A Deep Ensemble Approach for Recognition of Papaya Diseases using EfficientNet Models

Rashidul Hasan Hridoy, Mosammat Rokeya Anwar Tuli
Department of Computer Science and Engineering
Daffodil International University
Dhaka, Bangladesh
rashidul15-8596@diu.edu.bd, rokeya15-8925@diu.edu.bd

Abstract—Diseases of papaya impeded quality production and caused severe financial damages to growers. An efficient diagnosis approach for papaya diseases is enormously desired to control and prevent the spread of diseases. At first, using a dataset of 138980 images of affected and healthy leaves and fruits of papaya which was generated with image augmentation techniques from 13898 collected images, eight models of EfficientNet between B0 and B7 were trained via transfer learning technique to recognize eight diseases. Afterward, finetuned versions of the three best-performing models were selected for ensemble learning such as EfficientNet B5, B7, and B6, which achieved 98.13%, 96.93%, and 96.87% accuracy under the test set of 6931 images, respectively. The deep ensemble model showed more effective recognition performance than single models, and test accuracy increased by 1.61%. The experimental result demonstrates that the proposed ensemble model can recognize papaya diseases more efficiently than single models of EfficientNet.

Keywords—Papaya Disease Recognition, Deep Learning, Transfer Learning, Ensemble Learning, EfficientNet

I. Introduction

Papaya (scientific name: Carica papaya) is a widely consumed vegetable in several countries of Southeast Asia and South America which is an enormous source of different types of essential nutrition and provides significant health benefits. In Bangladesh, a remarkable number of grower cultivates papaya, which is locally known as Pepe [1]. Cultivation of papaya is immensely lucrative, but it is highly prone to several types of diseases, which causes intense economic losses to the agricultural economy. The existing methods of papaya disease recognition are principally based on visual recognition, which is a time-consuming and arduous task. An efficient and timely recognition approach of papaya diseases can assist growers in taking timely treatment on the papaya plants and thereby can decrease the pecuniary damages substantially.

In recent years, object recognition and computer vision have made remarkable advances. Deep learning-based classification approaches have exhibited great promise in image recognition, which is mainly based on the learning of features from a labeled training set of an image dataset. Moreover, in solving intricate problems, deep learning approaches have shown the significant ability of generalization than traditional approaches in computer vision. Recent improvements in convolutional neural networks (CNN) have significantly enhanced the image recognition ability of deep learning-based classification approaches. Till now, CNN is considered to be an efficient and optimal algorithm for pattern classification. In recent years, these

efficient deep learning-based disease diagnosis approaches are extensively applied in agriculture to recognize diseases of several types of crops. Training of large CNN architecture with a large dataset is highly time-consuming, on the other hand, the approach of using a pre-trained model is more convenient in terms of reducing training time and obtaining significant recognition accuracy than the approach of training a CNN architecture from scratch.

During cultivation, papaya is severely affected by various types of diseases such as powdery mildew, mosaic, leaf curl, ringspot, brown spot, black spot, anthracnose, and phytophthora blight. Using nine image augmentation techniques, a dataset of 138980 images was generated which contains one healthy leaf, five diseased leaves, one healthy fruit, and five diseased fruits classes. All images of this dataset were randomly divided into training, validation, and testing sets which contain 111198, 20851, and 6931 images, respectively.

Initially, eight models of the EfficientNet group between B0 and B7 were used via the transfer learning technique in this study [2]. For establishing more proficient recognition methods, EfficientNet was introduced which obtains more efficient performance by uniformly scaling depth, width, and resolution while scaling down the architecture. The classification performance of these models was examined using images of the test set. According to the result of performance evaluation, among eight models of EfficientNet, EfficientNet B5, B6, and B7 performed significantly better than others in this study and obtained 98.13%, 96.87%, and 96.93% test accuracy, respectively.

