Indonesian Journal of Electrical Engineering and Computer Science
Vol. 38, No. 3, June 2025, pp. 1489~1502
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v38.i3.pp1489-1502

Segmentation and classification of plant leaf disease using advanced deep learning approach and ensemble classifier

Suma S. Huddar[1], Jayashri Rudagi[2], Jagadish S. Jakati[3]
[1]Department of Electronics and Communication Engineering, S. G. Balekundri Institute of Technology, Affiliated to Visvesvaraya Technological University, Belagavi, India
[2]Department of Electronics and Communication Engineering, Veerappa Nisty Engineering College Hasanapur, Affiliated to Visvesvaraya Technological University, Belagavi, India
[3]School of Computer Science Engineering and Applications, D. Y. Patil International University Akurdi, Pune, Maharashtra, India

Article Info

Article history:

Received Mar 25, 2024
Revised Nov 6, 2024
Accepted Nov 24, 2024

Keywords:

Classification
CNN mask R-CNN
Ensemble
Plant disease
SVM

ABSTRACT

An essential component of maintaining global food production is plants. On other hand, a number of plant diseases can threaten agricultural output and cause large losses if left unchecked. Agricultural specialists and botanists physically track plant diseases in a labor-intensive, error-prone manner using a conventional method. AI can give evaluations that are quicker and more accurate than those made using conventional approaches by automating the identification and analysis of diseases. This technical development presents a viable way to lessen crop losses and lessen the severity of infections. As a result, we describe an ensemble machine learning strategy for plant disease classification in this study that is enabled by deep learning. Data augmentation is done in the first part of the study, and in the second step, we provide a modified mask region-based convolutional neural network (R-CNN) model for plant leaf segmentation. Afterwards, a model to extract the deep features based on CNN is shown. Lastly, the ensemble classifier is built using support vector machine classifier (SVM), random forest (RF), and decision tree (DT) with the aid of majority voting. The suggested method's effectiveness is tested on plant village, apple, maize, and rice, yielding overall accuracy values of 99.45%, 96.30%, 96.85%, and 98.25%, in that order.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Suma S. Huddar
Department of Electronics and Communication Engineering, S. G. Balekundri Institute of Technology
Affiliated to Visvesvaraya Technological University
S. G. Balekundri Institute of Technology, Belagavi-10, Karnataka, India
Email: sumavehaan18@gmail.com

1. INTRODUCTION

Plants have vital functions in preserving human welfare and are vital components of global biodiversity. Because of population expansion, there is a dramatic increase in demand for plant-based foods. Carefully understanding the life cycle and developmental mechanisms of plants is necessary. As a result, agriculture is extensively supported and accepted as a commercial sector in many nations. The demand for food is not being satisfied despite this rise, Global hunger has been on the rise since 2015, according to data from the food and agriculture organization (FAO) of the United Nations [1]. According to current estimates, 680 million people worldwide less than 9% of the world's population—are starving to death. This is an increase of over 120 million over ten years and 10 million within a year. Furthermore, agriculture supports more than 85% of the global population [2]. Eighty percent of the food produced worldwide is generated by farmers [3]. Regretfully, pests and plant diseases cause the loss of almost half of crop output every year [4]. Therefore, it is critical to identify and detect plant diseases as soon as possible with accuracy plant diseases have a serious influence on human existence. But identifying plant diseases in large areas is a difficult task that calls both skilled workers and optical leaf exams [4]. Farmers identify illnesses by using their knowledge and unaided observation of signs on plant leaves. This is a labor-intensive, time-consuming procedure that requires specific skills [5]. To enable non-pathologists and non-botanists detect and diagnose plant diseases, the goal of our effort is to design an automated approach [6]. Researchers have proposed a number of automated strategies where computer vision-based systems are frequently used to address these problems. Simple image pre-processing tasks including edge detection, image binarization, and image thresholding are performed using traditional techniques [7]. Still, it's a challenge to achieve the amazing precision of illness diagnosis. Furthermore, there is no cure for illness detection using these techniques. Researchers have suggested incorporating a ML based approach where activities such as picture pre-processing, image features extraction, and classification model training are carried out in order to overcome these challenges. Figure 1 illustrates a simple architecture for plant disease detection and classification using SVM classification.

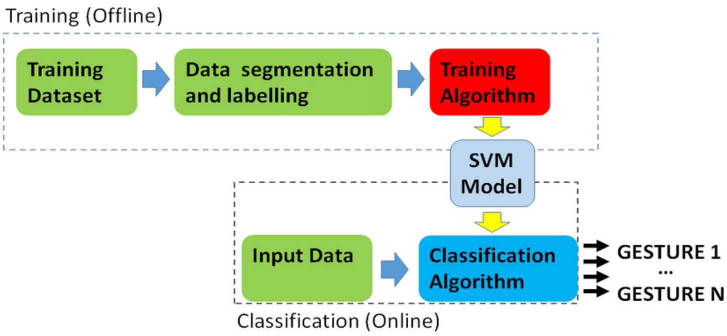

Figure 1. General architecture of plant disease classification

On the basis of this idea, several methods—including Shrivastava et al. introduced a trained support vector rice plant diseases may be identified using a color feature extraction-based technique and a machine-based classification model [8]. In their ML-based method, Harakannanavar *et al.* [9] offered K-means clustering and histogram equalization as picture pre-processing techniques to enhance image quality. Principal component analysis, the discrete wavelet transform (DCT), and Grey Level Co-occurrence Matrix descriptors are all used in the feature extraction process. To train various classifiers, these acquired characteristics are employed. Color histograms, Hu moments, Haralick patterns, and local binary patterns are used in the multiple feature extraction and fusion method developed by Basavaiah and Anthony [10] as feature extraction models. Then, the decision tree and random forest (RF) classifiers are trained using the acquired feature set. These approaches have made significant progress, but they still have problems with overfitting, class imbalance, model generalization, and training complexity. Deep learning-based techniques have been created for image segmentation and classification to address these problems. Upadhyay and Kumar [11] developed a DL-based model that considers the picture's size, shape, and color as important criteria. Additionally, it employs Otsu's global thresholding technique to produce binary pictures and remove background noise. An entirely connected convolutional neural network (CNN) is employed to finish the categorizing process. Having remarkable analytical and classification capabilities, the CNN is the most used classifier for image recognition [12]. Leaf vein patterns were the focus of early experiments with deep learning methods for plant picture identification [13]. The three leguminous plant species that the researchers successfully classified using a 3-to 6-layer CNN are white beans, red beans, and soybeans. By training a model to detect 14 crop species and 26 agricultural illnesses, Mohanty *et al.* [14] significantly expanded deep learning approaches. The model achieved an outstanding accuracy of 99.35% on the test dataset. With a deep CNN, Ma *et al.* [15] were able to detect 93.4% of the four cucumber diseases—downy mildew, anthracnose, powdery mildew, and target leaf spots—based alone on their symptoms. Additionally, the 94.9% accuracy in recognizing cucumber leaf disease was reported by Kawasaki *et al.* [16] using a CNN-based method. Principally, identification was done by leaf vein patterns [13]. Plant leaf segmentation has shown to be difficult for standard approaches, which causes uncertainty in data regarding the background and foreground of leaves. In this method, we use a DL-based approach to segment plant leaves in images. Afterwards, the features are extracted using CNN architecture based on deep learning. Ultimately, SVM, RF, and decision tree classifiers are combined to create an ensemble classifier, and majority voting techniques are used to determine the final classification result. The remaining portions of the article are arranged as follows: A brief review of the literature regarding current approaches to classifying plant diseases is presented in section 2. A detailed discussion of the proposed model is presented in section 3. The proposed approach and a comparison with current methods are presented in section 4. Finally, concluding remarks regarding the work are presented in section 5.

