Crop identification and disease classification using traditional machine learning and deep learning approaches

Aravind Krishnaswamy Rangarajan*, Raja Purushothaman*,***, Maheswari Prabhakar* and Cezary Szczepański**

School of Mechanical Engineering, SASTRA Deemed University, Thanjavur, Tamil Nadu, India

**Lukasiewicz Research Network, Institute of Aviation, Warsaw, Poland

*,*** Corresponding Author: raja_sastra@yahoo.com

ABSTRACT

Crop and disease classification is one of the important problems in automation of agricultural processes with multicropping method, where the field is cultivated with more than one crop. In order to solve this classification problem, a study has been carried out in the field cultivating eggplant (*Solanum melongena*) and tomato (*Solanum lycopersicum*) using the images obtained from a mobile phone camera. Textural descriptors, namely, contrast, correlation, energy, and homogeneity, were extracted from the gray-scale converted RGB image for crop identification, that is, tomato or eggplant, and the same descriptors were extracted from the gray-scale converted image from Hue Saturation Value (HSV) for disease classification (due to *Cercospora* leaf spot disease or two-spotted spider infestation). Discriminant analysis, Naive Bayes algorithm, support vector machine, and neural network were the classification algorithms used with a resulting best accuracy of 97.61%, 95.62%, 98.01%, and 98.94% for crop identification and 86.09%, 76.52%, 86.96%, and 86.04% for disease classification, respectively. Similarly, the application of algorithm with 6 histogram-based descriptors for health status detection resulted in an accuracy of 66.67%, 37.04%, 50%, and 72.9%, respectively. A deep learning algorithm, namely, AlexNet, was also evaluated, which resulted in an accuracy of 100% for crop identification, 89.36% for health status detection, and 81.51% for disease classification. Among the algorithms, AlexNet resulted in the best average accuracy of 90.29% for the above classification tasks.

Keywords: Crop identification; disease classification; machine learning algorithm; *Solanum melongena*; *Solanum lycopersicum*.

INTRODUCTION

Globally agricultural production is affected by the incidence of pest and diseases. The evolution of modern technology is providing new tools and techniques in disease classification of crops. In several parts of the world, two or more crops are grown in the same field (known as multicropping) to improve productivity and minimize disease incidence, weed growth, and pest infestation (Wang et al., 2014, Ehrmann & Ritz, 2014, Hussain et al., 2008, Anders et al., 1996). In a multicropping scenario, crop identification becomes necessary in order to determine the health status and disease classification in a particular crop. In this work, one such problem has been taken for the study, which involves crop identification in a multicropping scenario, health status detection, and classification of disease.

Crop identification

Many studies reported classification of plant species using images of the isolated leaf samples employing traditional machine learning algorithms such as neural networks and support vector machine, and recently, deep learning approaches such as deep convolutional neural network are also employed (Lee et al., 2017, Lukic et al.,

2017, Dyrmann et al., 2016, Aakif & Khan, 2015, Kalyoncu & Toygar, 2015, Olsen et al., 2015, Zhao et al., 2015, Fang & Li, 2014, Chaki & Parekh, 2012). Deep learning approaches learn features automatically, whereas features such as geometric, color, textural, or other suitable features have to be selected manually for studies with traditional machine learning algorithms based classification. In a field or a farm with multiple crops, classification without separation of parts (obtained from crop) during the usage of image processing technique is a challenging problem (Sun et al., 2017). Few studies have explored the identification of crop species by processing the crop image as a whole (Sun et al., 2017, Pahikkala et al., 2015). Sun et al. (2017) conducted an experiment with (Beijing Forestry University) BJFU100 dataset consisting of 10,000 images from 100 different ornamental crop species using a 26-layer deep learning model and reported an accuracy of 91.78%. Pahikkala et al. (2015) developed Rank Regularized Least Squares (RankRLS) learning algorithm for classification of *Taraxacum officinale,* a weed from *Avena sativa*. Texture-based features were used for classification with an error rate of 0.05% negative predictive value for a dataset of 20 images.

Precision crop protection in field condition due to biotic and abiotic stress is claimed to be one of the major problems in the implementation of precision in agriculture (Ampatzidis et al., 2017). In order to fulfill the precision crop protection, early diagnosis of diseases is required. Detection, classification, and severity estimation of disease on isolated sample leaf were carried out by some studies (Barbedo 2016a, Barbedo 2016b, Parikh et al., 2016, Pujari et al., 2014, Arivazhagan et al., 2013, Chaudhary et al., 2012, Bashish et al., 2011). Deep learning approaches are used for disease classification in other recent studies (Barbedo 2018a, Liu et al., 2018, Ferentinos, 2018, Brahimi et al., 2017, Mohanty et al., 2016). In some developing countries, on-site identification of the disease is carried out by experts. Also, some diseases require chemical analysis of the diseased samples (obtained from the crop) in a laboratory in order to confirm its existence. These techniques require skilled experts, and they are time-consuming (Pilli et al., 2014).

Raza et al. (2015) developed a system for the identification of tomato crop infected with Powdery Mildew (PM) disease. A thermal map was generated using a thermal camera, as well as the visible camera, while disparity map was created using a stereo camera. The various features obtained from this map were trained using support vector machine. The average detection accuracy was higher than 75% after 9 days of inoculation, whereas with upper and lower bound, the best detection accuracy achieved in the study was 68.40% on the 5th day and 89.60% on the 13th day. Another study by Schor et al. (2017) developed a robotic manipulator placed on a conveyor for its movement in a greenhouse. The robot was placed above the potted tomato crops. The manipulator was equipped with RGB camera and Near Infra-Red-Red-Green (NIR-R-G) multispectral camera. The system detected PM and Tomato Spotted Wilt Virus (TSWV) infection in tomato crop using principal component analysis. The accuracy for detection of PM and TSWV was 95% and 90%, respectively, using the RGB image, whereas with the NIR-R-G image, the resulting accuracy was 80% and 61%, respectively. These studies were conducted in a controlled or indoor lighting condition.

Classification of crop and disease from the images of the crop in the agricultural field (with varying lighting) demands a method for transforming the image to a common standard to cancel the effects of changes in luminance. Wspanialy and Moussa (2016) developed such a system, which consists of a mobile platform equipped with an augmented lighting system and a Single-Lens Reflex (SLR) camera for the detection of the PM in tomato crops cultivated in a greenhouse. Hough forest, which uses Hough transform for detection and random forest for learning, resulted in an accuracy of 85% for the detection of PM. A study by Johannes et al. (2017) explored the use of Naive Bayes classifier for the identification of diseases in wheat, namely, rust, septoria, and tan spot, with the images obtained from a mobile phone camera. The variance in the image was reduced using the color constancy transformation. The study classifies rust, septoria, and tan spot detected at an early stage with an accuracy of 78%, 76%, and 73%, respectively, while at the medium-late stage, it is 81%, 79%, and 82%, respectively. Fuentes et al. (2017) demonstrated the use of a deep neural network for the identification of 9 different types of diseases and pests in tomato crop. The detection technique utilizes fast region-based convolutional neural network, region-based fully convolutional network, and single-shot multibox detector, combined with deep feature extractors to minimize false positive. The technique was able to detect disease from the crop image even with complex background. Polder et al. (2014) developed a platform operated manually by the user for identification of tulip breaking virus disease in tulip (flower) field using NIR and RGB images. Fisher's linear discriminant algorithm was used for classification of the disease. Pilli et al. (2014) explored the use of a semiautomatic system for identifying the specific diseases in cotton and groundnut using the neural network and k-means clustering algorithm.

So far, the literature study focuses on the development of a system with an independent task of either crop identification or disease classification. Hence, a systematic approach is required to combine these independent tasks into a single system for effectively using it in a garden or a field with multiple crops. This study attempts on such an overall system for both crop and disease classification in order to analyze its feasibility (using an image captured

by a simple mobile camera). Presently, it can classify two crop species and classification of two major diseases in single species.

Statistical textural descriptors extracted from the segmented image have been utilized for classification using four machine learning algorithms. The images are used (directly) as input in the deep learning algorithm. These steps in classification have been explained in Section "Materials and Methods." The performance has been evaluated using these shallow machine learning algorithms (linear discriminant analysis, Naive Bayes algorithm, and support vector machine and neural network) and deep learning algorithm (AlexNet). The results are presented in Section "Results." The factors influencing the prediction were discussed in Section "Discussion."

MATERIALS AND METHODS

Crop selection and dataset creation

Based on the local cultivation, disease incidence, and production statistics, the study was conducted in a field cultivating eggplant (*Solanum melongena*) and tomato (*Solanum lycopersicum*). Regarding production quantity, tomato and eggplant contribute about 28.9 % and 8.1% of the total vegetable production (146.55 million tons) in India (Vanitha et al., 2013). Due to the wide-scale cultivation, the prevalence of pests and diseases is common among these crops.

Two important types of the problem affecting the production of eggplant, namely, *Cercospora* leaf spot and two-spotted spider infestation, were taken as the objective for classification using the proposed machine learning algorithms (discussed in Section "Selection and implementation of machine learning algorithm"). *Cercospora* leaf spot is identified by chlorotic lesion, which is angular with irregular in shape and becomes grayish-brown color in the later stage of the disease as shown in Fig. 1(a) (Rangaswami & Mahadevan, 2012). It is due to the deuteromycete fungus (*Cercospora solani-melongenae*) that is majorly confined to eggplant species. This crop pathogen only affects the leaves of eggplant and not the fruit. Two-spotted spider or red spider mite (*Tetranychus urticae* Koch) infestation is a serious pest issue affecting many horticultural crops such as tomato and beans including eggplant as shown in Fig. 1(b) (Srinivasan, 2009). The pest survives by consuming the crop cell, which results in the reduction of chlorophyll and dropping of leaves occur under severe infestation.

(a)   (b)

Figure 1. Diseases in eggplant: (a) *Cercospora* leaf spot; (b) two-spotted spider infestation.

The images (as shown in Figure 2) were captured in the field on a mid-day (without cloud cover) using a mobile phone camera (Xiaom Redmi Note 3) with a resolution of 16 Megapixels near Thanjavur (10°44 N, 79°03 E). The cameras were positioned approximately perpendicular to the ground.

| (a) | (b) |

Figure 2. Images of the crop specimen selected for study. (a) Sample image of tomato crop. (b) Sample image of eggplant.

