ECTI Transactions on Computer and Information Technology
Journal homepage: https://ph01.tci-thaijo.org/index.php/ecticit/
Published by the ECTI Association, Thailand, ISSN: 2286-9131

Ensemble Transfer Learning for Image Classification

Nayan Kumar Sarkar[1], Moirangthem Marjit Singh[2], Utpal Nandi[3] and Jyotsna Kumar Mandal[4]

ABSTRACT

The deep learning (DL) techniques used for image classification might not deliver the desired level of classification accuracy as some features belonging to some class of a dataset are missed during feature extraction. The ensemble learning (EL) based model improves classification accuracy by combining the strengths of individual classifiers. As a result, those features that were missed during feature extraction by a specific DL technique will be taken care of by another DL technique in an ensemble DL approach. In this paper, averaging EL (AENet), weighted averaging EL (WAENet), and stacking EL (StackedNet) approaches are proposed, considering the DenseNet201, EfficientNetB0, and ResNetRS101 as base models. The predictions of the base models are averaged to generate the AENet. The WAENet is constructed by assigning weights to each base model based on their prediction and then taking their average. Similarly, the StackedNet is developed by considering the DenseNet201, EfficientNetB0, and ResNetRS101 as base-learners and ResNetRS101 as meta-learner. Analysed performance of the considered pre-trained base models and the developed EL models on the standard and application-specific datasets such as MiniImageNet, CIFAR10, CIFAR100, Plant Village (PV), Tomato, Covid-19 and 9IndianFood. 80% of the datasets were used to train and 20% to test the base and proposed models. The models are trained for an epoch of 30, considering a learning rate of 0.001 and adam optimizer. The stackedNet delivered better results than others.

Article information:

Keywords: Ensemble Learning, Transfer Learning, DenseNet201, EfficientNetB0 and ResNetRS101

Article history:
Received: August 8, 2024
Revised: September 19, 2024
Accepted: October 10, 2024
Published: November 16, 2024 (Online)

DOI: 10.37936/ecti-cit.2025191.257836

1. INTRODUCTION

Image classification is a technique of identifying what class an image falls in. The applications of image classification have been overgrowing in recent times. The image classification techniques are essential for technical advancement [1]. The DL algorithms are primarily used in computer vision tasks. DL, a branch of machine learning (ML), mimics the working mechanism of the human brain. Various DL mechanisms have been developed for a variety of applications. Among the DL mechanisms, CNN is widely applied in various fields of image classification and addressed challenges. Transfer learning (TL) is an ML technique that solves the constraints of extensive data collection and reduces training costs. Educational psychology may have been the source of TL. According to psychologist C.H. Judd's generalization theory of transfer, learning to transfer is the outcome of experience being generalized. As long as any individual generalizes his experience, it is easy to realize the transfer from one particular scenario to another. In reality, someone who knows how to ride a bicycle can quickly learn to ride a motorbike faster. TL aims to enhance the result of a model by transferring the knowledge gained on the source domain to a related but different target domain [2]. It has been found that the TL has been a fascinating area of study ever since the ML revolution began [3]. From a DL perspective, TL involves utilizing a deep CNN model that was previously trained on a large dataset. A new dataset

[1,2] The author is with the Department of Computer Science Engineering, North Eastern Regional Institute of Science and Technology, Arunachal Pradesh, India, Email: nayankrsarkar@gmail.com and marjitm@gmail.com
[3] The author is with the Department of Computer Science, Vidyasagar University, West Bengal, India, Email: nandi.3utpal@gmail.com
[4] The author is with the Department of Computer Science Engineering, University of Kalyani, West Bengal, India, Email: jkm.cse@gmail.com
[2] Corresponding author: marjitm@gmail.com

with fewer training images is used to train further (fine-tune) the previously trained CNN technique [4].

The DL approach, especially developed based on CNN architecture, addressed various image classification problems. The DL models are generally designed to capture specific features. Still, in the complex task of image classification with imbalanced class or high variability, a DL model fails to learn all relevant features. For example, when a dataset is small or consists of intra-class variability, a single model fails to learn the essential features to define the class. As a result, the performance of the approach degrades. A DL model also suffers from the limitation of overfitting. Hence, there is still a scope for performance improvement by combining the prediction of multiple models, which is our motivation. Therefore, the EL concept is considered a solution to such a problem, as the EL technique enhances results by merging the strengths of individual approaches while compensating for their weaknesses. It is expected that the proposed EL techniques can address the limitations of DL techniques.

In literature, several EL techniques have been developed to address various challenges in image classification. However, hardly any EL technique has been developed considering the DenseNet201 [5], EfficientNet [6], and ResNetRS101 [7] as base models and ResNetRS101 as a meta-model. The considered base models also have certain advancements, such as the DenseNet201 mitigates the vanishing gradient problem, the EfficientNetB0 produces compound scaling by optimizing depth, width, and resolution for improving efficiency, and the ResNetRS101 provides better accuracy on large datasets by leveraging residual blocks, which makes the proposed EL models different from existing works. The major contributions of the paper are briefly discussed below:

- Proposed averaging, weighted averaging, and stacking EL approaches, namely AENet, WAENet, and StackedNet, considering the DenseNet201, ResNet152, and ResNetRS101 pre-trained CNN models as base models.
- Analysed performances of the three proposed approaches on standard (MiniImageNet, CIFAR10, CIFAR100) and application-specific (PV, Tomato, Covid-19, 9IndianFood) datasets.
- Computed results of the pre-trained DenseNet201, EfficientNetB0, and ResNetRS101 CNN models individually on the datasets.

The rest of the paper is organized as Section 2 briefly discusses the literature review, the proposed models are explained in section 3. The implementation and result analysis is discussed in section 4, section 5 discusses the conclusion and references are given at the end.

2. LITERATURE REVIEW

High importance is given to DL and TL methods in the development of image classification techniques. Recently, the EL technique has also been gearing up to develop such methods. Several approaches have been developed based on DL, TL, and EL techniques for categorizing healthy and diseased crop leaf images of rice, tomato, grapes, etc. Approaches have also been developed to categorize scene images, satellite images, medical images, food images etc. In the literature, numerous studies have been conducted to classify images of different domains. However, a few recent image classification methods developed based on DL and EL are briefly discussed in this section.

S. K. Upadhyay and A. Kumar [8] developed a deep CNN technique for diseased rice leaf image categorization. They considered Otsu's global thresholding method for removing background noise. E. Deniz *et al.* [4] suggested a TL approach to classifying histopathologic breast cancer images. They used pre-trained AleNet and VGG16 models for feature extraction, followed by SVM for classifying the extracted features, achieving an accuracy of 91.30%. Utilizing three compact CNNs O. Attallah [9] proposed a pipelined approach for classifying diseased tomato leaf images. The approach also used a hybrid feature selection technique for selecting a comprehensive set of features with lower dimensions. F. A. Shah et al. [10] suggested a deep CNN method for categorizing three rice leaf diseases, namely brown spot, bacterial blight, and blast. They applied image preprocessing, and samples were augmented to increase the dataset size, and as a result, their approach delivered 98.3% classification accuracy. N. K. Sarkar et al. [11] developed a DL approach for diseased crop leaf image classification that used network deconvolution operation to remove correlations from images and AReLU activation function for faster training. They considered PV, Tomato, and Grape datasets for results evaluation and obtained 99.27%, 99.10% and 100% classification accuracies on the datasets, respectively. To generate high-quality images, Z. Zhang *et al.* [12] developed an image augmentation method using the dual generative adversarial network (GAN) approach. They considered the VGG11 and ResNet18 approaches for analyzing the results. Rice leaf images were used in their experiments, and as a result, the ResNet18 delivered 4.57%, and VGG11 delivered 4.1% higher classification accuracy on the generated high-quality images than the original samples.

