dataset details and empirical

analysis of the D-ENN of the proposed approach. Finally, the findings and conclusion are discussed in section 4.

2. Methodology

2.1 C-GAN network as a data augmentation method

Data augmentation methods are employed to solve the problem of insufficient datasets. The most common classical data augmentation technique combines affine image transformations [35]. Various data augmentation methods were applied to the dataset, such as image rotation, flipping, transformation, zooming, shifting, and noise addition. These methods generate images with similar distributions as the original data. Therefore, the classification models cannot be trained with more information than the original. As a result, the model's ability to develop is constrained, necessitating the generation of different distributions to boost sample diversity.

Over recent years, the profound deep learning techniques used deep generative models as data augmentation technique such as Variational Auto-encoders (VAE), Generative Adversarial Network (GAN), and Conditional GAN (C-GAN). These models learn the feature distributions from the original images and generate images with similar distributions, adding additional information to the classification model. In this work, C-GAN [23] has been used as a data augmentation technique to contribute synthetic images from original images to prevent overfitting problems while training the network.

Formally, GANs are composed of two adversarial Convolutional neural networks models. The first model is the generator model that generates fake samples directly depending on the input vector from the random noise. In contrast, the adversary discriminator model differentiates the fake samples made by the generator from the real samples obtained from the original image dataset. Both the models are trained simultaneously, competing to outperform each other. Aside from the training, generator model strives to deceive discriminator model by generating progressively better samples. On the other hand, discriminator model trained to become more resilient in distinguishing images from fake and real distributions.

Conditional-GAN also comprises of generator G and discriminator D, both getting extra auxiliary conditional input data such as the current image class label or another attribute. Conditional input enables the network to resolve the overfitting problem and long convergence time. The overall layered structure of the C-GAN generator and discriminator is shown in figure 1.

Mathematically, X represents data samples, and 'C' denotes a class label data fed to both generator G and discriminator D. In addition, a random noise distribution

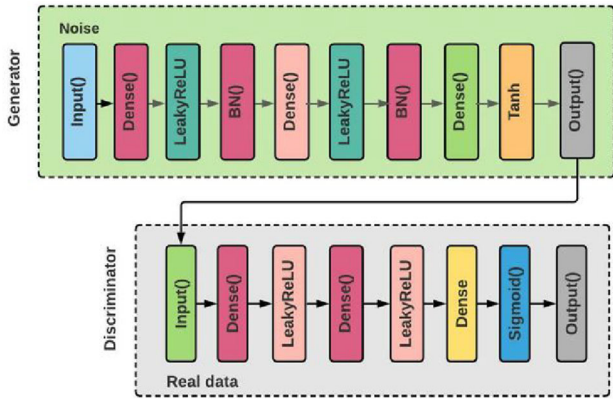


Figure 1. Structure of C-GAN.

given to G is indicated by $P_z(z)$. The goal of the D is to maximize the log-likelihood of correctly assigning class labels to the real and fake images generated by the G, given by $\log D(X/C)$, whereas G aims to maximize the generator loss denoted as $\log(1 - D(G(z/C)))$. The overall loss function of conditional-GAN is expressed as a two-player min-max game as given by equation (1)

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log D(X/C)] + \mathbb{E}_{z \sim p_z(z)} [\log [1 - D(G(z/C))]] \quad (1)$$

Initially, C-GAN trained with the original vegetable leaf dataset to generate sythetic diseased images. Suppose, a given leaf image dataset $\mathcal{LD} = \{(X^{(n)}, C^{(n)})\}_{n=1}^N$, where $X^{(n)}$ indicates a given leaf image and $C^{(n)} \in \{0, 1, 2, \dots, 12\}$, indicates its associated class label. A real leaf image $X^{(n)}$ with respective class label $C^{(n)}$ fed as input to the discriminator model to train C-GAN. At the same time, a random noise vector z and a class label $C^{(n)}$ is fed to the generator model. Afterward, the generator generates fake images F_i and passes them into the discriminator model. Finally, the discriminator attempts to differentiate the real and fake images. In this manner, C-GAN was trained on a custom vegetable leaves dataset as shown in figure 2. Finally, the synthetic images of each vegetable crop diseased leaves category are generated by C-GAN. All these real and synthetic images are integrated as input to the proposed D-ENN Network.