The acquired images include two different crops, and one of the crops, namely, eggplant, is affected with biotic stress. The images of eggplant with biotic stress were annotated with an expert in the plant pathology. The plant images, which have several infected leaves with visible symptoms, were considered for the study. An augmentation technique such as transformation (rotation and translation) and modification in intensity values was performed to increase the dataset size. The dataset created using the acquired images is shown in Table 1.

Table 1. Dataset.

Categories			Created dataset	Augmented dataset
Crop		Eggplant	135	810
		Tomato	56	448
Status	Eggplant	Healthy	39	234
		Unhealthy	96	576
Disease	Eggplant	Two-spotted spider mite	57	342
		Cercospora leaf spot	39	237

The first task in the study involves identification of eggplant from tomato crop using the machine learning algorithms. The second task deals with the health status detection and disease classification in eggplant using the same machine learning algorithms (shown in Figure 3).
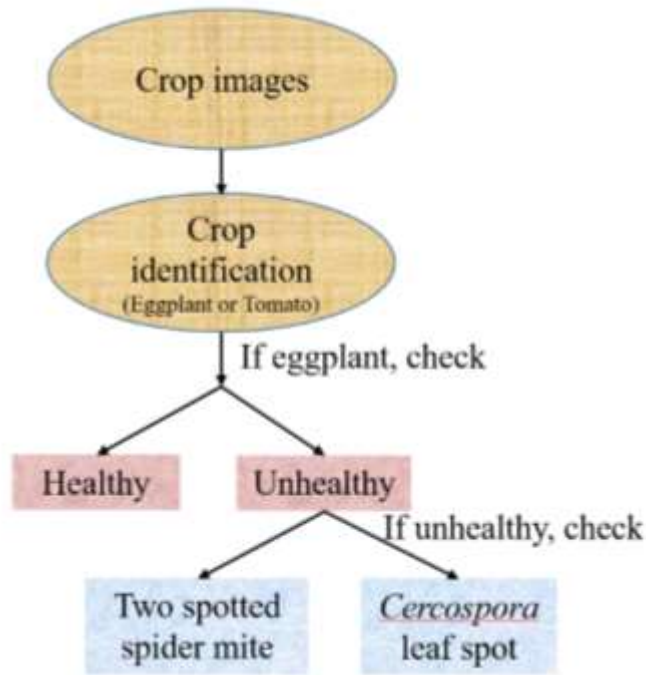
Figure 3. Classification tree of the entire task.

The entire second task was performed only in eggplant as this study is a preliminary pilot study for such a disease diagnosis system in a multicropping scenario. In the second part of the study, the same set of tasks was repeated using the deep learning algorithm (pretrained AlexNet), and the results were analyzed in Section "Results."

Feature selection

The plant physical characteristic features are influenced by both genetic information and environmental factors (Nap et al., 2017; Sano, 2010). This is evident from the images of tomato and eggplant as depicted in Fig. 4(a) and Fig. 4(b), which show a different arrangement of leaf, shape, and diameter of a crop after extraction of the crop from the background.

These factors provide a characteristic physical structure of crop when viewed in the field. In this case, the crop image is captured from the top view using a mobile camera as shown in Figure 2. The tomato crop has the compound leaf with odd leaflets, while the eggplant has distinctive large simple leaves as shown in Figure 4. The large leaves of eggplant are closely arranged with a minimum ground surface.


(a)   (b)

Figure 4. Segmented image of selected crop. (a) Sample image of tomato crop. (b) Sample image of eggplant.

Figure 5 shows the distribution of frequency of occurrence of gray-scale intensity values greater than zero of the images for a column width of 730 pixels. Fig. 5(a) shows that the distribution is not smooth as the frequency values vary widely, and there are many local minima. This is due to the presence of open spaces between the leaves. Fig. 5(b) shows a smooth broader distribution of the frequency values in eggplant compared to the tomato crop as it has minimum open spaces.
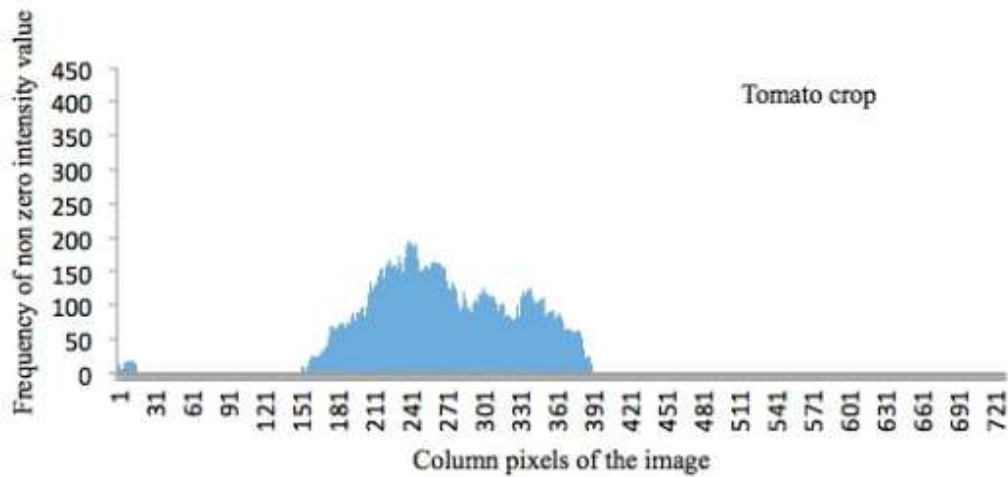
(a)
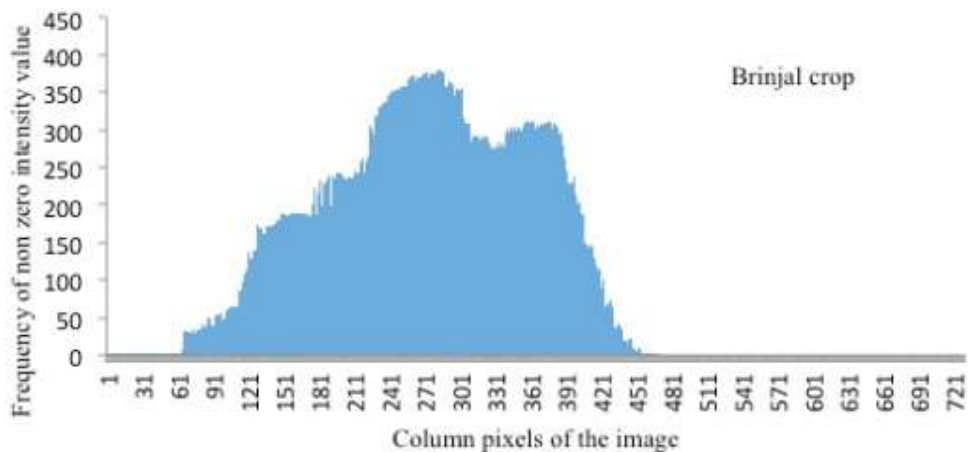
(b)

Figure 5. Frequency of non-zero intensity value distribution of tomato and eggplant.

(a) Frequency of non-zero intensity value distribution for tomato crop. (b) Frequency of non-zero intensity value distribution for eggplant.

For example, a frequency distribution of eggplant with *Cercospora* leaf spot is presented in Figure 6. The intensity value distribution of eggplant resulting from a disease shows near similar properties compared to the frequency distribution of healthy eggplant. These complex patterns of leaf arrangement and also the crops with foliage diseases must be classified for crop identification; hence, appropriate feature parameters must be selected for an effective
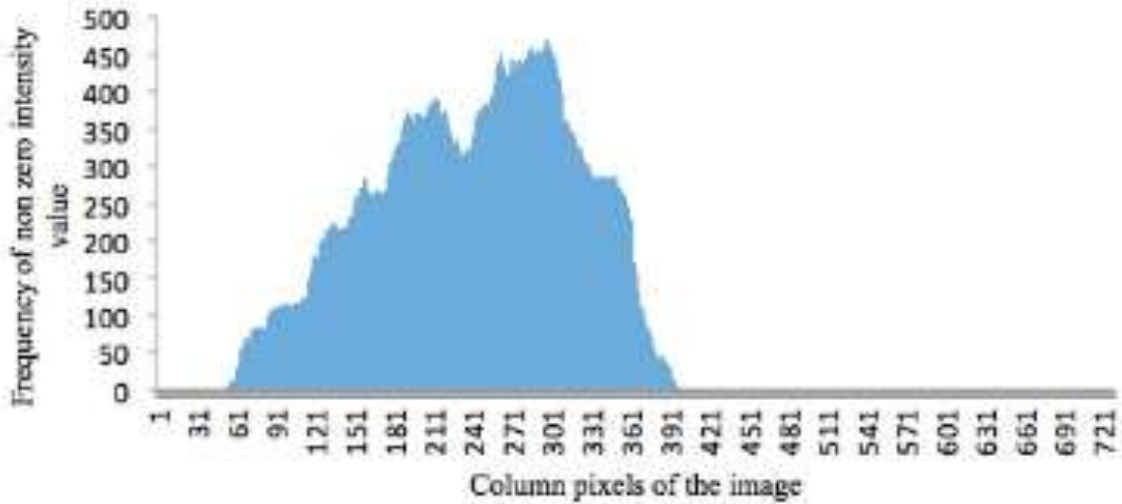
classification.

Figure 6. Gray-scale intensity value distribution of crop infected with *Cercospora* leaf spot.

Gray Level Cooccurrence Matrix (GLCM) based textural descriptors

GLCM based textural descriptors, namely, contrast, correlation, homogeneity, and energy, as shown in Table 2 (Gonzalez & Woods, 2012), are important feature parameters for pattern recognition. The contrast of an image with a random pattern tends to be higher when compared to an image with the regularized pattern. The arrangement of leaves in many crops has some regularized pattern unless some external factors such as the presence of an obstacle or other factor affect their growth. Each type of crop has a distinctive pattern, which affects the contrast parameter. The correlation parameter tends to be higher for the crop image with the repetitive pattern as it is sensitive to the mentioned patterns. Homogeneity descriptors calculate the closeness index of the pixel pairs in cooccurrence matrix of an image. The energy parameter increases for an image with constant intensity value over a large region.