Using resnet101, resnet152, VGG densenet-169 and 201, Q. Pan et al. [13] suggested an EL technique namely, WR-EL, to categorize rust-diseased leaf images of wheat. The developed approach delivered better results than each CNN model considered. Based on the stacking paradigm of ensemble T. Aboneh *et al.* [14] suggested another approach to classify multi-spectral images. The approach delivered 99.96% categorization accuracy on the Landsat image generated from Bishoftu town in Ethiopia. Using EfficientNetB0 and MobileNetV2 transfer learning models, H. T. Vo *et al.* [15] developed an Ensemble technique for the automatic categorization of diseased plant leaves. They used the PV dataset to assess the technique and obtained a classification accuracy of 99.77% on 20% of the test data. Considering five DL approaches, Z. Rahman *et al.* [16] proposed another ensemble approach to identify seven categories of skin cancers. Y. Zheng et al. [17] suggested another technique to categorize breast cancer. They considered the four pre-trained approaches as base models to construct the proposed EL approach that resulted in 98.90% categorization accuracy. D. Muller *et al.* [18] produced a medical image categorization pipeline for assessing the performances of EL techniques. The proposed pipeline consists of image augmentation and preprocessing techniques, along with nine CNN architectures. They considered four categories of medical image datasets to analyze results, and the stacking ensemble technique revealed the highest performance improvement comparatively.

Using chest XR images, A. K. Das *et al.* [19] developed an ensemble approach to identify covid19. Their approach considered the DenseNet201, Resnet50V2, and Inceptionv3 deep CNN approaches to construct the model that delivered 91.62% categorization accuracy. S. Balasubramaniam and K. S. Kumar [20] also proposed an optimal EL method to identify COVID-19 disease using chest XR images. Using a small training dataset, G. Batchuluun *et al.* [21] developed a D- based leaf image categorization technique called PI-CNN. The model's performance was evaluated using 70%, 50%, 30%, and 10% of the training data, and it delivered comparatively better results. S. M. Javidan *et al.* [22] suggested an ensemble technique for categorizing seven types of diseased and healthy Tomato leave images. They considered six ML approaches as base classifiers to generate the weighted ensemble approach, which resulted in 95.58% classification accuracy. J. Chen *et al.* [23] proposed a stacked EL approach to recognize plant leaf diseases. They considered the SE-MobileNet, Mobile-DANet, and MobileNetV2 CNN approaches as base models and delivered 99.61% average classification accuracy on the PV dataset.

Approaches have also been developed to classify food images. Using Efficientnetb0, G. VijayaKumari *et al.* [24] developed a TL-based technique for classifying of food images. To recognize various food items, the Food-101 dataset was used, achieving 80% accuracy. F. S. Konstantakopoulos et al. [25] conducted a detailed survey on food image recognition. To automatically detect and recognize chewable food items based on their eating sounds, Y. Kumar *et al.* [26] proposed an approach using signal processing and DL approaches. D. Xue *et al.* [27] developed a weighted EL approach to classify histopathology images. P. Khanarsa and S. Kitsiranuwat [46] developed DL-based EL techniques, namely the maximum occurrence of cervical cells and the maximum probability score of cervical cells for classifying pap smear images, and delivered more than 97% accuracy. N. N. Alabid [47] proposed a method for detecting spatial relationships between a human and an object considering 2-D and 3-D scenes. GMM, Viola-Jones, and KF algorithms were used in different phases of the implementation. The results showed that the approach could establish a relationship between a stationary and moving object. A. Dey *et al.* [48] developed a DL approach called ResDLCNN-GRU Attention for the identification and categorization of violence from video footage. They considered the Hockey Fights (HF), Movie Fights, and SCVD datasets for result assessment, achieving accuracies of 98.38%, 99.62%, and 90.57%, respectively.

3. PROPOSED MODELS

The DL and TL approaches developed for image categorization have performed better than several machine learning approaches. The individual image classification approaches developed using the DL concept may not adequately learn the features associated with all the labels in a dataset. A model M_1 might adequately learn the characteristics of a class C_1 but not C_2 of a dataset, whereas the model M_2 learns the characteristics of class C_2 well but not C_1. Therefore, to overcome the limitation, the EL concept has been introduced where classification performance can be improved by merging the strengths and weaknesses of several models, i.e., in ensemble technique, the results from several models (base-models) are combined to reduce bias and variance [28]. Generally, the EL techniques are of various types. However, we considered the averaging, weighted averaging, and stacking techniques to develop the proposed AENet, WAENet, and stackedNet methods, respectively, and they are briefly discussed below:

- **Averaging:** The averaging ensemble technique is very simple and mostly followed. In this technique, the predictions of the base models are averaged using the soft-max function to generate the final prediction. The output of the averaging ensemble approach is shown in equation (1).

$$Output_{avg} = \frac{1}{M} \sum_{i=1}^{M} O_i \tag{1}$$

  Where O_i is the output of model i, and M is the total models considered.

- **Weighted Averaging:** It is a development of the averaging ensemble approach where weight is assigned for each model considered. Weights are assigned based on the performance of base models, i.e., higher weight is assigned for a model that delivers higher accuracy individually; similarly, lower weight is assigned for a model that delivers lower accuracy. The mathematical formula is shown in equation (2).

$$Output_{weighted-avg} = \frac{1}{M} \sum\nolimits_{i=1}^{M} W_i O_i \qquad (2)$$

Where W_i and O_i are the weight and output of model i, respectively, and M is the total models considered.

- **Stacking:** In the stacking approach of EL, the predictions generated by the base-models are passed as input for a meta-model. It does not merely aggregate predictions; it actively learns the relationships between the outputs of different models and allows it to correct individual model errors more effectively.

The AENet results in robust prediction by reducing the variance of the individual models. It mitigates overfitting by combining the strengths and weaknesses of particular models. It is beneficial when the base models are trained on small or diverse datasets. Instead of treating all models equally, the WAENet allows the best-performing model to have the most influence on the final prediction. The WAENet performs better than simple averaging. Instead of averaging or manually assigning weights, the StackedNet uses the meta-model that learns the optimal way to combine base-model's predictions. The two level-structure of StackedNet reduces bias and variance. The meta-model learns patterns for errors made by base models and improves performance. StackedNet is particularly powerful in handling complex patterns in data, offering superior generalization in scenarios where individual models might suffer.

The inductive TL concept is considered in our proposed models, where a CNN model trained on a large dataset is again trained (fine-tuned) on a similar new dataset. The TL is mainly considered in DL applications as fine-tuning a previously trained approach is faster than training a newly constructed approach from scratch [4]. The considered TL approaches (base-models) are briefly discussed below:

- **DenseNet201:** A CNN architecture proposed by Huang *et al.* [5] and well known for its outstanding performance for object identification tasks. It uses a simple connectivity technique in which every layer is connected to each other in a feed-forward way that resulted in maximum flow of information in the network, reduces the vanishing gradient issue, increases feature propagation, and reduces the number of parameters.
- **EfficientNetB0:** It is a part of the EfficientNet family. EfficientNet [6] is a CNN and uses a scaling method. It uniformly scales all dimensions (width, depth, resolution) of a network, leading to improved performance. The EfficientNetB0 is based on the mobilenetv2 architecture and uses the uniform compound scaling method to improve accuracy. The EfficientNetB0 has fewer parameters compared to many other DL models, making it computationally efficient while maintaining high accuracy.
- **ResNetRS101:** It is a variant of the ResNet (Residual Network) architecture [7] that incorporates elements of random search optimization techniques during its training process. In ResnetRS, the RS stands for the "Revised/Scaled" and reflects the enhancements over the ResNet model.

