Deep CNN model for crops' diseases detection using leaf images

Yashwant Kurmi[1] · Prankur Saxena[2] · Bhupendra Singh Kirar[3] · Suchi Gangwar[4] · Vijayshri Chaurasia[5] · Aditya Goel[5]

Abstract

The agricultural yield of any country provides the base for the development of that nation. Sustainable growth needs to maintain crop production up to a certain level that depends on the research of their disease detection and treatment. The general approaches available in the literature follow attributes extraction and training a classifier model for leaf image classification that limits accuracy. The proffered technique eliminates the redundant information from the image dataset. We initially localize the region of interest in terms of the color attributes of leaf image based on the mixture model for region growing. The feature extraction is performed through a proposed deep convolutional neural network model followed by the classification of the leaf images. The deep learning model uses color images to learn the attributes that show different patterns that can be distinguished with the help of a convolutional neural network model. The execution measure of the proposed model is investigated using the PlantVillage dataset. The simulating replica outcomes show that the performance of the proposed model is far better as compared to the existing well-known methods of the domain with mean classifying accuracy and area under the characteristics curve of 95.35% and 94.7%, individually.

Keywords Leaf image preprocessing · Identification and classification · Computer-aided diagnosis · CNN model

✉ Yashwant Kurmi
  yashwantkurmi18@gmail.com

1   ECE Department, Madhyanchal Professional University, Bhopal, India

2   Indian Institute of Technology Gandhinagar, Ahmedabad, India

3   Indian Institute of Information Technology Bhopal, Bhopal, India

4   Agriculture Department, RKDF University, Bhopal, India

5   ECE Department, Maulana Azad National Institute of Technology, Bhopal, India

🙋 Springer

1 Introduction

The agricultural landmass is sufficient enough to feed today's world population. Mostly, the wealth of developing nations is based on agricultural production. Initial stage results of the leave image based-identification of disease in the field of agriculture play an important role to maintain their economy (Beucher and Meyer 1993).

Szegedy et al. (2013) analyzed the deep CNN for multi-object detection as localization and classification using a simplified model by developing the object mask. Tseng et al. (2014) invented a method for plant disease recognition using tone-based features. Ioffe and Szegedy (2015) introduced the batch normalization process that provides freedom to choose a higher learning rate with little caring initialization and eliminates the need for dropouts as well as requires less training time. Mahlein (2016) studied crop management termed as precision agriculture while plant phenotyping has been defined as the noninvasive analysis of the plant properties as physiological, biochemical, or anatomical for example data collection through optical, multispectral, and thermal sensors. Kasun et al. (2016) worked on the redundancy reduction of the data as extreme learning machine auto-encoder (ELM-AE) and sparse ELM-AE (SELM-AE) to make the system fast.

Mohanty et al. (2016) utilized the capabilities of smartphones along with computer perception through deep learning for the diagnosis of diseases. Wang et al. (2017) utilized the transfer learning for the automatic apple's leaf image for disease severity detection. Brahimi et al. (2017) introduced an automatic feature extraction through the Convolutional Neural Network (AFE-CNN) and trained it with the 14,828 images of tomato leaves that classified nine diseased and healthy imaginariums. He et al. (2017) presented a mask region-based CNN method (M-R-CNN) by adding a branch for predicting an object mask that makes the faster R-CNN more time saver.

The artificial intelligence-based methodologies have been utilized for the extraction of of visual features, followed by clustering and classification (Alexander et al. 2018). The complete automation-based features combined with handcrafted visual attributes (CDHVA) has been employed to learn the three fine-tuned layers of pre-trained deep convolutional neural networks (DCNNs) with handcrafted descriptors jointly in Zhang et al. (2018), Barbedo (2018). The color information and vector quantization learning along with CNN (CICQL-CNN) system was presented by Sardogan et al. for the detection and categorization of tomato leaf diseases (Sardogan et al. 2018).

The transfer learning using stacked sparse autoencoder (SSAE) subnetworks were employed to extract deep spatial and spectral features with one sequential SSAE subnetwork performs the smooth fusion of these deep attributes (Deng et al. 2019; Kurmi and Chaurasia 2020) to reduce the dependency of the system on a large labeled sample dataset. Apart from this, the transfer learning (TL) also helps to analyse the individual lesions and spots in images and identify multiple diseases on a single leaf and classification (Arnal Barbedo 2019). Bauer et al. (2019) presented an AirSurf a hybrid system as a combination of computer vision, machine learning, and software engineering (Kurmi and Chaurasia 2020) for measuring the yield-related phenotyping from the aerial imagery.

A large group of researchers (Aversano et al. 2020; Zeng et al. 2020) utilized the transfer learning on various standard deep learning models (AlexNet, Inception v3, DenseNet-169, SqueezeNet-1.1, ResNet-34, and VGG13) with dataset augmentation through the Generative Adversarial Networks (GAN) (Li et al. 2021; Liu and Wang 2020) for classifying citrus leaf images. An Alexnet based multi-scale CNN has been presented in Lv et al. (2020), with

batch normalization to prevent overfitting and increasing the systems robustness through the Adabound optimizer and parametric rectified linear unit function

Karthik et al. (2020) presented a cascaded connection of deep CNN models for significant feature extraction and context-relevant attribute attention-seeking. The first model gives equal importance for each feature while the second model decides the weight based on the relevant context.

Nagasubramanian et al. (2021) proposed a system that observes the crops' growth and leaf diseases continuously for advising farmers in need. the proposed framework uses machine learning techniques such as support vector machine and convolutional neural networks to provide analytical statistics on plant growth and disease patterns. It employed ensemble Nonlinear Support Vector Machine (ENSVM) for Ensemble Classification and Pattern Recognition for Crop Monitoring System (ECPRC) to identify plant diseases at the early stages.

The existing works still have limited performance for various multi-class categorization problems of diseased data along with limited accuracy. This work presents a novel leaf color information-based localization method for the region of interest segmentation and CNN-based categorization. The proffered technique leads at a higher level than the conventional approaches of entity localization in terms of accuracy and G-mean measure. To localize the disease confining region we employed the seed-based county growing approach (Callara et al. 2020). The mixture model-based region expansion is employed to refine the leaf area. Application of CNN model on the localized images provides the segregated attribute extraction. The compact delineation of these attributes has been found using the dropout method to bestows exceptional solutions for classification leaf images.

The contents of the article include materials and methods in Section 2, which covers the dataset and proffered localization approach. Part 3 provides an explanation of the developed convolutional neural network-based attribute extraction and categorization. In Section 4, we have explored the result obtained from various models and compared with the proposed model. Finally, Section 5 provides the final conclusion with future working research directions.

2 Materials and method

2.1 Dataset

We have taken the PlantVillage dataset that consists of a variety of 14 crops and their diseases, which comprises 54,309 labeled images. out of which we have taken three crops: bell pepper, potato, and tomato as shown in Fig. 1. The bell pepper dataset has 1478 healthy and 997 bacterial spot diseased images. The potato crop has three categories with 1000 early blight and 1000 late blight with 152 healthy images. Ten categories of tomato crop have been divided as target spot with 1404 images, 373 mosaic virus leaves, 3209 of yellow leaf curl virus, 2127 leaves affected by bacterial spot. There are 1000 early blight cases, 1591 healthy image cases, 1909 late blight, 952 leaf mold, 1771 septoria leaf spot, and 1676 spider mites affected tomato leaf images.
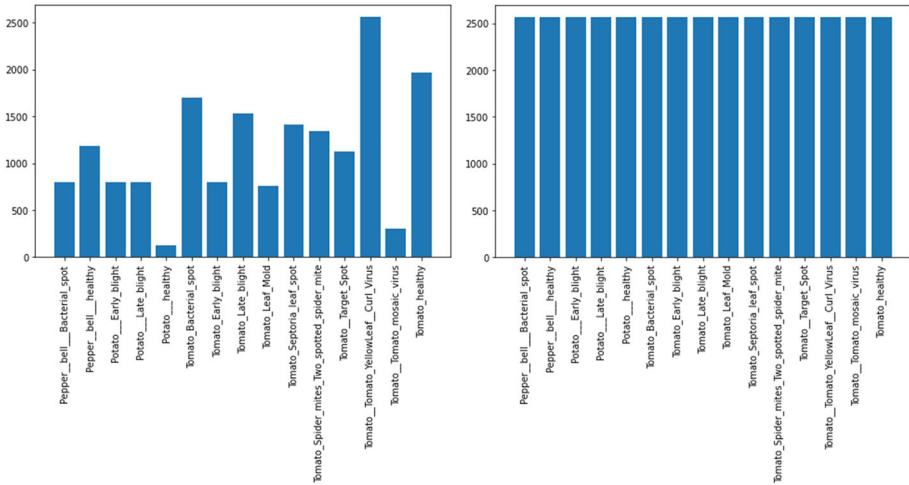
Fig. 1 The PlantVillage dataset description Left image shows the original number of images while the right image illustrates the imaginarium after augmentation for dataset balancing

2.2 Proposed method

The proffered system for plant leaf classification is depicted in Fig. 2, which comprises a couple of steps: 1) image localization using segmentation and 2) feature finding for classification.