S. No.	Texture analysis property	Objective	Formulae	Notation
1	Contrast	Determines contrast between a pixel and the specified neighborhood pixel	$$\sum_{i=1}^{K}\sum_{j=1}^{K}(i-j)^2 p_{ij}$$	p_{ij} - Probability of pixel pair i - Row index j - Column index
2	Correlation	Determines whether the pixel positively or negatively correlated with the specified neighborhood	$$\sum_{i=1}^{K}\sum_{j=1}^{K}\frac{(i-m_r)(j-m_c)p_{ij}}{\sigma_r \sigma_c}$$ $\sigma_r \neq 0;\ \sigma_c \neq 0$	m_r - Mean along the row m_c - Mean along the column

3	Homogeneity	Closeness value of elements in co-occurrence matrix with the diagonal element	$$\sum_{i=1}^{K}\sum_{j=1}^{K}\frac{p_{ij}}{1+|i-j|}$$	σ_r - Standard deviation along the row
4	Energy	Measure of equal gray level	$$\sum_{i=1}^{K}\sum_{j=1}^{K}p_{ij}^{2}$$	σ_c - Standard deviation along the column
				p_{ij} - Probability of pixel pair
				i - Row index
				j - Column index

Table 2. Textural descriptors.Histogram based textural descriptors consist of six different parameters, namely, average gray level, average contrast, measure of smoothness, third moment, measure of uniformity, and entropy, which are presented in Table 3 (Gonzalez and Woods, 2012).

Table 3. Textural descriptors for health status detection.

S. No.	Texture property	Objective	Formulae	Notation
1	Average gray level	Mean intensity value of image	$$\sum_{i=0}^{L-1} z_i p(z_i)$$	
2	Average contrast	Standard deviation of intensity values	$\sqrt{\sigma^2}$	
3	Measure of smoothness	0 for constant intensity image and 1 for wide variation in intensity	$$1-\frac{1}{(1+\sigma^2)}$$	z_i - Random variable for intensity
4	Third moment	Skewness index of histogram	$$\sum_{i=0}^{L-1}(z_i-m)^3 p(z_i)$$	$p(z_i)$- Histogram of intensity level
5	Measure of uniformity	Measure of equal gray level	$$\sum_{i=0}^{L-1} p^2(z_i)$$	

6	Entropy	Randomness index of an image	$$-\sum_{i=0}^{L-1} p(z_i) log_2 p(z_i)$$	L - Intensity value (0 – 255) σ - Standard deviation

These textural descriptors are based on the histogram of an image, and not based on the relative position. Average gray level and contrast denote the mean and standard deviation of the intensity values of an image, respectively. A measure of smoothness parameter explains distribution in terms of 0 (for constant intensity image) and 1 (for a wide variation in the intensity values). A measure of uniformity is an index for measuring the presence of equal gray level over an image. The third moment is the measure of the skewness of the histogram, whereas entropy is the measure of randomness.

In order to classify the disease, textural descriptors, namely, contrast, correlation, homogeneity, and energy, are used, whereas for the health status detection, textural descriptors like average gray level, average contrast, measure of smoothness, third moment, measure of uniformity, and entropy are selected based on the several trials. The classification methodology is described in Section "Selection and implementation of machine learning algorithm."

Implemented methodology for feature extraction

Initially, images were captured from the field using a mobile camera, and offline processing was performed in Matlab tool. Since weeds were present along with the crops, the object of interest was cropped using the cropping tool in Matlab. It is then placed as a foreground object on an image created with the background containing the uniform pattern of soil surface.

The images are processed for the extraction of textural descriptors in order to perform identification of crops (as shown in Figure 7) and classification of disease, which will be discussed later in this section. Preprocessing of an image involves the application of histogram equalization in three individual channels, that is, red, green, and blue. These histogram equalized (individual) channels of images are combined together to form an RGB image. The application of this equalization (to an image) broadens the intensity distribution over an entire range. This enhances the contrast, which differentiates the various regions of an image. Segmentation is done by using thresholding operation on an image I_0 as given in (1).

$$\begin{cases} I_0(i,j,R) = I_0(i,j,R) \\ I_0(i,j,G) = I_0(i,j,G) \\ I_0(i,j,B) = I_0(i,j,B) \end{cases} \quad \text{if}\big(I_0(i,j,G) > I_0(i,j,B)\big) \& \big(I_0(i,j,G) > I_0(i,j,R)\big)$$

$$\begin{cases} I_0(i,j,G) = 0 \\ I_0(i,j,G) = 0 \\ I_0(i,j,G) = 0 \end{cases} \quad \text{Otherwise}$$

$$(1)$$

The segmentation operation retains the pixels with green component higher than the blue and red components of an image. The segmented RGB image is converted into a gray-scale image (G_0) using (2) (Gonzalez & Woods, 2012).
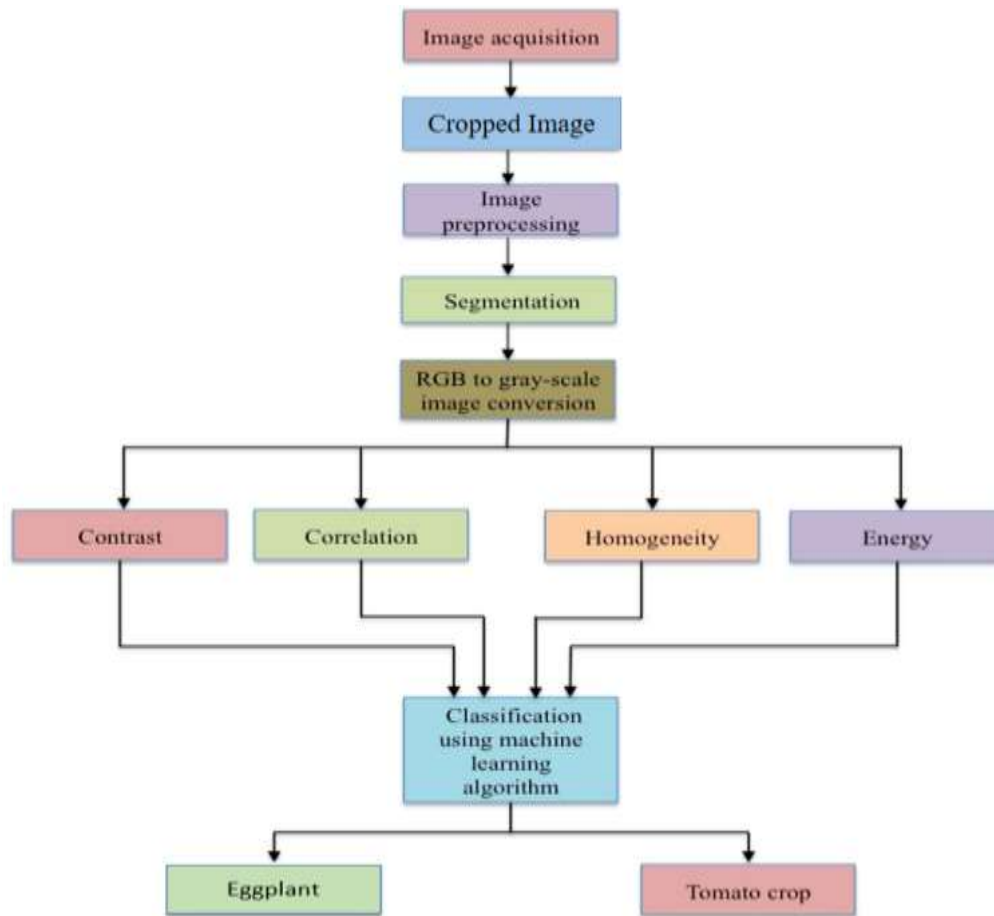
Figure 7. Steps of extraction of features for crop identification.

$$G_0 (i, j) = 0.2989R + 0.5870G + 0.1140B \qquad (2)$$

GLCM was estimated from the gray-scale image with an offset of 20 pixels along the rows of an image. It provides information regarding the number of times an intensity pair occurs with the pixel of interest and a pixel located 20 pixels apart along the row. The choice of 20 pixels for offset was selected, as it provided the best result upon numerous trials using different values.

As the gray-scale image has 256 intensity levels, the estimation of the cooccurrence matrix is computationally inefficient. Hence, the number of gray levels was reduced from 256 to 4 levels with each enclosing 64 levels in order to increase the computational efficiency. With an estimated compressed gray level matrix, co-occurrence matrix was calculated. A statistical approach was then used to estimate the textural descriptors namely contrast, correlation, energy and homogeneity with the co-occurrence matrix as explained in previous Section "Crop selection and dataset creation". These descriptors act as an input to the classification algorithm for classifying the crops. The process is also similar for health status detection but with histogram based textural parameters (like average gray level, average contrast, third moments, etc.,) are used as shown in Figure 8.
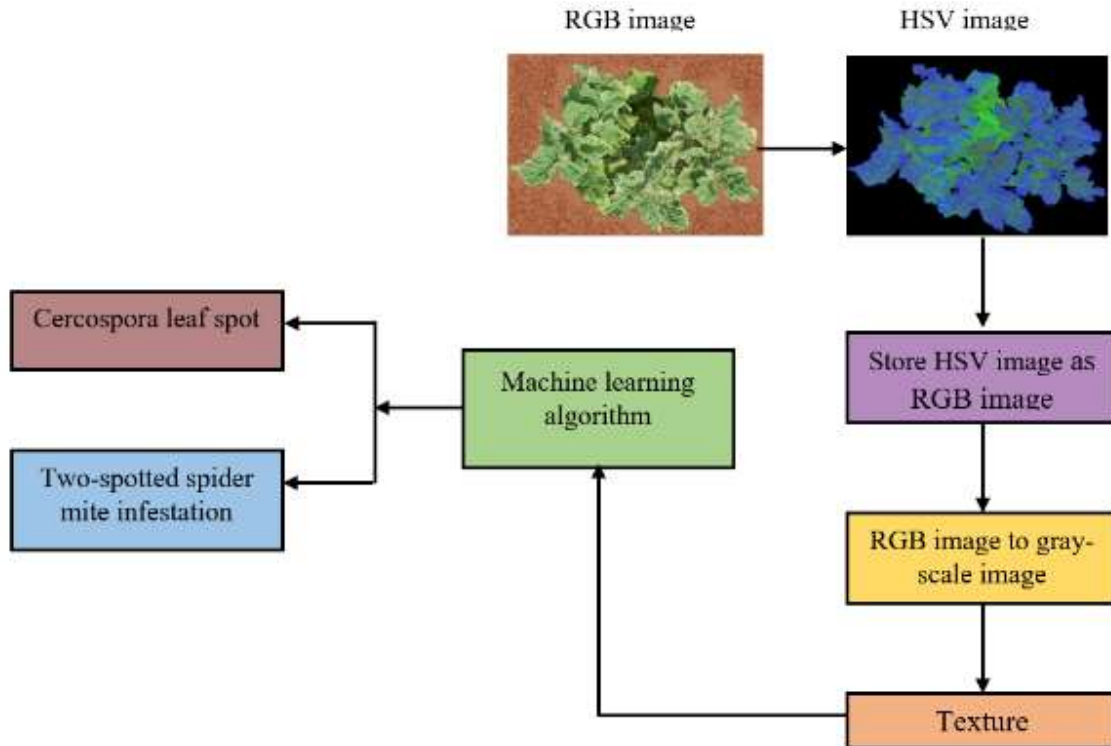
Figure 8. Textural descriptors for health status detection

