IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 24 January 2024, accepted 11 February 2024, date of publication 19 February 2024, date of current version 26 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3367443

RESEARCH ARTICLE

A Lightweight Meta-Ensemble Approach for Plant Disease Detection Suitable for IoT-Based Environments

RITESH MAURYA[1], **SATYAJIT MAHAPATRA**[2], **AND LUCKY RAJPUT**[3]
[1]Amity Centre for Artificial Intelligence, Amity University, Noida 201301, India
[2]Department of Information and Communication Technology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, India
[3]Amity School of Engineering and Technology, Amity University, Noida 201301, India

Corresponding author: Satyajit Mahapatra (satyajit.mahapatra@manipal.edu)

ABSTRACT By providing food to billions of people, agriculture contributes significantly to the global economy. Plant ailments, however, can reduce crop yields and result in financial losses. An automated artificial intelligence (AI)-based method for the automatic identification of plant diseases using resource-constrained Internet of Things (IoT) devices has been presented to solve this issue. However, the deployment of state-of-the-art convolution neural networks (CNNs) and Vision Transformers (ViT) on IoT devices is not feasible due to their large number of trainable parameters. To overcome this limitation, a meta-ensemble of lightweight MLP-Mixer and faster Long Short-Term Memory (LSTM) models has been proposed for plant disease detection on low-powered micro-controllers (MCUs) of IoT devices. The MLP Mixer model is based on a simple multi-layer perceptron network. The proposed meta-ensemble consists of two levels: predictions made by the trained models at the first level are used to train the machine learning classifier at the next level, resulting in further improvement of categorisation accuracy. The proposed meta-ensemble has been tested on three diverse datasets of varying sizes and plant species, including Maize, Cotton, and a dataset derived from the Plant Village(PV) dataset. On the Maize, Cotton, and derived PV datasets, respectively, experimental results demonstrate that the suggested technique obtained classification performance of 94.27%, 98.43%, and 97.45%. Moreover, prediction time of the proposed meta-ensemble is low, and it has considerably fewer trainable parameters than CNN and other transformer-based architectures. Therefore, the proposed meta-ensemble is an efficient and effective solution for plant disease detection with limited resources.

INDEX TERMS Convolution neural network, ensemble, artificial intelligence, deep learning, Internet of Things, disease detection.

I. INTRODUCTION

Plant diseases pose massive threats to agricultural productivity, necessitating their early detection utmost important for effective disease management. Manual analysis of plant by pathologists is time-consuming and subjective to the knowledge of domain expert, leading to the development of artificial intelligence-based automated systems for faster and more accurate identification of plant diseases. Conventionally, machine learning (ML)-based methods have been utilised, leveraging the features extracted from the diseased plant images though manual feature engineering process. In contrast to ML-based methods [1], [2] [3], deep learning methods, particularly convolution neural networks (CNNs), have shown promising results in learning relevant features automatically.

State-of-the-art CNNs like VGG16, ResNet50, DenseNet, InceptionNet, and MobileNet trained on ImageNet, have demonstrated significant improvement in performance across different domains etc. [4], [5], [6]. Customised CNNs [7],

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

© 2024 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License.
For more information, see https://creativecommons.org/licenses/by/4.0/

[8] and attention-based techniques [9], [10], [11] have also been deployed for plant disease classification. However, the deployment of such models on low-powered Internet of Things (IoT) devices with limited computational resources remains a challenge.

Recently, the lightweight MLP-Mixer architecture has gained attention due to its lesser architectural complexity and competitive performance on ImageNet dataset [12]. This architecture, which relies solely on multi-layer perceptrons(MLPs) without convolution and attention mechanism, present a promising solution for resource-constrained IoT environments.

A. MOTIVATION AND CONTRIBUTION

Despite existing literature on machine learning and deep learning algorithms for plant disease diagnostics, there is still a need for the development of lightweight solutions that can be easily implemented in resource-constrained environments with limited memory and computation power. This research presents a unique two-tier meta-ensemble approach to address the need for lightweight models that can be deployed in resource-constrained IoT-based situations for automated plant disease diagnosis. The proposed approach harnesses the benefits of the MLP-Mixer and Long Short Term Memory (LSTM) models to improve classification performance while staying appropriate for usage in resource-constrained contexts.

The adoption of the suggested meta-ensemble technique is supported by its lightweight nature, which makes it suited for deployment in resource-constrained situations such as IoT devices. Integrating MLP-Mixer and LSTM models into the proposed meta-ensemble allows for the use of their complimentary capabilities thereby enhancing the classification performance.

The rest of the article has been split up into the following sections: Section II details the related works. Section III provides details about the methods deployed in the proposed work. Section IV displays experimental results and provides detailed discussions of them. Section V provides a concrete outline of the proposed work.

II. RELATED WORKS

Some of the prior works related to the plant disease categorisation task have been discussed in this section. This paragraph describes some of the convolution neural network based models proposed by the different researchers. Zhao et al. [9] have proposed a method consisting of an inception module and residual connection for the identification of diseases related to the corn, potato and tomato plants. They also suggested the use of a web-based system for the real-time identification of plant diseases [9]. Pandey and Jain have proposed an attention-based dense CNN model for the detection of 44 diverse types of plant diseases using a dataset constructed from the 10,851 images captured from the field and achieved 97.33% categorisation accuracy [13]. Bedi and Gole proposed a convolutional autoencoder and CNN-based method for the categorisation of Bacterial Spot disease of the peach plants with 98.38% categorisation accuracy [14]. Ferentinos has tested different CNNs such as AlexNet, GoogleNet, and VGGNet for the categorisation of 17,548 images of 58 different classes of plant disease and obtained a categorisation accuracy of 99.53% [15]. The MobileNet model developed by Kamal et al. with deep separable convolution achieved 97.65% categorization accuracy on the PlantVillage dataset [16]. A customised CNN model has been suggested by Chohan et al. for the classification of illnesses in 15 distinct plants [17]. The InceptionResNet model was suggested by Hassan and Maji for the categorisation of 15 different plant disease types [18]. Atila et al. have proposed the EfficientNet model for the categorisation of 39 different diseases present in the PV dataset [19]. Amin et al. have proposed a method for corn leaf disease classification by combining the features extracted from the EfficientNetB0, and DenseNet121 deep CNN models and achieved 98.56% classification accuracy [20]. Maurya et al. have proposed a method for classification of diseases present in the PlantVillage dataset using pre-trained Vision Transformer network and interpreted the performance of the model using GradCAM algorithm [21].

Some of the works under miscellaneous category, proposed by different researchers for the plant disease categorisation have been summarised as follows: Abbas et al. have utilised generative adversarial networks to produce synthetic images of the diseased leaves of the tomato plant [22]. Five different types of potato plant diseases have been classified with the DenseNet121 model with 97.11% categorisation accuracy. Thakur et al. have utilised the ViT architecture for the categorisation of the images of plant diseases and achieved an average accuracy of more than 93% in the case of Apple, Maize, and Rice datasets [23]. For tomato leaf disease classification, Karthik et al. [24] proposed a strategy based on the use of the attention mechanism in a deep CNN. Their suggested model performed 98% categorization correctly when evaluated with 24001 photos [24]. Shah et al. suggested a teacher/student architecture for identifying 14 different plant diseases [25].

Most of the works discussed above either used convolution or attention mechanisms embedded with the CNN architecture. These models cannot be adapted to an IoT-based environment where there is a constraint of limited memory and computational power. Internet of things faces several challenges such as limited resources in terms of computing, power and memory capacity [26]. Therefore, in the proposed work, a lightweight approach has been presented which does not rely on convolution or attention mechanism, thereby, it is well suited for IoT-based deployment. The proposed model also utilises the multi-tier meta ensemble approach in which the prediction probabilities obtained from the trained models at the first level are used as a feature set to train the model at the second level. The meta-ensemble approach helps in further improving the categorisation performance of the proposed method.

III. DATASET USED

Three different publicly available datasets pertaining to various plants were used to test the proposed framework. Examples of the photos found in these datasets are shown in Fig. 1.

(a) Cotton (b) TPP (c) Maize

FIGURE 1. Images of the samples taken from each dataset (a) Cotton Dataset (b) Tomato/Potato/Pepper Dataset (c) Maize Dataset.

A. COTTON LEAF DISEASE DATASET (COTTON DATASET)

This dataset [27] consists of 1518 images of four different classes of cotton leaf disease images captured under real-world conditions and also from the internet. The images of four different leaf diseases such as 'Curl virus', 'bacterial blight', 'fusarium wilt' and 'healthy plant' leaves images were present in this dataset. The number of sample images present in each class of the Cotton Dataset has been presented in Table 1.