Ensemble learning provides more efficient classification performance by merging the features obtained from deep models into a classifier of good quality [3]. Ensemble model significantly decreases generalization error and variance of predictions, enhances the computational training, and provides remarkable performance with fewer training images. EfficientNet B5, B6, and B7 models were used to generate a deep ensemble model for recognizing diseases of papaya more efficiently which acquired 99.74% test accuracy. In this study, three single models of EfficientNet were combined to consolidate strengths, and the performance of single and ensemble models was examined with several experimental studies. The major contributions of this study are summarized as follows:

 An improved dataset of 138980 images of papaya leaves and fruits is created since no appropriate dataset is available for efficient recognition of papaya diseases till now.

- A deep ensemble model is proposed in this study for classifying diseases of papaya which obtained 99.74% test accuracy after evaluating the performance of single models and ensemble of three best-performing models of EfficientNet.
- To the best of our knowledge, this study is the first endeavor for recognizing five diseased leaves and five diseased fruits classes of papaya along with healthy leaf and fruit class using the ensemble approach with deep neural networks.

The structure of this paper is as follows: Section II represents related works. Section III describes the dataset and the proposed deep ensemble approach in detail. We describe experimental studies in Section IV. The results acquired and their interpretations are demonstrated in Section V. Lastly, the conclusion and future work are presented in Section VI.

II. RELATED WORKS

A remarkable number of researchers have introduced several image processing and computer vision-based recognition approaches in recent years for classifying diseases of crops using machine learning and deep learning algorithms. Md. Tarek Habib et al. addressed a machine vision-based disease recognition approach for classifying five diseases of papaya, used k-means clustering for segmenting the affected regions of images, and for classifying images decision tree (DT), Support vector machine(SVM), and Naive Bayes (NB) classifiers were used [4]. Among three classifiers, SVM showed more significant classification performance than others that acquired 95.2% accuracy with a 2.88% falsepositive rate. On the other hand, DT and NB acquired 86.67% and 77.78% accuracy with 8.0% and 13.33% false-positive rates, respectively. Sachin B. Jadhav et al. used AlexNet and GoogLeNet to identify plant diseases [5]. AlexNet performed better than GoogLeNet which acquired 98.75% accuracy, and GoogLeNet acquired 96.25% accuracy in their study. Mohit Agarwal et al. compared their proposed model with three pretrained models to detect tomato leaf diseases [6]. Their proposed model acquired 91.20% accuracy which was higher than three pre-trained models. VGG16 performed better than the other two pre-trained models and acquired 77.20% accuracy. The storage space of their proposed model was also smaller than used pre-trained models. Ramesh et al. used an optimized deep neural network for recognizing paddy leaf diseases [7]. The texture and color features of paddy leaves were used in their study. Their proposed model obtained 97.00% test accuracy which was higher than others. The Jaya optimization algorithm was played a crucial role in their study in achieving higher accuracy with five classes. To train their proposed model, Melike Sardogan et al. fed the learning vector quantization with the output feature vector in the convolution layer and achieved an average of 86.00% accuracy in classifying images of five classes of tomato leaves [8]. Norhalina Senan et al. introduced a classification approach to recognize diseases and pests of paddy leaf, and their proposed CNN architecture obtained 90.30%, and 96.60% accuracy with 50:50, and 70:30 ratio of training and test images [9]. Siti Zulaikha Muhammad Zaki et al. addressed a classification approach for recognizing four types of leaf diseases of tomato using MobileNet V2 and obtained 95.94% accuracy [10]. In their study, the performance of several optimizers was analyzed, and the adagrad optimizer performed significantly better than others. Dhiman Mondal et al. introduced an approach using an NB classifier for

classifying diseases of the leaf of okra and achieved 87.00% accuracy with four classes [11]. Sharada Prasanna Mohanty et al. addressed an automated approach for classifying diseases of several types of plants using deep learning, and the performance of five different ratios of training and test images was analyzed in their study [12]. AlexNet and GoogLeNet were trained in their study using three different types of datasets of 38 classes, and achieved 99.27% and 99.34% accuracy, respectively. Rashidul Hasan Hridoy et al. proposed a recognition approach to classify 20 different types of skin disorders, eight models of EfficientNet were trained with the transfer learning technique in their study and achieved 97.10% accuracy with the EfficientNet B7 model [13]. Khalid El Asnaoui introduced an ensemble deep learning model for classifying pneumonia disease using InceptionResNet-V2, ResNet50, and MobileNet-V2, and the ensemble model of ResNet50 and MobileNet-V2 obtained 95.17% accuracy [4].