In order to classify the disease present in the eggplant, the segmented RGB image is converted into Hue Saturation Value (HSV) image. The converted HSV image is stored in the memory as RGB image. The stored image is converted into a gray-scale image as shown in Figure 9. Then by using the 4 textural descriptors (i.e., contrast, correlation, energy and homogeneity) and by implementing the machine learning algorithms, the disease is classified into *Cercospora* leaf spot or two-spotted spider mite infestation.
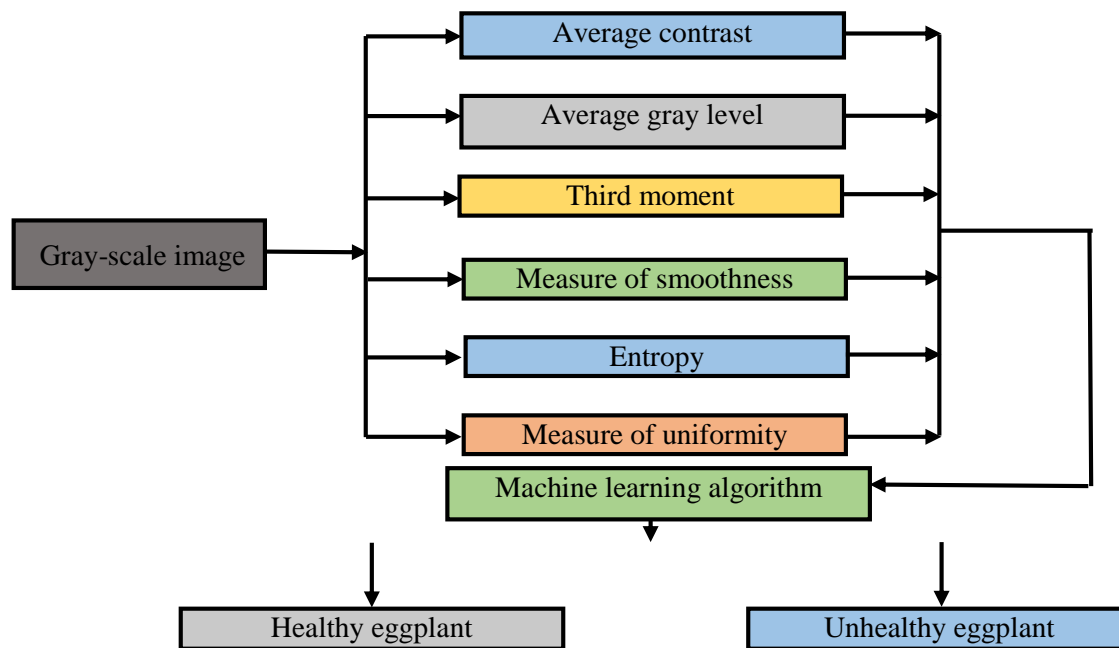
Figure 9. Disease classification methodology

Selection and implementation of machine learning algorithm

Five machine learning algorithms including a deep learning model (AlexNet) were implemented in Matlab as it has optimized execution time, availability of better data visualization tools, etc., for the classification tasks which will be discussed in this section. The class label for crop identification is tomato and eggplant while in the case of health status detection, the class label is either healthy or unhealthy. The class label for disease classification is *Cercospora* leaf spot or two-spotted spider mite infestation. Prior probabilities for each class is incorporated in all the three classification algorithms except neural network to improve the classification accuracy. A prior probability for a class is the ratio of number of dataset belonging to that class to the total number of a dataset. A 10 fold cross-validation is implemented where the dataset is randomly divided into 10 subsets. In each iteration, one of the 10 subsets will act as a test set while remaining will act as the training set.

Discriminant analysis

At first discriminant analysis (Guo et al., 2007) with a conditional probability is applied in the classification task with Bayes discriminant rule for classification (as shown in Figure 11) and the specifications are given in Table 4.
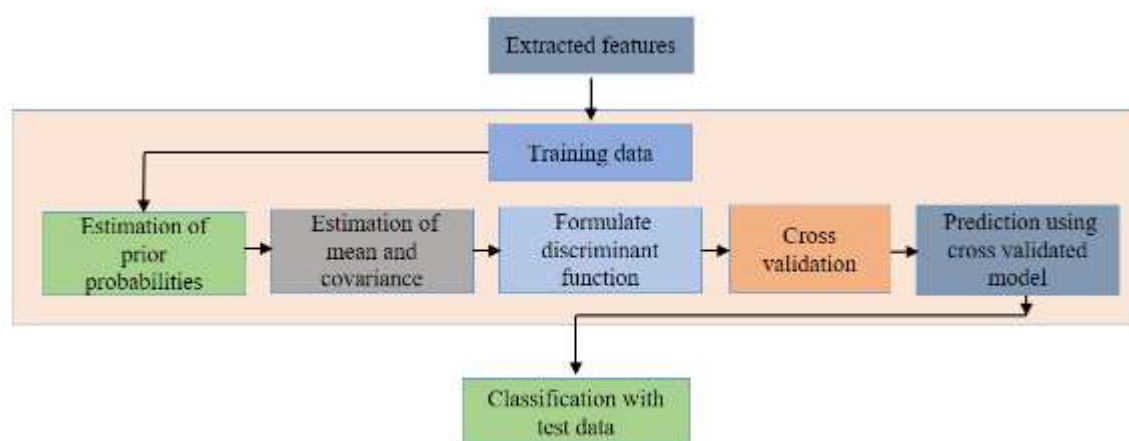

Figure 11. Discriminant analysis implemented for the classification task

Table 4. Specification for implementation of discriminant analysis

Parameters	classification
Splitting ratio	80:20
Prior probability	Crop (T - 0.64, E - 0.36) Unhealthy (H – 0.71, UH – 0.29) Disease (TSSM – 0.60, CLS – 0.40
Cross-validation	10 fold cross-validation
Discriminant type	Linear, diagonal linear, pseudo linear, quadratic, pseudo quadratic

T- Tomato, E-Eggplant, H- Healthy, UH – Unhealthy, CLS – *Cercospora* leaf spot, TSSM – Two Spotted Spider Mite

In general, for each classification task, the density function is estimated with the mean μ_k and covariance Σ_k of the textural parameter. Mean μ_k and co-variance Σ_k is estimated for each parameter from the training data. With the application of Bayes rule and classification cost for predicting the class labels of an observation, the predictor variable associates these observed parameters to a class such that the classification cost is minimum. The classification cost is the cost of classifying an observation to a class which may be its true or false class. The classification is 0 in the case of correct classification and 1 in the case of incorrect classification. Several variants of discriminant analysis are applied (as shown in Table 4) as each variant differs in the classification errors.

The discriminant variants applied are namely using linear, quadratic, pseudo-inverse linear, pseudo-inverse quadratic, diagonal linear and diagonal quadratic types (Paliwal & Sharma, 2012, Guo et al., 2007 & Regi et al., 2005). Linear methods are suitable when the model has same covariance matrix of textural parameter Σ_k for all the classes and quadratic methods for models with the heterogeneous covariance matrix. Pseudo-inverse linear and pseudo-inverse quadratic are used when the covariance matrix is singular (Guo et al., 2007). Diagonal linear and diagonal quadratic methods are used when the number of samples is low with unreliable input features (Regi et al., 2005). The performance for each classification task is explained in Section "Results".

Naive Bayes algorithm

The implementation of the Naive Bayes algorithm for this study is shown in Figure 12.
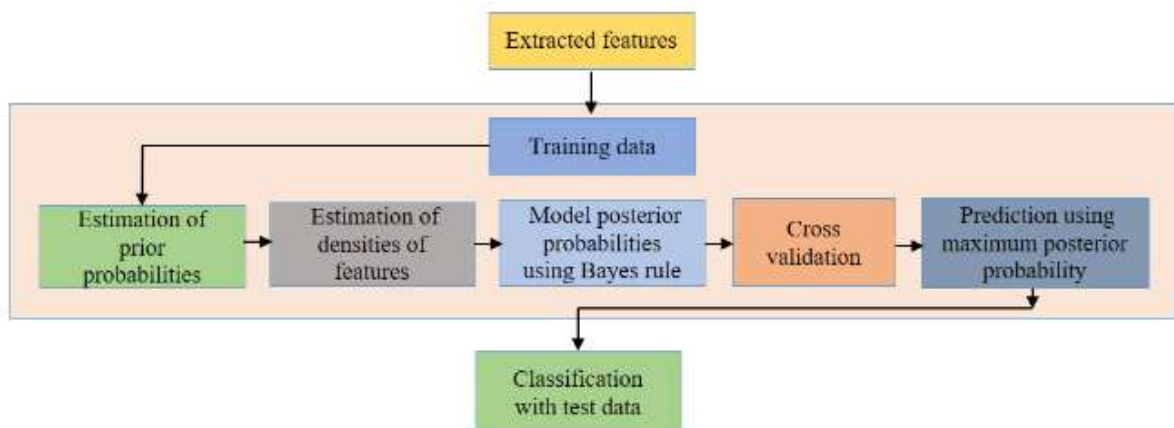

Figure 12. Naive Bayes algorithm implemented for the classification task

Naive Bayes uses Bayesian statistics with the Naive assumption stating that input feature variables are independent of each other. The class labels are same as it is used in the discriminant analysis for each classification task. The posterior probabilities of the class, given the input textural data, is calculated. 10 fold cross-validation and prior probabilities are applied similar to the discriminant analysis as shown in Table 5. Prediction of observed textural descriptors to a class is estimated by finding the maximum posterior probability implemented in the model.