TABLE 1. Number of sample images present in each class.

S.No.	Disease	#Sample Images
1	Fussarium Wilt (FW)	379
2	Curl Virus (CV)	385
3	Healthy (H)	411
4	Bacterial Blight (BB)	343
	Total	1518

B. MAIZE LEAF DISEASE DATASET (MAIZE DATASET)

This dataset [28], [29] has been derived from popular datasets such as PlantDoc and PV datasets. This dataset consists of 2529 images in '.jpg' format. This dataset includes photos of four different types of maize leaf illnesses, including "Common Rust," "Grey Leaf Spot," "Blight," and "Healthy" Plant Leaves. Total sample images present in each class of this dataset have been presented in Table 2.

TABLE 2. Number of sample images present in each class.

S.No.	Disease	#Sample Images
1	Common Rust (CR)	1192
2	Blight (BL)	870
3	Healthy (H)	21
4	Gray Leaf Spot (GLS)	446
	Total	2529

C. TOMATO/POTATO/PEPPER DATASET (TPP DATASET)

This dataset has been derived from the PV dataset [29] and consists of 20637 images of plant disease in '.jpg' format. This dataset consists of images of diseases belonging to three different types of plants such as tomato, potato and pepper. This dataset has been termed as 'TPP' (Tomato/Potato/Pepper) throughout the rest of the literature. The three classes of this dataset belong to the potato plant, two classes belong to the pepper plant and the rest of the classes belong to the tomato plant. The total sample images present in each class of the TPP dataset has been shown in Table 3.

TABLE 3. Number of sample images present in each class.

S.No.	Disease	#Sample Images
1	Pepper Bell Bacterial Spot (PBB)	997
2	Potato Healthy (PH)	152
3	Tomato Leaf Mold (TLM)	952
4	Tomato Yellow Leaf Curl Virus (TY)	3208
5	Tomato Bacterial Spot (TB)	2127
6	Tomato Septoria Leaf Spot (TSLS)	1771
7	Tomato Healthy (TH)	1590
8	Tomato Spider Mites Two-Spotted Spider Mite (TSMTSM)	1676
9	Tomato Early Blight (TEB)	1000
10	Tomato Target Spot (TTS)	1404
11	Pepper Bell Healthy (PBH)	1476
12	Potato Late Blight (PLB)	1000
13	Tomato Late Blight (TLB)	1756
14	Potato Early Blight (PEB)	1000
15	Tomato Tomato Mosaic Virus (TTMV)	373
	Total	20367

IV. PROPOSED METHODOLOGY

The methodology for the proposed meta ensemble framework for plant disease detection has been shown in Fig. 2. The whole methodology has been divided into four steps: (i) In the first step, the whole dataset has been split into training and test set (ii) then in the next step, pre-processed training set images were used to train the models (Mixer and LSTM) present at the level 1 (iii) After the level 1 models are trained, the level 2 support vector machine classifier is trained using the features that are extracted from these models (as an output of these models). (iv) After training the models present at both levels, the test set images were first given as input to the trained models present at level 1 to draw the features. Then drawn-out features of these models were concatenated and then given as an input to the trained SVM model present at level 2 to reach the final decision. Different component of the proposed methodology has been explained as follows:

A. SPLIT THE DATASET

Training and test sets have been created from the entire dataset. While the test set photos were used to gauge how well the proposed meta ensemble framework performed at categorising images, the training set images were utilised to train the models. The experimental findings section contains a description of the number of sample photos that were utilised for training and testing.

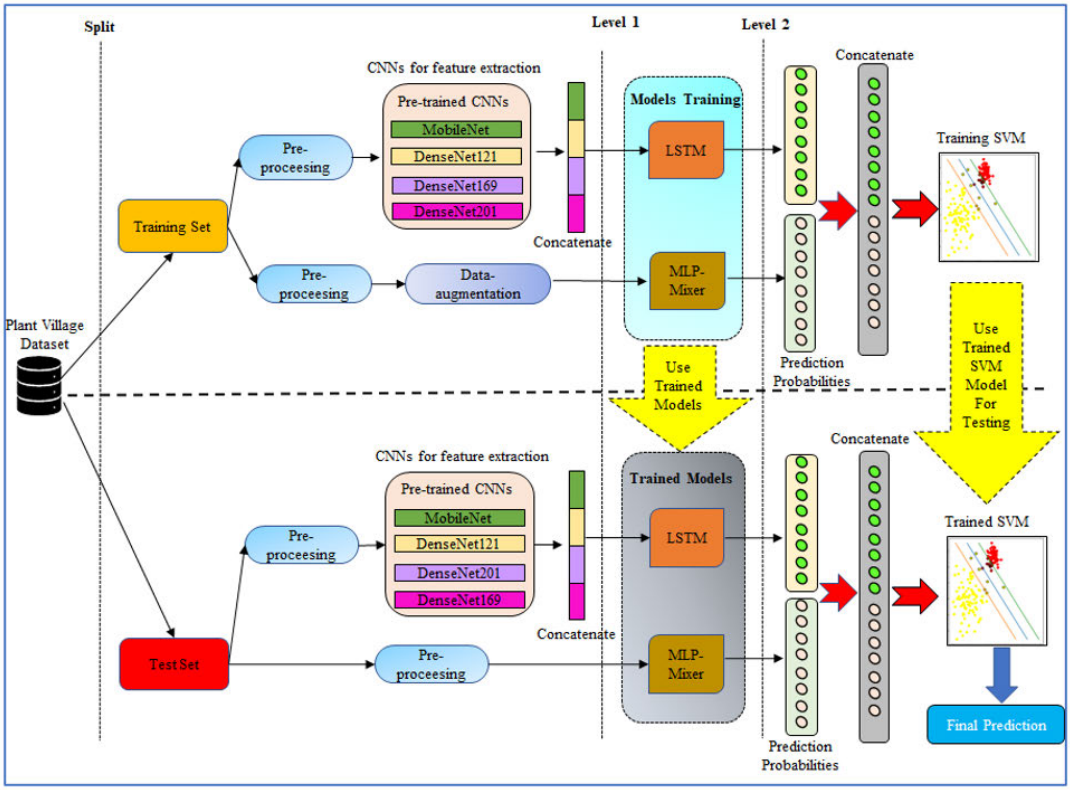


FIGURE 2. The suggested meta-ensemble framework for detecting plant diseases.

B. PRE-PROCESSING

Since learning the features from large shape images increases the computational burden of the models used in the proposed ensemble; therefore, considering the limited availability of computational resources, all the images were resized to an optimal shape before passing them as an input to these models. After several experiments, an optimal shape of input plant disease images for the proposed meta ensemble was found to be 64X64X3, therefore, all the images were reshaped to 64X64X3 before passing them to the models present in the proposed meta ensemble. After reshaping the images, all the images were normalised so that their pixel values come into the range 0 to 1. Normalisation helps in speeding up the convergence speed of the models used in the proposed meta ensemble.

C. DATA AUGMENTATION

Considering the small sample size of the training set images, the data augmentation technique in form of affine transformations has been quite useful in artificially increasing the number of training samples. After pre-processing the training set images, the size of the training set is increased using the data augmentation technique. Data augmentation helps in preventing the models present in a meta ensemble from overfitting the data; thus, increasing the generalisability of these models. Random flipping, random rotation, resizing, width- and zooming operations were used to artificially increase the training set samples.

D. THE COMPONENTS OF THE PROPOSED META ENSEMBLE

The architectural components of the proposed meta-ensemble can be explained as follows: at first, the details of the architectures used at each level of the proposed meta-ensemble (as shown in Fig. 2) have been described. Then, how these architectures were connected to give a final shape to the proposed meta-ensemble has been described. For better understanding, the design of the overall meta-ensemble has been divided into two levels: level 1 and level 2. The detail of architectures present at each level of the proposed meta ensemble has been provided as follows:

1) DESCRIPTION OF THE MODELS PRESENT AT LEVEL 1

a: MLP MIXER

The reason for choosing this MLP Mixer architecture [12] as one of the components of the proposed meta ensemble is that it is based on a simple multi-layer perceptron architecture and it does not use any kind of attention mechanism and convolution operations which makes the MLP Mixer model comparatively light-weight in comparison to the CNN and ViT architectures. The performance of the MLP Mixer is also commensurate to the state-of-the-art CNN and ViT architectures. The architecture of the proposed MLP-Mixer architecture has been shown in Fig. 3.