According to these related works, CNNs achieved remarkable results in classifying diseases of different types of crops. However, till now no study is conducted to recognize diseases of papaya using the deep ensemble approach with 12 classes. Hence, a deep ensemble approach of EfficientNet models is addressed in this study for efficient recognition of 12 classes of the used dataset.

III. MATERIALS AND METHODS

A. Dataset

A dataset of 138980 images of papaya leaves and fruits is created by utilizing image augmentation techniques from 13898 collected images. Figure 1 represents 12 representative images belonging to each class of the dataset.



Fig. 1. Sample of dataset: 1) healthy leaf 2) powdery mildew (leaf) 3) mosaic 4) leaf curl 5) ringspot (leaf) 6) brown spot 7) healthy fruit 8) powdery mildew (fruit) 9) black spot 10) Ringspot (fruit) 11) anthracnose 12) phytophthora blight.

To eradicate overfitting issue during the training phase of EfficientNet models, image augmentation techniques such as high brightness, low brightness, cropping, horizontal flip, 90-degree rotation, 180-degree rotation, 270-degree rotation, high contrast, and low contrast were utilized in this study for generating 125082 images from 13898 collected images. Figure 2 represents the image generation process. During the training phase, 111198 training and 20851 validation images were used to train and fit the models of EfficientNet and 6931 testing images were used for examining the classification performance of these models.

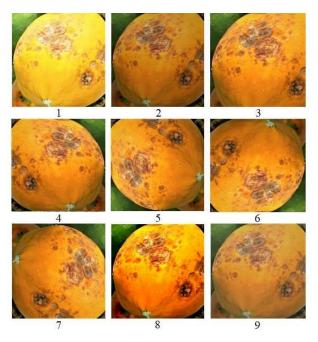


Fig. 2. Process of image generation using image augmentation: 1) high brightness, 2) low brightness, 3) cropping, 4) horizontal flip, 5) 90-degree rotation, 6) 180-degree rotation, 7) 270-degree rotation, 8) high contrast, and 9) low contrast.

B. Deep Ensemble Approach using Efficient Models

In this study, a deep ensemble model is introduced for efficient recognition of papaya diseases using models of EfficientNet. Based on recognition performance, three models were selected for ensemble learning.

EfficientNet is a group of CNN architectures, it scales all dimensions of depth, width, resolution using a set of fixed scaling coefficients uniformly, and inverted bottleneck residual blocks is the key building block of these architectures [3]. This group contains eight architectures from EfficientNet B0 to EfficientNet B7, and EfficientNet B0 is the base architecture of this group. These architectures use the Swish activation function while other state-of-the-art CNN architectures use ReLU, Swish activation function is a multiplication of sigmoid and linear activation functions [14]. These architectures use the depth separable convolution layer which decreases the computation of the network. As the architecture number of the EfficientNet group increases, the recognition performance improves noticeably, but the amount of parameters of architecture does not raise much. The architectures of the EfficientNet group obtain higher accuracy with better efficiency compared to existing state-of-the-art CNN architectures, also decrease the size of parameters and floating-point operations per second and EfficientNet-B7 acquired 84.4% top-1 and 97.1% top-5 accuracy on ImageNet. The diversity of EfficientNet architecture is presented below

in TABLE I, and the simplified representation of the EfficientNet B0 architecture is given in Figure 3.