2. LITERATURE SURVEY

This section covers current techniques for the recognition, forecasting, and classification of plant diseases. As was previously said, deep learning and machine learning techniques are extensively used in this industry to fully automate the process.

Guo *et al.* [17] employed a mathematical model based on deep learning for the diagnosis of plant diseases. Region proposal network (RPN) is used in the primary stage of this to locate the plant leaves in the surrounding complex. The leaves are then divided into segments and fed into a transfer learning model that has been trained using a small background dataset of sick leaves. The overall accuracy recorded by this approach is 83.57%, indicating the importance of the deep learning model. With the aid of the TensorFlow object detection framework, Saleem *et al.* [18] focused on the meta-architectures of deep learning approaches, such as the single shot multibox detector (SSD), faster region-based convolutional neural network (R-CNN), and region-based fully convolutional networks (RFCN). After being trained using the Adam optimizer, the SSD model has the greatest mean average precision of 73.07%.

Comparably, Roy and Bhaduri [19] presented a DL-based model for multi-class classification. This piece is predicated on the YoloV4 framework. The entire image is split up into many grids, after which bounding boxes are calculated, their confidence scores are calculated, and an estimate of the class probability map is made. This model makes use of Dense-CSPDarkNet53 to carry out these duties. The output of this module is then fed into the modified PA Net architecture, and the Head module generates detection results in the end. Chug *et al.* [20] designed k-nearest neighbors (KNN), AdaBoost, RF, logistic regression (LR), and stochastic gradient boosting are the five machine learning (ML) models used as classifiers in this hybrid deep learning model. Eight different pre-trained ML models, including Efficient Net B0 to B7 for feature extraction, are also used.

Numerous segmentation techniques have been created because plant leaf image segmentation is crucial to enhancing classification performance. The UNet model has demonstrated notable results in picture segmentation tasks, leading to its widespread adoption in a multitude of applications. Bhagat *et al.* [21] created a UNet-based model for plant leaf segmentation using the UNet technique. Additionally, the architecture extracts features using EfficientNet-B4 as an encoder model. Due to the fact that information degradation is a serious challenge for these models, the authors have updated the residual blocks and skip link in the decoder module. Reduces the computational complexity of the model. With the aid of the output layer, the low- and high-level features produced by the decoder model are integrated to enhance segmentation performance.

In order to separate leaf images from complicated backgrounds, Yang *et al.* [22] employed a deep learning-based approach. They also integrated a deep learning model for leaf categorization. This model makes use of the Mask R-CNN model, which is used for both feature map creation and bounding box regression classification. ResNet-FPN is used in the segmentation map creation phase to create the feature map. The generated feature map is then processed via the classification step, which includes ROI alignment, and bounding box regression and classification are used in the final stage.

The complex background has a significant influence on segmentation and classification, as stated by Yang *et al.* [23]. These problems include, but are not limited to, things like overlapping objects, noise interference, and uneven illumination. The authors offered Mask R-CNN and dual channel convolution neural network as solutions to these issues. It also uses a gentle non-maximum suppression technique to enhance detection performance for scenarios with overlapping items. Similar to this, the following step incorporates pooling techniques to help reduce loss during feature map and original picture alignment. The mask filer layer is ultimately used to conceal the intricate backgrounds. This model employs adaptive particle swarm optimization to enhance overall performance, replacing the SoftMax with the SVM.

3. PROPOSED METHOD

The suggested method for segmenting plant leaves using an ensemble model based on ML and DL is presented in this section. Using this method, loading the whole database with all of its labels comes first. The next stage is data augmentation, which involves procedures like picture blurring, rotation, and cropping. The deep learning-based segmentation module is used in the next step to separate the leaf pictures from complicated backgrounds. After the picture has been segmented, it is employed in the feature extraction step, where several feature extraction models are applied to produce a strong feature map. Eventually, to understand the intricate patterns and categorize them more accurately, an ensemble classifier is created. The subsections that follow provide a thorough explanation of each step.

3.1. Pre-processing and augmentation of data

The image data is resized to a 256×256 row and column matrix during pre-processing, and it is then converted to grayscale. Methods for data augmentation are also used. Image data augmentation methods involve transforming the source photos in different ways to expand the size and variety of the dataset. These improvements increase the machine learning models' durability and generalization. The following have been utilized in this work:

- Vertical and horizontal flipping: users may flip the picture vertically or horizontally. By doing so, the model may become more invariant to directions such as top-bottom or left-right.
- Rotation: rotate the image by a particular angle, often spontaneously picked between a specified range. This helps the model become less susceptible to fluctuations in object orientation.
- Scaling: to help the model learn to detect things at different sizes, resize the image by zooming in or out.
- Translation: move the image in a horizontal or vertical direction to assist the model become invariant to little changes in the location of objects.
- Brightness and contrast adjustment: modify the image's contrast and brightness. This makes the model more resilient to variations in illumination.
- Noise injection: add random noise to the image. This can help the model become more adaptable to noise in real-world images.

3.2. Deep learning for plant leaf segmentation

The suggested classification of deep learning model for plant leaf classification is shown in this section. According to approaches now in use, noise and lighting fluctuations are just two of the problems that plague leaf segmentation algorithms. As a result, we concentrated on creating a solid segmentation strategy. The segmentation model that has been suggested employs an instance segmentation technique called Mask R-CNN. Target identification and object segmentation at the pixel level are capabilities of this approach. Mask R-CNN 's design includes the region of interest alignment algorithm (ROIAlign) and the feature pyramid network (FPN) in addition to maintaining the fundamental architecture of faster R-CNN. The main architecture of Mask RCNN is shown in Figure 2. Its six main components are: input, backbone feature extraction network, FPN, RPN, ROIAlign, and the output for bounding box, class, and mask predictions (box, class, mask).