MLP Mixer model takes images as input in form of patches, therefore, before passing the pre-processed input images to the MLP Mixer model, each image has been

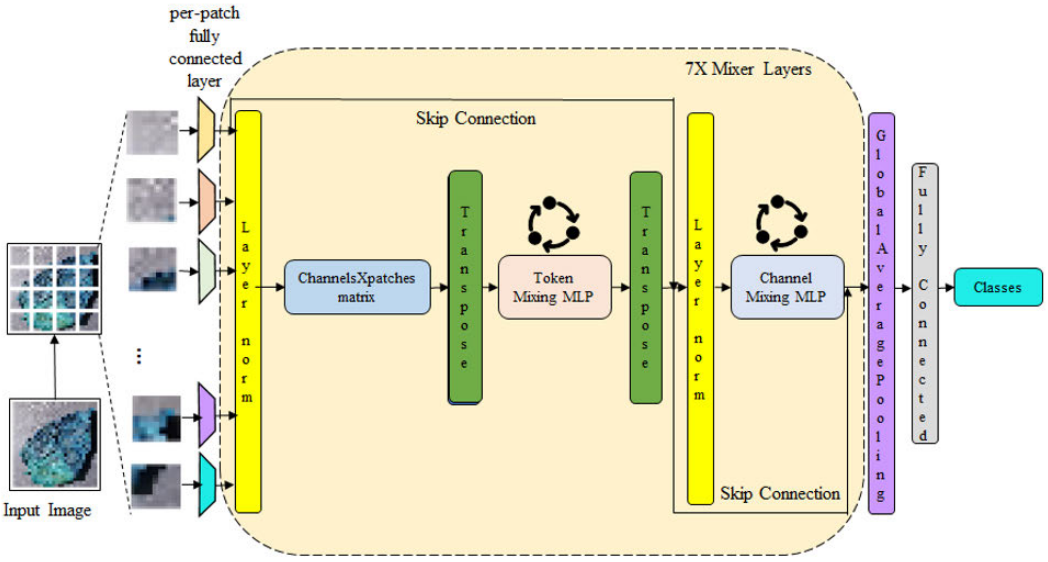


FIGURE 3. The architecture of the proposed MLP mixer architecture.

divided into several patches and each of these patches has been further projected into D dimensional space (here, D = 128) of fixed size, the projected embeddings are termed as 'tokens'. The core functionality of the Mixer architecture lies in its mixer layers. Mixer layers composed of MLPs which perform two different operations, i.e., mixing of the tokens and the channels. The token mixing allows the MLP Mixer architecture to learn the spatial relationship between the tokens (patch embeddings); whereas, the channel mixing MLP allows the model to learn the inter-relationship between the channels present in the single token itself.

Thus, in any mixer layer an input matrix of shape $(NXD, N = 9, D = 128)$, where N is the number of patches and D is the embedding dimension, passes through the token mixing and channel mixing MLP by transposing the input matrix accordingly. An MLP in the mixer layer consists of two fully-connected (FCN) layers with Gaussian Error Linear Unit (GELU) non-linearity. Thus, in any mixer layer: layer normalisation, GELU non-linearity and skip connections between the two MLPs are used for the smoother flow of the gradient among the layers. A series of mixer layers with the same form make up the MLP Mixer. Because increasing the number of mixer layers also makes the MLP Mixer more difficult, an ideal number of mixer layers (seven in this case) has been chosen for the current plant disease classification assignment. The output of the final mixer layer is passed through the normalisation layer first, the dropout layer (rate = 0.25), the global average pooling (GAP) layer, and then the categorisation layer with the activation function ''softmax'' for each row of an input matrix. Using Equations 1 and 2, the mixer layer in the MLP Mixer model can be represented.

$$T_{*,i} = X_{*,i} + W_2\sigma(W_1.LN(X)_{*,i}), \quad \text{for } i = 1, \ldots, C, \quad (1)$$

$$J_{j,*} = T_{j,*} + W_4\sigma(W_3.LN(T)_{j,*}), \quad \text{for } j = 1, \ldots, S, \quad (2)$$

where T and J denote the output of the first and the second FCN layers. LN denotes the layer normalisation operation. $W1$, $W2$, $W3$ and $W4$, denote the weight matrices. X denotes the input to the first FCN layer. C and S denote the number of channels and tokens respectively. The other hyperparameters related to the proposed MLP mixer have also been presented in Section V.

b: LONG SHORT TERM MEMORY (LSTM)

Long Short-Term Memory (LSTM) is a recurrent neural network (RNN) architecture that addresses the vanishing gradient problem in regular RNNs. LSTMs regulate information flow by using a memory cell and three gates (input, forget, and output). The cell stores information across the long sequences, allowing the network to capture and learn data dependencies more efficiently. Their capacity to handle long-range dependencies makes them ideal for a variety of sequential data applications. The reason for choosing LSTM architecture for the proposed meta-ensemble is that the LSTM model applies operations directly to the data and does not use the convolution concept as well as the attention mechanism. Therefore, LSTM is also lightweight in comparison to CNN and ViT architectures. Thus, LSTM has also been deployed for the development of the proposed meta-ensemble. Though it is not reasonable to train the LSTM directly on input images, considering the raw information present in an input image; therefore, the features drawn out from the last convolution layer of ImageNet-trained CNNs were used to train the proposed LSTM. Input images were resized to 64X64X3 before passing them as input to these ImageNet-trained CNNs for the extraction of features. These ImageNet-trained CNNs had their top-most FC layers removed, and the activations from that layer were then sent to the layer known as ''global average pooling'' to be used in feature extraction. The weights of the convolution

base in ImageNet-trained CNNs were kept frozen. The features drawn from four different pre-trained CNNs such as MobileNet, DenseNet121, DenseNet169 and DenseNet201 were concatenated to form features set having dimensionality, D = 5632. The dimensionality of the individual feature set drawn out from the MobileNet, DenseNet121, DenseNet201 and DenseNet169 was 1024, 1024, 1536, and 2048 consecutively. The concatenated features of shape (1, 5632) were given as input to the Long Short Term architecture. The value of the time-step chosen for the LSTM architecture was 1. The total number of cells chosen for the present categorisation task was 30. The number of trainable parameters present in the proposed LSTM was lesser than 0.3 million. The other hyperparameters used in the proposed LSTM have been described in Section V.

The combination of LSTM and MLP-mixer used at the level 1 of the proposed method helps in learning the patch-level, channel-level and feature-level dependencies present in an input image. LSTM model has been used to learn the distinguishing characteristics present in the one-dimensional combined feature vector obtained after combining the feature set obtained from the MobileNet, DenseNet121, DenseNet201 and DenseNet169 models. Both models when used together in the meta-ensemble at level 1, gives better classification performance with optimised run time, in comparison to the other combinations of models such as other variants of the vision transformer model.

2) DESCRIPTION OF THE MODEL PRESENT AT LEVEL 2

As shown in Fig. 2 the proposed meta ensemble is composed of models present at two different levels. The predictions made by the models (MLP Mixer and LSTM) present at level 1 are used as a feature set to train the ML model (support vector machine) present at level 2.

Support Vector Machine (SVM): SVM classifier is based on the theory of maximising the margin between separating hyperplanes [30]. SVM is well known for its better performance with a limited amount of training data [31]. Therefore, it has been chosen as the final classifier in the proposed two-level meta ensemble approach. SVM takes its input from the predictions made by the models present at level 1. SVM classifier has been chosen due to its better performance in contrast to the other ML classifiers such as Naïve Bayes, Random Forest and Nearest-Neighbor. The SVM classifier has also been proven to be superior to other classifiers in the context of the current categorization task, and the related experimental findings are presented in the results part of the current publication.

E. COMBINING THE MODELS PRESENT AT BOTH LEVELS OF THE PROPOSED META ENSEMBLE

Fig. 2 displays a graphic representation of the suggested meta-ensemble framework. The training and testing phases of the proposed meta-ensemble have been described separately to aid in understanding how it functions. Fig. 2 demonstrates how the suggested technique has been divided into two levels: level 1 of the proposed meta ensemble contains the MLP Mixer and LSTM models, while level 2 of the ensemble contains the SVM classifier.

The overall objective of this study is to build a lightweight framework that can be deployed on IoT-based devices, therefore, MLP-mixer and LSTM were found suitable to create meta ensemble since they are lightweight, more accurate and accelerates the prediction time. Pre-processed augmented training set images were used to train the MLP Mixer model present at level 2 while the LSTM model has been trained on the concatenated features, drawn out from the four different ImageNet-trained CNNs. All images were resized to 64×64 before passing them to ImageNet-trained CNNs for drawing out the features from them. The features drawn out from the pre-trained MobileNet, DenseNet121, DenseNet169 and DenseNe201 architectures were concatenated to form the combined feature vector of shape (1,5632) which is used to train LSTM. After training both the models present at level 1, the prediction probabilities of these trained models were recorded by providing training set images as input to them. The shape of the prediction probability matrix obtained from these models was No._of_training_images X num_classes. After concatenating the prediction probability vector obtained from both the trained models present at level 1, the combined feature representation matrix of shape (No. of training images X (2Xnum_classes)) was used to train the SVM classifier present at level 2.