The proposed preprocessing method expressed in three steps. First step environs the techniques for foreground segmentation i.e. leaf area extraction from the backdrop. Further step describes initial seed extrication purporting to the region growing approach. The workflow of the proposed method is given in Fig. 2.

2.2.1 Preprocessing

The image consists of some artifacts (noise) inside images of the leaves. The environmental noise that affects the sensors during image acquisition is called speckle noise. So, we should apply some sort of pre-processing mechanism to encounter this problem. This is done by analyzing every leaf's pixel through the histogram plot of intensity for the voxels of leaf area and employ auto-thresholding (Torr and Murray Sep. 1997). Voxel whose intensity is below the local threshold taken as background voxel.

2.2.2 Initial seed points selection

The leaf image color shows that the green ingredient is with greater values and the blue color with less value for the leaf region as compared to the background. The green to the blue color ratio (G/B) depicts seed region marking efficiently.

A global threshold has been utilized to separate the foreground and background (Al-Kofahi et al. April 2010) up to a certain level and it is not always acceptable, which needs a local-level analysis (Ridler and Calvard Aug 1978). The combined effect of local and global thresholding provides efficient leaf region extraction with empirically selected window size for local thresholding is at least 9x9.
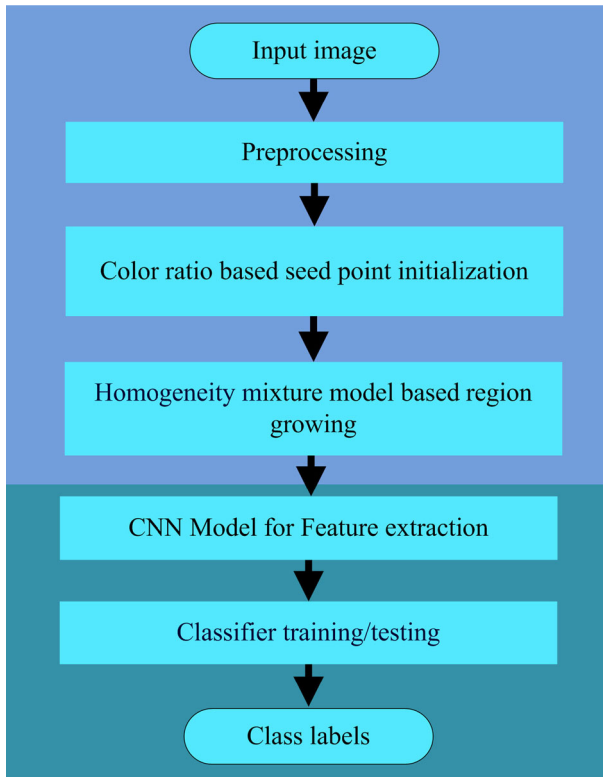
Fig. 2 The process flow of developed localization and classification technique

2.2.3 Region growing framework

The neighboring pixels of the initialized leaf region is chosen from seeds and are interfused for the region growing. Non-linear numeration of morphological attributes or image shapes is termed morphology operations. A decaying shape-based repetitive erosion (Haralick et al. 1989) was employed for shrinking the leaf to initialize seed. The region growing approach extracts the homogeneous county in the neighboring perimeter, which may not be an exact region of interest. To mitigate this, we have applied the mixture model-based region growing method (Callara et al. 2020). Global Gaussian distribution has been amalgamated with the prior information obtained from the region growing approach.

Consider an image Y with an associated K-class classification pattern X where a point l with intensity y_l is classified as belonging to class $x_l \in \{1, \dots K\}$. The Kth class model is a mathematical description of the conditional probability $P(y_l | x_l = k)$ (Calapez and Rosa Sep. 2010). The kth class of the model is described by the linear mixture model:

$$\psi_k(y) = \alpha_k \Psi_B(y - K_0) + (1 - \alpha_k) \Psi_{Sk}(y - K_0) \qquad (1)$$

where y represents the intensity level of the pixel, α_k denotes the mixture parameter, and K_0 denotes system offset. Ψ_B represents the distribution of the pixels of background generally possesses the normal distribution properties, with parameters mean K_0 and variance v_B and Ψ_{Sk} signifies the distribution of kth class pixels intensity, given by distributions as negative-

binomial having variance measure v_{Sk} and parameter mean μ_{Sk}. As per the Ref. (Calapez and Rosa Sep. 2010) the negative binomial distribution can be represented as given below:

$$p_K = \frac{\mu_{Sk}}{v_{Sk}} \tag{2}$$

and

$$r_K = \frac{\mu_{Sk}^2}{v_{Sk} - \mu_{Sk}} \tag{3}$$

The presence of single class k of pixels at the local level is assumed for the region growing utilities. To complete this task a broad model for a pixel y_l is defined by the 5-distribution parameters

$$
\begin{aligned}
\Psi_K(y_l; K_0, v_B, \alpha, r, p) = {} & \alpha \frac{1}{Z(v_B)} exp(-\frac{(y_l - K_0)^2}{2v_B}) \\
& + (1 - \alpha) \frac{\Gamma(y_l - K_0 + r)}{(y_l - K_0)!\Gamma(r)} p^r (1 - p)^{y_l - K_0}
\end{aligned}
\tag{4}
$$

where $\alpha \in [0, 1]$. $Z(v_B)$ is a normalizing parameter depending on variance v_B. The fitting of the model is performed using an expectation-maximization (EM) technique in which:

1. The parameter variables p and r are computed through the method of moments (Eqs. 2, 3)
2. Mean K_0 and variance v_B are calculated by maximizing the loglikelihood

$$L(\Theta|Y, X) = \sum_{y=\min(Y)}^{\max(Y)} \ln(\Psi) \tag{5}$$

3. α is given by the posterior density for N_s samples of pixel values $\{\alpha_y \, \forall \, y = 1, 2, \dots N_s\}$

$$\alpha = \frac{\sum_{y=\min(Y)}^{\max(Y)} \alpha_y}{N_s} \tag{6}$$

Homogeneity-based region growing is utilized by establishing local threshold levels for the confocal dataset. Basically, the background statistics with signal distributions are utilized as a linear mixture model (MM) to determine the likelihood with which a given pixel (voxel) can be considered as part of the foreground or not, as described below. The rule to grow regions is then designed from these probabilities. The initial seed is considered the central point of the region of interest. Further, the homogeneity property for the localized area is derived from an image volume centered on the seed. A generalized threshold using Otsu (Siddique et al. 2018) thresholding method for segmentation has been employed, which is an optimum solution for a multimodal distribution (Ng 2006). On the other hand, background with normal distributions comprising the negative binomial is fitted through an expectation-maximization (EM) method (Callara et al. 2020).