TABLE I. THE VARIANT OF EFFICIENTNET ARCHITECTURES

Model Name	Rate of Dropout	Coefficient of Depth	Coefficient of Width	Input Size
В0	0.2	1.0	1.0	224×224
B1	0.2	1.1	1.0	240×240
B2	0.3	1.2	1.1	260×260
В3	0.3	1.4	1.2	300×300
B4	0.4	1.8	1.4	380×380
В5	0.4	2.2	1.6	456×456
В6	0.5	2.6	1.8	527×527
В7	0.5	3.1	2.0	600×600

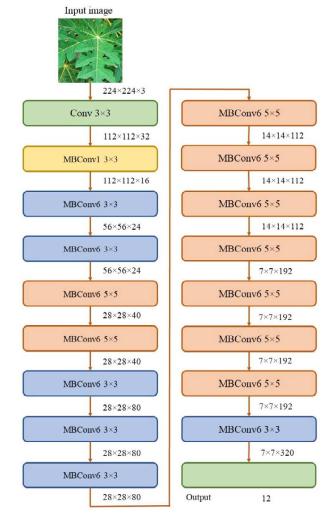


Fig. 3. The baseline architecture of EfficientNet.

Deep learning models built with the ensemble approach have the convenience of leveraging the learned knowledge from different classification models and concatenating these into a more efficient robust classification model. For generating a stacked ensemble model using three models of EfficientNet, outputs of these single models were concatenated in this study. Since the output of single models is a single node, a three-dimensional vector is used for

connecting these single nodes to a single neuron that is activated using the Softmax function. The proposed architecture of the ensemble model is given in Figure 4.

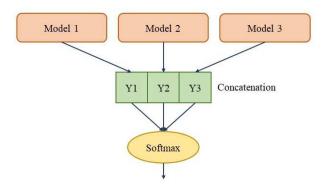


Fig. 4. The proposed architecture of the ensemble model, where Y1, Y2, and Y3 represent the outputs of three single trained models of EfficientNet.

IV. EXPERIMENTS

A. Training

Eight architectures of EfficientNet were used in this study using the approach of transfer learning. This approach of knowledge transfer helps to improve the baseline performance of the network and reduces the time required during the training phase to develop a model. During the training phase of single models, the final fully connected (FC) layer of EfficientNet models was replaced with 12 neurons as the used dataset contains 12 classes and all layers of these models were set as trainable. During the training of single models, Softmax was used in the FC layer, categorical cross-entropy was used as a loss function, Adam was used as an optimizer, the learning rate was set to 0.001 and mini-batch size was selected to 16. Single models were trained for thirty epochs using Google Colab. During the training phase of the ensemble model, feature layers were put frozen, FC layers were only available for training, and using He uniform initialization, parameters were randomly initialized. The ensemble model was also trained for thirty epochs.

B. Performance Metrics

For evaluating recognition performance of EfficientNet models and ensemble model, sensitivity (sen), specificity (spe), accuracy (acc), and precision (pre) were calculated using values of true positive (TP), true negative (TN), false positive (FP), and false-negative (FN) and equations of these performance metrics are presented below in between (1) to (4), where *pi* is the class number of the test set of the used dataset.

For a class pi,

$$Sen(pi) = \frac{TP(pi)}{TP(pi) + FN(pi)} \tag{1}$$

$$Spe(pi) = \frac{TN(pi)}{TN(pi) + FP(pi)}$$
 (2)

$$Acc(pi) = \frac{TP(pi) + TN(pi)}{TP(pi) + TN(pi) + FP(pi) + FN(pi)}$$
(3)

$$Pre(pi) = \frac{TP(pi)}{TP(pi) + FP(pi)}$$
(4)

Here, TP represents the summation of accurately classified images in each class, and TN represents the summation of correctly predicted images except the related class. On the other hand, FP represents the summation of incorrectly predicted images in other classes except for the related class and FN represents the summation of incorrectly predicted images in the related class [13].