The following procedures have been used to test the proposed method: first, unseen test photos were pre-processed in the same way that the training set images were. Then the models (LSTM and MLP-Mixer) trained during the training phase were used to obtain the prediction probabilities by giving test set images as an input to them. The prediction probabilities obtained from these models (LSTM and MLP-Mixer), were concatenated and then passed as input to the trained SVM classifier to make the final decision about the class of an input test set images. The testing phase of the proposed meta ensemble can be represented mathematically using Eq.3- 7.

$$X_{\text{test-LSTM}} = [F_{\text{MobileNet}}, F_{\text{DenseNet121}}, F_{\text{DenseNet169}}, F_{\text{DenseNet201}}] \tag{3}$$

$$P_{\text{LSTM}} = \text{LSTM}(X_{\text{LSTM}}, \Theta_{\text{LSTM}}) \tag{4}$$

$$P_{\text{Mixer}} = \text{MLP\_Mixer}(X_{\text{Mixer}}, \Theta_{\text{Mixer}}) \tag{5}$$

$$P_{\text{Concat}} = [P_{\text{Mixer}}, P_{\text{LSTM}}] \tag{6}$$

$$Y_{\text{Final}} = \text{SVM\_Predict}(P_{\text{Concat}}, \Theta_{\text{SVM}}) \tag{7}$$

$X_{\text{test-LSTM}}$ shown in Eq. 3 denotes the combined feature vector obtained after combining the feature vectors obtained from the MobileNet, DenseNet121, DenseNet169, and DenseNet201 models by giving test set images as input to these pre-trained deep CNN models. LSTM and MLP_Mixer, in Eq. 4 and Eq. 5, denote the trained LSTM and MLP-Mixer models, and P_{LSTM} and P_{Mixer} denote the

predicted probabilities of these models. After concatenating these probabilities into a vector, the final matrix denoted by P_{Concat} in Eq. 6 is used to test the SVM model trained during the training phase, as shown in Eq. 7.

V. EXPERIMENTAL RESULTS AND DISCUSSIONS

Python 3.6 was used to implement each experiment, and an Nvidia K80 GPU with 16GB of RAM was used. The effectiveness of the suggested meta-ensemble has been evaluated using a variety of assessment measures, including as precision, recall, F1 score, and accuracy. An ROC (Receiver operating characteristic) curve has also been plotted for each class represented in each dataset. The dataset has been split into the ratio of 0.8:0.2, 80% of the data was used for training and remaining 20% were used for the training. The division of the entire dataset into a training set and a test set is shown in Table 4.

TABLE 4. The number of training and test set images present in each dataset.

Dataset	Training Set	Testing Set	Total
Cotton	1214	304	1518
Maize	2023	506	2529
TPP	18573	2064	20637

The hyperparameters of the different architectures used in the proposed meta ensemble have been shown in Table 5 and Table 6. Table 5 shows the hyperparameters for the proposed MLP Mixer model and Table 6 shows the hyperparameters used in the proposed LSTM architecture used in designing the proposed meta-ensemble.

TABLE 5. Hyperparameters of the proposed MLP Mixer Architecture.

Hyperparameter	Value
Number of Patches	9
Number of Mixer layers	7
MLP Channel dimension	64
Hidden layer size	128
Input image shape	64×64
Epochs	50–100
Loss	sparse categorical crossentropy
Optimizer	Adam (learning rate = 0.001)
Batch Size	512

TABLE 6. Hyperparameters of the proposed LSTM architecture.

Hyperparameter	Value
Batch Size	256
Epochs	50–100
Time Step	1
Optimizer	rmsprop
Loss	'categorical crossentropy'
Validation Split	0.1

The SVM classifier used at the second level of the proposed meta-ensemble has been fine-tuned using a grid-search strategy. The grid-search has been performed using the following values: 'C': [0.1, 1, 10, 100, 1000], 'gamma': [1, 0.1, 0.01, 0.001, 0.0001] and 'kernel': ['linear', 'RBF']. The MLP-Mixer architecture consists of 7 mixer layers. The detailed overview of the single mixer layer including the name, output size and number of parameters have been provided in Table 7.

TABLE 7. Detailed overview of single Mixer layer of MLP-Mixer model.

Layer Name	Output Shape	No. of Parameters
Input (InputLayer)	(512, 64, 64, 3)	0
_augmentation (Sequential)	(512, 72, 72, 3)	7
conv2d_1 (Conv2D)	(512, 18, 18, 128)	6272
reshape_1 (Reshape)	(512, 324, 128)	0
layer_normalization_17 (LayerNormalization)	(512, 324, 128)	256
permute_1 (Permute)	(512, 128, 324)	0
dense_1 (Dense)	(512, 128, 128)	41600
tf.nn.gelu_16 (TFOpLambda)	(512, 128, 128)	0
dense_2 (Dense)	(512, 128, 324)	41796
permute_2 (Permute)	(512, 324, 128)	0
add_16 (Add)	(512, 324, 128)	0

Techniques such as dropout, layer normalization and advanced activation functions such as Gaussian error linear unit(GeLU) has been used to avoid local minima. Moreover, the performance of the proposed method has been analyzed on the validation set and the unseen test set to measure the correct generalizability of the proposed model. To test the generalisation of the MLP Mixer and LSTM models used in the proposed meta ensemble, training and validation accuracy and loss curves have also been plotted for each dataset as shown in Fig. 4 and Fig. 5 respectively. It can be analysed from the training and validation accuracy curves that trained models have neither the high bias nor the high variance and both the models (MLP Mixer and LSTM) have achieved convergence. It can also be observed from Fig. 4 and Fig. 5 that the convergence in the case of LSTM architecture is faster than the convergence of MLP Mixer architecture.

The confusion matrices obtained for the final SVM classifier for each dataset have been shown in Fig. 6(a), 6(b), and 6(c) for TPP, Maize and Cotton datasets respectively. The performance metrics calculated from these confusion matrices have also been presented in Tables 8(a), 8(b) and 8(c) for each dataset. As shown in Table 8(a), in the case of the Maize dataset worst f1 score of 0.83 has been obtained for the 'healthy' class whereas the best f1 score of value 1 has been obtained for the 'blight' and 'grey leaf spot' disease class. The average categorisation accuracy of 94.27% has been obtained in the case of the Maize dataset. As shown in Table 8(b), for the 'Corn' dataset, the best f1 score of 0.99 has been obtained for the 'curl_virus' and 'healthy' classes. The average categorisation accuracy of 98.43% has been obtained in the case of the Cotton dataset. It can be analysed from Table 8(c) that for the 'TPP' Dataset, the lowest f1 score of 0.89 has been obtained for the 'early blight' disease class and the lowest precision and recall have been achieved for 'healthy' and 'spider mite' diseased class of tomato plant. The highest f1 score of 0.99 has been obtained in the case of two different classes of tomato plant named 'late blight'

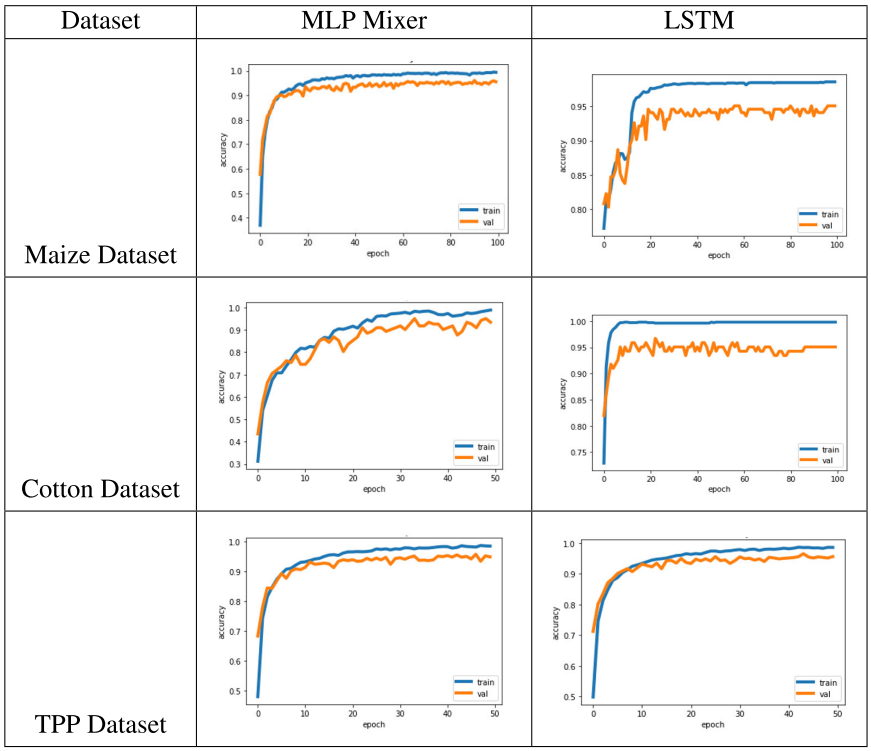


FIGURE 4. The training and validation accuracy curves of the MLP mixer and LSTM models used in the proposed meta ensemble.