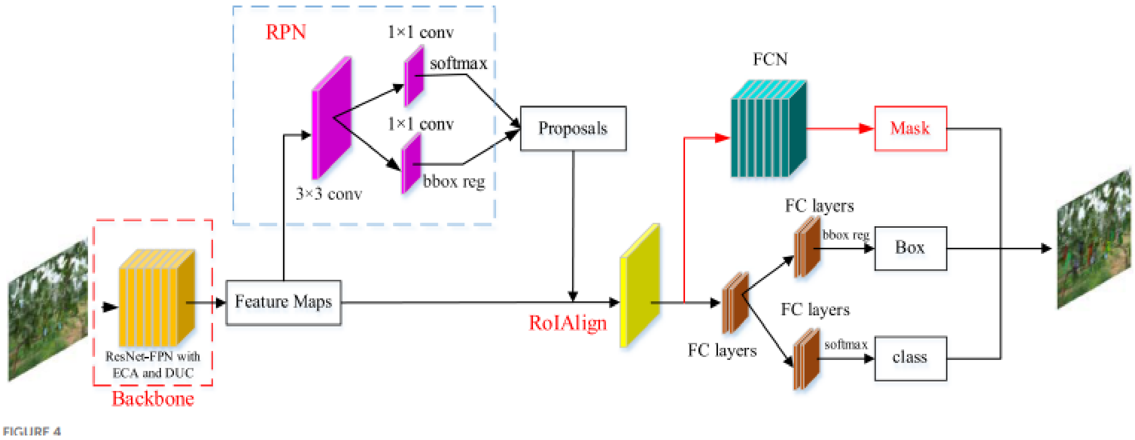

Figure 2. Overall architecture of Mask R-CNN

Partitioning is crucial in this context of identifying plant diseases. This Mask R-CNN process feeds the input image with plant leaves into the ResNet50 + FPN network model in order to extract pertinent features and produce feature maps that are unique to the leaves. Via the RPN, these feature maps help identify possible regions of interest (ROIs), which in this case correspond to individual plant leaves. To differentiate between leaf and non-leaf areas within the ROIs, a SoftMax classifier is used after the ROIs have been identified. Techniques for frame regression are used to improve the accuracy of leaf boundaries. For a more efficient selection of pertinent leaf regions, non-maximum suppression is also used to prune redundant ROIs. The feature maps then, in addition to along with the refined ROIs undergo processing in the RoIAlign layer. This layer makes it easier to create uniform feature maps for every leaf region, allowing for reliable segmentation even with a variety of leaf forms and sizes. At this point, the flow splits into two directions: one leads to a fully connected layer for border refinement and leaf classification, while the other feeds into a complete convolutional network (FCN) with pixel-level segmentation as its primary function. With its two branches, this method guarantees thorough leaf segmentation for both high-level classification and pixel-by-pixel delineation.

Typically, this design's core module makes use of the 101-layer ResNet101 architecture. Nonetheless, we have examined plant leaf photos in our study, where the use of ResNet101 can significantly lower the operating speed. As a result, ResNet50 was employed in this project. Furthermore, because diverse plant leaf kinds cause the considered data to vary greatly in size, a basic convolution network is unable to extract the precise features from these photos. In order to enhance the feature extraction procedure, we have taken into consideration the feature pyramid network (FPN). By using a hierarchical structure with horizontal connections, the FPN makes it easier to create a network feature pyramid from an input that is only available at one scale. This method minimizes the number of parameters required while addressing the difficult task of extracting target objects from photos at various sizes. The Feature Map used in SSD's pyramid structure served as the model for FPN (single shot multibox detector). Nevertheless, in contrast to SSD, FPN also combines shallow Feature Maps in addition to deep feature maps from networks like VGG. By utilizing a blend of top-down, lateral, and bottom-up connections, FPN maximizes the integration of various feature maps. By ensuring effective information flow across many levels of abstraction, this all-encompassing strategy improves detection accuracy without appreciably lengthening processing times. FPN successfully addresses the multi-scale object detection problem by utilizing both deep and shallow Feature Maps in a coherent way, which makes it an excellent tool for a variety of computer vision applications.

However, Mask R-CNN models face a number of difficulties because to the intricate structure of plant leaf pictures. To address this problem, we concentrate on implementing an attention mechanism that allows the model to focus on and enhance important feature information while removing unimportant characteristics. By narrowing its focus, the model becomes more resilient to changes in field circumstances, which in turn improves its performance and adaptability in a variety of scenarios. In this study, we have implemented an upsampling convolution channel attention method in the Mask R-CNN backbone model. Figure 3 shows the modified design of the backbone model.

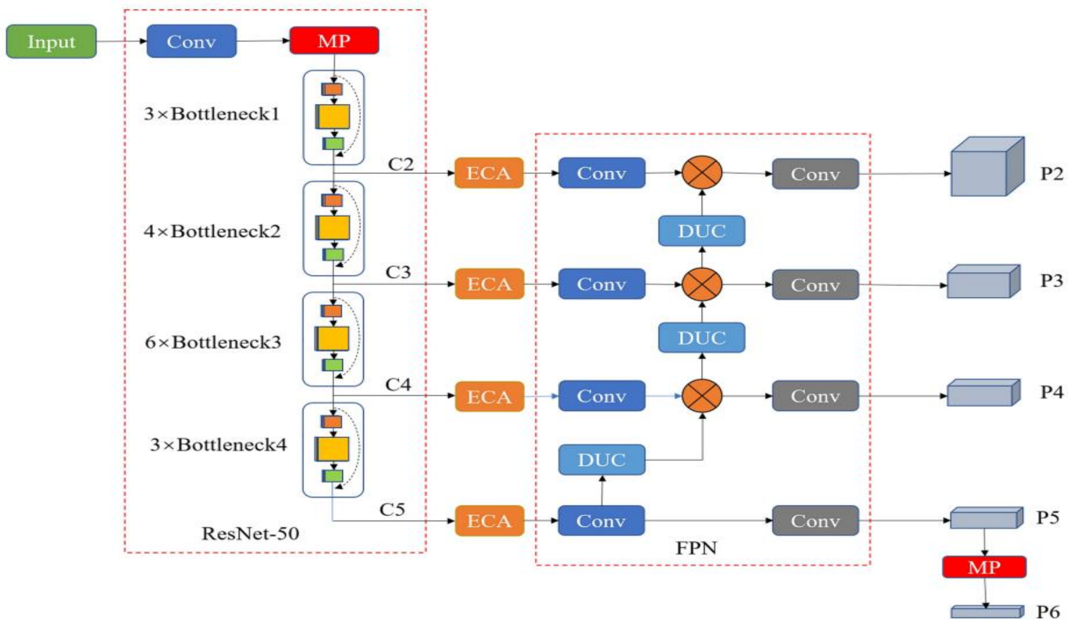

Figure 3. Improved backbone architecture

Figure 4 shows the architecture of the channel attention model. The fundamental concept is to propose a local cross-channel interaction strategy without dimensionality reduction. By comparing each channel to its KNN after the global average pooling (GAP) of channels, this method seeks to locally record interaction information among channels. To cover a global receptive field, the ECA module first uses GAP to compute the input feature map with dimensions H*W*C, where C is the number of feature channels. This produces a feature vector of size 1*1*C. Next, data on cross-channel interactions is obtained using 1D convolution with a convolution kernel size of k. The number of input channels affects the convolution kernel k's size, which may be chosen as follows:

$$\psi(C) = \left|\frac{log_2(C)}{\gamma} + \frac{b}{\gamma}\right|_{odd} \quad (1)$$

where γ and b are set to 2 and 1, respectively, and C is the number of input channels. Obtaining the coverage of local cross-channel exchanges is helpful. The architecture of the attention mechanism is shown in Figure 4.