DenseNet201, EfficientNetB0, and ResNetRS101 are base models because of their distinct architectural strengths, such as the DenseNet201 is strongly connected and mitigates the vanishing gradient problem. EfficientNetB0 produces compound scaling by optimizing depth, width, and resolution to improve efficiency. It delivers higher accuracy than models like MobileNet in real-time and resource constraint environments. The ResNetRS101 provides better accuracy on large datasets by leveraging residual blocks and regularization. These advantages ensure that the models not only learn the features but also improve the ensemble's overall performance.

Each of the base models is frozen to utilize the pre-trained parameters. The DenseNet201 model was frozen at layer 200, and a new layer was added with global average pooling and ReLU activation function before the classification layer. Similarly, EfficientNetB0 and ResNetRS101 were frozen at layers 236 and 100, respectively, before adding the new layer, as was done with DenseNet201 and the classification layer. To reduce the overfitting, a dropout rate of 20% was applied before the classification layer of each base model. The proposed models are constructed by combining the predictions of the three considered base models. In Fig. 1 prediction1, prediction2, and prediction3 represent the generated predictions of DenseNet201, EfficientNetB0, and ResNetRS101 on the dataset. The predictions are then averaged to construct the AENet model. The proposed WAENet is developed based on the weighted averaging ensemble concept, where different weights are assigned manually against each base model based on classification accuracy delivered by them. The base model that offers the highest accuracy is assigned with the maximum weight, while the model with the lowest accuracy is assigned the least weight, and so on. The average weighted sum of the accuracies of the base models is calculated as the accuracy of the WAENet. Similarly, the StackedNet model is constructed, where the predictions of the three base models are passed as input to the meta-model. The ResNetRS101 is considered a meta-model. The diagrammatic view of the proposed EL models is shown in Fig. 1.

4. IMPLEMENTATION AND RESULT ANALYSIS

The approaches are implemented in Python on a Google Colab environment configured with Nvidia T4 GPU, 12.67GB RAM, etc. The results of the methods are assessed in accuracy (Acc.), precision (Prec.), recall (Rec.) and, F1 score (F-1). Generally, the result of a classification approach falls into four states such as true positive (TP), true negative (TN), false Positive (FP), and false Negative (FN). Using these four states of classification, the classification metrics are defined in equations 3, 4, 5, and 6, respectively.

$$Acc. = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$Prec. = \frac{TP}{TP + FP} \tag{4}$$

$$Rec. = \frac{TP}{TP + FN} \tag{5}$$

$$F-1 = \frac{2 \times Prec \times Rec.}{Prec. + Rec.} \tag{6}$$

4.1 DATASETS

The performances of the approaches are assessed on both standard and application-specific datasets. The considered standard datasets are CIFAR10, CIFAR100 and, MiniImageNet. The datasets are briefed below.

- **CIFAR10** [29]: Consisting of 60,000 tiny images of 10 different types. Each type comprised 6000 images.
- **CIFAR100** [29]: Consisting of 60,000 images of 100 different types. Each type consists of 600 samples.
- **MiniImageNet** [30]: It is a reduced type of the ImageNet dataset. It is consisting 100 categories, each with 600 samples of various dimensions.

All samples of the datasets are resized to 112 × 112 dimensions. The datasets are divided into 80% to train and 20% to test the models. Some of the test samples of the CIFAR10 dataset are shown in Fig. 2. The models were trained for an epoch of size 30, larning-rate of 0.001 with Adam optimizer. The performances obtained by the approaches on the MiniImageNet dataset are shown in Table 1. Similarly, the performances delivered on the CIFAR10 and CIFAR100 datasets are shown in Tables 2 and 3, respectively.

Table 1: *Performances delivered by the approaches on the MiniImageNet dataset.*

Approach	Acc.	Prec.	Rec.	F-1
DenseNet201	64.05%	65.24%	64.55%	64.69%
EfficientNetB0	78.26%	78.87%	79.10%	79.55%
ResNetRS101	76.26%	77.71%	76.15%	76.92%
AENet	79.47%	80.25%	79.94%	79.89%
WAENet	79.60%	80.75%	80.14%	79.78%
StackedNet	80.10%	80.77%	80.21%	79.90%

Table 2: *Performances delivered by the approaches on the CIFAR10 dataset.*

Approach	Acc.	Prec.	Rec.	F-1
DenseNet201	87.62%	87.51%	87.68%	87.33%
EfficientNetB0	92.58%	93.18%	92.68%	92.51%
ResNetRS101	91.76%	92.11%	92.35%	92.29%
AENet	92.98%	92.25%	92.78%	93.19%
WAENet	92.1%	92.67%	91.91%	92.4%
StackedNet	93.24%	93.44%	93.14%	94.06%

It is seen from Table 1 that the pre-trained EfficientNetB0 delivered 14.21% more classification accuracy on the MiniImageNet dataset than DenseNet201 and 2.00% more accuracy than the ResNetRS101 model. Similarly, it is seen in Tables 2 and 3

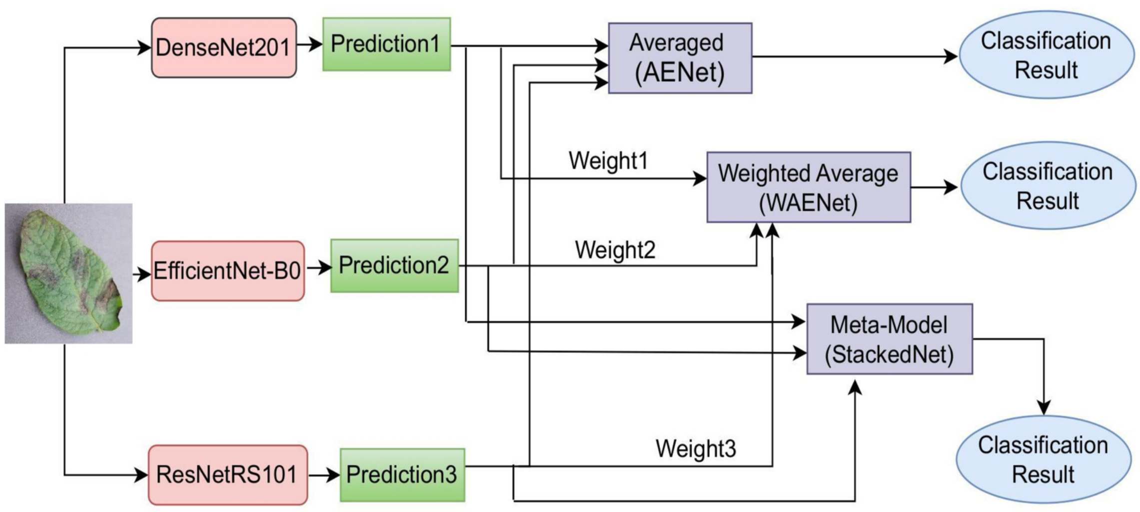

Fig.1: *Proposed EL models.*

Table 3: *Performances delivered by the approaches on the CIFAR100 dataset.*

Approach	Acc.	Prec.	Rec.	F-1
DenseNet201	61.42%	63.63%	61.42%	61.49%
EfficientNetB0	74.84%	75.97%	74.84%	74.93%
ResNetRS101	73.46%	74.72%	73.46%	73.48%
AENet	79.19%	80.20%	79.82%	79.99%
WAENet	80.45%	81.27%	80.81%	81.34%
StackedNet	80.75%	80.56%	80.86%	81.14%

that the EfficientNetB0 model delivered better results on CIFAR10 and CIFAR100 datasets than the DenseNet201 and ResNetRS101 models. It is also found from Table I that the proposed StackedNet model delivered 0.63% and 0.5% more classification Acc., 0.52% and 0.02% more Prec., 0.27% and 0.07% more Rec. and 0.01% and 0.12% more F-1 than the AENet and WAENet models respectively. The StackedNet model also delivered better results than the AENet and WAENet in Tables 2 and 3 on the CIFAR10 and CIFAR100 datasets.