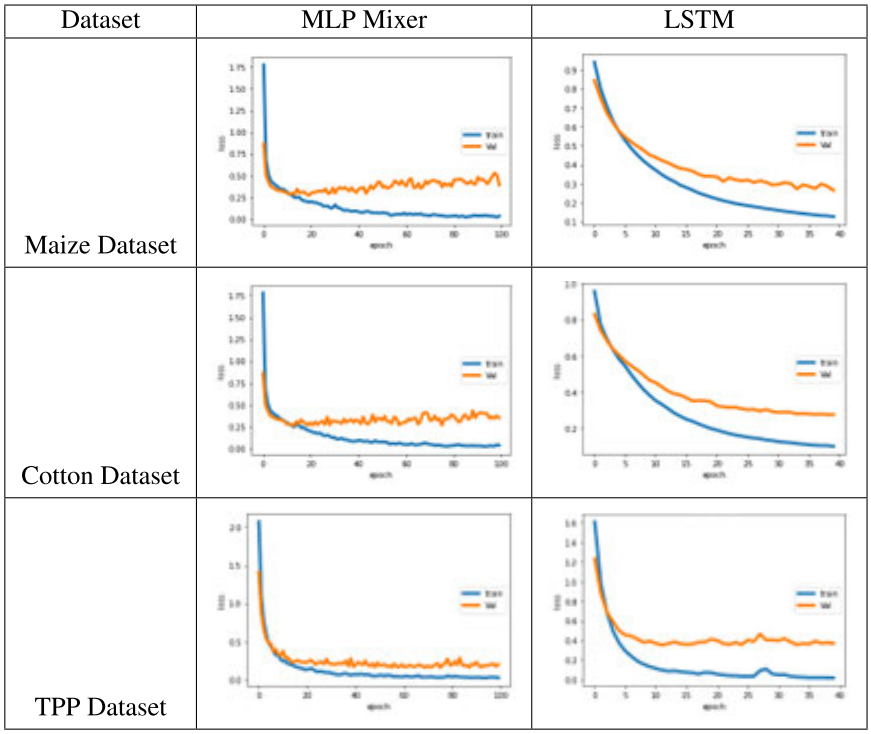


FIGURE 5. The training and validation loss curves of the MLP mixer and LSTM models used in the proposed meta ensemble.

and 'mosaic virus' whereas, in the case of the potato plant, the 'early blight' class has obtained the highest f1 score. The average categorisation accuracy of 97.45% has been obtained in the case of the 'TPP' dataset.

As presented in Table 9, the number of parameters in the proposed meta ensemble is near about a million. The time taken by the proposed meta ensemble, i.e., the time required in getting output from level 1 models (MLP Mixer

TABLE 8. Performance metrics for (a) Maize dataset (b) Cotton dataset (c) TPP Dataset.

(a)

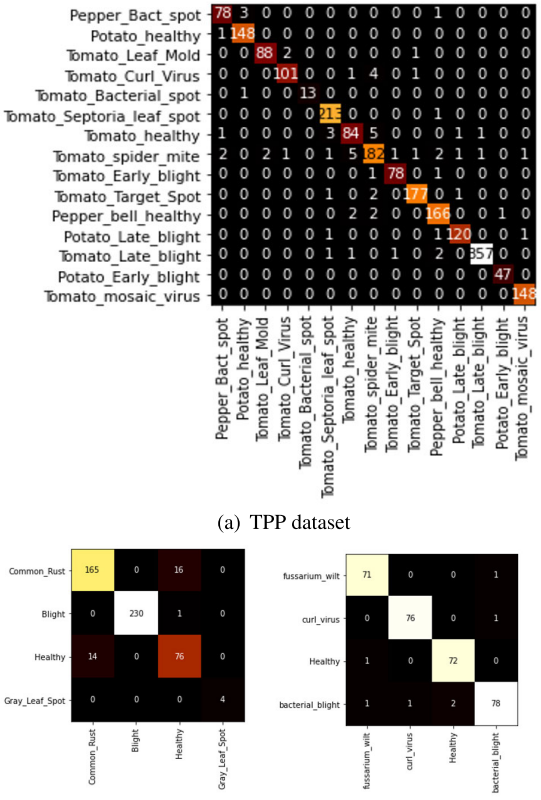
Class Names (Short Form)	Precision	Recall	F1-score	Support
CR	0.89	0.96	0.93	181
B	1.00	1.00	1.00	231
H	0.90	0.77	0.83	90
GLS	1.00	1.00	1.00	4

(b)

Class Names (Short Form)	Precision	Recall	F1-score	Support
FW	0.97	0.99	0.98	72
CV	1.00	0.99	0.99	77
H	0.99	0.99	0.99	73
BB	0.98	0.98	0.98	82

(c)

Class Names (Short Form)	Precision	Recall	F1-score	Support
PBB	0.95	0.95	0.95	82
PH	0.97	0.99	0.98	149
TLM	0.98	0.97	0.97	91
TY	0.97	0.94	0.96	107
TB	1.00	0.93	0.96	14
TSLS	0.97	1.00	0.98	214
TH	0.90	0.88	0.98	95
TSMTSM	0.93	0.91	0.92	200
TEB	0.95	0.94	0.89	80
TTS	0.93	0.91	0.92	181
PBH	0.97	0.97	0.97	171
PLB	0.98	0.98	0.98	123
TLB	0.99	0.99	0.99	362
PEB	0.98	1.00	0.99	47
TTMV	0.99	1.00	0.99	148



(a) TPP dataset

(b) Maize dataset (c) Cotton dataset

FIGURE 6. Confusion matrix for (a) TPP dataset, (b) Maize dataset, and (c) Cotton Dataset.

and LSTM) and the time required in the categorisation of the outputs obtained from these models by the SVM classifier at level 2, has also been shown in Table 9.

It can be observed from Table 9 that the time required by the proposed meta ensemble in the case of the Maize, Cotton and TPP datasets is 0.812 seconds, 0.691 seconds and 1.91 seconds respectively. It can also be analysed from Table 9 that when the outputs from MLP Mixer and LSTM models in the form of prediction probabilities are utilised as a feature set for training and then testing of the SVM classifier; it results in the overall improvement of the categorisation performance of the proposed meta ensemble. The number of parameters and test time taken by the component models present in the proposed meta-ensemble has also been shown in Table 9. The proposed LSTM model is more time efficient with a smaller number of trainable parameters contrast to MLP Mixer; however, the proposed MLP Mixer model is more accurate in terms of %categorisation accuracy.

The comparison of the proposed meta ensemble with other architectures has also been made based on different criteria such as %categorisation accuracy, prediction time and total count of trainable parameters as shown in Table 10. The different variants of the ViT architectures such as ViTL32 [32] and ViTB16 [32] models have been compared with the proposed meta ensemble. The last categorization layer of these ViT models has been replaced with a categorisation layer whose number of neurons matches the number of classes contained in that dataset. The pre-trained version of these ViT models has been used for comparison. Table 10 shows that there are around 300 million and 86 million parameters in the ViTL32 and ViTB16 models, respectively, but only about a million parameters in the suggested meta ensemble. However, the categorisation accuracy obtained by the proposed meta-ensemble in the case of each dataset is higher than the accuracy obtained with both the variants (ViTL32 and ViTB16) of the ViT network. As displayed in Table 10, the testing time of the proposed meta-ensemble is also far lesser than that of the testing time required in heavy-weight ViT architectures.