All original images are an RGB image within the range of 0 to 255, the maximum value is 255 but these values would be too large for the proposed model to process as showin in Fig. 3. Therefore, values are targeted in between 0 and 1. Each pixel value is rescaled from the [0,255] to [0, 1] range by 1./255.
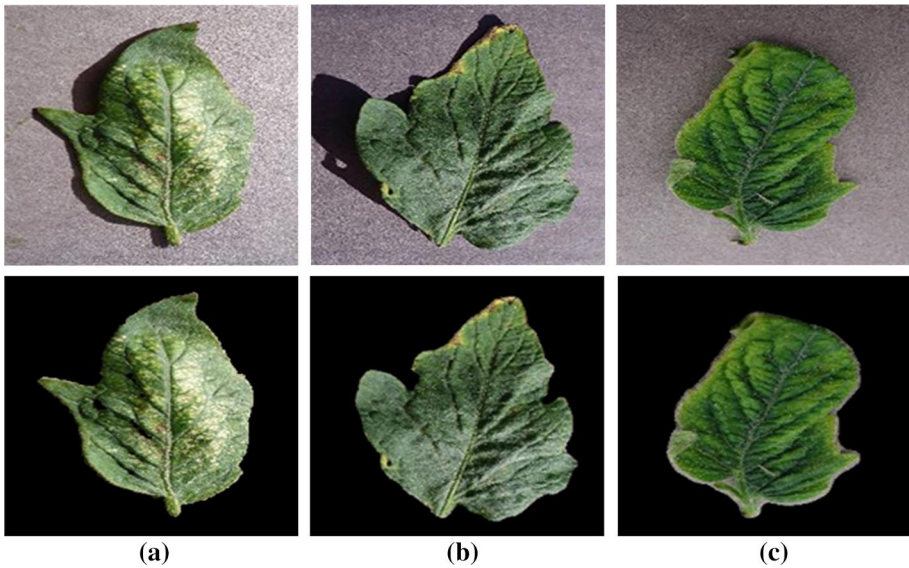
Fig. 3 The sample images for original image in top row and the corresponding background separated image in bottom row

2.3 Convolutional neural network model

2.3.1 CNN architecture

CNN's are feed-forward neural networks and a group of neural networks that have been shown to be very effective in the recognition and classification of images and are made up of several layers. CNN's comprise kernels, neurons, and filters that have learnable weights, parameters, and biases. This filter receives inputs, transforms them, and optionally continues them with nonlinearity (Uçar et al. 2017). Figure 4 demonstrates CNN architecture. It comprises the Convolutional, Pooling, Fully Connected, and Rectified Linear Unit (ReLU) layers.

The input image size of dimension $256 \times 256 \times 3$ RGB image is converted into gray image as in Fig. 5, 16 convolution filters of 3×3 size $= 128 \times 128 \times 16$ followed by the ReLU activation function as illustrated in Fig. 6. The width of Conv. layers (the number of channels) initially
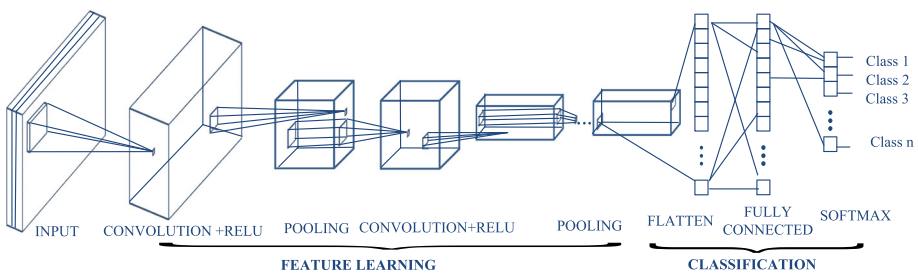

Fig. 4 The proposed CNN architecture for the attributes retrieval and classification of various diseases for different crops
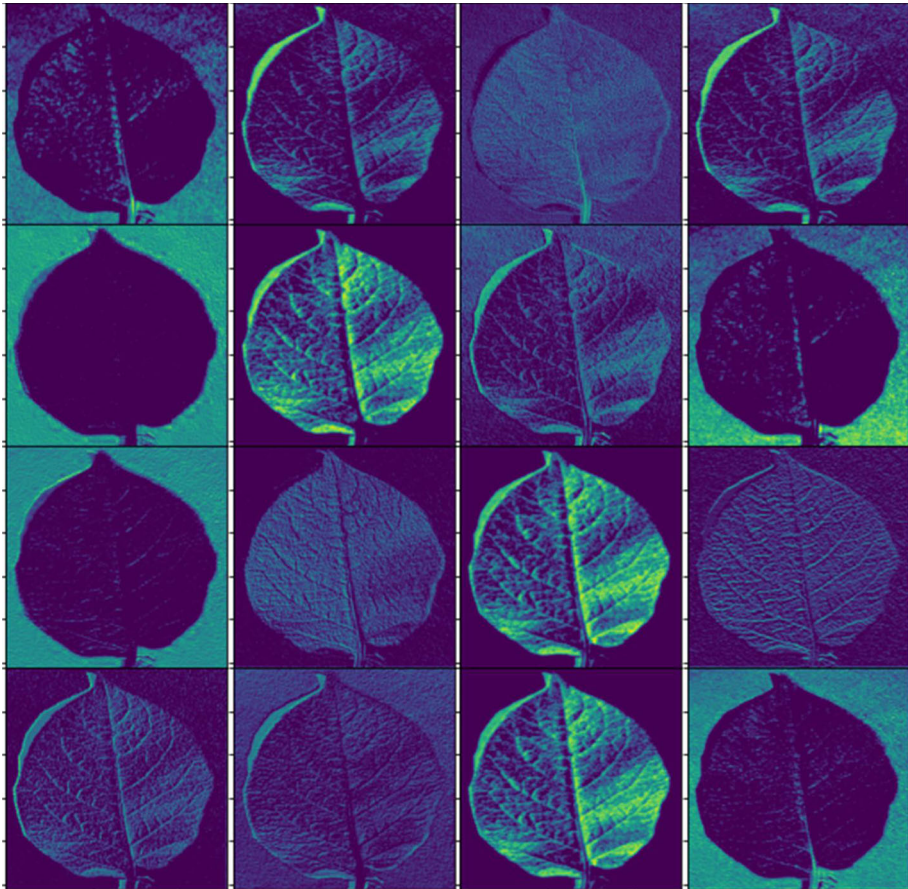
Fig. 5 The image at first layer of size 256×256

is 16 and increased by twice for each convolution layer. After pooling $64 \times 64 \times 16$, image patches are obtained as depicted in Fig. 7.

Further, the number of channels is taken as 32 (convolution filters) of 3×3 sizes that provide $32 \times 32 \times 32$. The pooling provides the down sampled images of dimension $16 \times 16 \times 32$. Furthermore, the number of channels is taken as 64 (convolution filters) sizes that provide $16 \times 16 \times 64$ as shown in Fig. 8.

The pooling provides the downsampled images of dimension $8 \times 8 \times 64$. Further, the number of channels is taken as 128 (convolution filters) of 3×3 sizes that provide $8 \times 8 \times 128$.

Find out the vertical and horizontal center lines of each patch to intact the information and the max pooling for the remaining $8 \times 8 \times 256$ provides the down sampled images of dimension $4 \times 4 \times 256$.