4.2 APPLICATION OF THE PROPOSED APPROACHES ON DIFFERENT FIELDS:

The results of the proposed approaches are also carried out on different applications, such as identifying diseased and healthy crop leaves, types of Indian foods and covid19. Therefore, the PV, Tomato, 9IndianFood, and Covid19 datasets are considered to analyze the performances in agriculture, food, and healthcare. Some of the test samples of Tomato, 9IndianFood, and PV datasets are shown in Fig. 3, 4, and 5 accordingly. The datasets are briefly discussed below:

- **PV** [31]: The PV is a widely used dataset consisting 54305 images of 38 categories of diseased and healthy plant leaves of 14 types of crops.
- **Tomato**: Generated from PV with 18160 images of 10 types of diseased and healthy tomato leaves.
- **9IndianFood** [32]: The 9IndianFood dataset consists of a total 2755 of 9 types popular Indian food images.
- **Covid19** [33]: Consists of 10192 healthy and 3616 covid positive chest X-ray images.

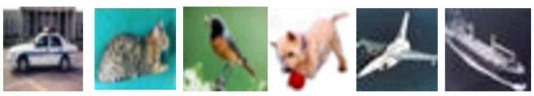

Fig.2: *Some test images of the CIFAR10 dataset.*

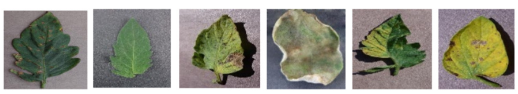

Fig.3: *Some test images of the Tomato dataset.*

Fig.4: *Some test images of the 9IndianFood dataset.*

Fig.5: *Some test images of the PV dataset.*

Results delivered by the models on PV, Tomato, 9IndianFood, and Covid19 datasets are shown in Tables 4, 5, 6 and 7, respectively.

Table 4: *Results delivered by the approaches on the PV dataset.*

Approach	Acc.	Prec.	Rec.	F-1
DenseNet201	98.40%	97.87%	98.07%	97.94%
EfficientNetB0	98.79%	98.58%	97.98%	98.21%
ResNetRS101	99.60%	99.41%	99.26%	99.33%
AENet	99.66%	99.48%	99.34%	99.41%
WAENet	99.78%	99.72%	99.57%	99.63%
StackedNet	99.80%	99.56%	99.37%	99.82%

Table 5: *Results delivered by the approaches on the Tomato dataset.*

Approach	Acc.	Prec.	Rec.	F-1
DenseNet201	95.26%	97.20%	92.82%	94.80%
EfficientNetB0	98.51%	98.24%	98.01%	98.12%
ResNetRS101	98.68%	98.52%	98.36%	98.40%
AENet	99.39%	99.30%	99.39%	99.35%
WAENet	99.63%	99.33%	99.44%	99.48%
StackedNet	99.71%	99.42%	99.69%	99.58%

Table 6: *Results delivered by the approaches on the 9IndianFood dataset.*

Approach	Acc.	Prec.	Rec.	F-1
DenseNet201	86.39%	86.24%	85.55%	85.69%
EfficientNetB0	90.93%	90.77%	90.46%	90.55%
ResNetRS101	94.01%	93.71%	94.35%	93.92%
AENet	94.19%	94.25%	93.84%	93.99%
WAENet	95.74%	95.67%	95.71%	95.64%
StackedNet	95.77%	95.70%	95.79%	95.74%

Table 7: *Results delivered by the approaches on the Covid-19 dataset.*

Approach	Acc.	Prec.	Rec.	F-1
DenseNet201	98.28%	97.70%	97.14%	97.42%
EfficientNetB0	98.99%	98.89%	98.44%	98.66%
ResNetRS101	98.59%	98.44%	97.85%	98.14%
AENet	99.17%	99.11%	98.70%	98.90%
WAENet	99.28%	99.42%	99.68%	99.04%
StackedNet	99.35%	99.47%	99.72%	99.10%

It is found from Table 4 that the ResNetRS101 delivered 1.20% more classification accuracy than