As shown in Table 10, the suggested meta ensemble approach's categorization accuracy has also been contrasted with that of other pre-trained CNNs, including MobileNet, DenseNet121, DenseNet201, DenseNet169, VGG16, VGG19, and ResNet50. The output from the final convolution base of the ImageNet-trained CNNs has been transmitted to the global average pooling (GAP) layer, and this is how the pre-trained CNNs have been trained. The output from the GAP layers is further passed to two FCN layers consisting of 512 neurons each, in order to measure the performance of the pre-trained CNN and compare it with the proposed meta ensemble. Finally, these CNNs have a categorization layer with a ''softmax'' activation function. It can be analysed from Table 10 that the performance of

TABLE 9. % Accuracy, prediction time, and total count of trainable parameters present in the models used in the proposed meta ensemble. The prediction time of the proposed meta ensemble includes the time required to extract the features from LSTM and Mixer models present at level 1 and the prediction time of the SVM classifier present at level 2. The number of trainable parameters includes the number of trainable parameters of LSTM and Mixer models; SVM cannot be compared with other neural network-based models in terms of the number of parameters; therefore, no parameter has been shown in the case of the SVM classifier.

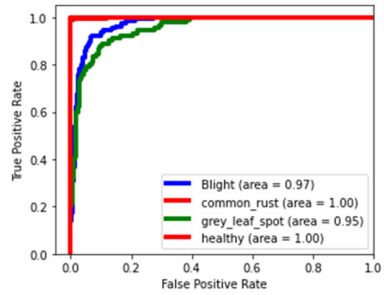
Method	Maize			Cotton			TPP		
	% Accuracy	Test Time	#Trainable Parameters	% Accuracy	Test Time	#Trainable Parameters	% Accuracy	Test Time	#Trainable Parameters
Proposed	**94.27**	**.812**	**923,698**	**98.43**	**.691**	**1,041,316**	**97.45**	**1.91**	**1,041,316**
LSTM	89.35	.12	112,782	94.98	.10	225,885	89	0.29	225,885
MLPMixer	93	.69	810,916	95.53	0.59	815,431	96.02	1.5	815,431
SVM	94.27	.002	-	98.43	.002	-	97.45	0.12	-

the proposed meta ensemble has surpassed the performance of the different DCNNs used for the comparison purpose and the number of parameters in the models (MLP Mixer and LSTM) used in the proposed meta ensemble is also far less than the number of parameters present in these CNN models.

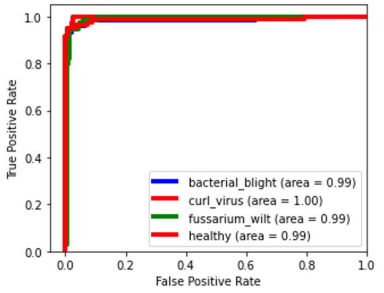
The total count of trainable parameters present in the proposed meta ensemble is near to a million; whereas, the best performing CNN model (other than the proposed one) in the case of Maize, Cotton and TPP dataset is having near about 3 million, 19 million and 7 million parameters respectively. Thus, it establishes that the proposed meta-ensemble is lightweight in comparison to the other ViT and CNN-based models. The % categorisation accuracy obtained by the proposed meta ensemble is also the best among other models used for comparison purposes. The prediction time of the proposed meta ensemble in comparison to the other Transformer and deep CNN-based models is the lowest and the classification accuracy of the proposed meta-ensemble is also the highest, as shown in Table 10. Thus, it establishes that the proposed meta ensemble is accurate, time efficient as well as lightweight and therefore, it is best suited for IoT-based deployment.

Considering the limited memory capacity and limited computing power of IoT-enabled devices the proposed model takes only 18.02 Kilobytes of memory and the number of FLOPS required were 1.88e+04. The ROC (receiver operating characteristic) curve for the proposed meta-ensemble has also been drawn for each class present in these datasets. It can be observed from Fig. 7 that the average AUC (area under the curve) value of 0.98, 0.995 and 0.9993 has been obtained for the Cotton, Maize and TPP datasets respectively. The performance of the proposed model has also been analysed on the Raspberry Pi 4 Model B micro-controller with 4GB RAM and it has been observed that the proposed model has achieved the test time of 1.05 seconds, 0.89 seconds and 2.5 seconds on the test set, for Maize, cotton and TPP datasets respectively.

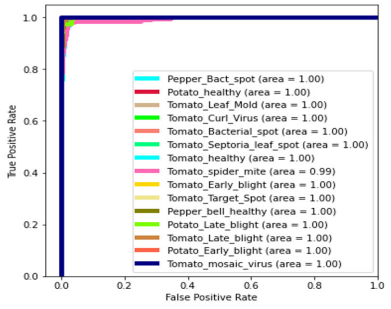
To support the usage of the proposed SVM classifier over other ML classifiers like random forest, Naive Bayes, and K-nearest neighbour, performance comparisons with the other classifiers have also been done. The experimental findings for this comparison are shown in Table 11.



(a) Maize dataset



(b) Cotton Dataset



(c) TPP dataset

FIGURE 7. ROC curves for (a) Maize, (b) Cotton and (c) TPP dataset.

It can be discerned from Table 11 that the in case of all the datasets, the best categorisation accuracy has been obtained with the SVM classifier. Therefore, SVM has been chosen as a level 2 classifier in the proposed meta-ensemble. However, at level 1, MLP Mixer and LSTM models have been chosen due to their lightweight nature and, when these models are used in synchronisation with the level 2 model; it results in further improvement of the categorisation performance of the proposed meta-ensemble approach.

TABLE 10. Comparison of the proposed method with other state-of-the-art vision transformer and convolutional neural networks.

Method	Maize			Cotton			TPP		
	% Accuracy	Time	#Trainable Parameters	% Accuracy	Time	#Trainable Parameters	% Accuracy	Time	#Trainable Parameters
Proposed	**94.27**	**.812**	**923,698**	**98.43**	**.691**	**1,041,316**	**97.45**	**1.91**	**1,041,316**
ViTL32	80.02	6.7	305,468,420	87.02	5.7	305,479,695	80.05	6.9	305,479,695
ViTB16	83.24	3.3	85,663,492	90.12	2.9	85,671,951	83.66	3.9	85,671,951
MobileNet	93.33	1.5	3,996,484	75.43	1.2	3,996,484	93.14	1.7	4,002,127
EfficientNetB0	89.23	0.91	5,288,548	70.34	0.88	5,288,548	88.14	0.91	5,288,618
SqueezeNet	90.37	0.85	1,248,424	73.55	0.72	1,248,424	86.55	0.85	1,248,704
DenseNet121	91.59	6.7	7,743,364	79.51	6.6	7,743,364	93.19	7.02	7,749,007
DenseNet169	92.92	10.05	13,601,668	85.99	9.98	13,601,668	89.90	10.67	13,607,311
DenseNet201	87.20	11.05	19,341,188	93.01	10.99	19,341,188	87.14	12.38	24,907,151
VGG16	93.05	1.22	15,242,052	75.42	1.21	15,242,052	93.09	2.33	15,247,695
VGG19	93.09	1.35	20,551,748	75.00	1.29	20,551,748	93.57	2.35	20,557,391
ResNet50	92.79	2.7	24,848,388	61.19	2.1	24,848,388	92.71	2.9	24,854,031

TABLE 11. Comparison of the SVM classifier used in the proposed meta ensemble with other classifiers.

Classifiers	Hyperparameters	Corn	Maize	TPP
SVM	C=2, kernel='rbf'	98.27	94.27	97.45
KNN	Number of neighbors=5	97.36	94.15	96.22
RF	#estimators=100,max-depth=4	95.02	93.38	92.73
NB	Kernel='gaussian'	95.02	93.22	92.73

TABLE 12. Performance comparison of different methods on Cotton, Maize, and TPP datasets.

Dataset	Work	Method	Dataset Size	#Classes	% Accuracy
Cotton	**Proposed meta ensemble**	**Two-level ensemble, lightweight, MLP Mixer, LSTM, SVM**	**1518**	**4**	**98.45**
	Rai et al. [33]	Customised CNN	2293	5	97.98
	Azath et al. [34]	VGG19	2400	4	96.4
	Liang et al. [35]	S-DenseNet	150	2	92
	Dong et al. [36]	CBR with fuzzy logic technique	1600	4	89.4
Maize	**Proposed meta ensemble**	**Same as above**	**2529**	**4**	**94.27**
	Mishra et al. [37]	Customised CNN	4382	3	88.46
	Waheed et al. [38]	XceptionNet, NasNet	12,332	4	93.52, 91.9
	Arvind et al. [39]	Bag of Features and SVM	2000	4	83.7
	Yin et al. [40]	Dilated-inception module, embedded attention	1268	4	94.25
TPP	**Proposed meta ensemble**	**Same as above**	**20267**	**15**	**97.45**
	Abbas et al. [22]	C-GAN, DenseNet121	16012	10	97.11%
	Agarwal et al. [41]	CNN Network	17,500	10	91.20
	Elhassouny & Smarandache [42]	Customised MobileNet	7178	10	90.3
	Widiyanto et al. [43]	CNN model	1000	5	96.60
	Oppenheim & Shani [44]	Fine-tuned VGGNet	2465	5	95.85

Other related methods, using different sizes of a dataset, and different numbers of classes for similar plant disease categorisation tasks have been compared with the proposed meta ensemble as presented in Table 12.