2.2 Ensemble Network architecture - D-ENN

Figure 3 depicts the ensemble architecture of the proposed model. The model comprised two simultaneous networks: ensemble network E_N and prediction network P_N . Initially, the training dataset, $S = \{S_1, S_2, \dots, S_K\}$ contains both original and synthetic images of vegetable plant diseased images

with class labels as ground truth. Let O_k is the ground truth of class label S_k . Each sample image S_k is fed into the ensemble network E_N . Ensemble network contains the M distinct Convolution Neural Networks (CNNs). Each CNN is composed of softmax and Fully Connected (FC) layers. For a given sample image S_k , each CNN generate the individual features $[f_1(S_k), f_2(S_k), \dots, f_N(S_k)]$. These features are combined to generate the final ensembled deep features, which functions as input to the prediction network P_N . The prediction Network P_N comprises three Fully connected (FC) layers and one softmax layer. This network predicts the class label for each sample image S_k according to the ensemble features $f(S_k)$, i.e., $\widehat{O}_k = o(f(S_k))$. The diseased vegetable leaf is identified when each image S_k is fed to the proposed D-ENN network.

Earlier deep learning ensemble approaches are fine-tuned after training the ensemble model. In contrast to these prior methods, a unique ensemble framework including two losses - ensemble loss (\mathcal{L}_{Ens}) and prediction loss (\mathcal{L}_{Pre}) to upgrade the ensemble network and prediction network concurrently while training the model is designed in this work. Two losses are added together and propagated back to the ensemble and prediction networks during each epoch. The proposed end-to-end D-ENN system can learn bigoted fine-grained features and tackle the issues of vegetable crop disease. The overall loss (\mathcal{L}_{Total}) function to optimize the proposed D-ENN network is represented using equation (2).

$$\mathcal{L}_{Total} = \mathcal{L}_{Ens} + \lambda \mathcal{L}_{Pre} \quad (2)$$

where \mathcal{L}_{Ens} and \mathcal{L}_{Pre} indicate the ensemble loss and prediction loss, respectively. λ refers to a constant weighted loss value. The ensemble loss \mathcal{L}_{Ens} directs the ensemble network to extract the fine-grained features generated by each CNN based on the class labels. This loss function is based on the categorical cross-entropy of all CNNs obtained in the equation. Each CNN has the last softmax layer to act as an activation function. The ensemble loss \mathcal{L}_{Ens} for m^{th} CNN is measured by equation (3).

$$\mathcal{L}_{Ens,m} = -\frac{1}{K} \sum_{k=1}^K O_k \log o_m(f_m(S_k)) \quad (3)$$

where $o_m(f_m(S_k))$ indicates the softmax prediction of m^{th} CNN of m^{th} feature.

Similarly, equation (4) gives the prediction loss \mathcal{L}_{Pre} on the training dataset S . w_e indicate the total parameters of all CNNs. The prediction network employs the deep features supplied by $f(S_k)$. This characteristic drives the prediction network to classify the diseased leaves images with a correct class label from deep learned features. Also, the prediction loss boosts the ensemble network to update the CNNs, as each CNN generate consistent, fine-grained features with that of prediction network P_N .