It is further a fully connected layer to come as $1 \times 1 \times 4096$ followed by the second fully connected layer with $1 \times 1 \times 256$ to the softmax containing a number of nodes equal to the number of output classes.
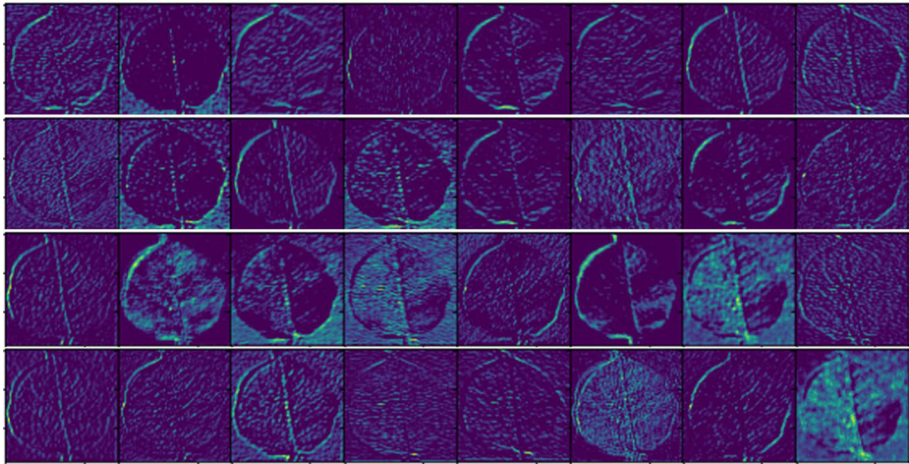
Fig. 6 The original image at second convolutional layer of size 128×128
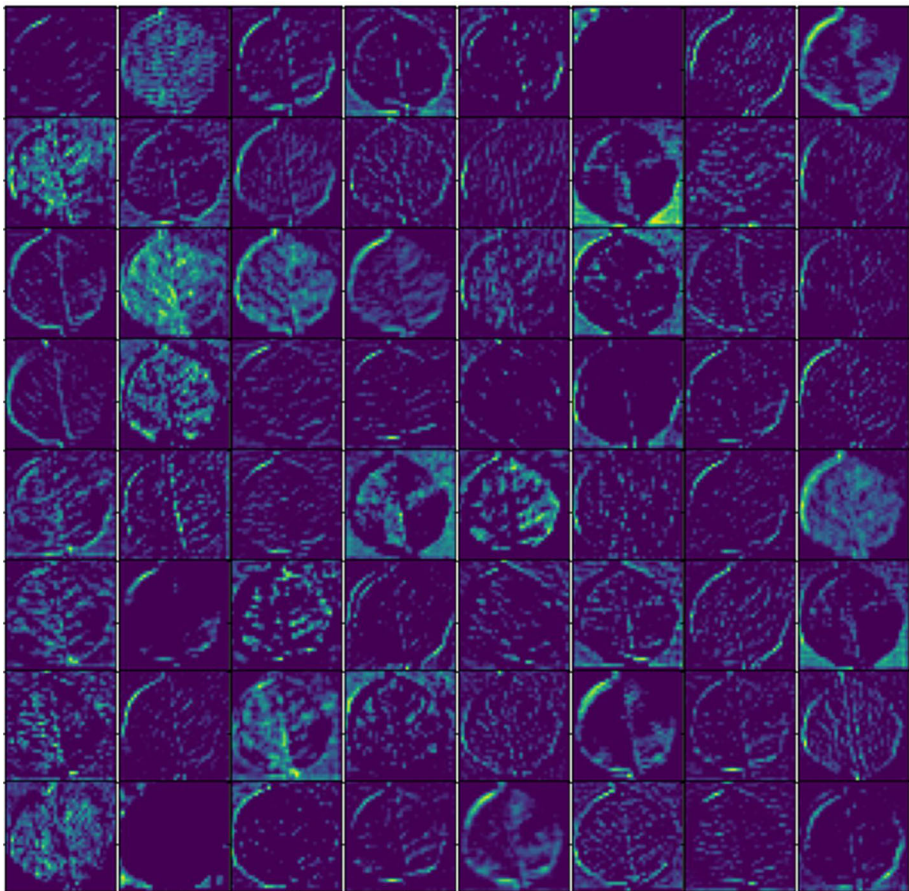
Fig. 7 The original image at the third convolutional layer of size 64×64
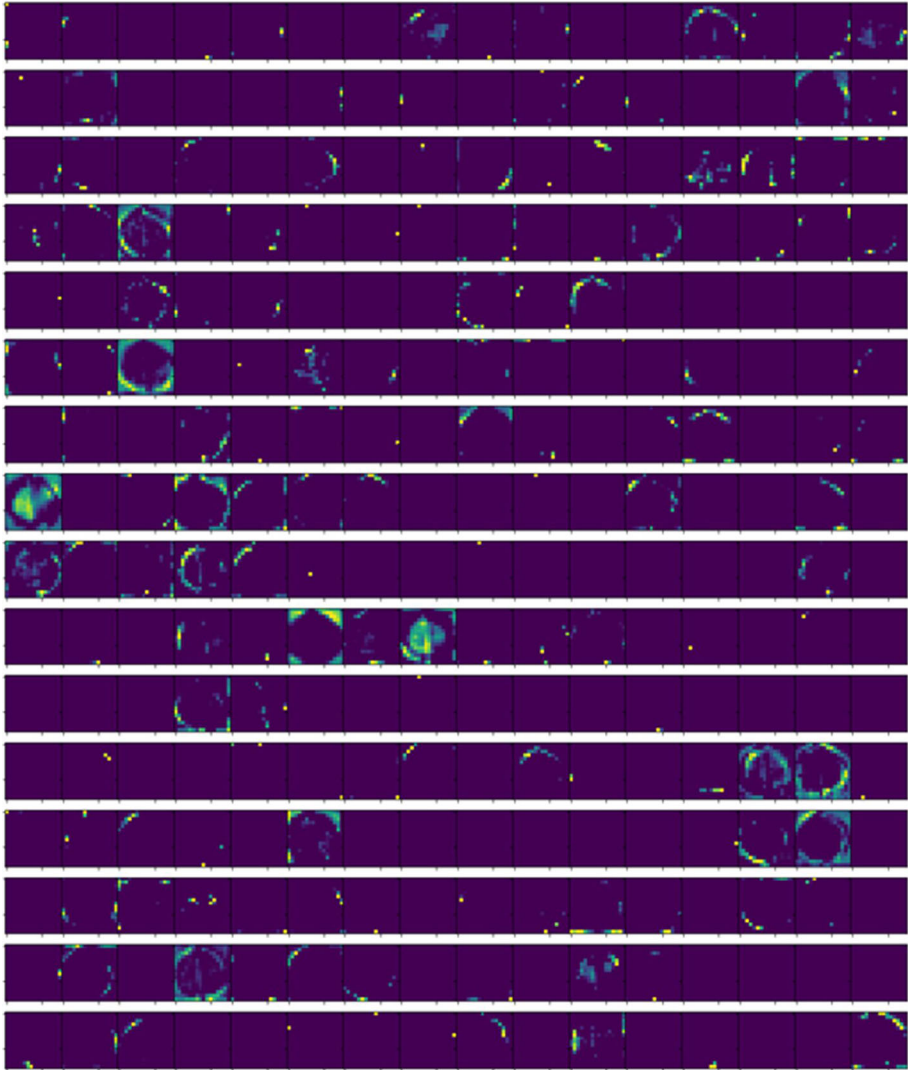
Fig. 8 The original image at fifth convolutional layer of size 16×16

2.3.2 Convolutional layer

The convolutional layer is the main building block for a convolutional network and performs several computational task. The primary purpose is, get characteristics from the input data of the image form. By learning image attributes, it manages and maintains the spatial connection among pixels. It uses small squares of the given image. The provided input image is convoluted by the use of a group of detectable neurons. This generates an activation or feature map in the output frame, and subsequently, the feature maps are inserted into the next convolution layer as input data.

2.3.3 Pooling layer

It decreases the dimensionality of every activation map. Nevertheless, the most relevant data remains available. The image input is slitted into a group of rectangles that are not overlapping. A non-linear operation like average/maximum wills down-sample each region. A pooling or a sub-sampling layer in CNN layers is added after a convolution layer once getting the function maps. This is to reduce the computing power needed for the data processing by reducing the dimensionality. Pooling shortens training time and prevents over- controls. The max-pooling layer most of the time follows rectified linear unit (ReLU) activation layer. Here, we utilize max-pooling of window size 2×2 pixels. Followed by the pooling the feature map has been obtained by employing the ReLU activation function.

2.3.4 ReLU layer

It is a wise non-linear element procedure that comprises rectifier-employing units. It is applied per pixel, and all negative values are reconstituted by zero in the feature map. The ReLU activation function is defined as,

$$F(x) = \max(0, x) \tag{7}$$

where, x is the weighted sum of inputs.

2.3.5 Fully connected layer

When each filter of the preceding layer is linked in the next layer of each filter then it is called a fully connected layer (FCL). The results of all the layers like pooling, convolutional, and ReLU are instances of the high-level input image features. The purpose of using the FCL is to identify the input image into different classes by using all the features, depending on the training set. FCL is known as the final pooling layer which uses Softmax, an activation function to feed the features to a classifier. At the output layer softmax, the neuron is used for binary classification. A softmax activation function is a form of logistic regression that normalizes an input value into a vector of values that follows a probability distribution whose total sums up to 1. The softmax activation function is defined as:

$$\sigma(x_j) = \frac{e^{x_j}}{\sum_{k=1}^{n} e^{x_k}} \tag{8}$$

where x is a vector of the inputs to the output layer and j indexes the output unit.

2.3.6 Optimization

For DL models, the right option for optimization algorithm could significantly improve both declines in training time and progress in precision. Adaptive moment estimation (ADAM) was first reported in 2014, as an optimization algorithm to train deep neural networks (DNN) with adaptive learning. ADAM optimizer is gaining enormous popularity in DL applications such as computer vision. This algorithm is an improved and updated version of the traditional stochastic gradient descent algorithm. ADAM optimizer shows finer results and performance as compared to classical stochastic gradient descent. Adam optimizer calculates the individual adaptive learning rate for each parameter from estimates of the first and second moments of the gradients. The intuition behind the Adam is that we don't want to roll fast because we

can jump over the minimum, we want to decrease the velocity a little bit for careful search. The equations for weight up gradation using adam can be given by,

$$m_t = B_1 m_{t-1} + (1 - B_1) g_t \tag{9}$$

$$v_t = B_2 v_{t-1} + (1 - B_2) g_t^2 \tag{10}$$

where m_t and v_t are estimates of first and second order moment respectively.

$$m_t' = \frac{m_t}{1 - B_1^t} \tag{11}$$

$$v_t' = \frac{v_t}{1 - B_2^t} \tag{12}$$

where m_t' and v_t' are bias corrected estimates of first and second moment, respectively. Finally, we update the parameter as shown below,

$$W_n = W_0 - \frac{n m_t'}{\sqrt{v_t' + \epsilon}} \tag{13}$$

where, W_n is updated weight, W_0 is old weight and n is learning rate. B_1, B_2, ϵ are hyper parameters.