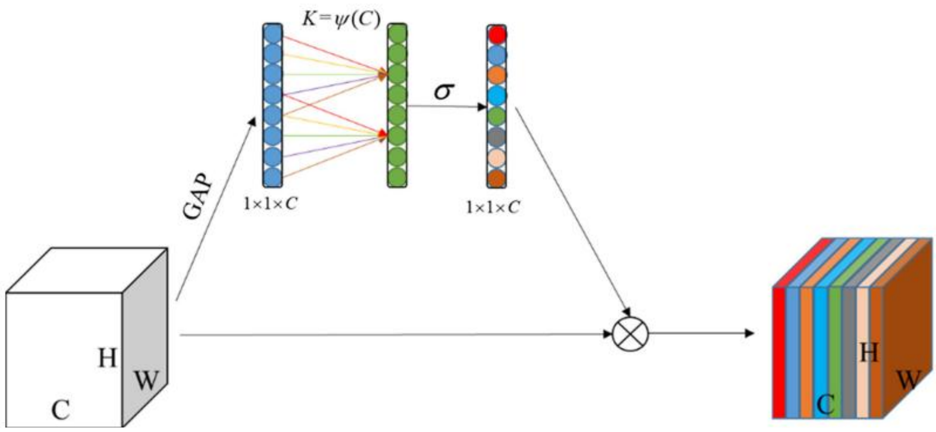

Figure 4. Channel attention mechanism

The weights of each feature are subsequently determined using the sigmoid activation function. The output feature channel weight vector is multiplied by the original input feature map in the next step to finish the original feature calibration along the channel dimension. By suppressing any faulty or inefficient feature channels, this technique improves the directed character of the retrieved features. As a result, this refinement increases the extraction of useful and efficient characteristics.

During the Mask R-CNN procedure, the FPN fuses the feature map from the bottom up. However, the resolution of the feature map changes with each stage, and the depth parameters of the final feature map also change. Thus, in order to get feature maps of the same size, upsampling procedures must be carried out. Then, pixel by pixel, the acquired feature set is combined to create the feature fusion needed to build the pyramid structure. The traditional Mask R-CNN upsamples by employing nearest neighbor interpolation; nevertheless, this may lead to the loss of accurate data for important aspects. We provide a unique upsampling strategy to overcome this issue, feeding the resulting feature maps (H/r*W/r*C) as 3x3 convolutions for learning. The process of convolution yields a feature map of size H/r*W/r*(C*r^2), which is then reshaped as H×W*C, where r is the ratio of the upsample feature to the original feature map size. The upsampling procedure is shown in Figure 5.

The segmentation and classification model is trained using the final feature map. While the segmentation model creates the segmented output, the classification model classifies the plant leaf category. Loss function modeling is needed during the training phase to enhance learning and improve segmentation results. The loss function is expressed as:

$$L = L_{cls} + L_{box} + L_{mask} \quad (2)$$

where L_{cls} represents the classification loss, L_{box} represents the regression loss of bounding box and L_{mask} is the mask loss.

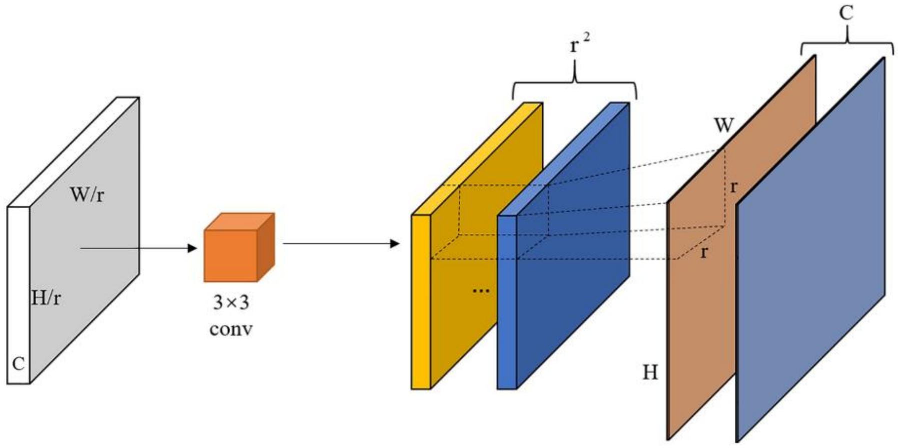

Figure 5. Upsampling model for feature extraction

3.3. Feature extraction

The approached feature extraction model is explained in this part. The suggested feature extraction model receives the segmented leaf image as input. We employ CNN-based deep learning architecture for feature extraction in place of the conventional manual feature extraction method. Convolutional neural networks are commonly utilized in deep learning models for the extraction of features from images CNN. CNNs are frequently utilized for jobs involving images because of their capacity to automatically extract pertinent characteristics from unprocessed pixel data. The convolution layer, activation function, and pooling layer process the picture data in the CNN feature extraction process. Let's write X for the input picture and H^l for the l^{th} convolutional layer's output feature map. The l^{th} convolutional layer has the following mathematical formulation:

$$H_{i,j,k}^{(l)} = f^{(l)}\left(\sum_{m=1}^{M^{(l-1)}} \sum_{p=1}^{P^{(l)}} \sum_{q=1}^{Q^{(l)}} W_{p,q,m,k}^{l} . X_{(i+p-1),(j+q-1),m}^{(l-1)} + b_k^{(l)}\right) \tag{3}$$

where $H_{i,j,k}^{(l)}$denotes the activation at position (I,j) in the k^{th} feature map of the l^{th} layer, $W_{p,q,m,k}^{l}$ represents the weight parameter of the k^{th} filter for the m^{th} input channel in the l^{th} layer, $X_{(i+p-1),(j+q-1),m}^{(l-1)}$ denotes the activation at position (i+p-1), (j+q-1) in the m^{th} input channel of the $(l\text{-}1)^{th}$ layer, $b_k^{(l)}$ represents the bias term for the k^{th} filter in the l^{th} layer, and $f^{(l)}$ enotes the activation function applied element-wise. In next step we apply ReLU activation function which is expressed as:

$$f^{(l)}(x) = \max(0, x) \tag{4}$$

further, pooling operations performed which reduces the spatial dimensions of the feature maps. It can be expressed as:

$$H_{i,j,k}^{(l)} = \max_{p=1}^{P^{(l)}} \max_{q=1}^{Q^{(l)}} \left(H_{(i-1)\times S^{(l)}+p,(j-1)\times S^{(l)}+q,k}^{(l-1)}\right) \tag{5}$$

where $S^{(l)}$ denotes the stride of pooling operation. After performing these operations, the features are flattened and fed into fully connected layer which can be expressed as:

$$H^{(l)} = f^{(l)}\left(W^{(l)} . H^{(l-1)} + b^{(l)}\right) \tag{6}$$

where $W^{(l)}$ represents the weight matrix of the l^{th} fully connected layer, $b^{(l)}$is the bias vector. The obtained vector is considered as final feature vector.