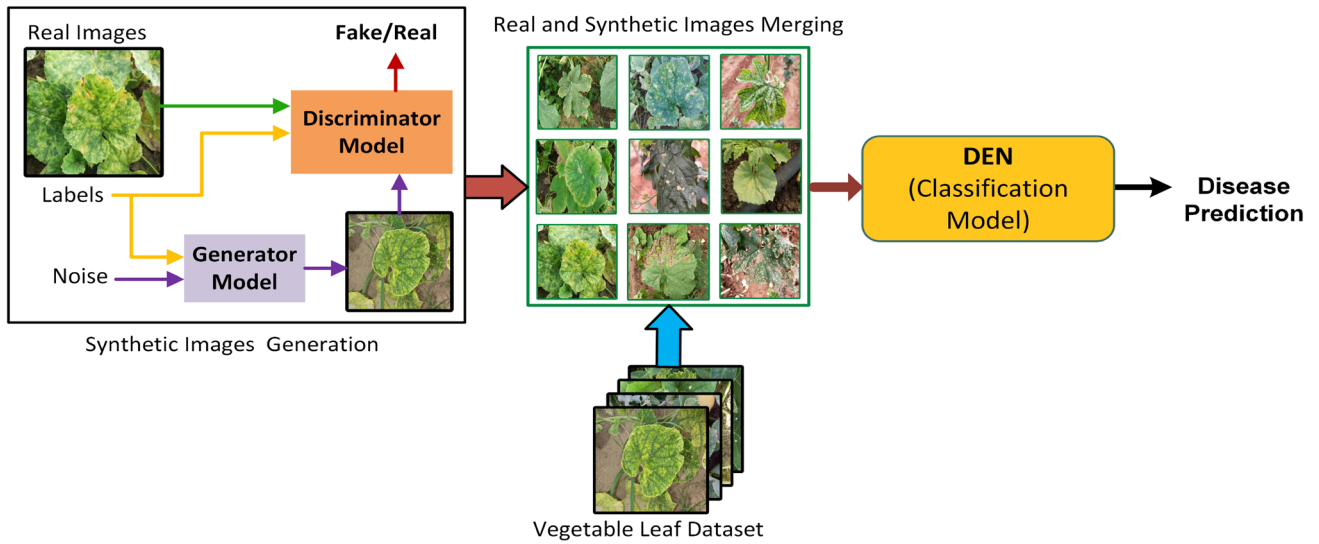


Figure 2. Synthetic Image Generation using Conditional- GAN.

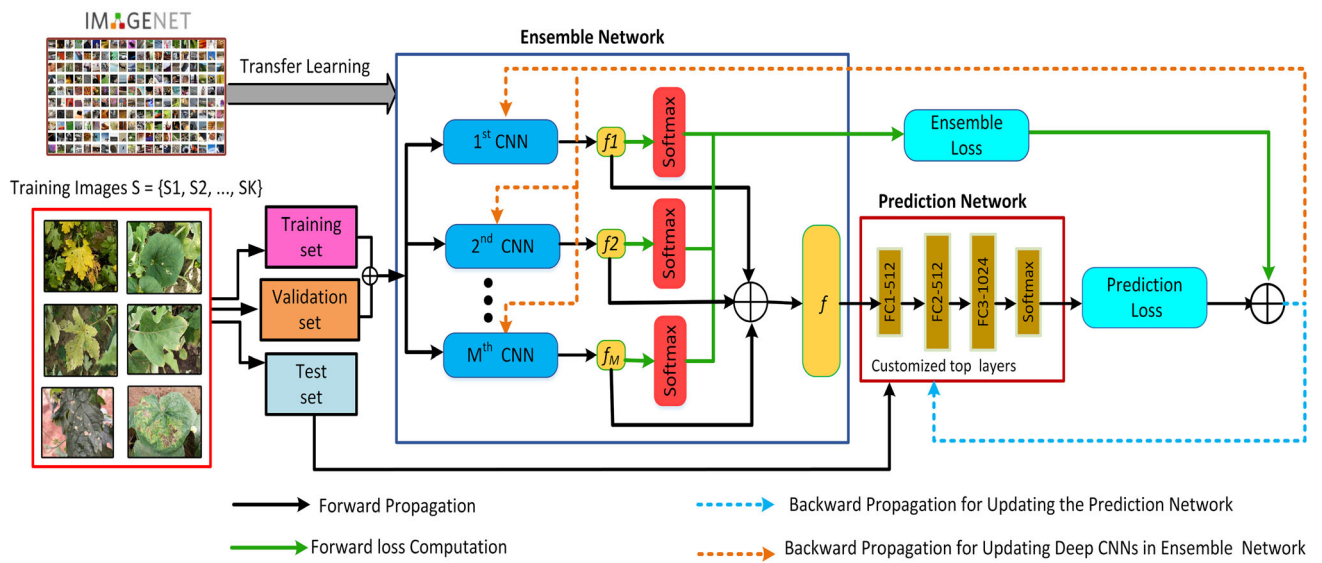


Figure 3. The Overall Pipeline of the Deep Ensemble Neural-(D-ENN) Framework.

$$\mathcal{L}_{Pre}(S, w_e) = -\frac{1}{K} \sum_{k=1}^K O_k \log o(f(S_k)) \quad (4)$$

2.3 Implementation details D-ENN network

In this work, as part of the proposed (D-ENN) design, four different benchmark CNN models includes Inceptionv3 [24], MobileNetv2 [25], DenseNet121 [26], and EfficientNetB0 [27] are used. Each CNN in the ensemble network (EN) generates fine-grained features. These

features are combined and, given as input to the prediction network (PN). The present work employs the approach of transfer learning to tackle the small or limited dataset issues of diseased leaf identification studies. The transfer learning technique is a method that optimizes pre-trained model parameters by training them with the target domain dataset [36]. The source domain is the autonomous plant disease scenario, and the target domain is the vegetable disease detection. Given that the CNN models have been well trained on an extensive dataset in the source domain, fine-tuning these pre-trained models with a specific dataset in the target domain might yield significant results. An added