The stochastic gradient descent (SGD) method with a learning rate of 0.01 and the weight update equation can be given by,

$$W_n = W_0 - n \nabla J(W) \tag{14}$$

where, W_n is the New weight, W_0 is the Initial weight, n is the Learning Rate, $\nabla J(W) =$ represents gradient with respect to parameter w. After that, to convert all the pooled images through flattening into a continuous vector a Flatten function has been used. In this additional parameters are not required as Keras can understand that the object classifier already holds pooled image pixels so they need to be flattened. In the next step, two dense functions have been used which are an FCL, in the first dense layer, Keras used the vector as the input for the NN which has been obtained above, and provided the output of 4 classes by using ReLU AF. In the next dense layer, a softmax function has been used to determine specific target output results. The ADAM optimizer has been used for better results.

2.4 Performance parameters

The success rate evaluation of the segmentation system is computed and compared in terms of F_1-score, modified Hausdorff distance (MHD), and Dice similarity coefficient (DC) (Kurmi et al. 2019; Dubuisson and Jain 1994). The performance of the classification method is defined in terms of accuracy (Ac) (Kurmi and Chaurasia Aug. 2018; Chaurasia and Chaurasia 2016) and receiver operating characteristic (ROC) curve (Fawcett 2006; Kurmi and Gangwar 2021) along with the evaluation of the area under the characteristic curve (AUC) (Fawcett 2006). For a better classification system the AUC should be of high value (Kurmi et al. 2021). The detail of the evaluation measure is

$$True\ positive\ rate\ (TPR) = (Sensitivity) = \frac{Tr_{Po}}{Tr_{Po} + Fa_{Na}}, \quad TNR = \frac{Tr_{Na}}{Tr_{Na} + Fa_{Po}} \tag{15}$$

$$False\ positive\ rate\ (FPR) = (1 - Specificity) = \frac{Fa_{Po}}{Fa_{Po} + Tr_{Na}} \tag{16}$$

where Tr_{Po} signifies the correctly identified positive number of samples, Tr_{Na} indicates the appropriately classified negative entities, Fa_{Po} represents falsely marked negative samples, and Fa_{Na} is the measure of incorrectly classified positive samples. For major class the true negative rate (TNR) and for minor class, the false positive rate (FPR) are defined as:

$$Precision = \frac{Tr_{Po}}{Tr_{Po} + Fa_{Po}} \ \ and \ \ Recall = \frac{Tr_{Po}}{Tr_{Po} + Fa_{Na}} \tag{17}$$

$$G - mean = (\Pi_{k=1}^{m} Recall_k)^{\frac{1}{m}} \tag{18}$$

where m is the number of classes. The G-mean metric is given by the ratio of a number of items from the minority to the majority class. The overall accuracy does not provide a true score when there is an imbalance in the dataset among the number of classes in each category. This imbalance is corrected by the G-mean by enhancing the accuracy of skewed class distribution (He and Garcia Sep. 2009). The ROC curve plots the true positivity rate vs the false positivity rate for the model at different cutoff points, to calculate the accuracy of the system. The area under the curve (AUC) (Fawcett 2006; Huang and Ling Mar. 2005) represents how well the binary classes (one vs. all in case of multiple classes) can be separated, with an ideal point at (0,1) where there are no misclassifications. Higher AUC implies the better performance of the model. The accuracy is given by:

$$Ac = (Tr_{Po} + Tr_{Na})/(Tr_{Po} + Fa_{Po} + Fa_{Na} + Tr_{Na}) \tag{19}$$

3 Result and discussion

The performance analysis of the proposed localization-based classification technique is done using three existing PlantVillage datasets of leaf images of bell pepper, potato, and tomato crops; The simulation has been performed using Python and its supporting packages as Tensorflow backend (Team 2018), Keras API (Keras 2018), and Scikit-learn (Pedregosa et al. 2011) library. The personal desktop (CPU: Core i5 processor 2.30 GHz, RAM: 16 GB) with google colab was used to train and test the network.

3.1 Evaluation of segmentation work

The performance of the proposed leaf region extraction has been compared with three state-of-the-art methods. A set of image taken for segmentaion performance anaysis was 150 image 10 from each class. The LSSC (Soares and Jacobs 2013) and SFAT (Sharma et al. 2017) approaches offered 0.862 and 0.868 F_1-score, respectively. On the other hand, the DC (MHD) values for LSSC and SFAT methods are 0.725 (10.58) and 0.769 (10.32), individually. The F_1-score provided by the proposed method is 0.908 with 0.817 DC and 7.54 MHD values that are far better than the state-of-the-art methods as shown in Table 1.

The complexity analysis of SFCC (Biswas et al. 2014) and SFAT (Sharma et al. 2017) approaches is approximated to $O(N^3)$ order. On the other hand, the computations of the proposed segmentation system are at par with LSSC (Soares and Jacobs 2013) of $O(N^2)$ order.

Table 1 Segmentation performance measures comparison with existing approaches

Techniques	F_1-Score	Dice coefficient	Modified Hausdorff distance	Computatonal complexity
LSSC Soares and Jacobs (2013)	0.862	0.725	10.58	$O(N^2)$
SFAT Sharma et al. (2017)	0.868	0.769	10.32	$O(N^3)$
SFCC Biswas et al. (2014)	0.779	0.786	11.52	$O(N^3)$
Proffered approach	**0.908**	**0.817**	**7.54**	$O(N^2)$

Best performing values are bold faced

Table 2 A comparative analysis of the performance measures as Average accuracy (Ac) as well as area under the characteristics curve (AUC) on considered dataset for proposed method

Datasets → Methods ↓	Bell pepper dataset		Potato dataset		Tomato dataset	
	Ac	AUC	Ac	AUC	Ac	AUC
PDDD Ferentinos (2018)	0.906	0.825	0.926	0.928	0.898	0.874
DLLA Bharali et al. (2019)	0.931	0.928	0.917	0.908	0.916	0.922
DLDIC Hang et al. (2019)	0.904	0.916	0.941	0.878	0.921	0.862
SELM-AE Singh and Misra (2017)	0.924	0.899	0.914	0.882	0.887	0.918
AFE-CNN Qin et al. (2016)	0.926	0.918	0.886	0.896	0.912	0.884
M-R-CNN Islam et al. (2017)	0.918	0.912	0.896	0.936	0.887	0.895
CDHVA Schor et al. (Jan 2016)	0.907	0.879	0.865	0.901	0.858	0.872
CLIQL-CNN Kaur et al. (May 2018)	0.948	0.898	0.908	0.933	0.904	0.898
Without Segmentation	0.924	0.908	0.917	0.912	0.896	0.884
Proposed method	**0.958**	**0.948**	**0.941**	**0.949**	**0.926**	**0.934**

Best performing values are bold faced

3.2 Classification work evaluation

For the classification work evaluation from the total images 20% are reserved for testig and remaining image are utilized for the traing with 10 fold cross validation. The analysis of the proposed classification work with existing models is provided in Table 2. For the classification of the tomato dataset, the SELM-AE (Singh and Misra 2017) method shows 0.887 Ac and 0.918 AUC. The categorization of potato images has been performed by SELM-AE with 0.914 Ac and 0.882 AUC.

The DLLA (Bharali et al. 2019) method offered accuracy for tomato datasets, potato, and bell pepper are 0.916, 0.917, and 0.931, respectively with AUC values of 0.922, 0.908, and 0.928, respectively. The accuracy performance of the CLIQL-CNN (Kaur et al. May 2018) approach is 0.904, 0.908, and 0.948, for tomato, potato, and pepper datasets, respectively.

One additional analysis has also been performed using all classes dataset for training and testing of the model. The analysis of the confusion matrix for all the 15 classes is shown in Fig 9. The clear illustration of all categories with their diagonal elements as correct categorized classes and off-diagonal entities as wrongly classified values.