3.4. Ensemble classification

To categorize the illness kind, the final feature map is given into the ensemble classifier model. To increase overall prediction accuracy, an ensemble classifier combines many separate classifiers. Assume for the moment that the ensemble classifier is represented by the notation E(x), where x stands for the input data. Consider the following scenario: We have N basic classifiers, C_1, C_2, ..., C_N. For a given input, each base classifier returns its prediction, which is expressed as C_i (x), where i is the base classifier's index. The final prediction of the ensemble classifier is formed by combining the predictions of the basic classifiers using a variety of techniques. Majority voting is a popular technique in which the base classifiers' majority vote determines the final prediction. In terms of, Mathematically, the ensemble classifier $E(x)$ can be represented as:

$$E(x) = argmax_y \sum_{i=1}^{N} \delta(C_i(x), y) \tag{7}$$

Where y stands for the potential classes and δ(.) for the Kronecker delta function, which is 1 in the case of equal arguments and 0 in the absence of one, is represented. We employed the RF, decision tree, and SVM classifiers in this study. E(x), the combined classifier, may be written as follows:

$$E(x) = argmax_y \left(\sum_{i=1}^{N} \delta\left(C_{SVM_i}(x), y\right) + \sum_{i=1}^{N} \delta\left(C_{DT_j}(x), y\right) + \sum_{i=1}^{N} \delta\left(C_{RF_k}(x), y\right)\right) \tag{8}$$

where the final prediction, E(x), is the obtained class that gained the greatest number of votes from the SVM, DT, and RF classifiers, and N is the number of base classifiers. Figure 6 illustrates the general ensemble categorization procedure.

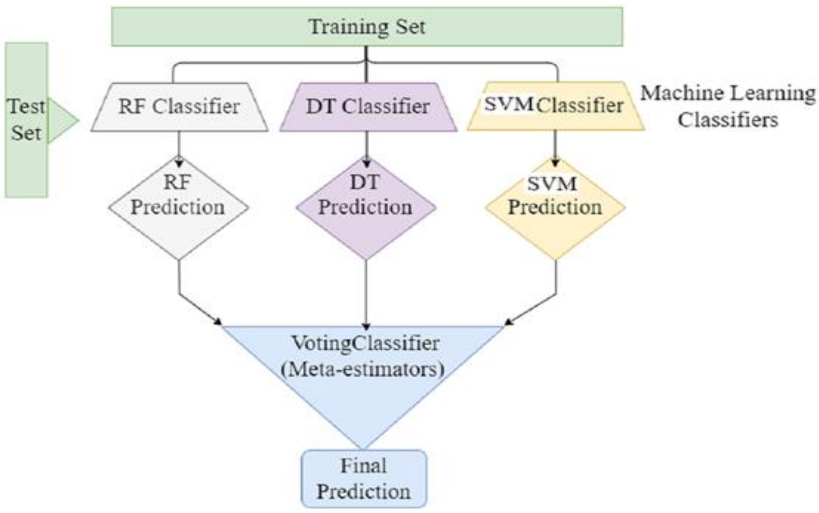

Figure 6. Ensemble classifier

4. RESULTS AND DISCUSSION

This chapter presents the suggested model's results and compares them with the most advanced models for classifying plant leaves and diseases. The first paragraph provides a brief overview of the dataset utilized in this study, the second subsection explains the parameters used to gauge the classification performance, and the third and final chapter provides a comparison analysis to demonstrate the stability of the suggested model.

4.1. Dataset details

Images related to different plant species and the diseases they are linked to are compiled into the PlantVillage dataset as shown in Figure 7. The number of photos in this collection is 54,305, which represents 14 different species and 38 distinct disease classifications. A sampling of the PlantVillage dataset's photos is shown in Figures 7(a)-(h), representing the apple, cherry, maize, potato, tomato, grape, orange, and peach, respectively. Table 1 shows the information pertaining to images in each category.

Apple dataset-there are 3,642 photos in all, split up into four categories in the collection. Out of these, only 1,822 have labels, with the rest photos being label-free. The annotated photos have been carefully selected to address several categories, including different illnesses, rust, scab, and healthy. Table 2 shows the distribution of photos by category. Interestingly, the photos were taken in the field in uncontrollable lighting circumstances.

Maize dataset-there are four different kinds of maize illnesses included in the 400 photos that make up the Maize collection. Table 3 provides the overall count of photos for each category, including the photographs in the test set. Notably, no background removal was used when taking these photos in the actual field. A subset of the photographs in each category is reserved for testing, and 100 images within each category are designated for training.

Rice dataset: the Xiamen, China-based Fujian Institute of Subtropical Botany provided this dataset, which includes rice illnesses grouped into five categories. There are 100 photos in each category, for a total of 500 photographs. In addition, different numbers of examples are included with test photos. Table 4 provides particular information on the number of photos in each category.

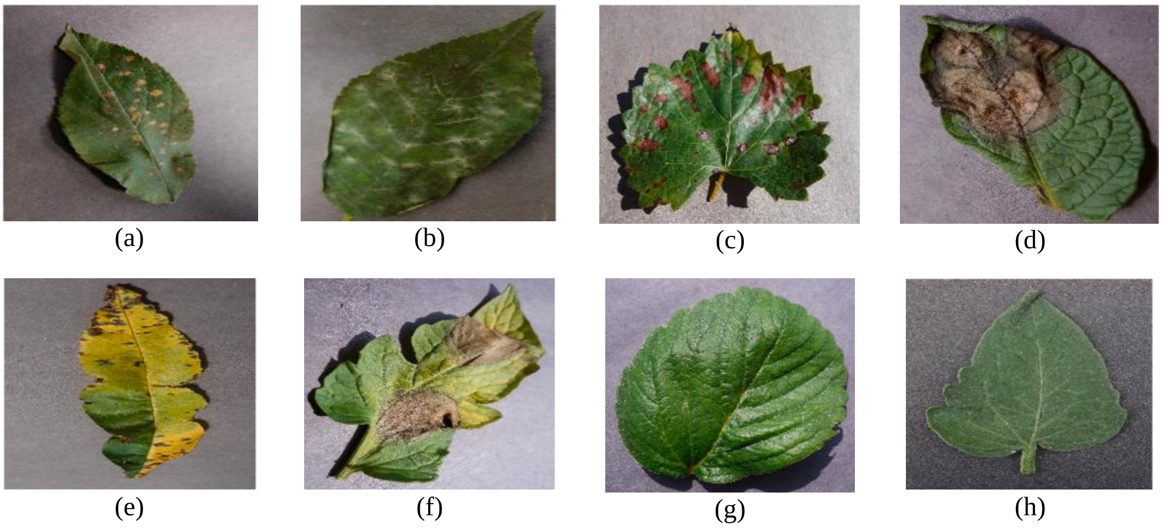

Figure 7. Sample images of PlantVillage dataset (a) apple, (b) cherry, (c) maize, (d) potato, (e) tomato, (f) grape, (g) orange, and (h) peach