benefit of these techniques helps the proposed model to improve detection performance in less time. In this work, All the CNNs are pretrained by large-scale ImageNet [28]. These CNNs have tremendous image analysis accomplishments and can classify 1000 wide range of objects. So, the top prediction layers of the pre-trained CNN models are removed and incorporated in the and incorporate the custom prediction top layers to classify the diseased leaves images. Each CNN is combined with one fully connected layer (each layer with 512 nodes) and a softmax layer. Then, the softmax layer helps to compute the final ensemble loss. Lastly, the prediction network is used to produce the classification task. The deep features of all the CNNs are integrated and passed to a prediction network. It contains three fully connected layers; each layer contains 512 (FC-1), 512 (FC-2), and 1024 (FC-3) nodes. Here, the prediction loss is computed by the softmax layer of the prediction network.

3. Experimental results and analysis

The experiments are conducted on Intel (R) Xeon (R) E5-2698v4 processors, NVIDIA TeslaV100, and 32GB memory per GPU. The proposed D-ENN framework is implemented using deep neural network libraries, using the python Keras Tensorflow framework.

3.1 Dataset

Custom field vegetable leaves images were taken from different cultivated vegetable fields with wide variations in illumination and weather conditions in this study. The original vegetable dataset is imbalanced and has 1,986 images included with 12 total classes. Among the twelve classes, 11 classes are about diseases of vegetable plant leaves, and 1 class pertains to all healthy leaves. The disease classes (OYMV, RGALS, OCLF, CALB) contain a limited number of images and are heavily biased to specific disease classes. Synthetic images are generated using C-GAN for specific diseased classes. Adding the synthetic images generated by C-GAN to the original data set, the volume of the data increased by nearly five times. The Vegetable dataset with class-wise distribution of images is presented in table 1. Acronyms were given for each class, and all the images were resized to 256x256 to speed up computations. The total vegetable dataset has been divided into three subsets: training set, validation set, and test set, in the ratio of 70:10:20 to mitigate the overlapping conflicts between the sets. The training and validation sets were used to train the D-ENN model, and the test set was used to assess the model performance.

Table 1. Dataset statistics.

Crop_Disease_Name	Label_Name	Total_Images
Cucumber_Alternaria leaf blight	CALB	590
Cucumber_Herbicide burn	CHB	1706
Cucumber_Mosaic virus	CMV	1086
Okra_Cercospora Leaf Spots	OCLF	576
Okra_Yellow Vein Mosaic Virus	OYMV	370
Bitter Gourd_yellow mosaic virus	BGYMV	786
Bitter Gourd_Herbicide burn	BGHB	582
Eggplant_Alternaria leaf spot	EALS	1100
Ridge gourd_Alternaria Leaf Spot	RGALS	470
Ridge gourd_Aphid damage	RGAD	1008
Ridge gourd_Mosaic disease	RGMD	690
Healthy	HEALTHY	950

3.2 Evaluation metrics

Two series of evaluation metrics were carried out in this study. The first series of metrics belong to evaluate the performance of C-GAN. C-GAN model is trained to generate synthetic images of training set with 250 epochs. At each epoch, the weights are updated to both generator and discriminator models to produce the synthetic images near actual images. After training, the C-GAN model generates a total of 3250 vegetable leaves images. The metrics Peak-Signal-to-Noise-Ratio (PSNR) and Inception score (IS_{gen}) measure the quality of the generated images by C-GAN respectively.

$$PSNR = 20 \cdot \log_{10}(MAX_I) - 10 \cdot \log_{10}(MSE) \quad (5)$$

$$MSE = \frac{1}{mn} \sum_{i=0}^{p-1} \sum_{j=0}^{q-1} [S(i,j) - T(i,j)]^2 \quad (6)$$

where, MAX_I denotes the maximum level of pixel intensity, generally 255 and MSE denotes the mean squared error.

$$IS_{C-GAN} = \exp(E_{s \sim p_g} D_{KL}(p(t|s) || p(t))) \quad (7)$$

where, $s \sim p_g$ is the sample image s taken from g , $p(t|s)$ refers the conditional label distribution, $p(t)$ is marginal label distribution. Higher inception scores indicate a better outcome, whilst $D_{KL}(p(t|s) || p(t))$ indicates the Kullback Leiblerm (KL) - divergence corresponding to larger value between the two distributions.