Table 5. Specification for implementation of Naïve Bayes algorithm

Parameters	classification
Splitting ratio	80:20
Prior probability	Crop (T - 0.64, E - 0.36) Unhealthy (H – 0.71, UH – 0.29) Disease (TSSM – 0.60, CLS – 0.40
Distribution	Multivariate multinomial
Kernel	Gaussian
Cross-validation	10 fold cross-validation

Support Vector Machine (SVM)

The implementation of the support vector machine for this study is shown in Figure 13. The two algorithms discussed above is based on the Bayes rule and posterior probability. Contrastingly support vector machine classifies using decision boundary and hyperplane (Bishop, 2006). The input vector of textural descriptors from the observation is trained using the support vector machine. The different configurations for implementing the support vector machine is shown in Table 6.
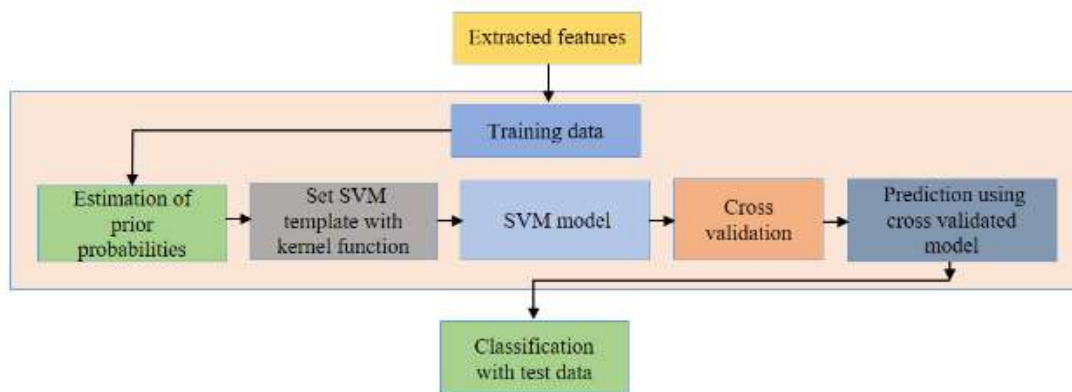

Figure 13 SVM implemented for the classification task

Table 6. Specification for implementation of Naïve Bayes algorithm

Parameters	classification
Splitting ratio	80:20
Prior probability	Crop (T - 0.64, E - 0.36) Unhealthy (H – 0.71, UH – 0.29) Disease (TSSM – 0.60, CLS – 0.40
Kernel function	Linear, polynomial and Gaussian
Cross-validation	10 fold cross-validation

When the input data of each class is linearly separable, a simple line or hyperplane is sufficient to separate each class. In the case of non-separable descriptors, it has to be converted to separable form using kernel function. The kernel function maps the input descriptors to kernel space which is a higher-dimensional space for better class prediction.

In this study, various kernel functions namely Gaussian, linear and polynomial are used and the accuracies are compared in Section "Results". The data points on the boundaries which form the support vectors are nearer to the separating hyperplane. The classification parameter determines the class label of the give textural input descriptors. If the value is less than 0, it belongs to one class, and if it is greater than 0, then it belongs to the other class. The solver used is sequential minimum optimization method. Similar to the previous methods 10 fold cross-validation and prior probabilities are applied to the algorithm.

Artificial neural network

Discriminant analysis and Naive Bayes classification algorithm are based on the underlying assumption of probability density function and the classification accuracy is bounded to this decision function. The performance of the support vector machine depends on the choice of the kernel function for non-separable input textural descriptors. Pattern recognition neural networks are independent of the decision function and able to classify non-separable classes (Gonzalez & Woods, 2012). The network used in this study (as shown in Figure 14) is a simple feed-forward network with the tan-sigmoid transfer function in each neuron (Dorofki et al., 2012).
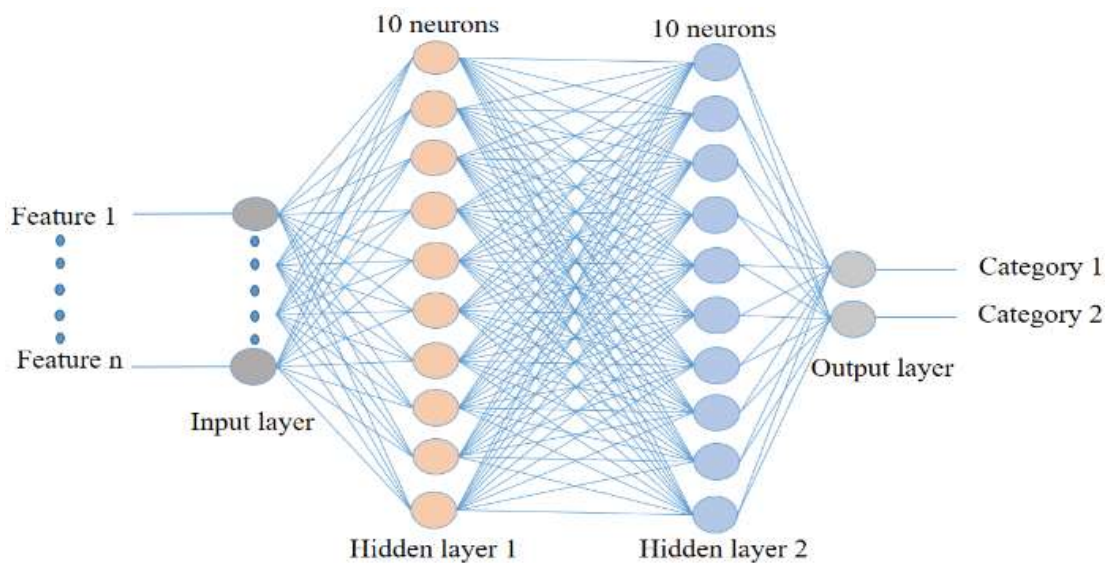

Figure 14. Neural network implementation for different classification task

Tan-sigmoid transfer function acts as a soft transition function where the value varies between -1 to +1 instead of high and low. Tan-sigmoid transfer function produces faster results compared to log sigmoid and purelin functions in each neuron (Dorofki et al., 2012). The network is initialized with random weights and bias for each input feature. The first step is to train a neural network with the given input features which updates weight and bias parameter. It optimizes the network for classification. The specification tuned for the implementation of the neural network is shown in Table 7.

Table 7. Specification for implementation of Neural network

Parameters	classification
Splitting ratio	80:20
Number of hidden layers	2
Number of neurons in hidden layers	10
Activation function	Tan-sigmoid
Training algorithm	Scaled conjugate gradient

The training algorithm used in this study is scaled conjugate gradient method which constructs a training rule to minimize the error between the desired output and the actual output. The selected training algorithm acts as a better classifier of patterns in the pattern recognition problem at a faster computational speed (Beale et al., 2015). The neural network has input layer with 4 inputs for crop identification and disease classification while health status detection has 6 inputs. The number of hidden layer and neuron in each layer is varied depending on the performance of the network. The output layer basically has 2 outputs for all the classification task. The trained network is provided with the test data for classification. The performance of the algorithm is discussed in Section "Results".

Deep convolutional neural network (AlexNet)

In this study, a pre-trained AlexNet model (as shown in Figure 15) in Matlab was used for crop identification, health status detection and disease classification. AlexNet is one of the deep network architectures and was successful in classifying 1000 categories of the object in the event ImageNet Large Scale Visual Recognition challenge (ILSVR -2010) surpassing the previous benchmark. The AlexNet pre-trained with 1.2 million images of the ImageNet database (Krizhevsky et al., 2012) was used for this study. The AlexNet consist of several convolution layers followed by Rectified Linear Unit (ReLU) with optional normalization and maxpooling layers. Finally, the architecture ends with three fully connected, softmax and a classification layer. The image dataset used for the different classification task with the previous four algorithms were used as the input to the AlexNet. In the case of the traditional machine learning algorithm, textural features were extracted from the images for the different classification task. Contrastingly AlexNet learns features automatically and utilizes these features for the classification task
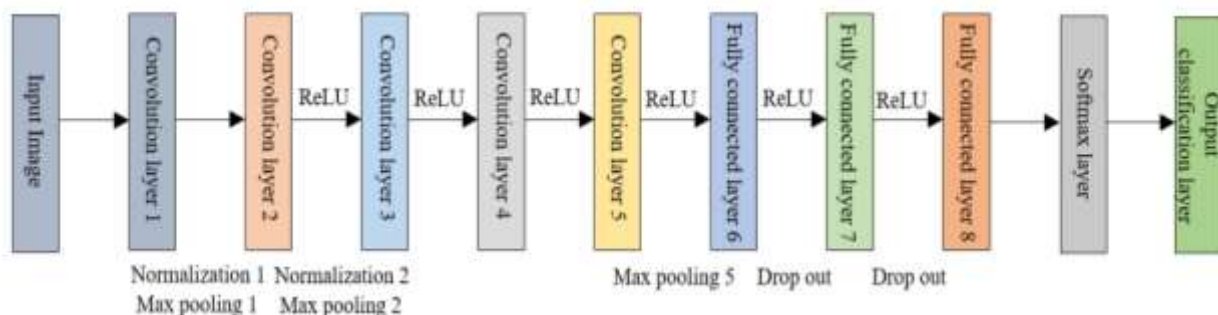

Figure 15. AlexNet deep learning architecture

Table 8.

Hyperparameters	Value
Splitting ratio	80:20
Initial learning rate	0.0001
Momentum	0.9
Learning rate of weight for fully connected layer 8	0.003
Weight L2 factor of fully connected layer 8	1
Learning rate of bias for fully connected layer 8	0.002
Bias L2 factor of fully connected layer 8	0
Minibatch size	32/64
Epoch	10-50

Tuned

hyperparameters for implementation of AlexNet

The hyperparameters setting for the model is tuned which is shown in Table 8. Fully connected layer 6 and 7 have 4096 neurons connected to each other followed by a drop out layer. The number of neurons in the last fully connected layer is based on the number of classes; where in this case is 2. The weight learning rates of all the previous layer are not modified except the weights in the last three layers. It is boosted to 30 times the global learning rate of the model to take large steps in the update of the weight parameters. The maximum epoch was varied from 10 to 50 epochs with minibatch size configured as 32 for training and 64 for testing.

RESULTS

The experiment was carried out with the augmented dataset of 1258 sample crops consisting of tomato and eggplant from the local field among which 576 are diseased. The analysis carried out with the developed dataset and different machine learning algorithm are provided in this section.

Crop identification

The extracted descriptors (i.e., contrast, correlation, homogeneity and energy) from the processed image for crop identification were evaluated for its significance in identification. From equation in Table 1, the average contrast value of tomato crop was found to be 0.47 which is greater than the eggplant where its value was 0.26. Similarly, the average correlation value (0.59) of tomato crop was less than the eggplant value (0.80) including the diseased crops. The mean correlation and contrast prove to be a potential descriptor for successful classification. The average energy and homogeneity descriptors for both the crop types do not show any significant difference.