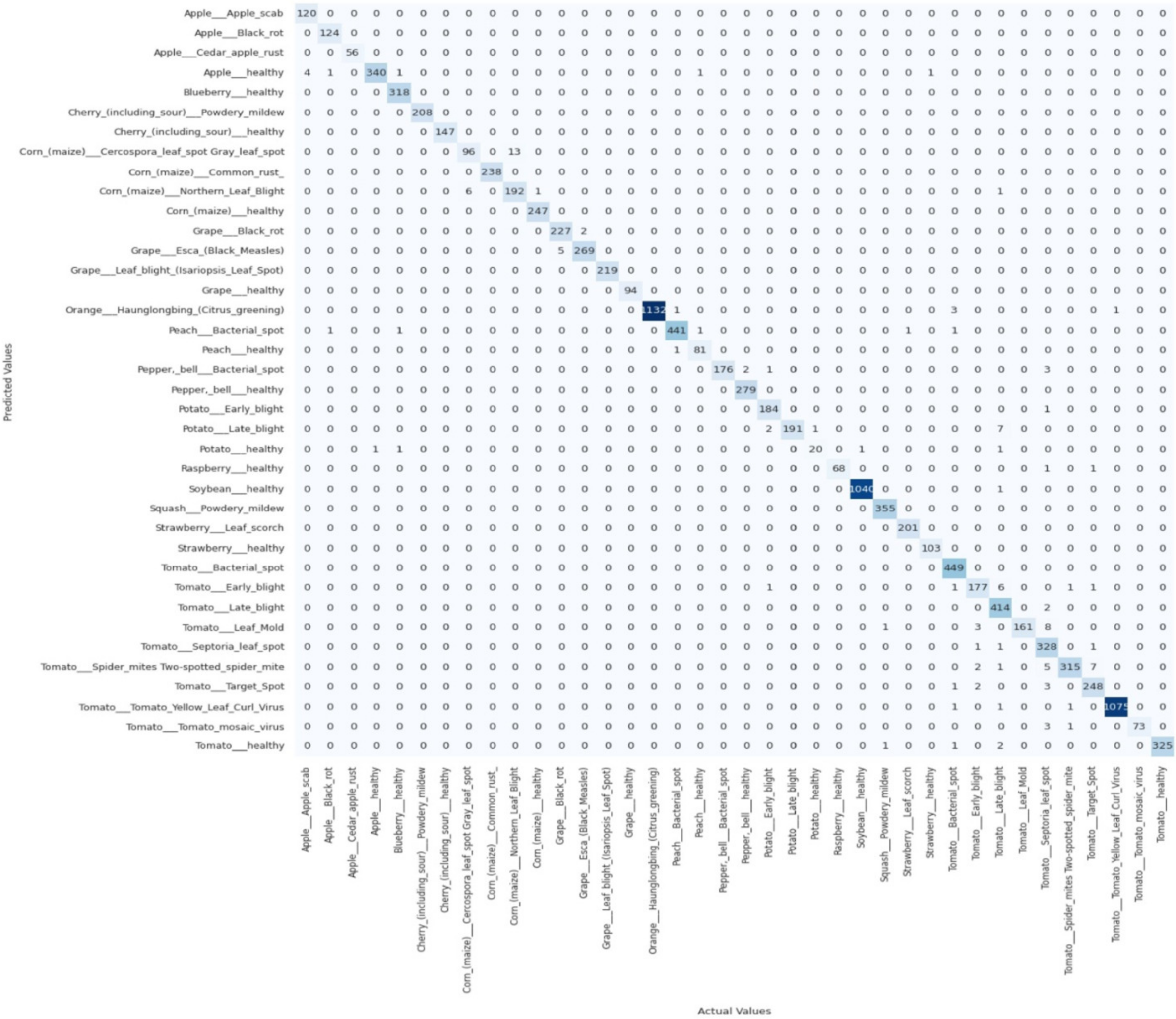

Fig.6: *Confusion matrix generated by the StackedENet on the PV dataset.*

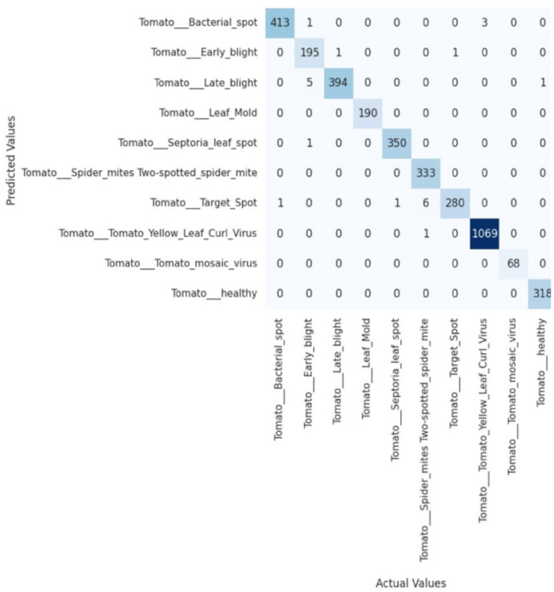

Fig.7: *Confusion matrix generated by the StackedENet on the Tomato dataset.*

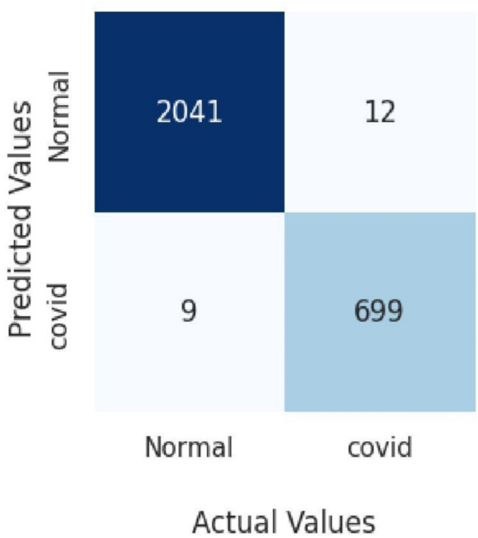

Fig.8: *Confusion matrix generated by the StackedENet on the Covid-19 dataset.*

DenseNet201 and 0.81% more accuracy than the EfficientNetB0 model on the PV dataset. The StackedNet also delivered 0.14% and 0.02% more accuracy than the AENet and WAEnet models on the PV dataset. It is also observed from Tables 5 and 6 that the ResNetRS101 model delivered better results on Tomato and 9IndianFood datasets than the DenseNet201 and EfficientNetB0 models. However, in Table 7, the EfficientNetB0 delivered slightly better results than the DenseNet201 and ResNetRS101 models on the Covid19 dataset. It is also seen from Tables 5, 6, and 7 that the StackedNet model delivered better results than the AENet and WAENet on the Tomato, 9IndianFood, and Covid19 datasets, respectively. The confusion matrices generated for the PV, Tomato, Covid19, and 9IndianFood datasets are shown in Fig. 6, 7, 8, and 9, respectively. The ROC-AUC curve for the PV, Tomato, 9IndianFood, and COVID-19 datasets are shown in Fig. 10, 11, 12, and 13, respectively. It is seen that except for 2 classes, all other classes delivered nearly 1 AUC value for PV, all classes delivered an AUC value of 1 for Tomato and two classes delivered an AUC value 1 and the rest delivered an AUC nearly 0.99 for 9IndianFood. For COVID19 dataset the delivered AUC value is 0.98. These observations indicate that the overall model performance is good. The result comparison of various EL-based approaches with the proposed models is shown in Table 8.

From result analysis, it is found that among the three pre-trained CNN models, the EfficientNetB0 delivered better results on the considered standard datasets. In the same way, the ResNetRS101 delivered higher results than the DenseNet201 and EfficientNetB0 on the considered domain-specific datasets. It is also seen that the developed stackedNet approach delivered better results than AENet and WAENet.

The advantages of the proposed models are that the base models deliver better results individually than other CNN-based models. Due to the individual performance of base models, the proposed EL models also delivered better results than others. Generally, EL techniques have demerits such as complex architecture, as they require training multiple models, which increases both training time and the need for computational resources. To mitigate these issues, simple base models can be considered that will reduce computational complexity and weighted averaging techniques or selecting a subset of high-performing models can also be considered.

In real-time applications, the ensemble models may be prohibited because of their computational overhead, and single models could be considered in such applications. EL models are also not suitable for devices with limited memory and computation power, such as mobile devices. Though, the EL models deliver better results, there, is a trade-off between per-

Table 8: *The result comparison of the proposed approaches with other EL approaches.*

Author	Year	Dataset	Classes	Acc.
H. Tu *et al.* [15]	2023	PV	38	99.77%
J. Chen *et al.* [23]	2022	PV	38	99.61%
S. Jain *et al.* [44]	2023	PV	38	97.7%
H. T. Vo *et al.* [34]	2023	PV	38	99.77%
Ulutas & Aslantas [35]	2023	Tomato	10	99.60%
P. Kaur *et al.* [36]	2023	Tomato	11	97.24%
M. Astani *et al.* [37]	2022	Tomato	13	95.98%
E. Saraswathi *et al.* [38]	2023	Tomato	14	96%
Pandiyaraju *et al.* [45]	2024	Tomato	9	98.7%
M. R. Ahmed *et al.* [39]	2024	Tomato	10	99.51%
S. M. Javidan *et al.* [22]	2023	Tomato	7	95.58%
J Patel & K Modi [40]	2023	9Indian Food	9	95.3%
Eldawoudy *et al.* [41]	2023	Covid19	2	99.05%
M. Azam *et al.* [42]	2022	Covid19	2	82.29%
T. H. Bui *et al.* [43]	2023	Covid19		98.09%
A. K. Das *et al.* [19]	2021	Covid19	2	91.62%
AENet (proposed)		PV	38	99.66%
AENet (proposed)		Tomato	10	99.39%
AENet (proposed)		9Indian Food	9	94.19%
AENet (proposed)		Covid19	2	99.29%
WAENet (proposed)		PV	38	99.78%
WAENet (proposed)		Tomato	10	99.63%
WAENet (proposed)		9Indian Food	9	95.74%
WAENet (proposed)		Covid19	2	99.34%
StackedNet (proposed)		PV	38	99.80%
StackedNet (proposed)		Tomato	10	99.71%
StackedNet (proposed)		9Indian Food	9	95.77%
StackedNet (proposed)		Covid19	2	99.35%