The second series of metrics is used to evaluate the proposed D-ENN system trained on original and synthetic images. Since the classification task for the plant disease detection problem is associated with 12 classes, the performance metrics are computed using equations (8), (9), (10), and (11) respectively.



Figure 4. Sample instances of original and synthetic images of vegetable dataset.

$$Accuracy = \frac{\sum_{cls=1}^{12} TP_{cls} + \sum_{cls=1}^{12} TN_{cls}}{\sum_{cls=1}^{12} TP_{cls} + \sum_{cls=1}^{12} TN_{cls} + \sum_{cls=1}^{12} FP_{cls} + \sum_{cls=1}^{12} FN_{cls}} \quad (8)$$

$$Precision = \sum_{cls=1}^{12} \frac{TP_{cls}}{TP_{cls} + FP_{cls}} \quad (9)$$

$$Recall = \sum_{cls=1}^{12} \frac{TP_{cls}}{TP_{cls} + FN_{cls}} \quad (10)$$

$$F1 - Score = \sum_{cls=1}^{12} \frac{2 * Precision_{cls} * Recall_{cls}}{Precision_{cls} + Recall_{cls}} \quad (11)$$

Table 2. PSNR Analogy.

Label_Name	PSNR_Real (dB)	PSNR_(Real + Syntehtic) (dB)
Cucumber_Alternaria leaf blight	27.147	27.107
Cucumber_Herbicide burn	27.124	27.612
Cucumber_Mosaic virus	27.016	27.605
Okra_Cercospora Leaf Spots	28.947	29.104
Okra_Yellow Vein Mosaic Virus	28.101	28.524
Bitter Gourd_yellow mosaic virus	27.894	27.920
Bitter Gourd_Herbicide burn	27.192	27.706
Eggplant_Alternaria leaf spot	27.194	27.095
Ridge gourd_Alternaria Leaf Spot	28.861	28.349
Ridge gourd_Aphid damage	28.106	27.916
Ridge gourd_Mosaic disease	27.610	27.101
Healthy	27.107	27.741

Table 3. Mean Inception score.

Image_category	Inception score
Real Images	2.642 ± 0.146
Synthetic Images	2.968 ± 0.203

3.3 Optimization and hyper-parameter selection

The overall vegetable leaves images are divided into 12 classes throughout the training. With the batch size of 16 images, momentum as 0.8 and weight decay as 0.0001 considered. Adam optimizer [29], is used for quick convergence. After 200 epochs, the model training is completed. Additionally, examine the λ and learning rate to analyze the correlation between ensemble loss (\mathcal{L}_{Ens}) and prediction loss (\mathcal{L}_{Pre}) in the D-ENN model. Initially, λ is set to 0, and then the ensemble loss (\mathcal{L}_{Ens}) is considered when back propagating the model. With an increase in the value of λ , the significance of prediction loss (\mathcal{L}_{Pre}) also increases. As part of the evaluation, the value of λ was varied from 0 to 2 to observe the impact of λ while training

the D-ENN model. In addition to λ , the learning rates also influence model convergence. The learning rate tweaked from 10^{-3} to 10^{-5} to examine whether the network converged. In the proposed D-ENN model, loss \mathcal{L}_{Ens} reduces all CNN losses to gain in-depth fine-grained features, loss \mathcal{L}_{Pre} reduces the prediction losses to achieve accurate prediction outcomes. In this way, appropriately integrating \mathcal{L}_{Ens} and \mathcal{L}_{Pre} allows the D-ENN network to be optimized more effectively. Finally, the empirical findings indicate that the D-ENN model with a learning rate = 10^{-4} and $\lambda = 0.5$ has the best F-measure value.

3.4 Performance analysis of C-GAN and D-ENN Network

Figure 4 shows original vegetable leaf images and synthetic leaf images generated by C-GAN. It exhibits that synthetic images closely resemble the original images. Table 2 presents the relation between real and synthetic images regarding PSNR values. The PSNR values reveal that the images generated by C-GAN are almost similar to the

Table 4. Comparison of D-ENN framework results and execution time with different benchmark methods.