The performance measure of the proffered approach as compared with existing approaches in terms of AUC is illustrated in Fig. 10.
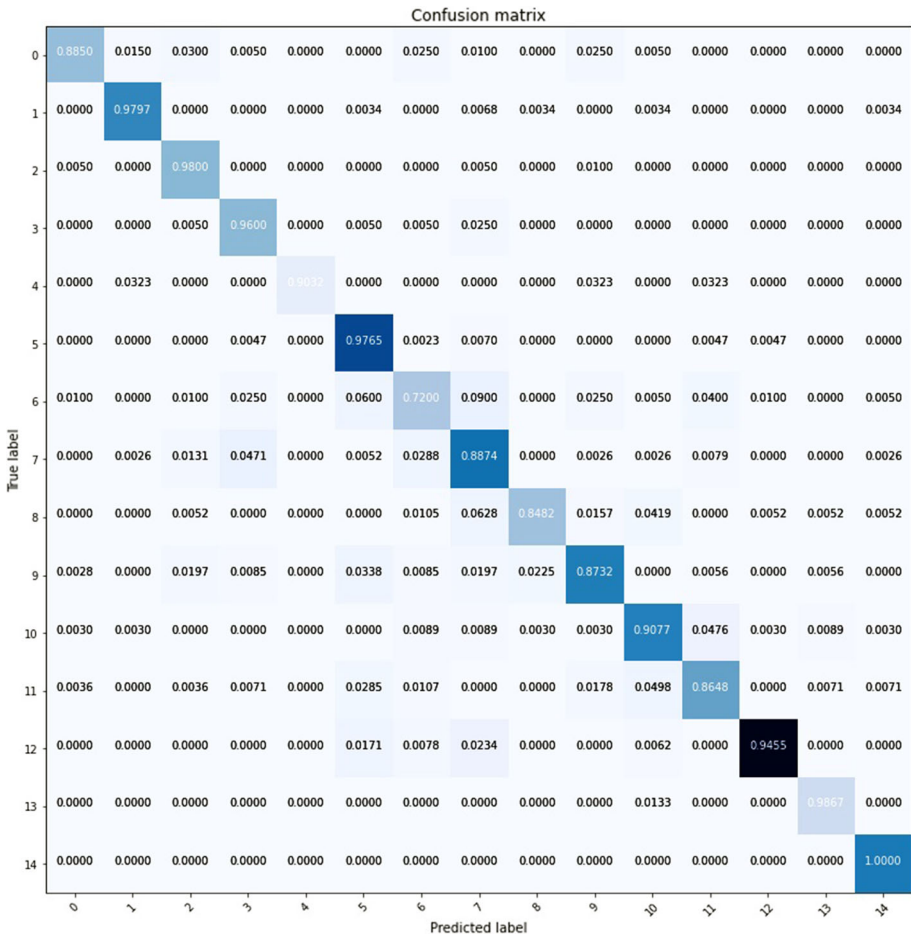
Confusion matrix

True label \ Predicted	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	0.8850	0.0150	0.0300	0.0050	0.0000	0.0000	0.0250	0.0100	0.0000	0.0250	0.0050	0.0000	0.0000	0.0000	0.0000
1	0.0000	0.9797	0.0000	0.0000	0.0000	0.0034	0.0000	0.0068	0.0034	0.0000	0.0034	0.0000	0.0000	0.0000	0.0034
2	0.0050	0.0000	0.9800	0.0000	0.0000	0.0000	0.0000	0.0050	0.0000	0.0100	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0000	0.0000	0.0050	0.9600	0.0000	0.0050	0.0050	0.0250	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	0.0000	0.0323	0.0000	0.0000	0.9032	0.0000	0.0000	0.0000	0.0000	0.0323	0.0000	0.0323	0.0000	0.0000	0.0000
5	0.0000	0.0000	0.0000	0.0047	0.0000	0.9765	0.0023	0.0070	0.0000	0.0000	0.0000	0.0047	0.0047	0.0000	0.0000
6	0.0100	0.0000	0.0100	0.0250	0.0000	0.0600	0.7200	0.0900	0.0000	0.0250	0.0050	0.0400	0.0100	0.0000	0.0050
7	0.0000	0.0026	0.0131	0.0471	0.0000	0.0052	0.0288	0.8874	0.0000	0.0026	0.0026	0.0079	0.0000	0.0000	0.0026
8	0.0000	0.0000	0.0052	0.0000	0.0000	0.0000	0.0105	0.0628	0.8482	0.0157	0.0419	0.0000	0.0052	0.0052	0.0052
9	0.0028	0.0000	0.0197	0.0085	0.0000	0.0338	0.0085	0.0197	0.0225	0.8732	0.0000	0.0056	0.0000	0.0056	0.0000
10	0.0030	0.0030	0.0000	0.0000	0.0000	0.0000	0.0089	0.0089	0.0030	0.0030	0.9077	0.0476	0.0030	0.0089	0.0030
11	0.0036	0.0000	0.0036	0.0071	0.0000	0.0285	0.0107	0.0000	0.0000	0.0178	0.0498	0.8648	0.0000	0.0071	0.0071
12	0.0000	0.0000	0.0000	0.0000	0.0000	0.0171	0.0078	0.0234	0.0000	0.0000	0.0062	0.0000	0.9455	0.0000	0.0000
13	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0133	0.0000	0.0000	0.9867	0.0000
14	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

Fig. 9 The confusion matrix shows the comparison of the proposed system for internal classes categorization accuracy

The DLDIC (Hang et al. 2019) approach illustrates the lowest value of AUC 0.879 and the PDDD technique offers 0.893. The proposed method shows 0.942 AUC that performs 5% better measure than existing methods. The performance measures sometimes do not provide fair comparison through accuracy and it needs another measure G-mean as shown in Fig. 11.

The PDDD approach offers 0.934 G-mean while the DLLA technique provides 0.943 and the DLDIC method gave 0.948 average G-mean. The proffered approach provides a better G-mean of 0.952 than the existing approaches.

The time required for the training has also been analyzed (in minutes) for the different methods and is given by the graph in Fig. 12. The PDDD technique is most efficient in terms of time taken to train the model, at 140 min while the SELM-AE is most costly. The DLLA (Bharali et al. 2019) method requires 150 min as average training time while DLDIC (Hang et al. 2019) approach needs 143 training minutes. The average time required to train the proposed method is 480 min The DLDIC (Hang et al. 2019) classification accuracy performance is at par with the proposed method, but the computational complexity more than the proposed
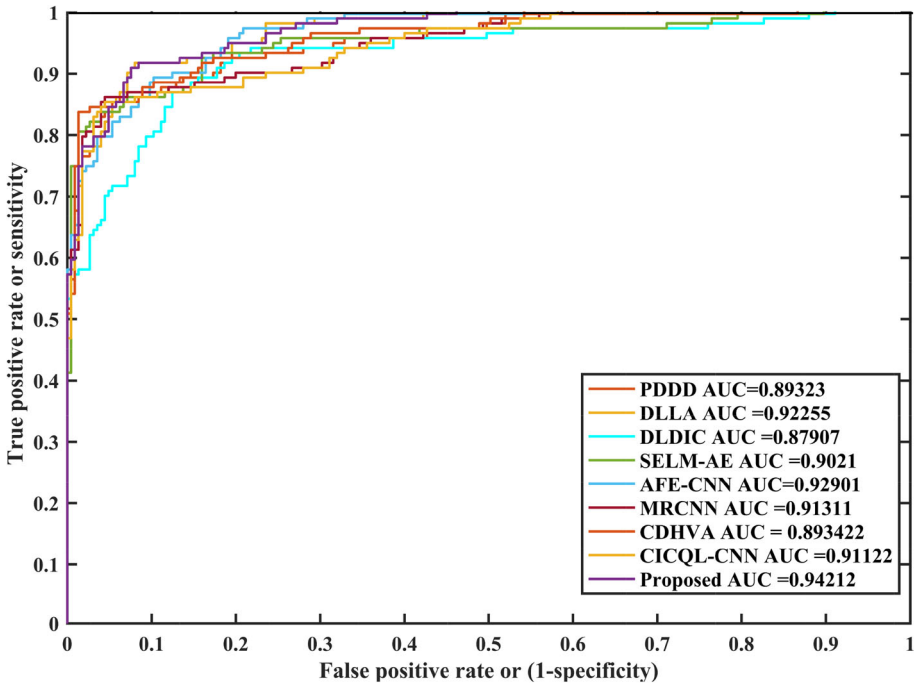
Fig. 10 The comparison of ROC curves for the proposed system with different existing approaches
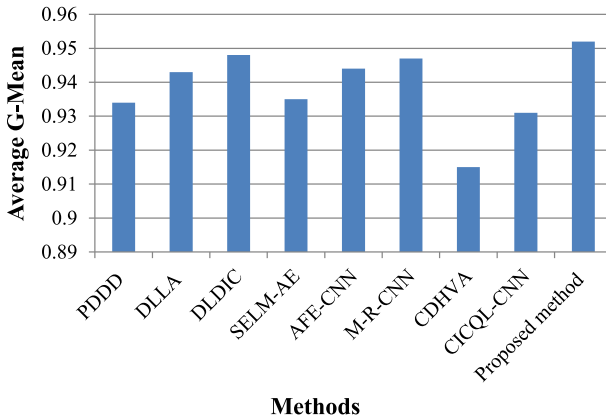

Fig. 11 The G-mean comparison of the proposed technique with different existing classification methods

method. hence, by comparing the accuracy, AUC, and time complexity the proposed method provides better performance measures than the existing approaches.