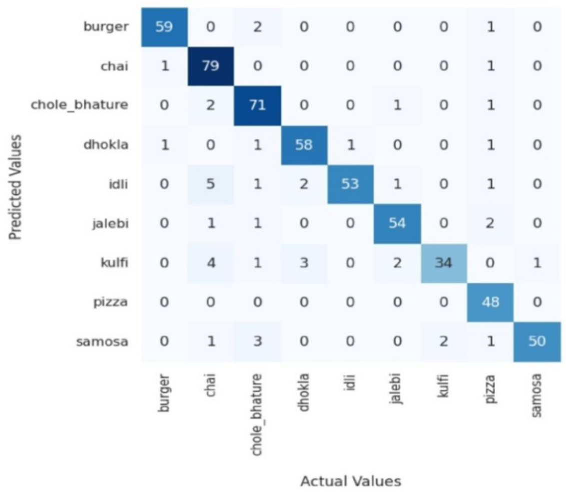

Fig.9: *Confusion matrix generated by the StackedENet on the 9IndianFood dataset.*

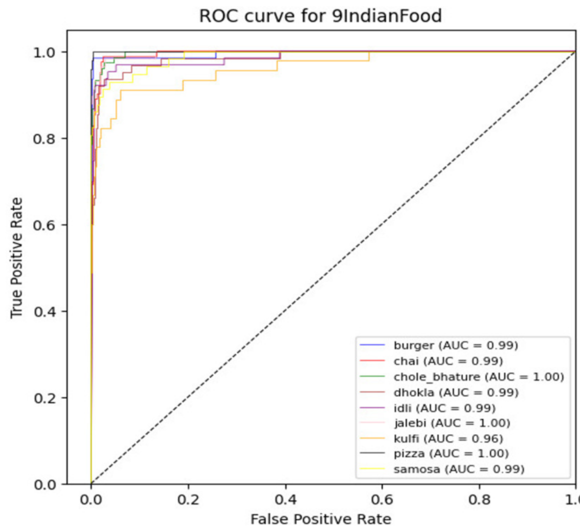

Fig.12: *ROC curve (9IndianFood dataset).*

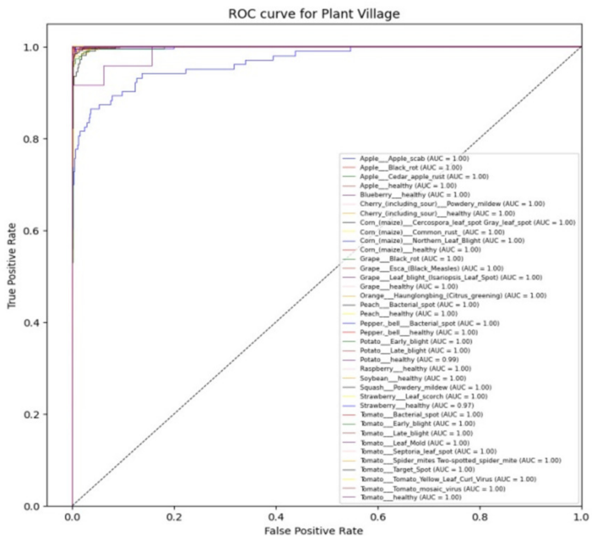

Fig.10: *ROC curve (PV dataset).*

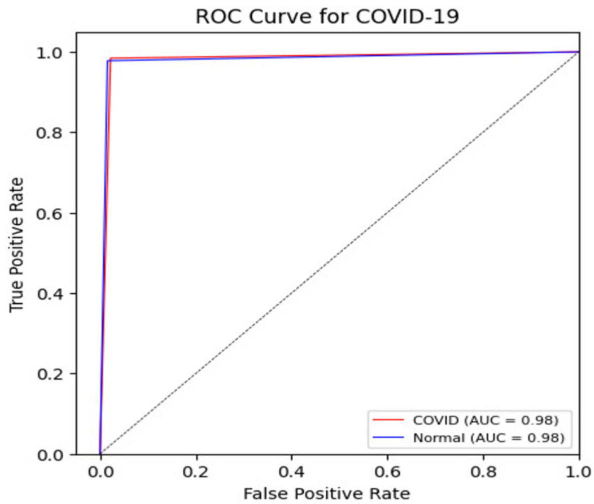

Fig.13: *ROC curve (Covid-19).*

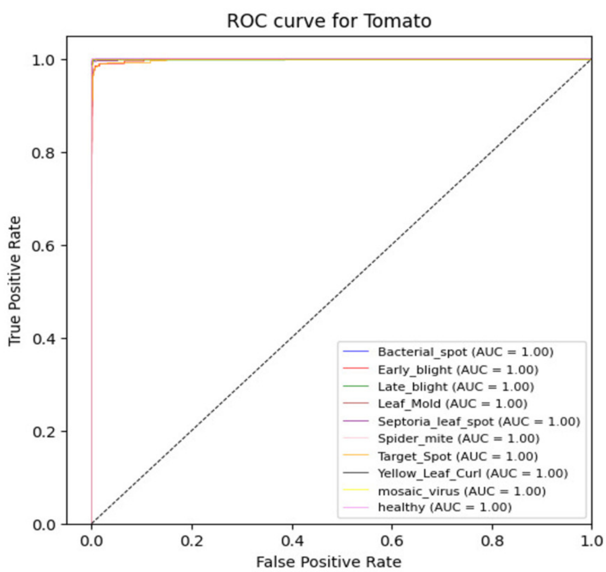

Fig.11: *ROC curve (Tomato dataset).*

formance and computational complexity when compared with single models.

5. CONCLUSION

Proposed AENet, WAENet, and StackedNet EL approaches for image classification purposes in the paper. DenseNet201, EfficientNetB0, and ResNetRS101 TL-based CNN models were considered base models to construct the proposed models. A dropout rate of 20% was applied before fine-tuning each base model. Among the pre-trained models, the EfficientNetB0 and ResNetRS101 delivered the highest individual accuracy on the considered standard and application-specific datasets. The StackedNet delivered better results than the AENet and WAENet, with 99.80%, 99.71%, 95.77%, and 99.35% accuracies on the PV, Tomato, 9IndianFood, and Covid19 datasets, respectively. The proposed methods can be applied in healthcare, agriculture, bio-metric authentication, object detection, facial recognition, surveillance systems, traffic monitoring, driver-less cars, disaster response, climate change monitoring, etc. Although the approaches delivered better results on the considered datasets compared to several other recent approaches, the computational complexity of the EL-based approach is high. Hence, further research can focus on developing new classification approaches based on parallel and distributed ensemble learning.

AUTHOR CONTRIBUTIONS

Conceptualization, Implementation and Drafting, Sarkar, N. K.; Investigation, Methodology, Analysis and Supervision, Singh, M. M.; Validation and Editing, Nandi, U.; Review, Mandal, J. K. All authors have read and agreed to the published version of the manuscript.

Corresponding author: Moirangthem Marjit Singh Department of Computer Science & Engineering, North Eastern Regional Institute of Science and Technology, Arunachal Pradesh, India.
Email: marjitm@gmail.com

References

[1] N. K. Sarkar, M. M. Singh and U. Nandi, "Recent researches on image classification using deep learning approach," *International Journal of Computing and Digital Systems*, vol. 12, no. 1, pp. 1357–1374, 2022.