Model	Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Average Time (sec)
VGG [31]	Method_VGG19	82.41	82.62	81.41	82.01	0.005
ResNet [32]	Method_ResNet101	83.40	84.21	82.11	83.14	0.006
Inception [33]	Method_InceptionV3	84.21	85.32	83.91	84.60	0.007
MobileNet [25]	Method_MobileNetV2	85.90	86.20	82.71	84.41	0.006
DenseNet [26]	Method_DenseNet169	82.61	84.21	83.92	84.06	0.005
Xception [34]	Method_Xception	87.01	86.10	88.91	86.49	0.006
EfficeintNet [27]	Method_EfficeintNetB0	88.42	88.25	87.52	87.88	0.005
AV_CNN [30, 31]	Method_AV_CNN	89.64	88.65	88.92	88.78	0.021
DR_CNN [26, 32]	Method_DR_CNN	90.15	90.40	89.91	90.15	0.034
Proposed	D-ENN	96.02	95.71	95.32	95.51	0.045

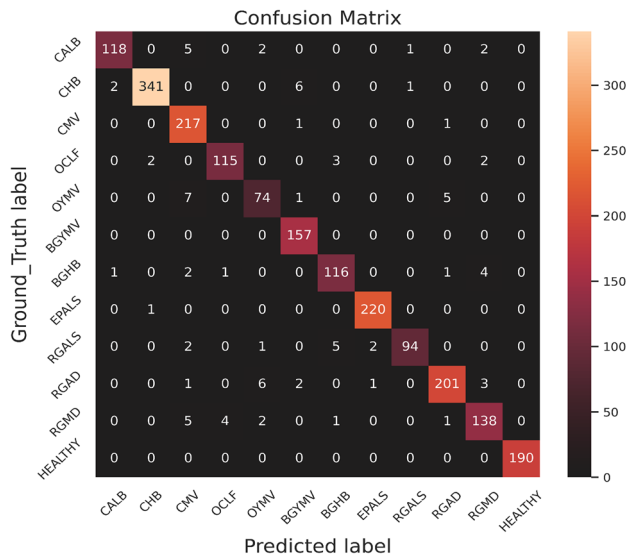


Figure 5. Confusion matrix for D-ENN framework.

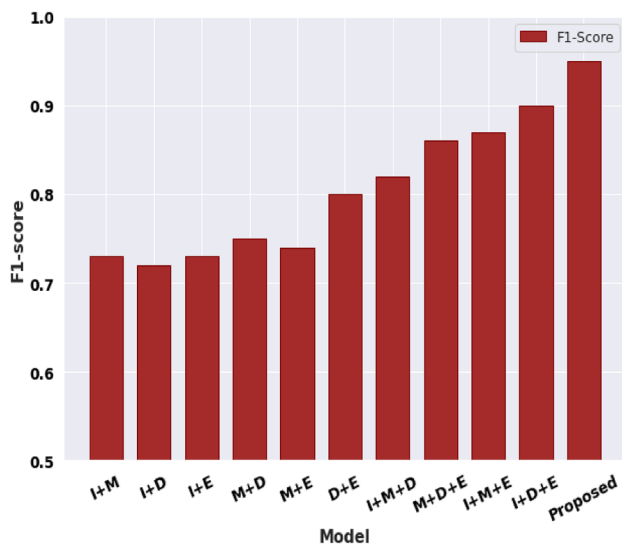


Figure 6. Accuracy comparison of proposed framework with various CNN combinations.

original images. The total mean inception scores were calculated on real and synthetic images, as shown in table 3. Synthetic image inception score is quite close to that of real images inception score.

The proposed D-ENN model is compared with several existing benchmark deep learning models and ensemble models. Five recent CNNs are compared with proposed model including ResNet50 [32], Inceptionv3 [24], MobileNetv2 [25], DenseNet121 [26], and EfficientNet-B0 [27]. Furthermore, two well known ensemble based deep learning models were examined, namely AV-CNN and DR-CNN. AV-CNN include the ensemble of Alexnet [30] and VGG [31], while DR-CNN includes a combination of

DenseNet169 [26] and ResNet [32]. These models are initially pretrained by ImageNet [28] and fine-tuned by a custom vegetable training dataset. The results and average computational time of all the above state-of-the-art methods are represented in table 4. EfficientNetB0 [27] achieves the best accuracy compared to other deep learning methods due to its compound scaling approach. Despite this, ensemble-based deep learning methods yield better results than deep learning methods. These findings suggest that ensemble deep learning models can learn fine-grained features for plant disease classification tasks. As shown in table 4, when compared with DR-CNN, AV-CNN give worse results since DR-CNN have deeper residual and dense block interconnections in the network.