Table 1. Dataset details: PlantVilage

Species	Category	Image count	Species	Category	Image count
Apple	Scab	630	Grape	Black rot	1180
Apple	Black rot	621	Grape	Black measles	1383
Apple	Cedar apple rust	275	Grape	Isariopsis leaf spot	1076
Apple	Athletic	1645	Grape	Athletic	423
Cherry	Athletic	854	Orange	Citrus greening	5507
Cherry	Powdery Mildew	1052	Peach	Athletic	360
Corn	Gray Leaf Spot	513	Peach	Bacterial spot	2297
Corn	Common Rust	1192	Pepper	Bacterial spot	997
Corn	Athletic	1162	Pepper	Athletic	1478
Corn	Northern leaf blight	985	Pepper	Early blight	100
Potato	Athletic	152	Raspberry	Athletic	371
Potato	Early Blight	1000	Strawberry	Athletic	456
Potato	potato blight	1000	Strawberry	Leaf scorch	1109
Tomato	Bacterial spot	2127	Tomato	Yellow leaf curl virus	5357
Tomato	Early blight	1000	Tomato	Mosaic virus	373
Tomato	Athletic	1591	Tomato	Target spot	1404
Tomato	Tomato blight	1909	Tomato	Two spot spider mite	1676
Tomato	Leaf mold	952	Tomato	Blueberry	1502

Table 2. Apple dataset description

Category	Image count
Athletic	518
Rust	535
Scab	590
Multiple disease	93
Total	1821

Table 3. Maize dataset description

Category	Image count
Ppecks	108
Gross's Bacetrial Wilt	115
Pyricularia grisea	115
Phaeosphaeria leaf Spot	118
Total	480

Table 4. Rice dataset description

Category	Image count
Bacterial leaf streak	110
Sun scorch	113
White smut	115
Stackburn	105
white gradient	117
Total	560

4.2. Performance measurement parameters

Four metrics—accuracy, sensitivity, specificity, and F-measure—are used to assess the classification performance using a confusion matrix. The confusion matrix is displayed in Table 5, and further parameters are calculated using the information in the equations (number).

Table 5. Confusion matrix representation

	Positive	Negative	Total
Positive	T_p	F_p	$T_p + F_p$
Negative	F_N	T_N	$F_N + T_N$
Total	$T_P + F_N$	$T_P + T_N$	

The values acquired as stated in the confusion matrix, such as true positive, false positive, false negative, and true positive, are used to determine the detection accuracy. The precision may be calculated as (9):

$$Accuracy = \frac{T_P+T_N}{T_P+T_N+F_P+F_N} \tag{9}$$

The confusion matrix is also used to calculate the sensitivity performance. One way to express the sensitivity is (10):

$$Sensitivity = \frac{T_P}{T_P+F_N} \tag{10}$$

the specificity can be computed as (11):

$$Specificity = \frac{T_N}{T_N+F_P} \tag{11}$$

and, F-measure is can be computed as (12).

$$F - measure = \frac{2\times T_P}{2\times T_P+F_N+F_P} \tag{12}$$

4.3. Comparative analysis

Using models for data augmentation and pre-processing is the initial step in this process. Figure 8 illustrates the result of this step. This stage indicates that the "Pepper Bell" class is afflicted with the "Bacterial Spot" sickness:

On this data, we also use a CNN-based feature extraction technique, in which activation functions, pooling layers, and convolution layers carry out specific tasks to produce a feature set. Figure 9 shows the characteristics that correlate to the activation functions.

The suggested ensemble classification technique is used in the next step to categorize the leaf type and the associated illness. Table 6 shows the aggregate classification performance for the PlantVillage, apple, maize, and rice datasets in terms of accuracy and loss.

EfficientNet B0, while more efficient than traditional architectures like ResNet or Inception, may still have relatively high computational requirements compared to lighter structures like MobileNet or ShuffleNet. It also tends to have higher memory requirements compared to other light weight architectures. MobileNetv2 has limited capacity for capturing complex features and its performance is sensitive to hyper parameters and training arrangements. The comparative analysis reveals that the proposed method has reported highest accuracy as 99.45%, 96.30%, 96.85%, and 98.25% for PlantVillage, apple, maize, and rice database.