[2] F. Zhuang *et al.*, "A Comprehensive Survey on Transfer Learning," in *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43-76, Jan. 2021.

[3] M. Iman, H. R.Arabnia and K. Rasheed, "A review of deep transfer learning and recent advancements," *Technologies* vol. 11, no. 2, 2023.

[4] E. Deniz *et al.*, "Transfer learning based histopathologic image classification for breast cancer detection," *Health information science and systems*, vol. 6, no.1:18, pp. 1–7, 2018.

[5] G. Huang, Z. Liu, L. V. D. Maaten and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4700–4708, 2017.

[6] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning*, pp. 6105–6114, 2019.

[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*, pp. 770–778, 2016.

[8] S. K. Upadhyay and A. Kumar, "A novel approach for rice plant diseases classification with deep convolutional neural network," *International Journal of Information Technology*, vol. 14, no. 1, pp. 185–199, 2022.

[9] O. Attallah, "Tomato leaf disease classification via compact convolutional neural networks with transfer learning and feature selection," *Horticulturae*, vol. 9, no.2:149, 2023.

[10] F. A. Shah *et al.*, "Towards intelligent detection and classification of rice plant diseases based on leaf image dataset," *Computer Systems Science & Engineering*, vol. 47, no. 2, 2023.

[11] N. K. Sarkar, M. M. Singh and U. Nandi, "A novel deep neural network model using network deconvolution with attention based activation for crop disease classification," *Multimedia Tools and Applications*, vol. 83, no. 6, pp. 17025–17045, 2023.

[12] Z. Zhang, Q. Gao, L. Liu and Y. He, "A High-Quality Rice Leaf Disease Image Data Augmentation Method Based on a Dual GAN," in *IEEE Access*, vol. 11, pp. 21176-21191, 2023.

[13] Q. Pan, M. Gao, P. Wu, J. Yan and M. A. AbdelRahman, "Image classification of wheat rust based on ensemble learning," *Sensors*, vol. 22, no. 16:6047, 2022.

[14] T. Aboneh, A. Rorissa and R. Srinivasagan, "Stacking-based ensemble learning method for multi-spectral image classification," *Technologies*, vol. 10, no. 1:17, 2022.

[15] H.-T. Vo, L.-D. Quach and T. N. Hoang, "Ensemble of deep learning models for multi-plant disease classification in smart farming," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 5, 2023.

[16] Z. Rahman, M. S. Hossain, M. R. Islam and M. M. Hasan and R. A. Hridhee, "An approach for multiclass skin lesion classification based on ensemble learning," *Informatics in Medicine Unlocked*, vol. 25, no. 100659, 2021.

[17] Y. Zheng *et al.*, "Application of transfer learning and ensemble learning in image-level classification for breast histopathology," *Intelligent Medicine*, vol. 3, no. 2, pp. 115–128, 2023.

[18] D. Müller, I. Soto-Rey and F. Kramer, "An Analysis on Ensemble Learning Optimized Medical Image Classification With Deep Convolutional Neural Networks," in *IEEE Access*, vol. 10, pp. 66467-66480, 2022.

[19] A. K. Das, S. Ghosh, S. Thunder, R. Dutta, S. Agarwal and A. Chakrabarti, "Automatic covid-19 detection from x-ray images using ensemble learning with convolutional neural network," *Pattern Analysis and Applications*, vol. 24, pp. 1111–1124, 2021.

[20] S. Balasubramaniam and K. S. Kumar, "Optimal ensemble learning model for covid-19 detection using chest x-ray images," *Biomedical Signal Processing and Control*, vol. 81, no. 104392, 2023.

[21] G. Batchuluun, S. H.Nam and K. R.Park, "Deep learning-based plantimage classification using a

small training dataset," *Mathematics*, vol. 10, no. 17:3091, 2022.

[22] S. M. Javidan, A. Banakar, K. A. Vakilian and Y. Ampatzidis, "Tomato leaf diseases classification using image processing and weighted ensemble learning," *Agronomy Journal*, vol. 116, no. 3, pp. 1029-1049, 2023.

[23] J. Chen, A. Zeb, Y. Nanehkaran and D. Zhang, "Stacking ensemble model of deep learning for plant disease recognition," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 9, pp. 12359–12372, 2023.

[24] G. VijayaKumari, P. Vutkur and P. Vishwanath,: "Food classification using transfer learning technique," *Global transitions proceedings*, vol. 3, no. 1, pp. 225–229, 2022.

[25] F. S. Konstantakopoulos, E. I. Georga and D. I. Fotiadis, "A review of imagebased food recognition and volume estimation artificial intelligence systems," *IEEE Reviews in Biomedical Engineering*, vol. 14, pp. 136-152, 2024.

[26] Y. Kumar *et al.*, "Automated detection and recognition system for chewable food items using advanced deep learning models," *Scientific Reports*, vol. 14, no. 6589, 2024.

[27] D. Xue *et al.*, "An Application of Transfer Learning and Ensemble Learning Techniques for Cervical Histopathology Image Classification," in *IEEE Access*, vol. 8, pp. 104603-104618, 2020.

[28] D. Arpit, H. Wang, Y. Zhou and C. Xiong, "Ensemble of averages: Improving model selection and boosting performance in domain generalization," *Advances in Neural Information Processing Systems*, vol. 35, pp. 8265–8277, 2022.

[29] [Online]. Available:`https://www.cs.toronto.edu/kriz/cifar.html`[Accessed: May 10, 2024].

[30] [Online]. Available:`https://www.kaggle.com/datasets/arjunashok33/miniimagenet`[Accessed:].

[31] [Online]. Available:`https://www.kaggle.com/datasets/mohitsingh1804/plantvillage.`[Accessed: April 25, 2024].

[32] [Online]. Available:`https://www.kaggle.com/datasets/jigarsharp/indian-food-9-class.`[Accessed: May 13, 2024].

[33] [Online]. Available:`https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiographydatabase.`[Accessed: May 13, 2024].

[34] H.-T. Vo, L.-D. Quach and H. T. Ngoc, "Ensemble of Deep Learning Models for Multi-plant Disease Classification in Smart Farming," *International Journal of Advanced Computer Science and Applications(IJACSA)*, vol. 14, no. 5, 2023.

[35] H. Ulutaş and V. Aslantaş, "Design of efficient methods for the detection of tomato leaf disease utilizing proposed ensemble cnn model," *Electronics*, vol. 12, no. 4:827, 2023.

[36] P. Kaur *et al.*, "DELM: Deep Ensemble Learning Model for Multiclass Classification of Super-Resolution Leaf Disease Images," *Turkish Journal of Agriculture and Forestry*, vol. 47, no. 5:12, pp. 727-745, 2023.

[37] M. Astani, M. Hasheminejad and M. Vaghefi, "A diverse ensemble classifier for tomato disease recognition," *Computers and Electronics in Agriculture*, vol. 198, no. 107054, 2022.

[38] E. Saraswathi and J. FarithaBanu, "A Novel Ensemble Classification Model for Plant Disease Detection Based on Leaf Images," *2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF)*, Chennai, India, pp. 1-7, 2023.

[39] M. R. Ahmed *et al.*, "Towards Automated Detection of Tomato Leaf Diseases," *2024 6th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT)*, Dhaka, Bangladesh, pp. 387-392, 2024.