DR-CNN relies on a majority voting technique for label prediction; our proposed scheme employs a prediction network rather than a majority voting scheme leading to accurate prediction label outcomes. Furthermore, our framework trained end-to-end way by concurrently updating the parameters of both ensemble and prediction networks, thus achieving the best performance metrics compared to AV-CNN and DR-CNN ensemble models and all other recent benchmark methods. Generally, ensemble based deep learning methods incorporate multiple CNNs; it takes longer to compute than approaches that employ one CNN. Even so, the proposed framework mean frame rate (seconds) is above 25, indicating that the proposed method is a practical solution to aid in plant diagnosis. Figure 5 presented the confusion matrix of the proposed D-ENN framework. As shown in figure 5, the x-axis represents the predicted outcome, and the y-axis represents the ground truth class labels of the plant diseased images.

3.5 Ablation studies

The ablations experiments were undertaken to exemplify the effectiveness of the proposed D-ENN framework by considering various combinations of Convolutional Neural Networks (CNNs). To overcome the GPU card and memory storage constraints, the four light-weight CNN models at a time, namely, Inceptionv3 [24], MobileNetv2 [25], DenseNet121 [26], and EfficientNetB0 [27] are simultaneously trained. And also considered the same dataset (train and test), and the total number of epochs for these experiments. For the sake of keeping things simple, each CNN abbreviated with specific notations. For example, I+M+E indicates that the proposed model is trained with Inceptionv2, MobileNetv3, and EfficientNetB0 to extract the fine-grained features. Ablations studies were determined by taking various combinations of CNNs as represented in terms of F1-Score as shown in figure 6. The union of CNNs, M+D gives the second-best F1-score. And, the union of D+E CNNs owned the first-best F1 score. Furthermore, the F1 score of I+D+E was the highest when three CNNs were combined. These results suggest that the

union of InceptionNetv2, MobileNetv3, and EfficientNetB0 CNNs can get better fine-grained features for plant disease classification tasks. Despite this, the proposed approach with four CNNs had the highest F1-score compared to all the above mentioned experiments.

4. Conclusion

In the present study, a novel ensemble-based D-ENN framework to solve the plant disease detection problems using real cultivated field images is developed. The proposed method also uses a Conditional Generative Adversarial network to generate synthetic image samples for specific limited disease classes. This kind of proposed augmentation approach enhances network generalizability and prevents overfitting problems. Afterward, the proposed model was trained on original and synthetic images. In contrast to the other existing deep ensemble models, the proposed model can learn in-depth features from different CNNs and successful prediction accuracy at the same time in an effective end-to-end way. To adequately train the parameters of the D-ENN network, both ensemble loss and the prediction loss are presented. Four CNNs are used in the architecture: InceptionNetv2, MobileNetv3, DenseNet, and EfficientNetB0. Ablation experiments suggest that combining several CNNs and losses for plant disease categorization is beneficial. The experiments have proven that the proposed technique is significantly superior to the competing benchmark models. In Future, the proposed approach will be integrated with diagnostic methods to analyze pathological parameters and examine the infected portions in plant diseased images.

Acknowledgements

The authors would like to thank High-Performance Computing Lab, Department of Computer Applications, National Institute of Technology, Tiruchirappalli, India, for providing computing facilities for conducting the experiments.