Among the dataset of 1258 crop images, 20% of the samples were used for testing the classification model generated using the discriminant method. The classifier model was trained with the 4 textural descriptors which generate model rules for classifying the test data. Initially, the algorithm was tested without applying cross-validation approach. Several combinations of 4 textural descriptors which yield 6 combinations with classification boundaries based on the classifier model are shown in Figure 16. The data plots of contrast versus homogeneity and energy versus homogeneity show the least separation as depicted in Figure 16.

The accuracy of the algorithm with linear and pseudo linear discriminant method resulted in the least accuracy of 94.42%, while with diagonal linear, quadratic and pseudo quadratic resulted in the accuracy of 94.82%, 97.61% and 97.61% respectively. When the confusion matrix using linear and pseudo linear discriminant method were assessed, several images were misclassified as tomato where their true class was eggplant which resulted in lower accuracy. The re-substitution error was ranging from 0.027 to 0.049 for the different discriminant method. When the model was validated with 10 fold cross-validation method, there was no significant difference in the accuracy. Exclusion of prior probability resulted in an increase in cross-validation error as prior information about a class is not available. The lowest cross-validation error of 0.027 was obtained when quadratic and pseudo quadratic discriminant type were used. The re-substitution and cross-validation error remained low and approximately close to each other which shows that the model is generalizable and can be applied to new test data.

The second study consist of applying the textural descriptors in Naive Bayes algorithm for testing the classification feasibility. Mean and standard deviation for each textural variables are calculated based on the category of the class label with the Gaussian density function. Gaussian contours are plotted based on the input descriptors as shown in Figure 17. According to the above figure, the contrast versus correlation, correlation versus energy and correlation versus homogeneity shows a clear distinction for tomato and eggplant. These above features have a significant influence on discriminating eggplant from tomato.
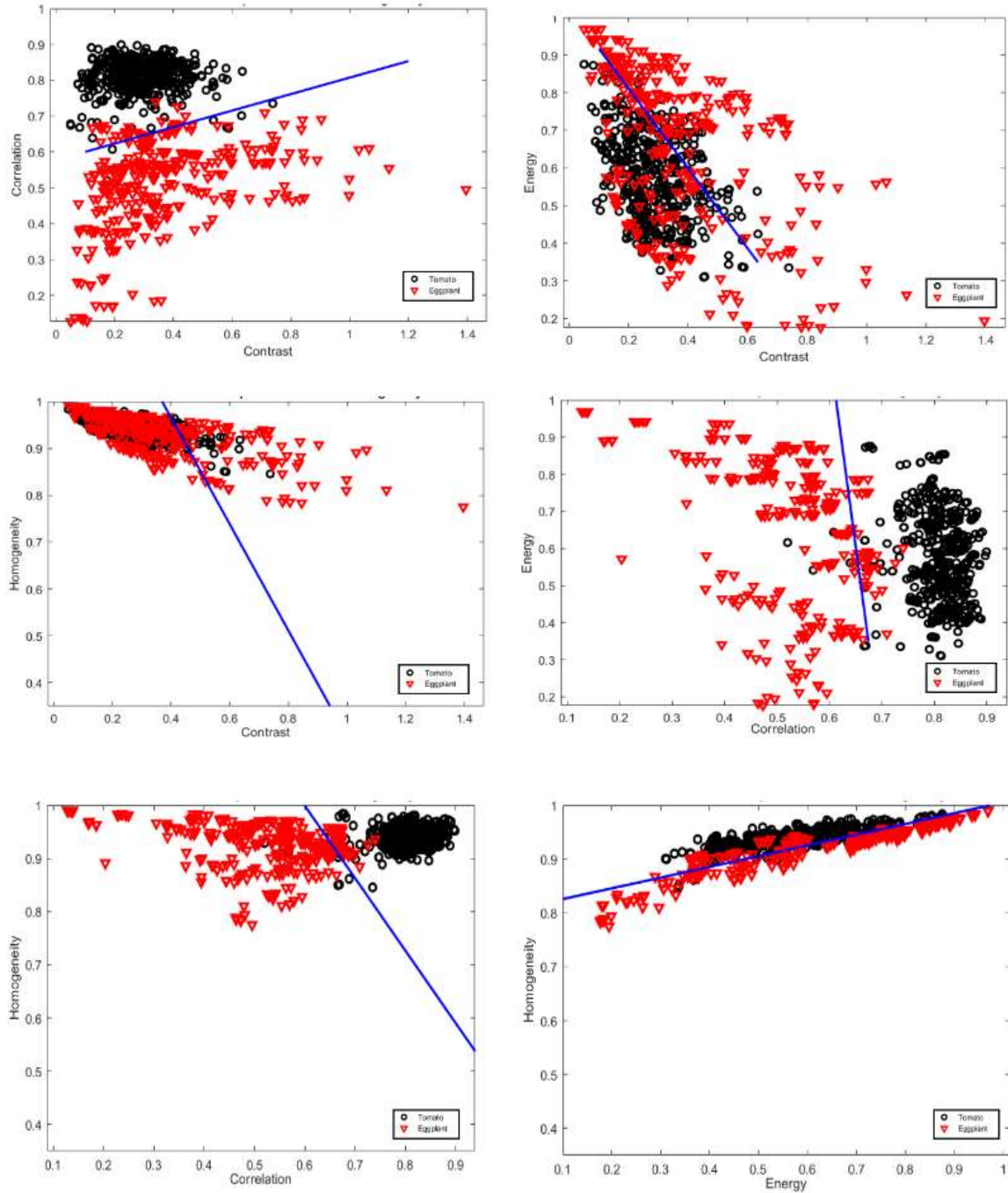

Figure 16. Sample classification plot for crop identification using textural descriptors

The algorithm without application of 10 fold cross-validation resulted in an accuracy of 95.62% with a re-substitution error of 0.24. In contrast to the confusion matrix of discriminant analysis, some of the tomato crops were misclassified as eggplant. When cross-validation, prior probabilities were introduced in the model, there was no change in the accuracy.
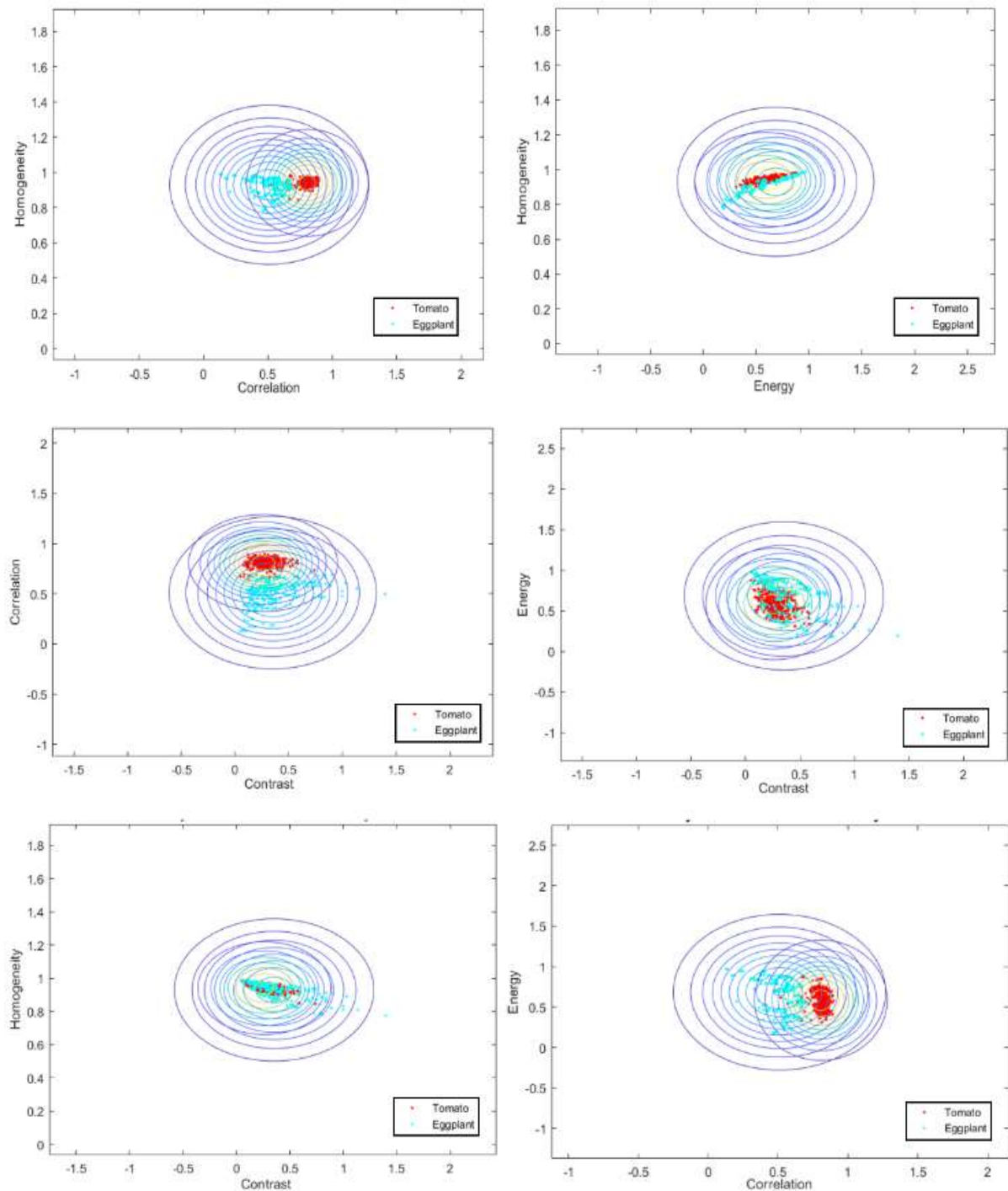

Figure 17. Demo scattered plot with contours generated using Naive Bayes algorithm

Similarly, the textural descriptors were given to support vector machine with 1006 crop sample for training and 252 sample observation for testing. The non-cross validated classifier accuracy with linear, polynomial and Gaussian kernel function were 80.87%, 86.96%, 84.35% respectively. The accuracy did not change when 10 fold cross-validation was performed and the cross-validation error was minimum. This shows that the performance of the support vector machine was not affected with different kernel function. The re-substitution errors were 0.010 (Gaussian), 0.014 (linear) and 0.010 (polynomial) which is lower than the discriminant analysis and Naive Bayes

algorithm. Although discriminant method yielded approximately similar accuracy, the lower re-substitution error suggests that support vector machine is better for the classification of new test data.