V. RESULTS AND DISCUSSIONS

This study is conducted for developing an efficient recognition approach for classifying papaya disease, and the recognition performance of EfficientNet models and proposed ensemble model was evaluated using 6931 images of the test set. In this study, EfficientNet B5 performed significantly better compared to others which acquired 99.37% training and 98.13% test accuracy. In addition, EfficientNet B1 obtained less training and test accuracy than others, 88.18%, and 87.03%, respectively. However, the proposed deep ensemble model was built using fine-tuned versions of EfficientNet B5, B6, and B7 which are the top three models of this study, and obtained 98.13%, 96.87%, and 96.93% test accuracy, respectively. On the other hand, the proposed deep ensemble model obtained 99.74% recognition accuracy on the test set. Recognition performance of EfficientNet models on the training and test set of the dataset is below illustrated in Figure

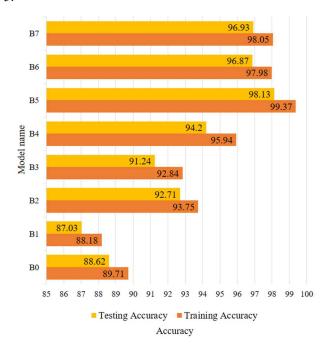


Fig. 5. Training and Test accuracies of EfficientNet models.

During the training phase, eight models of EfficientNet from B0 to B7 were consumed 1534, 1926, 2177, 2431, 3128, 4482, 5648, and 6794 seconds, respectively, to complete one epoch. For demonstrating the recognition efficiency of single and ensemble models more perceptibly, the total number of false predictions is also discussed in this study. EfficientNet B5, B7, and B6 incorrectly predicted 130, 213, and 217 images, respectively, of the test set which contains 6931 images. EfficientNet B1 misclassified 910 images which was the highest false prediction number. On the other hand, the deep ensemble model wrongly predicted 18 images of the test set which represent the effective recognition efficiency of the

proposed model, and the confusion matrix of the proposed ensemble model is given below in Figure 6.

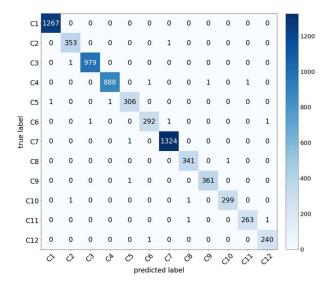


Fig. 6. Confusion matrix of the deep ensemble model: C1) healthy leaf C2) powdery mildew (leaf) C3) mosaic C4) leaf curl C5) ringspot (leaf) C6) brown spot C7) healthy fruit C8) powdery mildew (fruit) C9) black spot C10) Ringspot (fruit) C11) anthracnose C12) phytophthora blight.

In this study, the class-wise recognition performance of the deep ensemble model is checked using sensitivity, specificity, accuracy, and precision which are calculated using TP, TN, FP, and FN values obtained from the confusion matrix given in Figure 6. The class-wise recognition performance of each class was very close to each other. In the healthy leaf class, the deep ensemble model obtained a sensitivity value of 100.00%. On the other hand, the deep ensemble model acquired the highest specificity value in healthy leaf, mosaic, leaf curl, black spot, ringspot (fruit), and anthracnose classes, 99.98%. In addition, in the healthy leaf class, the proposed deep ensemble model acquired higher accuracy and precision value compared to other classes, 99.99%, and 99.92%, respectively. The class-wise classification performance of the proposed ensemble model is given below in TABLE II.

TABLE II. CLASS-WISE CLASSIFICATION PERFORMANCE OF THE PROPOSED ENSEMBLE MODEL

Class Name	Sen (%)	Spe (%)	Acc (%)	Pre (%)
Healthy leaf	100.00	99.98	99.99	99.92
Powdery mildew (leaf)	99.72	99.97	99.96	99.44
Mosaic	99.90	99.98	99.97	99.90
Leaf curl	99.66	99.98	99.94	99.89
Ringspot (leaf)	99.35	99.97	99.94	99.35
Brown spot	98.98	99.97	99.93	99.32
Healthy fruit	99.92	99.96	99.96	99.85
Powdery mildew (fruit)	99.71	99.97	99.96	99.42
Black spot	99.72	99.98	99.97	99.72
Ringspot (fruit)	99.34	99.98	99.96	99.67
Anthracnose	99.25	99.98	99.96	99.62
Phytophthora blight	99.59	99.97	99.96	99.17

The total false prediction numbers of the three best performing models of the EfficientNet such as B5, B6, and B7, and the deep ensemble model (DEM) for all classes of the test set are given in TABLE III. The number of false prediction images of the introduced model for twelve classes was less than single models of EfficientNet.