[40] J. Patel and K. Modi, "Indian food image classification and recognition with transfer learning technique using mobilenetv3 and data augmentation," *Engineering Proceedings* vol. 56, no. 1:197, 2023.

[41] H. H. Eldawoudy, M. A. Mohamed and E. AbdElhalim, "An Ensemble DNN Model for Automatic Detection of COVID-19 from CXR Scans," *Mansoura Engineering Journal*, vol. 49 , no. 1:10, 2023.

[42] M. Azam, A. Yousaf, F. Zafar, M. S. Munir and M. I. Saeed, "A Realtime Data Analysis Approach for Predicting COVID-19 Outcomes using Heterogeneous Ensemble Learning," *2023 International Conference on Engineering and Emerging Technologies (ICEET)*, Istanbul, Turkiye, pp. 1-6, 2023.

[43] T. H. Bui, K. Hamamoto, L. K. Bui and M. P. Paing, "Multi-Disease Classification of COVID-19 in Chest Radiographs using Ensemble of Optimized Deep Learning Models," *2023 15th Biomedical Engineering International Conference (BMEiCON)*, Tokyo, Japan, pp. 1-5, 2023.

[44] S. Jain, P. Jaidka and V. Jain, "Deep Learning Ensemble Method for Plant Disease Classification," *2023 International Conference on Communication, Security and Artificial Intelligence (ICCSAI)*, Greater Noida, India, pp. 383-387, 2023.

[45] V. Pandiyaraju *et al.*, "Improved tomato leaf disease classification through adaptive ensemble models with exponential moving average fusion and enhanced weighted gradient optimization," *Frontiers in Plant Science*, vol. 15, 2024.

[46] P. Khanarsa and S. Kitsiranuwat, "Deep

Learning-based Ensemble Approach for Conventional Pap Smear Image Classification," *ECTI-CIT Transactions*, vol. 18, no. 1, pp. 101–111, Feb. 2024.

[47] N. Alabid, "Interpretation of Spatial Relationships by Objects Tracking in a Complex Streaming Video," *ECTI-CIT Transactions*, vol. 15, no. 2, pp. 245–257, May 2021.

[48] A. Dey, S. Biswas and L. Abualigah, "Efficient Violence Recognition in Video Streams using ResDLCNN-GRU Attention Network," *ECTI-CIT Transactions*, vol. 18, no. 3, pp. 329–341, Jul. 2024.

Nayan Kumar Sarkar received his M.Sc. degree in Information Technology from Gauhati University, Guwahati, India in 2013 and the M.Tech. degree in Information Technology from the Tezpur University, Napaam, India in 2016. He is currently pursuing his Ph.D. degree in Computer Science and Engineering from North Eastern Regional Institute of Science and Technology (NERIST), Arunachal Pradesh, India. Mr. Sarkar has published some research papers in international journals and conferences. He received the best paper award in ICEAI-2023 held at Taylor's university, Malaysia. His research area of interest includes machine learning, deep learning and image classification.

Moirangthem Marjit Singh is currently an Associate Professor in Computer Science & Engineering Department at North Eastern Regional Institute of Science & Technology (NERIST), Arunachal Pradesh, India. He received B.Tech. & M.Tech. degrees from NERIST and PhD degree from University of Kalyani, India in 2001,2010 and 2017 respectively. He was the Head of Department Computer Science & Engineering, NERIST (2018 – 2022), founder Honorary Joint Secretary of the IE(I), Arunachal Pradesh State Centre, India (2019-2021) and founder member Unnat Bharat Abhiyan NERIST Cell (2017-2024). Currently, he is In-charge of Educational Technology Cell at NERIST. He is a Fellow of IETE India, Fellow of IE (I) and senior member IEEE, USA. He was honoured with "Academic Excellence Award" by Taylor's University, Malaysia in recognition of his outstanding academic performance on 13 September 2023. He received the "IE(I) Young Engineers Award 2014–2015" in Computer Engineering Division from Institution of Engineers, India. He received the "Best Paper Awards" at international conferences namely the ICEAI 2023(Taylors' University, Malaysia) and the ICACCT 2016, (APIIT, India). Dr. Marjit secured First Position in X and Second Position in XII Examinations conducted by CBSE, New Delhi, India, amongst the candidates sent up from Jawahar Navodaya Vidyalayas (JNVs) of North Eastern region states of India, in 1995 and 1997, respectively. He did his schooling at JNVSA Kakching, Thoubal District, Manipur (1990-1997). He was Gold Medallist in the M.Tech.(CSE) program. He has a patent granted for 20 years by the Patents office Govt. of India with effect from 30 July, 2021. He has published several papers in international journals, book chapters and conferences of repute. He has been associated with many technical conferences held in India and abroad. Dr. Marjit has delivered many technical/invited talks as well. His research interests include MaNet, WSN, Security, ML, DL, and Image Classification.

Utpal Nandi received his M.Sc. degree in computer science from Vidyasagar University, West Bengal, India, in 2006 and secured 2 nd position (Silver medallist). He earned his M.Tech. degree in computer science and engineering and secured 1 st position (Gold medallist) from the University of Kalyani, West Bengal, India in 2009. He completed the Ph.D. (Engg.) degree in 2018 from the same university. He is currently working as an Assistant Professor in the Department of Computer Science, Vidyasagar University. He has 15 years of teaching and research experience. He has published more than 55 articles in International journals, book chapters, and conferences. His research interests include deep learning based sign language recognition, plant disease identification, hyperspectral image classifications and band selection, design of optimizers of deep learning models, data and image compression techniques, image processing, multimedia technology, computer vision, and artificial intelligence. Dr. Nandi has three patents also. He was awarded by Vidyasagar University for outstanding contribution in research activity (Research and Publication Award) for the years 2021, 2022, and 2023. He also delivered lectures in different conferences and seminars invited as resource person. He involved in different organizing and technical program committees of conferences and seminars.

Jyotsna Kumar Mandal M. Tech. in Computer Science from University of Calcutta in 1987, awarded Ph. D. (Engineering) in Computer Science and Engineering by Jadavpur University in 2000. Working as Professor of Computer Science & Engineering, University of Kalyani. Former Vice Chancellor, Raiganj University, West Bengal, Former Dean, Faculty of Engineering, Technology & Management, KU for two consecutive terms during 2008-2012. Former Director, IQAC and Chairman CIRM Kalyani University. Served as Professor Computer Applications, Kalyani Government Engineering College for two years. He was Associate Professor Computer Science for eight years at North Bengal University and Assistant Professor Computer Science North Bengal University for seven years. He also served as lecturer at NERIST, Itanagar for one year. 36 years of teaching and research experience in Coding Theory, Data and Network Security and authentication; Remote Sensing & GIS based Applications, Data Compression, Error Correction, Visual Cryptography and Steganography. Awarded 30 Ph. D. Degrees and 8 are pursuing. Supervised 03 M. Phil, more than 80 M. Tech and more than 130 M.C.A. Dissertations. Published more than 450 research articles. Recently he has published a text book on Reversible Steganography and Authentication via Transform Encoding from Springer (`https://link.springer.com/book/10.1007/978-981-15-4397-5`). This book has been translated into Chinese and republished from China by Springer. Organized more than 60 International Conferences and Corresponding Editors of edited volumes and conference publications of Springer, IEEE and Elsevier etc. and edited 60 volumes as volume editor. Received "Shiksha Ratna" Award from Government of West Bengal, India for outstanding teaching and research work in 2018. ISO world Convenor of ISO/IEC JTC 1/SC36/WG7. Governing Council(GC) Member of IETE, India.