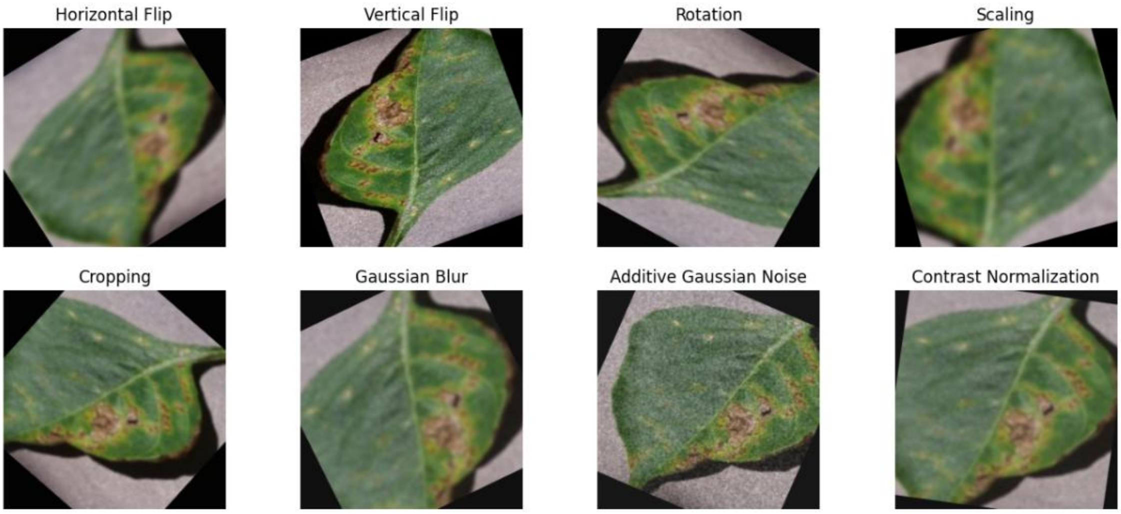

Figure 8. Data augmentation

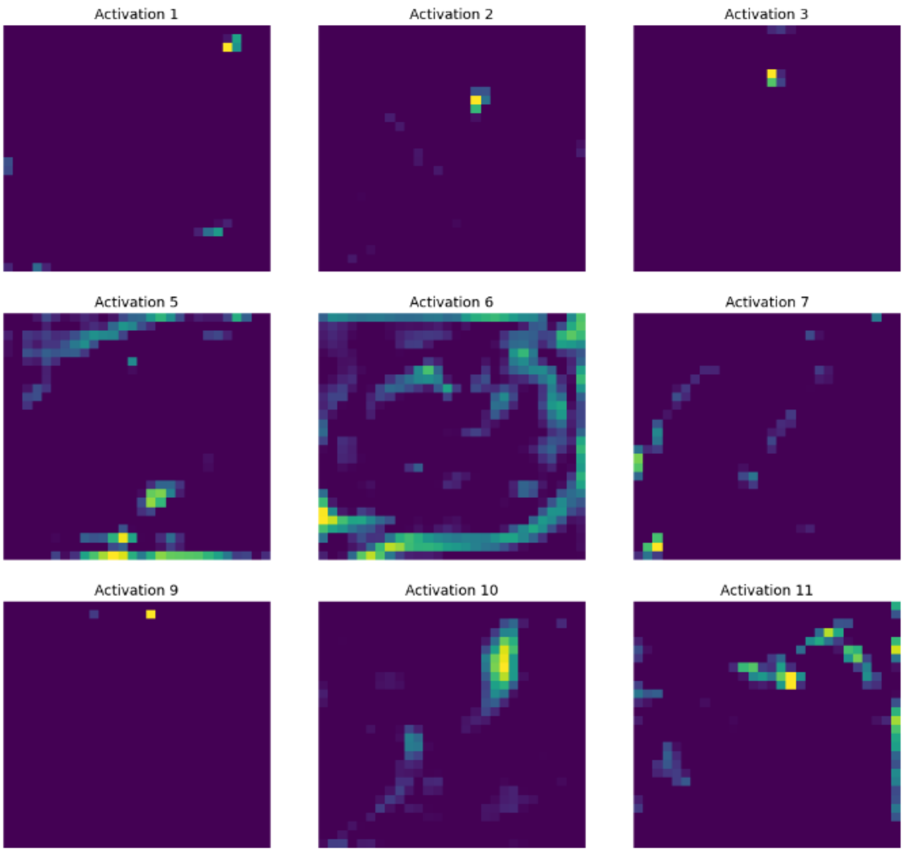

Figure 9. Outcome of CNN based feature extraction

Table 6. Comparative analysis with existing methods for different dataset

Classification method	Evaluation parameter	Dataset			
		PlantVillage	Apple	Maize	Rice
EfficientNet B0 [24]	Loss	0.2121	2.10	0.399	0.420
	Accuracy	95.69	64.51	85.18	86.67
MobileNet v2 [24]	Loss	0.040	0.652	0.422	0.817
	Accuracy	99.25	81.72	93.82	95.00
ShuffleNet v2 [24]	Loss	0.096	2.62	1.20	1.17
	Accuracy	97.96	41.93	77.78	75.00
GhostNet [24]	Loss	0.169	3.93	2.67	2.32
	Accuracy	96.18	39.78	39.50	43.33
Residual CNN with attention [25]	Loss	0.506	1.01	2.45	0.8945
	Accuracy	92.79	69.31	62.96	80.00
Teacher student multitask CNN [26]	Loss	0.1252	0.548	0.961	0.5798
	Accuracy	98.10	89.78	87.13	93.33
Multi-crop CNN [27]	Loss	0.9301	1.165	1.54	0.1057
	Accuracy	91.25	87.10	86.42	95.00
Deep transfer learning [28]	Loss	0.105	0.301	1.557	0.2933
	Accuracy	96.82	75.89	80.38	92.00
VGG-ICNN [24]	Loss	0.0497	0.0078	0.825	0.0542
	Accuracy	99.16	94.24	91.36	96.67
Proposed Model	Loss	0.085	0.0056	0.251	0.0351
	Accuracy	99.45	96.30	96.85	98.25

5. CONCLUSION

In this proposed algorithm, we have concentrated on developing an advanced deep learning-based method for classifying and detecting plant diseases. In order to solve the problem of data imbalance and take into account the variety of the dataset, the initial part of this study involves data augmentation techniques. The following phase presents a feature extraction model based on Mask R-CNN, whereby the backbone architecture is altered through the addition of channel attention and upsampling procedures. To get the final feature set, the segmented picture is processed through a CNN-based feature extraction step. Ultimately, the majority voting strategy is employed to build the ensemble classifier using SVM, DT, and RF classifiers. When compared with the current method, the experimental result analysis demonstrates that the suggested strategy improved classification accuracy and decreased training loss.

ACKNOWLEDGMENTS

Our heartfelt thanks go to D Y Patil International University Pune and S G Balekundri Institute of Technology Belagavi. We also acknowledge the Prof (Dr). Prabhat Ranjan, vice chancellor DYPIU Pune and Dr. B R Patagundi, SGBIT Belagavi for making this research possible.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

Name of author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Suma S. Huddar	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	
Jayashri Rudagi		✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Jagadish S. Jakati	✓		✓	✓	✓	✓			✓	✓	✓	✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The authors confirm that the data supporting the findings of this study are available within the article.

REFERENCES

[1] "The state of food security and nutrition in the World 2023," in FAO; IFAD; UNICEF; WFP; WHO; eBooks, 2023, doi: 10.4060/cc3017en.

[2] Deshpande, T, "State of agriculture in India," Prsindia.org. [Online]. Available: https://prsindia.org/files/policy/policy_analytical_reports/State%20of%20Agriculture%20in%20India.pdf

[3] I. Z. Mukti and D. Biswas, "Transfer learning based plant diseases detection using ResNet50," in *2019 4th International Conference on Electrical Information and Communication Technology (EICT)*, Dec. 2019, doi: 10.1109/eict48899.2019.9068805.

[4] J. Arunnehru, B. S. Vidhyasagar, and H. A. Basha, "Plant leaf diseases recognition using convolutional neural network and transfer learning," *In Lecture Notes in Electrical Engineering*, 2020, pp. 221–229, doi: 10.1007/978-981-15-2612-1_21.

[5] S. Verma, A. Chug, and A. P. Singh, "Prediction models for identification and diagnosis of tomato plant diseases," 2018 *International Conference on Advances in Computing, Communications and Informatics (ICACCI)* Sep. 2018, doi: 10.1109/icacci.2018.8554842.

[6] V. Singh, N. Sharma, and S. Singh, "A review of imaging techniques for plant disease detection," *Artificial Intelligence in Agriculture*, vol. 4, pp. 229–242, Jan. 2020, doi: 10.1016/j.aiia.2020.10.002.

[7] G. Liu, "An image combination segmentation method based on clustering analysis and edge detection," *In Proceedings of the 2021 4th International Conference on Digital Medicine and Image Processing*, Nov. 2021, doi: 10.1145/3506651.3506975.

[8] V. K. Shrivastava and M. K. Pradhan, "Rice plant disease classification using color features: a machine learning paradigm," *Journal of Plant Pathology*, vol. 103, no. 1, pp. 17–26, Oct. 2020, doi: 10.1007/s42161-020-00683-3.

[9] S. S. Harakannanavar, J. M. Rudagi, V. I. Puranikmath, A. Siddiqua, and R. Pramodhini, "Plant leaf disease detection using computer vision and machine learning algorithms," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 305–310, Jun. 2022, doi: 10.1016/j.gltp.2022.03.016.