TABLE III. CLASSIFICATION NUMBERS OF MODELS FOR ALL CLASSES

Class Name	В6	В7	B5	DEM
Healthy leaf	13	11	7	0
Powdery mildew (leaf)	19	23	9	1
Mosaic	22	18	13	1
Leaf curl	21	16	16	3
Ringspot (leaf)	18	17	13	2
Brown spot	12	19	9	3
Healthy fruit	22	15	8	1
Powdery mildew (fruit)	16	21	10	1
Black spot	20	13	11	1
Ringspot (fruit)	23	17	14	2
Anthracnose	17	24	12	2
Phytophthora blight	14	19	8	1

The accuracy and loss curves acquired on training and validation sets for the EfficientNet B5 model which performed significantly better than other single models in this study are given in Figure 7 and Figure 8.

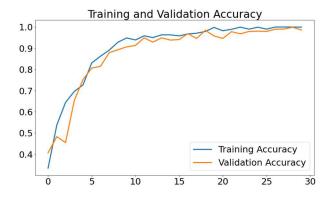


Fig. 7. Training and validation accuracy curve of EfficientNet B5.

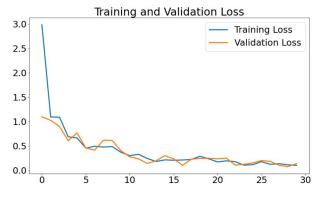


Fig. 8. Training and validation loss curve of EfficientNet B5.

The remarkable recognition performance obtained in classifying papaya diseases by the deep ensemble approach which was built using fine-tuned versions of EfficientNet B5, B6, and B7 was compared in this study with other relevant studies which were addressed in the literature for several crops diseases recognition, and presented in TABLE IV.

TABLE IV. COMPARISON OF SEVERAL METHODS FOR CROP DISEASE RECOGNITION

Study	Method	Number of Classes	Accuracy
Md. Tarek Habib et al. [5]	SVM	5	95.20%
Sachin B. Jadhav et al. [6]	AlexNet	4	98.75%
Agarwal et al. [7]	CNN	10	91.20%
Ramesh et al. [8]	CNN	5	97.00%
Melike Sardogan et al. [9]	CNN	5	86.00%
Norhalina Senan et al. [10]	CNN	4	93.00%
Siti Zulaikha Muhammad Zaki et al. [11]	MobileNet V2	4	95.94%
Dhiman Mondal et al. [12]	NB	4	87.00%
Sharada Prasanna Mohanty et al. [13]	CNN	38	99.35%
Our study	Ensemble model	12	99.74%

VI. CONCLUSION

Crop diseases are a major threat to the sustainable development of an agricultural-based economy, and initial identification of these diseases is decisive for continuing the trend of development in agriculture. The combination of recent developments in computer vision through deep learning approaches and growing global smartphone infiltration has laid the way for smartphone-based disease recognition. This study addressed a deep ensemble approach for efficient recognition of papaya diseases using EfficientNet models. At first, eight models of EfficientNet from Bo to B7 were trained using the transfer learning technique for thirty epochs, and the recognition efficiency of these CNNs was evaluated with 6931 images of the test set. The recognition performance of EfficientNet B5 was more effective compared to other models which acquired the highest training and test accuracy in this study, 99.37% and 98.13%, respectively. In addition, EfficientNet B6 and B7 also showed significant performance on the test set, obtaining 96.87% and 96.93% accuracy, respectively. However, based on the recognition performance of models on the test set, three best performing models such as EfficientNet B5, B6 and B7 were selected for the ensemble learning. The proposed deep ensemble model obtained 99.74% accuracy on the test set which was more than single models. The proposed model misclassified 18 images of the test set, where EfficientNet B5, B6, and B7 misclassified 130, 217, and 213 images, respectively. The results of several experimental studies demonstrate that the introduced deep ensemble model attains end-to-end recognition of papaya disease and provides a significant solution and a reference for the computer visionbased methods in the efficient identification of crop diseases.