4 Conclusion

An image signal processing system has broad application in almost each and every domain of science, engineering, and management. here we are discussing the application in agro-
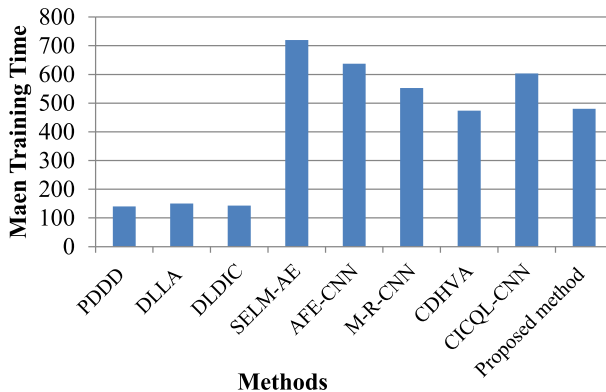
Fig. 12 Analysis of training time (in minutes) for the proposed system with existing techniques

economic growth systems like vegetation measurement, vigor diagnosis, phenotyping, etc. A proffered region localization-based deep CNN learning system offers discriminatory attributes to identify the crops as well as the diseases. The leaf region fixation was carried out using the leaf color properties and region growing approach. The segmentation measure of the localization system depicts 0.916 F_1- score with 0.824 and 7.29, Dice coefficient, and modified Hausdorff distance, respectively. The classification performance with mean Ac and AUC curves are 0.942 and 0.948, respectively, with 0.952 G-mean scores and a training time of 480 min. Hence, the proposed system gives better performance measures than the state of art techniques. the proposed work can also be employed for other crop identification and disease classification.

Declarations

Conflict of interests The authors declare that they have no conflict of interest.

References

Al-Kofahi, Y., Lassoued, W., Lee, W., & Roysam, B. (2010). Improved automatic detection and segmentation of cell nuclei in histopathology images. *IEEE Transactions on Biomedical Engineering, 57*(4), 841–852.

Alexander, J., Eggers, T., Picon, A., Alvarez-Gila, A., Ortiz Barredo, A. M., & Diez-Navajas, A. M. (2018). *System and method for detecting plant diseases*. United States patent, WO2017194276A1.

Arnal Barbedo, J. G. (2019). "Plant disease identification from individual lesions and spots using deep learning," *Biosystems Engineering,180*, 96–107. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1537511018307797

Aversano, L., Bernardi, M. L., Cimitile, M., Iammarino, M., & Rondinella, S. (2020). "Tomato diseases classification based on vgg and transfer learning," In *IEEE international workshop on metrology for agriculture and forestry (MetroAgriFor),2020*, 129–133.

Barbedo, J. G. (2018). Factors influencing the use of deep learning for plant disease recognition. *Biosystems Engineering, 172,* 84–91.

Bauer, A., Bostrom, A., Ball, J., Applegate, C., Cheng, T., Laycock, S., et al. (2019). Combining computer vision and deep learning to enable ultra-scale aerial phenotyping and precision agriculture: a case study of lettuce production. *Horticulture Research, 6,* 06.

Beucher, S., & Meyer, F. (1993). *The morphological approach to segmentation: the watershed transformation*, 01, (Vol. 34, pp. 433–481).

Bharali, P., Bhuyan, C., & Boruah, A. (2019). "Plant disease detection by leaf image classification using convolutional neural network," in *Information and Communications Technology,* (pp. 194–205). Springer, Singapore.

Biswas, S., Jagyasi, B., Singh, B. P., & et al. (2014). "Severity identification of potato late blight disease from crop images captured under uncontrolled environment," In *Canada Internship Humanity Technical Conference - (IHTC),* (pp. 1–5).

Brahimi, M., Kamel, B., & Moussaoui, A. (2017). Deep learning for tomato diseases: Classification and symptoms visualization. *Applied Artificial Intelligence, 31,* 1–17.

Calapez, A., & Rosa, A. (2010). A statistical pixel intensity model for segmentation of confocal laser scanning microscopy images. *IEEE Transactions on Image Processing, 19*(9), 2408–2418.

Callara, A. L., Magliaro, C., Ahluwalia, A., et al. (2020). A smart region-growing algorithm for single-neuron segmentation from confocal and 2-photon datasets. *Frontiers in Neuroinformatics, 14,* 9.

Chaurasia, V., & Chaurasia, V. (2016). Statistical feature extraction based technique for fast fractal image compression. *Journal of Visual Communication and Image Representation, 41,* 87–95.

Deng, C., Xue, Y., Liu, X., Li, C., & Tao, D. (2019). Active transfer learning network: a unified deep joint spectral-spatial feature learning model for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing, 57*(3), 1741–1754.

Dubuisson, M., & Jain, A.K. (1994). "A modified Hausdorff distance for object matching," In *Proceedings of 12th international conference on pattern recognition,* (vol. 1, pp. 566–568).

Fawcett, T. (2006). An introduction to roc analysis. *Pattern Recognition Letters, 27*(8), 861–874.

Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture, 145,* 311–318.

Hang, J., Zhang, D., Chen, P., Zhang, J., & Wang, B. (2019). Classification of plant leaf diseases based on improved convolutional neural network. *Sensors, 19,* 4161.

Haralick, R. M., Zhuang, X., Lin, C., & Lee, J. S. J. (1989). The digital morphological sampling theorem. *IEEE Transactions on Acoustics, Speech, and Signal Process., 37*(12), 2067–2090.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. B. (2017). "Mask R-CNN," *CoRR*, vol. arxiv:1703.06870.

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering, 21*(9), 1263–84.

Huang, J., & Ling, C. X. (2005). Using auc and accuracy in evaluating learning algorithms. *IEEE Transactions on Knowledge and Data Engineering, 17*(3), 299–310.

Ioffe, S., & Szegedy, C. (2015). "Batch normalization: Accelerating deep network training by reducing internal covariate shift," In F. Bach and D. Blei, (Eds.) *Proceedings of the 32nd international conference on machine learning,* Proceedings of Machine Learning Research, (vol. 37. pp. 448–456). Lille, France: PMLR, 07–09 Jul.

Islam, M., Wahid, Anh Dinh, K., & et al. (2017). "Detection of potato diseases using image segmentation and multiclass support vector machine," In *Canadian Conference on Electrical and Computer Engineering (CCECE),* (pp. 1–4).

Karthik, R., Hariharan, M., Anand, S., Mathikshara, P., Johnson, A., & Menaka, R. (2020). Attention embedded residual cnn for disease detection in tomato leaves. *Applied Soft Computing, 86,* 105933.

Kasun, L. L. C., Yang, Y., Huang, G.-B., & Zhang, Z. (2016). Dimension reduction with extreme learning machine. *IEEE Transactions on Image Processing, 25*(8), 3906–3918.

Kaur, S., Pandey, S., & Goel, S. (2018). Semi-automatic leaf disease detection and classification system for soybean culture. *IET Image Processing, 12*(6), 1038–1048.

Keras. (2018). "Keras Documentation," https://keras.io, [Online; accessed 2-Feb-2018].

Kurmi, Y., & Chaurasia, V. (2020). Classification of magnetic resonance images for brain tumour detection, *IET Image Processing,14*(12), 2808–2818. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-ipr.2019.1631

Kurmi, Y., & Gangwar, S. (2021). "A leaf image localization based algorithm for different crops disease classification," *Information Processing in Agriculture*, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S221431732100024X

Kurmi, Y., Chaurasia, V., & Ganesh, N. (2019). "Tumor malignancy detection using histopathology imaging," *Journal of Medical Imaging and Radiation Sciences*, *50*(4), 514–528.