The textural descriptors for crop identification were tested with a multi-layer neural network. A two-layer neural network with 10 neurons in each hidden layer was used for recognizing the pattern in the training data of textural descriptors. The input data were randomly divided into training and test data. 1006 crop sample data were used for training and validation while 252 samples were used for testing. Initially, the classification accuracy was tested by increasing the number of neurons in the single layer which yielded poor results. The number of the hidden layers was increased to 2 layers with 10 neurons in each layer which yielded the best result beyond which the performance of the network decreased. The attributed reason for the increase in performance is due to the ability of the neural network to learn tiny details with the fine-tuned architecture. The network performance decreases as more neurons and layers are added as other additional fine change in features are not significant for classification. The neural network was able to classify with 98.94% accuracy at 38th epoch with best validation result as shown in Figure 18. The network stops training when the validation performance decreases for 6 consecutive epochs hence the training ceases at 44th epoch. The gradient was also minimum at 38th epoch. When confusion matrix was examined, 2 images of tomato crop were misclassified as eggplant.

Finally, the image dataset was trained and validated with AlexNet architecture. The resized images of dimension 227 x 227 were given as input to the network. The training was carried out for 10 epochs and convergence occurred within the 1st epoch and 10 iterations. It resulted in the best validation accuracy of 100% which is greater than all the above algorithm. The reported result is limited to the images in the test dataset. The obtained accuracy may vary when images with variation in background, illumination and other anomalies occur. Several other factors are discussed in the discussion section.
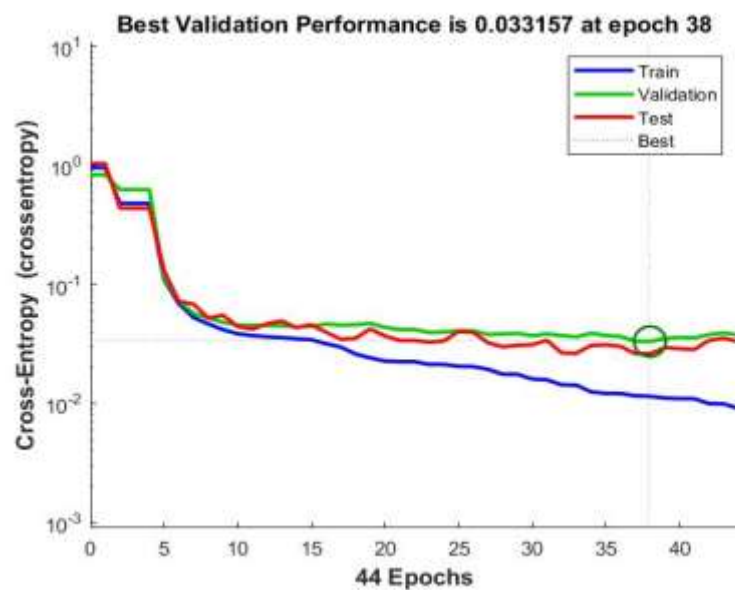

Figure 18. Performance plot of crop identification using neural network

Health status detection and disease classification in eggplant

As crop identification was successfully performed using the above algorithms, the second task involves health status detection. The reason for using 6 textural descriptors was the fact that all the algorithms resulted in poor classification accuracy with 4 parameters. The machine learning algorithms used 80% of the input data to train the model and 20% as test data. The prior probabilities for health status detection were set as 0.29 for healthy class and 0.71 for an unhealthy class which is based on the input data. The best result of 66.67% with a re-substitution error of 0.27 was obtained using linear and pseudo linear discriminant method. The quadratic and pseudo quadratic method also resulted in an equivalent accuracy of 66.05%. The re-substitution error was higher compared to crop identification task as misclassification rate in test data is higher and hence generalized model will result in lower accuracy with the new test data. With 10 fold cross-validation, quadratic and pseudo quadratic type yielded the accuracy of 61.11% with an error of 0.24. Naive Bayes algorithm resulted in a poor accuracy of 37.04% with cross-validation error of 0.24 for prediction of the test data associated with the class label. Similarly support vector

machine resulted in an accuracy of 50% with polynomial kernel function. When confusion matrix for above algorithms was analysed, unhealthy class was majorly affecting the classification accuracy. The cross-validation error for health status detection was also higher compared to crop identification.

In the case of neural networks, the number of hidden layers and neurons in each layer were same as the network implemented for crop identification. The neural network obtains the best classification accuracy of 72.9% which is better than the previous two algorithms. Several trials were performed by increasing the number of neurons and layers. These variations did not improve the accuracy and in several cases, the accuracy dropped to a lower value. AlexNet resulted in the best classification accuracy of 89.36%. In the case of AlexNet, unhealthy class was misclassified as healthy class and reported an accuracy of 78.72%. The increase in number of epoch or change in minibatch size did not improve the accuracy. The accuracies obtained for all the classification task and different algorithms are shown in Table 9.

Classification	No of textural parameter	Accuracy					Table 9.
		Discriminant analysis	Naive Bayes method	Support vector machine	Neural network	AlexNet	
Crop identification	4	97.61%	95.62%	98.01%	98.94%	100%	
Health status detection	6	66.67%	37.04%	50%	72.9%	89.36%	
Disease classification	4	86.09%	76.52%	86.96%	86.04%	81.51%	
Average accuracy of classification	-	83.45%	69.72%	78.32%	85.96%	90.29%	

Accuracy of machine learning algorithms for various classification problems

In the case of disease classification, 576 crop sample data including 115 test sets were used to evaluate the performance of the algorithm. The test accuracy with support vector machine was 86.96% which is the best accuracy among the 5 algorithms. In all the cases, the disease class namely *Cercospora* leaf spot was affecting the classification accuracy. Surprisingly, AlexNet which is a deep learning model showed a lower accuracy in the case of disease classification. Several trials were also conducted with various number of epochs, minibatch size, learning rate, etc. Despite these variations there was no significant alteration in the performance of the AlexNet. The attributed reasons and factors affecting the accuracy of the above classification task will be discussed in the next section.

DISCUSSION

According to Table 4, the average classification accuracy (90.29%) using AlexNet was higher compared to the other machine learning-based method despite the limited dataset. In the case of crop species classification, most of the literature reported a method of using isolated samples and acquisition of image in laboratory-controlled condition (Lee et al., 2017, Lukic et al., 2017, Dyrmann et al., 2016, Aakif & Khan, 2015, Kalyoncu & Toygar, 2015 & Zhao et al., 2015). Despite the development in deep learning approaches, studies were carried out using the isolated leaf images (Lee et al., 2017). In our study, texture features were used in the case of four shallow machine learning algorithm and a deep learning architecture namely AlexNet for identification of crop species with the images of entire plant. All the above algorithm reported an accuracy greater than 95%. The presence of weed in the images was not evaluated as the entire plant was manually segmented and placed on the background with red soil. The intensity values were varied randomly (within certain limit) while applying augmentation process which reflects changes in illumination property to some extent. Also the other factors such as the presence of other soil color, plants with multiple disorder were not evaluated. The presence of other factors will have a significant effect on the extracted textural features. Although deep learning model can counteract these effects, if the background has strong features or occupies significant portion of the images, it may influence the classification accuracy (Barbedo, 2018a).

In the case of health status detection, the similarity of textural features resulted in lower accuracy. From the results, it is evident that the deep learning model, AlexNet was able to discriminate with higher accuracy compared to the shallow machine learning algorithms. As the unhealthy crop may have the symptoms of these two diseases, the textural features vary for these diseases. Hence the machine learning model will have difficulty in the learning the features of these unhealthy crop class which is evident from the classification accuracy. As AlexNet is a deeper architecture and has the ability to learn the complex features, the performance was better for health status detection. Similar to crop species identification, presence of weed, variation in sunlight will significantly affect the accuracy.

Similar to the crop classification, many literature reported classification of disease with an accuracy of greater than 98% (Barbedo 2018b, Ferentinos, K.P., 2018 & Mohanty et al., 2016) using the isolate leaf samples which is difficult to be implemented in a robotic system. Also, such a system fails to identify correctly with the images from field (Ferentinos, K.P., 2018 & Mohanty et al., 2016). Hence in our study, images of entire plant were used for evaluating the disease classification. In this evaluation, the estimated average values of the textural features showed a clear distinction between the two diseases. The accuracy with all the algorithm was higher compared to the health status detection. AlexNet reported a lower accuracy compared to other machine learning model as it requires more images for the *Cercospora* leaf spot.

As automation systems are increasingly being introduced in the agricultural process, automatic identification of crop species will be beneficial. As many diseases are possible in different crops, species identification will narrow down the classification of the disease in a particular crop.

CONCLUSION

Many studies discussed in Section "Introduction" shows that crop identification and disease classification have been done independently. In this paper, a combined task namely, crop identification (namely eggplant and tomato), health status detection and disease classification in eggplant (namely *Cercospora* leaf spot and two-spotted spider infestation) have been performed. All classification mentioned above was implemented using machine learning algorithms namely linear discriminant analysis, Naive Bayes algorithm, support vector machine and neural network with the textural descriptors as the input parameters. In addition to that a deep learning architecture namely AlexNet was used in this study. A statistical textural descriptor with relative location information was explored for classification of crop type and disease, while histogram-based descriptors were experimented for health status detection.

It has been found that the performance of AlexNet was better than the other machine learning algorithm which yielded combined average classification accuracy of 90.29%. This accuracy may change with the addition of new images in the database for all the algorithms in the testing phase. Further evaluation is needed by increasing the number of training and test data with varying environmental condition. The above-discussed method is suitable for implementation in an automated system (like a robot) using a visible camera for the evaluation of crop with reasonable accuracy. This provides an opportunity for real-time identification of diseases in the crops. Another scope of the study will be using multi and hyperspectral heads in order to better identify and analyze the crop foliage diseases (like little leaf disease, sooty mold, thrips infestation, etc.,) using the deep convolutional neural network.

REFERENCES

Aakif, A., & Khan, M.F. 2015. Automatic Classification of plants based on their leaves. Biosystems Engineering. 139: 66-75. https://doi.org/10.1016/j.biosystemseng.2015.08.003

Ampatzidis, Y., Bellis, L.D., & Luvisi, A. 2017. iPathology: Robotic applications and management of plants and plant diseases. Sustainability. 9(6): Article ID 1010. https://doi.org/10.3390/su9061010

Anders, M.M., Potdar, M.V., & Francis, C.A. 1996. Significance of intercropping in cropping systems. International Research Center for Agricultural Sciences, Tokyo.