As shown in Table 12, the proposed meta-ensemble consisting of two-level ensemble composed of MLP-Mixer, LSTM and SVM, has demonstrated superior categorisation accuracy across all the datasets used for the comparison purpose with its lightweight architecture. It has achiever, 98.45%, 94.26% and 97.45% categorisation accuracy with Cotton, Maize and 'TPO' datasets respectively. The proposed framework is able to achieve this performance with a smaller size dataset and fewer trainable parameters, highlighting its efficiency and suitability and efficiency for its deployment in resource constrained IoT environments.

In contrast to Rai et al. [33] who utilised customised CNN on the Cotton dataset, the proposed meta-ensemble surpasses their categorisation accuracy of 97.98% by achieving 98.45% accuracy even with the smaller size of dataset. Similarity, on Maize dataset, the proposed work has outperformed Mishra et al. [37], Waheed et al. [38], Arvind et al. [39] even with smaller size dataset. On the 'TPO' dataset, the proposed meta-ensemble achieves the highest accuracy surpassing Abbas et al. [22] utilizing conditional generative adversarial networks for data augmentation and DenseNet121 for classification purpose. It has also surpassed the performance of the other methods discussed in Table 12.

The best performance of the proposed method among the performance of all the other methods with limited size of dataset and model parameters, makes the proposed method useful for its deployment with the resource constrained IoT devices.

VI. CONCLUSION

This paper aims at building a lightweight framework for plant disease categorisation that can easily be deployed in a resource-constrained IoT-based environment. To meet this goal, a meta-ensemble approach has been proposed in this work, which is composed of lightweight state-of-the-art architectures such as MLP Mixer and LSTM. The modular and lightweight nature of proposed model ensures scalability of use by integrating it with the resource constrained IoT devices. The proposed model has been trained on three diverse datasets; therefore, it tends to learn common features across diverse types of plant diseases. In addition to that the use of features extracted from the multiple CNN models further ensures the generalisability of the proposed solution. The proposed meta ensemble approach has achieved the categorisation accuracy of 98.43%, 94.27% and 97.45% in the case of the Corn, Maize and TPP dataset respectively. Due to the lightweight nature of the proposed meta ensemble, the proposed meta ensemble also accelerates the prediction time. Thus, considering the overall benefits of the suggested method in terms of its accuracy, lesser number of trainable parameters and fast processing capability made the proposed meta ensemble the obvious choice for its deployment on Internet of Things-based platforms.

In future, the proposed method can be advanced so that it can be utilised in precision agriculture, by enhancing the capabilities of the model by training it using multimodal data, including soil information, real-time weather conditions that affect the plant health. This will help in improving the adaptability of the model and in automatically adjusting its predictions based on the real-time changes in environmental conditions, promoting, accurate and real-time response.

We see the necessity for a more thorough investigation of the policy implications to further enhance the conversation. Through expanding the conclusion to discuss prospective investment possibilities and strategic considerations for policymakers, we want to offer insightful information that closes the knowledge gap between cutting edge research and real-world applications. This improvement will serve the needs of investor and legislators, presenting our lightweight framework as a valuable resource in the field of agricultural technology.

DECLARATION OF COMPETING INTEREST

The authors affirm that they do not have any competing financial interests or personal relationships that could have influenced the reported work in this paper.

FUNDING

This research did not receive any particular funding from public, commercial, or not-for-profit organizations.

DATA AVAILABILITY

The authors have also stated that data will be accessible upon request.

REFERENCES

[1] S. Deepa and R. Umarani, "Steganalysis on images using SVM with selected hybrid features of Gini index feature selection algorithm," *Int. J. Adv. Res. Comput. Sci.*, vol. 8, no. 5, p. 1503, 2017.

[2] S. Zhang and Z. Wang, "Cucumber disease recognition based on global–local singular value decomposition," *Neurocomputing*, vol. 205, pp. 341–348, Sep. 2016.

[3] S. Zhang, X. Wu, Z. You, and L. Zhang, "Leaf image based cucumber disease recognition using sparse representation classification," *Comput. Electron. Agricult.*, vol. 134, pp. 135–141, Mar. 2017.

[4] A. Loddo, M. Loddo, and C. Di Ruberto, "A novel deep learning based approach for seed image classification and retrieval," *Comput. Electron. Agricult.*, vol. 187, Aug. 2021, Art. no. 106269, doi: 10.1016/j.compag.2021.106269.

[5] C. Qian, M. Tong, X. Yu, and S. Zhuang, "CNN-based visual processing approach for biological sample microinjection systems," *Neurocomputing*, vol. 459, pp. 70–80, Oct. 2021, doi: 10.1016/j.neucom.2021.06.085.

[6] R. Maurya, V. K. Pathak, and M. K. Dutta, "Deep learning based microscopic cell images classification framework using multi-level ensemble," *Comput. Methods Programs Biomed.*, vol. 211, Nov. 2021, Art. no. 106445, doi: 10.1016/j.cmpb.2021.106445.

[7] S. Huang, G. Zhou, M. He, A. Chen, W. Zhang, and Y. Hu, "Detection of peach disease image based on asymptotic non-local means and PCNN-IPELM," *IEEE Access*, vol. 8, pp. 136421–136433, 2020.

[8] S. Yadav, N. Sengar, A. Singh, A. Singh, and M. K. Dutta, "Identification of disease using deep learning and evaluation of bacteriosis in peach leaf," *Ecolog. Informat.*, vol. 61, Mar. 2021, Art. no. 101247.

[9] Y. Zhao, C. Sun, X. Xu, and J. Chen, "RIC-Net: A plant disease classification model based on the fusion of inception and residual structure and embedded attention mechanism," *Comput. Electron. Agricult.*, vol. 193, Feb. 2022, Art. no. 106644, doi: 10.1016/j.compag.2021.106644.

[10] R. Maurya, R. Burget, R. Shaurya, M. Kiac, and M. K. Dutta, "Multi-head attention-based transfer learning approach for porato disease detection," in *Proc. 15th Int. Congr. Ultra Modern Telecommun. Control Syst. Workshops (ICUMT)*, Oct. 2023, pp. 165–169, doi: 10.1109/ICUMT61075.2023.10333272.

[11] X. Chen, G. Zhou, A. Chen, J. Yi, W. Zhang, and Y. Hu, "Identification of tomato leaf diseases based on combination of ABCK-BWTR and B-ARNet," *Comput. Electron. Agricult.*, vol. 178, Nov. 2020, Art. no. 105730.

[12] I. Tolstikhin, N. Houlsby, A. Kolesnikov, L. Beyer, X. Zhai, T. Unterthiner, J. Yung, A. Steiner, D. Keysers, J. Uszkoreit, M. Lucic, and A. Dosovitskiy, "MLP-mixer: An all-MLP architecture for vision," 2021, *arXiv:2105.01601*.

[13] A. Pandey and K. Jain, "A robust deep attention dense convolutional neural network for plant leaf disease identification and classification from smart phone captured real world images," *Ecolog. Informat.*, vol. 70, Sep. 2022, Art. no. 101725, doi: 10.1016/j.ecoinf.2022.101725.

[14] P. Bedi and P. Gole, "Plant disease detection using hybrid model based on convolutional autoencoder and convolutional neural network," *Artif. Intell. Agricult.*, vol. 5, pp. 90–101, Jan. 2021.

[15] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Comput. Electron. Agricult.*, vol. 145, pp. 311–318, Feb. 2018, doi: 10.1016/j.compag.2018.01.009.

[16] K. Kc, Z. Yin, M. Wu, and Z. Wu, "Depthwise separable convolution architectures for plant disease classification," *Comput. Electron. Agricult.*, vol. 165, Oct. 2019, Art. no. 104948, doi: 10.1016/j.compag.2019.104948.

[17] M. Chohan, A. Khan, R. Chohan, S. H. Katpar, and M. S. Mahar, "Plant disease detection using deep learning," *Int. J. Recent Technol. Eng. (IJRTE)*, vol. 9, no. 1, pp. 909–914, May 2020, doi: 10.35940/ijrte.a2139.059120.

[18] S. M. Hassan and A. K. Maji, "Plant disease identification using a novel convolutional neural network," *IEEE Access*, vol. 10, pp. 5390–5401, 2022.