Kurmi, Y., Gangwar, S., Agrawal, D., Kumar, S., & Srivastava, H. S. (2021). "Leaf image analysis-based crop diseases classification," *Signal, Image and Video Processing*, 2021. [Online]. Available: https://doi.org/10.1007/s11760-020-01780-7

Kurmi, Y., & Chaurasia, V. (2018). Multifeature-based medical image segmentation. *IET Image Process, 12*(8), 1491–1498.

Liu, J., & Wang, X. (2020). Tomato diseases and pests detection based on improved yolo v3 convolutional neural network. *Frontiers in Plant Science, 11,* 898.

Li, L., Zhang, S., & Wang, B. (2021). Plant disease detection and classification by deep learning-a review. *IEEE Access, 9,* 56-683-56–698.

Lv, M., Zhou, G., He, M., Chen, A., Zhang, W., & Hu, Y. (2020). Maize leaf disease identification based on feature enhancement and dms-robust alexnet. *IEEE Access, 8,* 57 952-57 966.

Mahlein, A. .-K. (2016). Plant disease detection by imaging sensors – parallels and specific demands for precision agriculture and plant phenotyping. *Plant Disease, 100*(2), 241–251 (**pMID: 30694129**).

Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science, 7,* 1419.

Nagasubramanian, G., Sakthivel, R. K., Patan, R., Sankayya, M., Daneshmand, M., & Gandomi, A. H. (2021). "Ensemble classification and iot based pattern recognition for crop disease monitoring system," *IEEE Internet of Things Journal*, pp. 1–1.

Ng, H. F. (2006). Automatic thresholding for defect detection. *Pattern Recognition Letters, 27,* 1644–1649.

Pedregosa, F., Varoquaux, G., & Gramfort, A. (2011). Scikit-learn: machine learning in Python. *Journal of Machine Learning Research, 384*(12), 2825–2830.

Qin, F., Liu, D., Sun, B., et al. (2016). Identification of alfalfa leaf diseases using image recognition technology. *PLOS ONE, 11*(12), 1–26.

Ridler, T. W., & Calvard, S. (1978). Picture thresholding using an iterative selection method. *IEEE Transactions on Systems, Man, and Cybernetics, 8*(8), 630–632.

Sardogan, M., Tuncer, A., & Ozen, Y. (2018). "Plant leaf disease detection and classification based on cnn with lvq algorithm," In *2018 3rd international conference on computer science and engineering (UBMK),* (pp. 382–385).

Schor, N., Bechar, A., Ignat, T., et al. (2016). Robotic disease detection in greenhouses: combined detection of powdery mildew and tomato spotted wilt virus. *IEEE Robotics and Automation Letters, 1*(1), 354–360.

Sharma, Aparajita, R., Singh, A., & et al. (2017). "Image processing based automated identification of late blight disease from leaf images of potato crops," In *2017 40th International Conference on Telecomm. and Signal Processing (TSP),* (pp. 758–762).

Siddique, M. A. B., Arif, R. B., & Khan, M. M. R. (2018). Digital image segmentation in matlab: a brief study on otsu's image thresholding, In *2018 international conference on innovation in engineering and technology (ICIET),* (pp. 1–5).

Singh, V., & Misra, A. (2017). Detection of plant leaf diseases using image segmentation and soft computing techniques. *Information Processing in Agriculture, 4*(1), 41–49.

Soares, J. . a. V., & Jacobs, D. . W. . (2013). Efficient segmentation of leaves in semi-controlled conditions. *Machine Vision and Applications, 24*(8), 1623–1643.

Szegedy, C., Toshev, A., & Erhan, D. (2013). *Deep neural networks for object detection*, 01, (pp 1–9).

Team, G. B.. (2018). "TensorFlow," https://www.tensorflow.org/, [Online; accessed 2-Feb-2018].

Torr, P. H. S., & Murray, D. W. (1997). The development and comparison of robust methodsfor estimating the fundamental matrix. *International Journal of Computer Vision, 24*(3), 271–300.

Tseng, S.-M., Su, J.-H., Chang, W.-Y., Peng, Y.-H., & Chen, W.-C. (2014). "Method and system for recognizing plant diseases and recording medium," *United States patent*, vol. US8781174B2, 07.

Uçar, A., Demir, Y., & Güzeliş, C. (2017). Object recognition and detection with deep learning for autonomous driving applications. *Simulation, 93*(9), 759–769.

Wang, G., Sun, Y., & Wang, J. (2017). Automatic image-based plant disease severity estimation using deep learning. *Computational Intelligence and Neuroscience, 2017,* 1–8.

Zeng, Q., Ma, X., Cheng, B., Zhou, E., & Pang, W. (2020). Gans-based data augmentation for citrus disease severity detection using deep learning. *IEEE Access, 8,* 172-882-172–891.

Zhang, J., Xia, Y., Xie, Y., Fulham, M., & Feng, D. D. (2018). Classification of medical images in the biomedical literature by jointly using deep and handcrafted visual features. *IEEE Journal of Biomedical and Health Informatics, 22*(5), 1521–1530.

Yashwant Kurmi received B.E. degree in Electronics and Communication Engineering (ECE) from JEC Jabalpur, RGPV University of Bhopal, in the year 2006. He acquired his Masters of Digital communication in 2013 from Maulana Azad National Institute of Technology (MANIT), Bhopal. Presently he is a Research Fellow in the department of ECE of MANIT, Bhopal, India. He has more than 10 journals and 5 international of international repute. His research interest includes signal and image processing, image localization and classification, and pattern recognition, object detection, computational intelligence, machine learning, and data mining. The supreme research objective is to move along the rudimentary concepts with scientific utensils to evolve structured strategies for solving the real-world complex problems.

Prankur Saxena received Masters from IIT Gandhinagar, India. Bhupendra Singh Kirar received BE degree in Electronics and Communication Engineering from Samrat Ashok Technological Institute, Vidisha, India in 2004. He received M.Tech. and Ph.D. degree from Maulana Azad National Institute of Technology, Bhopal, India in 2009 and 2019, respectively. He has over 13 years of teaching experience in the field of Electronics and Communication Engineering. Also, he has co-authored a text book on "Basic Electrical & Electronics Engineering", by M/s. B.S. Publications; Hyderabad in the year 2011. He has published more than 10 papers in national and international journals/ conferences under SCI/Scopus. He is an active reviewer of IET Image Processing, Biomedical Signal Processing & Control. He is Assistant Professor in the department of Electronics and Communication Engineering, Indian Institute of Information Technology, Bhopal, India. His research interest includes area of signal processing, image processing, computer-aided medical diagnosis and machine learning.

Suchi Gangwar received the Bachelor's Degree in Agriculture Engineering from the AKA agriculture college, Sehore, India, the M.Sc. Degree in agronomy and crop from JNKVV, Jabalpur, India. She carried out her research in agronomy and crop science and awarded Ph.D. degree in 2015 from JNKVV, Jabalpur, India. Presently she is with RKDF University, Bhopal, India. She has published more than 50 research paper of national and international repute. Her research interests include agronomy, Agricultural Biodiversity analysis, image pattern recognition and data classification.

Vijayshri Chaurasia received the Bachelors Degree in ECE from the Pt. Ravishankar Shukla University, Raipur, India, the M.Tech. Degree in Digital Communication from MANIT, Bhopal, India, she carried out her research in image compression and awarded Ph.D. degree in 2011 from MANIT, Bhopal, India. Presently she is with MANIT, Bhopal, India. She has published more than 50 research paper of national and international repute and associated with professional bodies IETE, ISTE and IEEE. Her research interests include image processing, signal processing, pattern recognition and feature extraction.

Aditya Goel received the B.E. (Hons.) degree in electronics and communication and the M.Tech degree in communication engineering from the Indian Institute of Technology, Bombay, India. He received the Ph.D. degree in electronics and communication engineering in 2000. He has teaching and research experience of more than 24 years in various capacities. He is currently a Professor in the Department of Electronics and Communication Engineering, Maulana Azad National Institute of Technology, Bhopal, India. He had been a Professor and the Head of the Department of Electronics and Communication Engineering from November 2011 to November 2013 for a tenure of two years. He has published more than 100 research papers in reputed international journals and conferences including the IEEE. His areas of interests include optical communication, digital signal processing, optical networks, computer communication, etc. He has guided several M.Tech dissertations and Ph.D. thesis on these areas. He is a Life Fellow Member of the Institution of Electronics and Telecommunication Engineers, Life Fellow member of the Institution of Engineers, and Senior Member of the Computer Society of India.