Arivazhagan, S., Shebiah, R.N., Ananthi, S., & Varthini, S.V. 2013. Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features, CIGR Journal. 15(1): 211-217.

Barbedo, J.G.A. 2016a. A review on the main challenges in automatic plant disease identification based on visible range images. Biosystems Engineering. 144: 52-60. http://doi.org/10.1016/j.biosystemseng.2016.01.017

Barbedo, J.G.A., Koenigkan, L.V., & Santos, T.T. 2016b. Identifying multiple plant diseases using digital image processing. Biosystems Engineering. 147: 104-116. http://doi.org/10.1016/j.biosystemseng.2016.03.012

Barbedo, J.G.A. 2018a. Factors influencing the use of deep learning for plant disease recognition. Biosystems Engineering. 172: 84-91. https://doi.org/10.1016/j.biosystemseng.2018.05.013

Barbedo, J.G.A. 2018b. Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. Computer and Electronics in Agriculture 153: 46-53. https://doi.org/10.1016/j.compag.2018.08.013

Brahimi, M., Boukhalfa, K., & Moussaoui, A. 2017. Deep learning for tomato diseases: classification and symptoms visualization. Applied Artificial Intelligence. 31: 299-315.

Bashish, D.A., Braik, M., & Ahmad, S.B. 2011. Detection and classification of leaf diseases using K-means based segmentation and neural networks based classification. Information Technology Journal. 10(2): 267-275. https://doi.org/10.3923/itj.2011.267.275

Beale, M.H., Hagan, M.T., & Demuth, H.B. 2015. Neural network toolbox™ user's guide. The MathWorks, Inc, Natick, Massachusetts.

Bishop, C.M. 2006. Pattern Recognition and Machine Learning, Springer, New York.

Chaki, J., & Parekh, R. 2012. Plant leaf recognition using Gabor filter. International Journal of Computer Applications. 56(10): 26-29. https://doi.org/10.5120/8927-3000

Chaudhary, P., Chaudhari, A.K., Cheeran, A.N., & Godara, S. 2012. Color transform based approach for disease spot detection on plant leaf. International Journal of Computer Science and Telecommunications. 3(6): 65-70.

Dorofki, M., Elshafie, A.H., Jaafar, O., Karim, O.A., & Mastura, S. 2012. Comparison of artificial neural network transfer function to simulate extreme runoff data. International Conference on Environment, Energy and Biotechnology, Singapore, pp. 39-44.

Dyrmann, M., Karstoft, H., & Midtiby, H.S. 2016. Plant species classification using deep convolutional neural network. Biosystems Engineering 151: 72-80. https://doi.org/10.1016/j.biosystemseng.2016.08.024

Ehrmann, J., & Ritz, K. 2014. Plant: soil interactions in temperate multi-cropping production systems. Plant and Soil. 376(1-2): 1-29. https://doi.org/10.1007/s11104-013-1921-8

Fang, H., & Li. H. 2014. Plant leaves recognition and classification model based on images features and neural network. International Journal Computer Science. 11(2): 100-104.

Ferentinos, K.P. 2018. Deep learning models for plant disease detection and diagnosis. Computer and Electronics in Agriculture 145: 311-318. https://doi.org/10.1016/j.compag.2018.01.009

Fuentes, A., Yoon, S., Kim, S.C., & Park, D.S. 2017. A Robust deep – learning – based detector for real-time tomato plant diseases and pests recognition. Sensors. 17(9): s17092022.

Gonzalez, R.C., & Woods, R.E. 2012. Digital Image Processing, Third ed., Dorling Kindersley Private Limited, Noida.

Guo, Y., Hastie, T., & Tibshirani, R. 2007. Regularized linear discriminant analysis and its application in microarrays. Biostatistics. 8(1): 86-100. https://doi.org/10.1093/biostatistics/kxj035

Hussain, S.A., Ali, N., Rab, A., & Shah, M. 2008. Yield and economic dynamics of intercropping in summer vegetables. Sarhad Journal Agriculture. 24(1): 31-36.

Johannes, A., Picon, A., Gila, A.A., Echazarra, J., Vaamonde, S.R., Navajas, A.D., & Barredo, A.O. 2017. Automatic plant disease diagnosis using mobile capture devices, applied on a wheat use case. Computers and Electronics in Agriculture. 138: 200-209.

Kalyoncu, C., & Toygar, O. 2015. Geometric leaf classification. Computer Vision and Image Understanding 133: 102-109. https://doi.org/10.1016/j.cviu.2014.11.001

Krizhevsky, A., Sutskever, I., & Hinton, G.E. 2012. Imagenet classification with deep convolutional neural network. Advances in Neural Information Processing Systems, Lake Tahoe, United States of America, pp.1097-1105.

Lee, S.H., Chan, C.S., Mayo, S.J., & Remagnino, P. 2017. How deep learning extracts and learns leaf features for plant classification. Pattern Recognition. 71: 1-13. https://doi.org/10.1016/j.patcog.2017.05.015

Liu, B., Zhang, Y., He, D.J., & Li, Y. 2018. Identification of apple leaf diseases based on deep convolutional neural networks. Symmetry. 10(11). https://doi.org/10.3390/sym10010011

Lukic, M., Tuba, E., & Tuba, M. 2017. Leaf recognition algorithm using support vector machine with Hu moments and local binary patterns. Proceedings of IEEE International Symposium on Applied Machine Intelligence and Informatics, Herl'any, Slovakia, pp. 486-490. https://doi.org/10.1109/SAMI.2017.7880358

Mohanty, S.P., Hughes, D.P., & Salathe, M. 2016. Using deep learning for image-based plant disease detection. Frontiers in Plant Science. Article ID 1419. https://doi.org/10.3389/fpls.2016.01419

Nap, J.P., Perez, G.F.S., & Dijk, A.D.J.V. 2017. Similarities between plant traits based on their connection to underlying gene functions. PLoS ONE. 12(8): Article ID e0182097.

Olsen, A., Han, S., & Calvert, B. 2015. In situ leaf classification using histograms of oriented gradients. IEEE International Conference on Digital Image Computing: Techniques and Applications, Adelaide, Australia. https://doi.org/10.1109/DICTA.2015.7371274

Pahikkala, T., Kari, K., Mattila, H., Lepisto, A., Teuhola, J., Nevalainen, O.S., & Tyystjari, E., 2015. Classification of plant species from images of overlapping leaves. Computers and Electronics in Agriculture. 118: 186-192. https://doi.org/10.1016/j.compag.2015.09.003

Paliwal, K.K., & Sharma, A. 2012. Improved pseudoinverse linear discriminant analysis method for dimensionality reduction. International Journal Pattern Recognition and Artificial Intelligence 26(1), Article ID 1250002.

Parikh, A., Raval, M.S., Parmar, C., & Chaudhary, S. 2016. Disease detection and severity estimation in cotton plant from unconstrained images. IEEE International Conference on Data Science and Advance Analytics, Montreal, Canada, pp. 594-601.

Pilli, S.K., Nallathambi, B., George, S.J., & Diwanji, V, 2014. eAGROBOT – A robot for early crop disease detection using image processing. IEEE International conference on Electronics and Communication Systems, Coimbatore, India, pp. 1-6.

Polder, G., Heijden, G.W.A.M.V.D., Doorn, J.V., & Baltissen, T.A.H.M.C. 2014. Automatic detection of tulip breaking virus in tulip fields using machine vision. Biosystems Engineering. 117: 35-42. https://doi.org/10.1016/j.biosystemseng.2013.05.010

Pujari, J.D., Yakkundimath, R., & Byadgi, A.S. 2014. Identification and classification of fungal disease affected on agriculture/horticultural Crops using image processing techniques. IEEE International Conference on Computational Intelligence and Computing Research, Coimbatore, India. https://doi.org/10.1109/ICCIC.2014.7238283

Rangaswami, G., & Mahadevan, A. 2012. Diseases of Crop Plants in India, fourth ed. PHI Learning Private Limited, New Delhi.

Raza, S.A., Prince, G., Clarkson, J.P., & Rajpoot, N.M. 2015. Automatic Detection of Diseased Tomato Plants Using Thermal and Stereo Visible Light Images. PLoS ONE. 10(4): Article ID e0123262. https://doi.org/10.1371/journal.pone.0123262

Regi, R.P., Ortega, A., & Asgharzadeh, S. 2005. Sequential diagonal linear discriminant analysis (SeqDLDA) for microarray classification and gene identification. IEEE Computational Systems Bioinformatics Conference Workshops, Stanford, USA, pp. 112-113.

Sano, H. 2010. Inheritance of Acquired Traits in Plants. Plant Signal. Behav. 5(4): 346-348.

Schor, N., Berman, S., Dombrovsky, A., Elad, Y., Ignat, T., & Bechar, A. 2017. Development of a robotic detection system for greenhouse pepper plant diseases. Precision Agriculture 18: 394-409. https://doi.org/10.1007/s11119-017-9503-z

Srinivasan, R. 2009. Insect and mite pests on egg plants: A field guide for identification and management, AVRDC Publication No 09-729, Shanhua.

Sun, Y., Liu, Y., Wang, G., & Zhang, H. 2017. Deep learning for plant identification in natural environment. Computational Intelligence Neuroscience. Article ID 7361042.

Vanitha, S.M., Chaurasia, S.N.S., Singh, P.M., & Naik, S.P. 2013. Vegetable statistics. Technical Bulletin No.51. Indian Institute of Vegetable Research, Indian Council Agricultural Research, Varanasi, India.

Wang, Z.G., Jin, X., Bao, X.G., Li, X.F., Zhao, J.H., Sun, J.H., Christie, P., & Li, L. 2014. Intercropping enhances productivity and maintains the most soil fertility properties relative to sole cropping. PLoS ONE. 9(12): Article ID e113984. https://doi.org/10.1371/jounral.pone.0113984

Wspanialy, P., & Moussa, M. 2016. Early powdery mildew detection system for application in greenhouse automation. Computers and Electronics in Agriculture. 127: 487-494. https://doi.org/10.1016/j.compag.2016.06.027

Zhao, C., Chan, S.S.F., Cham, W.K., & Chu, L.M. 2015. Plant identification using leaf-shapes – a pattern counting approach. Pattern Recognition. 48: 3203-3215. https://doi.org/10.1016/j.patcog.2015.04.004