[10] J. Basavaiah and A. A. Anthony, "Tomato leaf disease classification using multiple feature extraction techniques," *Wireless Personal Communications*, vol. 115, no. 1, pp. 633–651, Jun. 2020, doi: 10.1007/s11277-020-07590-x.

[11] 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, Oct. 2021, doi: 10.1007/s41870-021-00817-5.

[12] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Computers and Electronics in Agriculture*, vol. 147, pp. 70–90, Apr. 2018, doi: 10.1016/j.compag.2018.02.016.

[13] G. L. Grinblat, L. C. Uzal, M. G. Larese, and P. M. Granitto, "Deep learning for plant identification using vein morphological patterns," *Computers and Electronics in Agriculture*, vol. 127, pp. 418–424, Sep. 2016, doi: 10.1016/j.compag.2016.07.003.

[14] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers in Plant Science*, vol. 7, Sep. 2016, doi: 10.3389/fpls.2016.01419.

[15] J. Ma, K. Du, F. Zheng, L. Zhang, Z. Gong, and Z. Sun, "A recognition method for cucumber diseases using leaf symptom images based on deep convolutional neural network," *Computers and Electronics in Agriculture*, vol. 154, pp. 18–24, Nov. 2018, doi: 10.1016/j.compag.2018.08.048.

[16] Y. Kawasaki, H. Uga, S. Kagiwada, and H. Iyatomi, "Basic study of automated diagnosis of viral plant diseases using convolutional neural networks," *in Lecture notes in computer science*, 2015, pp. 638–645, doi: 10.1007/978-3-319-27863- 6_59.

[17] Y. Guo *et al.*, "Plant disease identification based on deep learning algorithm in smart farming," *Discrete Dynamics in Nature and Society*, vol. 2020, pp. 1–11, Aug. 2020, doi: 10.1155/2020/2479172.

[18] M. H. Saleem, S. Khanchi, J. Potgieter, and K. M. Arif, "Image-based plant disease identification by deep learning meta-architectures," *Plants*, vol. 9, no. 11, p. 1451, Oct. 2020, doi: 10.3390/plants9111451.

[19] A. M. Roy and J. Bhaduri, "A deep learning enabled multi-class plant disease detection model based on computer vision," *AI*, vol. 2, no. 3, pp. 413–428, Aug. 2021, doi: 10.3390/ai2030026.

[20] A. Chug, A. Bhatia, A. P. Singh, and D. Singh, "A novel framework for image-based plant disease detection using hybrid deep learning approach," *Soft Computing*, vol. 27, no. 18, pp. 13613–13638, Jun. 2022, doi: 10.1007/s00500-022-07177-7.

[21] S. Bhagat, M. Kokare, V. Haswani, P. Hambarde, and R. Kamble, "Eff-UNet++: A novel architecture for plant leaf segmentation and counting," *Ecological Informatics*, vol. 68, p. 101583, May 2022, doi: 10.1016/j.ecoinf.2022.101583.

[22] K. Yang, W. Zhong, and F. Li, "Leaf segmentation and classification with a complicated background using deep learning," *Agronomy*, vol. 10, no. 11, p. 1721, Nov. 2020, doi: 10.3390/agronomy10111721.

[23] X. Yang *et al.*, "Instance segmentation and classification method for plant leaf images based on ISC-MRCNN and APS-DCCNN," *IEEE Access*, vol. 8, pp. 151555–151573, Jan. 2020, doi: 10.1109/access.2020.3017560.

[24] P. S. Thakur, T. Sheorey, and A. Ojha, "VGG-ICNN: a lightweight CNN model for crop disease identification," *Multimedia Tools and Applications*, vol. 82, no. 1, pp. 497–520, Jun. 2022, doi: 10.1007/s11042-022-13144-z.

[25] R. Karthik, M. Hariharan, S. Anand, P. Mathikshara, A. Johnson, and R. Menaka, "Attention embedded residual CNN for disease detection in tomato leaves," *Applied Soft Computing*, vol. 86, p. 105933, Jan. 2020, doi: 10.1016/j.asoc.2019.105933.

[26] M. Brahimi, S. Mahmoudi, K. Boukhalfa, and A. Moussaoui, "Deep interpretable architecture for plant diseases classification," In *2019 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, Sep. 2019, doi: 10.23919/spa.2019.8936759.

[27] A. Picon, M. Seitz, A. Alvarez-Gila, P. Mohnke, A. Ortiz-Barredo, and J. Echazarra, "Crop conditional convolutional neural networks for massive multi-crop plant disease classification over cell phone acquired images taken on real field conditions," *Computers and Electronics in Agriculture*, vol. 167, p. 105093, Dec. 2019, doi: 10.1016/j.compag.2019.105093.

[28] J. Chen, J. Chen, D. Zhang, Y. Sun, and Y. A. Nanehkaran, "Using deep transfer learning for image-based plant disease identification," *Computers and Electronics in Agriculture*, vol. 173, p. 105393, Jun. 2020, doi: 10.1016/j.compag.2020.105393.

BIOGRAPHIES OF AUTHORS

Suma S. Huddar received the B.Eng. degree in electronics and communication engineering from Gogte Institute of Technology which is affiliated to Visvesvaraya Technological University, Belagavi, India, in 2011 and the M.Tech. degree in digital electronics and communication systems, in 2013 from Adichunchanagiri Institite of Technology Chickmagaluru, which is Affiliated to Visvesvaraya Technological University, Belagavi, India. Currently, she is a research scholar at the department of electronics and communication engineering, Visvesvaraya Technological University. Her research interests include data science, deep learning algorithms, machine learning algorithms, time series forecasting, computer vision, and natural language processing. She can be contacted at email: sumavehaan18@gmail.com.

Dr. Jayashri Rudagi received her Ph.D. degree from Shivaji University, Maharashtra in 2019. She has more than 32 years of teaching experience. She worked as a professor in the department of ECE at S. G. Balekundri Institute of Technology, Belagavi. She is guiding 3 scholars toward their Ph.D. She is currently associated with S. G. Balekundri Institute of Technology, Belagavi, India as a research supervisor. Her research interests include image processing, artificial intelligence, deep learning, machine learning, and computer vision. She is an executive member of NKSS IEEE, IEI, and ISTE. She is also a member of the VTU Board of Studies. She can be contacted at email: jmrudagi@gmail.com.

Dr. Jagadish S. Jakati he received B.E. in electronics and communication engineering in 2010, M. Tech in 2012 and Ph.D. in 2022 from VTU, Belagavi presently he is working as assistant professor at school of computer science engineering and application, D. Y. Patil International University Akurdi, Pune, Maharashtra 411044, India. He Holds a Ph.D. degree in Electrical engineering with specialization in Speech processing. His research areas are image/signal processing, biometrics, medical image analysis and pattern recognition. Dr Jagadish has filed a number of patents and industrial designs on his innovative ideas and has been published with four Indian patents and more than 20 international journals. His research interests include image/signal processing, biometrics, medical image and analysis, and pattern recognition. He can be contacted at email: jagadishjs30@gmail.com or jagadish.jakati@dypiu.ac.in.