This proposed model can only recognize diseases of papaya and is unable to recognize pests, which also cause severe damage to papaya farming. In future work, we plan to extend the dataset by increasing crop diversity and the number of images and classes to generate a more efficient and rapid recognition approach for crop diseases.

REFERENCES

- [1] Ventura J.A., Costa H., and Tatagiba J.S., "Papaya Diseases and Integrated Control," In: Naqvi S.A.M.H. (eds) Diseases of Fruits and Vegetables, Springer, Dordrecht, 2004, vol. 2.
- [2] Tan, Mingxing and Le, Quoc, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," 2019, arXiv:1905.11946.
- [3] El Asnaoui, K., "Design ensemble deep learning model for pneumonia disease classification," Int J Multimed Info Retr 10, 2021, pp. 55-68.
- [4] Tarek Habib, Anup Majumdar, A.Z.M. Zakaria, Morium Akter, Mohammad Shorif Uddin, and Farruq Ahmed, "Machine Vision Based Papaya Disease Recognition," Journal of King Saud University -Computer and Information Sciences, 2018.
- [5] Jadhav, S.B., Udupi, V.R. and Patil, S.B., "Identification of plant diseases using convolutional neural networks," Int. j. inf. tecnol., 2020.
- [6] Mohit Agarwal, Abhishek Singh, Siddhartha Arjaria and Amit Sinha, "ToLeD: Tomato Leaf Disease Detection using Convolution Neural Network," Procedia Computer Science, 2020, vol. 167, pp. 293-301.
- [7] S. Ramesh and D. Vydeki, "Recognition and classification of paddy leaf diseases using Optimized Deep Neural network with Jaya algorithm," Information Processing in Agriculture, 2020, vol. 7, pp. 249-260.
- [8] M. Sardogan, A. Tuncer, and Y. Ozen, "Plant Leaf Disease Detection and Classification Based on CNN with LVQ Algorithm," 3rd International Conference on Computer Science and Engineering (UBMK), 2018, pp. 382-385.
- [9] Norhalina Senan, Muhammad Aamir, Rosziati Ibrahim, N. S. A. M. Taujuddin, and W.H.N Wan Muda, "An Efficient Convolutional Neural Network for Paddy Leaf Disease and Pest Classification," International Journal of Advanced Computer Science and Applications, 2020.
- [10] Siti Zulaikha Muhammad Zaki, Mohd Asyraf Zulkifley, Marzuraikah Mohd Stofa, Nor Azwan Mohammed Kamari and Nur Ayuni Mohamed, "Classification of tomato leaf diseases using MobileNet V2," IAES International Journal of Artificial Intelligence (IJ-AI), 2020, vol. 9, pp. 290-296.
- [11] D. Mondal, A. Chakraborty, D. K. Kole and D. D. Majumder, "Detection and classification technique of Yellow Vein Mosaic Virus disease in okra leaf images using leaf vein extraction and Naive Bayesian classifier," 2015 International Conference on Soft Computing Techniques and Implementations (ICSCTI), 2015, pp. 166-171.
- [12] Mohanty SP, Hughes DP and Salathé M, "Using Deep Learning for Image-Based Plant Disease Detection," Front. Plant Sci. 7:1419, 2016.
- [13] R. H. Hridoy, F. Akter and A. Rakshit, "Computer Vision Based Skin Disorder Recognition using EfficientNet: A Transfer Learning Approach," 2021 International Conference on Information Technology (ICIT), 2021, pp. 482-487.