References

- [1] Tai, Amos P K, Maria Val Martin and Colette L Heald 2014 Threat to future global food security from climate change and ozone air pollution. *Nature Climate Change* 4.9 : 817–821
- [2] Singh Vijai, Namita Sharma, and Shikha Singh 2020 A review of imaging techniques for plant disease detection. *Artificial Intelligence in Agriculture*
- [3] Yang Xin, and Tingwei Guo 2017 *Machine learning in plant disease research.*: 1
- [4] Wani Javaid Ahmad *et al* 2021 Machine Learning and Deep Learning Based Computational Techniques in Automatic Agricultural Diseases Detection: Methodologies, Applications, and Challenges. *Archives of Computational Methods in Engineering*: 1–37
- [5] Bock C H *et al* 2010 Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging. *Critical reviews in plant sciences* 29.2 : 59–107
- [6] Pagola Miguel *et al* 2009 New method to assess barley nitrogen nutrition status based on image colour analysis: comparison with SPAD-502. *Computers and electronics in agriculture* 65.2 : 213–218
- [7] Sannakki Sanjeev S *et al* 2011 Leaf disease grading by machine vision and fuzzy logic. *Int J* 2.5 : 1709–1716
- [8] Farabet Clement *et al* 2012 Learning hierarchical features for scene labeling. *IEEE transactions on pattern analysis and machine intelligence* 35: 1915–1929
- [9] Krizhevsky Alex, Ilya Sutskever and Geoffrey E Hinton 2017 ImageNet classification with deep convolutional neural networks. *Communications of the ACM* 60: 84–90
- [10] Coulibaly Solemane *et al* 2019 Deep neural networks with transfer learning in millet crop images. *Computers in Industry* 108: 115–120
- [11] Selvaraj Michael Gomez *et al* 2019 AI-powered banana diseases and pest detection. *Plant Methods* 15: 1–11
- [12] Abbas Amreen *et al* 2021 Tomato plant disease detection using transfer learning with C-GAN synthetic images *Computers and Electronics in Agriculture* 187: 106279
- [13] Wang, Yibin, Haifeng Wang and Zhaohua Peng 2021 Rice diseases detection and classification using attention based neural network and Bayesian optimization. *Expert Systems with Applications* 178: 114770
- [14] Sun Henan *et al* MEAN-SSD: 2021 A novel real-time detector for apple leaf diseases using improved light-weight convolutional neural networks. *Computers and Electronics in Agriculture* 189: 106379
- [15] Xie Xiaoyue *et al* 2020 A deep-learning-based real-time detector for grape leaf diseases using improved convolutional neural networks. *Frontiers in plant science* 11 : 751
- [16] Mohanty, Sharada P, David P Hughes and Marcel Salathé. 2016 Using deep learning for image-based plant disease detection. *Frontiers in plant science* 7 : 1419
- [17] Atila UÜmit *et al* 2021 Plant leaf disease classification using EfficientNet deep learning model. *Ecological Informatics* 61: 101182
- [18] Xu Yingxue *et al* 2021 Multiple attentional pyramid networks for Chinese herbal recognition. *Pattern Recognition* 110: 107558
- [19] Mukhopadhyay Somnath *et al* 2021 Tea leaf disease detection using multi-objective image segmentation. *Multi-media Tools and Applications* 80: 753–771
- [20] Pantazi, Xanthoula Eirini, Dimitrios Moshou, and Alexandra A. Tamouridou 2019 Automated leaf disease detection in different crop species through image features analysis and One Class Classifiers. *Computers and electronics in agriculture* 156: 96–104
- [21] Gadekallu Thippa Reddy *et al* 2020 “A novel PCA-whale optimization-based deep neural network model for classification of tomato plant diseases using GPU.” *Journal of Real-Time Image Processing*: 1–14
- [22] Gonzalez-Huitron Victor *et al* 2021 Disease detection in tomato leaves via CNN with lightweight architectures

- implemented in Raspberry Pi 4. *Computers and Electronics in Agriculture* 181: 105951
- [23] Mirza M and Osindero, S 2014 Conditional generative adversarial nets, arXiv preprint [arxiv:1411.1784](https://arxiv.org/abs/1411.1784)
- [24] Xia, Xiaoling, Cui Xu, and Bing Nan 2017 Inception-v3 for flower classification. *2nd International Conference on Image, Vision and Computing (ICIVC)*. IEEE
- [25] Sandler Mark *et al* 2018 Mobilenetv2: Inverted residuals and linear bottlenecks. *Proceedings of the IEEE conference on computer vision and pattern recognition*
- [26] Huang Gao *et al* 2017 Densely connected convolutional networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*
- [27] Tan Mingxing, and Quoc Le 2019 Efficientnet: Rethinking model scaling for convolutional neural networks. *International Conference on Machine Learning*. PMLR
- [28] Deng Jia *et al* 2009 Imagenet: A large-scale hierarchical image database. *IEEE conference on computer vision and pattern recognition*
- [29] Kingma, Diederik P and Jimmy Ba 2014 Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*
- [30] Alom Md Zahangir, *et al* The history began from alexnet: A comprehensive survey on deep learning approaches. arXiv preprint [arXiv:1803.01164](https://arxiv.org/abs/1803.01164)
- [31] Simonyan, Karen and Andrew Zisserman 2014 Very deep convolutional networks for large-scale image recognition.” arXiv preprint [arXiv:1409-1556](https://arxiv.org/abs/1409.1556).
- [32] He Kaiming *et al* 2016 Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*
- [33] Szegedy, Christian *et al* 2015 Going deeper with convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*
- [34] Chollet François 2017 Xception: Deep learning with depth-wise separable convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*
- [35] Mikołajczyk, Agnieszka and Michał Grochowski 2018 Data augmentation for improving deep learning in image classification problem. *International interdisciplinary PhD workshop (IIPhDW)*
- [36] Zhuang Fuzhen *et al* 2020 A comprehensive survey on transfer learning. *Proceedings of the IEEE* 109: 43–76