[19] Ü. Atila, M. Uçar, K. Akyol, and E. Uçar, "Plant leaf disease classification using EfficientNet deep learning model," *Ecolog. Informat.*, vol. 61, Mar. 2021, Art. no. 101182, doi: 10.1016/j.ecoinf.2020.101182.

[20] H. Amin, A. Darwish, A. E. Hassanien, and M. Soliman, "End-to-end deep learning model for corn leaf disease classification," *IEEE Access*, vol. 10, pp. 31103–31115, 2022, doi: 10.1109/ACCESS.2022.3159678.
[21] R. Maurya, N. N. Pandey, V. P. Singh, and T. Gopalakrishnan, "Plant disease classification using interpretable vision transformer network," in *Proc. Int. Conf. Recent Adv. Electr., Electron. Digit. Healthcare Technol. (REEDCON)*, May 2023, pp. 688–692, doi: 10.1109/REEDCON57544.2023.10151342.
[22] A. Abbas, S. Jain, M. Gour, and S. Vankudothu, "Tomato plant disease detection using transfer learning with C-GAN synthetic images," *Comput. Electron. Agricult.*, vol. 187, Aug. 2021, Art. no. 106279, doi: 10.1016/j.compag.2021.106279.
[23] P. S. Thakur, P. Khanna, T. Sheorey, and A. Ojha, "Explainable vision transformer enabled convolutional neural network for plant disease identification: PlantXViT," 2022, *arXiv:2207.07919*.
[24] R. Karthik, M. Hariharan, S. Anand, P. Mathikshara, A. Johnson, and R. Menaka, "Attention embedded residual CNN for disease detection in tomato leaves," *Appl. Soft Comput.*, vol. 86, Jan. 2020, Art. no. 105933, doi: 10.1016/j.asoc.2019.105933.
[25] D. Shah, V. Trivedi, V. Sheth, A. Shah, and U. Chauhan, "ResTS: Residual deep interpretable architecture for plant disease detection," *Inf. Process. Agricult.*, vol. 9, no. 2, pp. 212–223, Jun. 2022.
[26] B. B. Gupta and M. Quamara, "An overview of Internet of Things (IoT): Architectural aspects, challenges, and protocols," *Concurrency Comput., Pract. Exper.*, vol. 32, no. 21, Nov. 2020, Art. no. e4946, doi: 10.1002/cpe.4946.
[27] S. K. Noon, M. Amjad, M. A. Qureshi, and A. Mannan, "Computationally light deep learning framework to recognize cotton leaf diseases," *J. Intell. Fuzzy Syst.*, vol. 40, no. 6, pp. 12383–12398, Jun. 2021.
[28] D. Singh, N. Jain, P. Jain, P. Kayal, S. Kumawat, and N. Batra, "PlantDoc: A dataset for visual plant disease detection," in *Proc. 7th ACM IKDD CoDS 25th (COMAD)*, Jan. 2020, pp. 249–253.
[29] G. Geetharamani and J. A. Pandian, "Identification of plant leaf diseases using a nine-layer deep convolutional neural network," *Comput. Elect. Eng. J.*, vol. 76, pp. 323–338, Jun. 2019. [Online]. Available: https://doi.org/10.1016/j.compeleceng.2019.08.010
[30] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer, 2000, doi: 10.1007/978-1-4757-3264-1.
[31] A. A. Nurhanna and M. F. Othman, "Multi-class support vector machine application in the field of agriculture and poultry: A review," *Malaysian J. Math. Sci.*, vol. 11, pp. 35–52, Feb. 2017.
[32] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16×16 words: Transformers for image recognition at scale," 2021, *arXiv:2010.11929*.
[33] C. K. Rai, "Automatic categorisation of real-time diseased cotton leaves and plants using a deep convolutional neural network," Res. Square, Durham, NC, USA, 2022, doi: 10.21203/rs.3.rs-1440994/v1.
[34] A. M., M. Zekiwos, and A. Bruck, "Deep learning-based image processing for cotton leaf disease and pest diagnosis," *J. Electr. Comput. Eng.*, vol. 2021, pp. 1–10, Jun. 2021, doi: 10.1155/2021/9981437.
[35] X. Liang, "Few-shot cotton leaf spots disease classification based on metric learning," *Plant Methods*, vol. 17, no. 1, p. 114, Dec. 2021, doi: 10.1186/s13007-021-00813-7.
[36] Y. Dong, Z. Fu, S. Stankovski, Y. Peng, and X. Li, "A cotton disease diagnosis method using a combined algorithm of case-based reasoning and fuzzy logic," *Comput. J.*, vol. 64, no. 1, pp. 155–168, Nov. 2019, doi: 10.1093/comjnl/bxaa098.
[37] S. Mishra, R. Sachan, and D. Rajpal, "Deep convolutional neural network based detection system for real-time corn plant disease recognition," *Proc. Comput. Sci.*, vol. 167, pp. 2003–2010, Jan. 2020, doi: 10.1016/j.procs.2020.03.236.
[38] A. Waheed, M. Goyal, D. Gupta, A. Khanna, A. E. Hassanien, and H. M. Pandey, "An optimized dense convolutional neural network model for disease recognition and classification in corn leaf," *Comput. Electron. Agricult.*, vol. 175, Aug. 2020, Art. no. 105456, doi: 10.1016/j.compag.2020.105456.
[39] K. R. Aravind, P. Raja, K. V. Mukesh, R. Aniirudh, R. Ashiwin, and C. Szczepanski, "Disease classification in maize crop using bag of features and multiclass support vector machine," in *Proc. 2nd Int. Conf. Inventive Syst. Control (ICISC)*, Jan. 2018, pp. 1191–1196, doi: 10.1109/icisc.2018.8398993.
[40] C. Yin, T. Zeng, H. Zhang, W. Fu, L. Wang, and S. Yao, "Maize small leaf spot classification based on improved deep convolutional neural networks with a multi-scale attention mechanism," *Agronomy*, vol. 12, no. 4, p. 906, Apr. 2022, doi: 10.3390/agronomy12040906.
[41] M. Agarwal, A. Singh, S. Arjaria, A. Sinha, and S. Gupta, "ToLeD: Tomato leaf disease detection using convolution neural network," *Proc. Comput. Sci.*, vol. 167, pp. 293–301, Jan. 2020, doi: 10.1016/j.procs.2020.03.225.
[42] A. Elhassouny and F. Smarandache, "Smart mobile application to recognize tomato leaf diseases using convolutional neural networks," in *Proc. Int. Conf. Comput. Sci. Renew. Energies (ICCSRE)*, Jul. 2019, pp. 1–4, doi: 10.1109/ICCSRE.2019.8807737.
[43] S. Widiyanto, R. Fitrianto, and D. T. Wardani, "Implementation of convolutional neural network method for classification of diseases in tomato leaves," in *Proc. 4th Int. Conf. Informat. Comput. (ICIC)*, Oct. 2019, pp. 1–5, doi: 10.1109/icic47613.2019.8985909.
[44] D. Oppenheim and G. Shani, "Potato disease classification using convolution neural networks," *Adv. Animal Biosci.*, vol. 8, no. 2, pp. 244–249, 2017, doi: 10.1017/s2040470017001376.

RITESH MAURYA received the B.Tech. degree in computer science and engineering, the M.Tech. degree in computer science and engineering from the ABV-Indian Institute of Information Technology and Management, and the Ph.D. degree in computer science and engineering from the Centre for Advanced Studies, Dr. A.P.J. Abdul Kalam Technical University, with a focus on machine learning and deep learning applications in biology and medicine. He is currently an Associate Professor with the Amity Centre for Artificial Intelligence, Amity University, Noida, India. He has published research in esteemed journals, including IEEE, Wiley, Springer, and Elsevier. His research interests include machine learning, and deep learning and its applications in diverse domains.

SATYAJIT MAHAPATRA received the B.Tech. degree in electronics and telecommunication engineering from the Biju Patnaik University of Technology, the M.Tech. degree in electronics and communication engineering from Siksha 'O' Anusandhan University, and the Ph.D. degree in electronics and communication engineering from the Birla Institute of Technology, Mesra, with a focus on machine learning and signal processing for genomic data analysis. He is currently an Assistant Professor with the Department of Information and Communication Technology, Manipal Institute of Technology, MAHE. He has published research in esteemed journals, including IEEE, Oxford University Press, and Wiley. His research interests include applied machine learning, image processing, and genomic signal processing.

LUCKY RAJPUT received the B.Sc. degree in physics and mathematics, in 2018, and the master's degree in physics, in 2021. She is currently pursuing the M.Tech. degree in data science with Amity University, Noida. Her research interests include machine learning and deep learning.