https://doi.org/10.56093/ijas.v90i2.98996

Indian Journal of Agricultural Sciences **90** (2): 249–57, February 2020/Review Article

Plant disease identification using Deep Learning: A review

SAPNA NIGAM[1] AND RAJNI JAIN[2]

ICAR-Indian Agricultural Statistics Research Institute, New Delhi 110 012, India

Received: 03 June 2019; Accepted: 18 July 2019

ABSTRACT

The paper reviews various classification techniques exclusively used for plant disease identification. Early stage plant disease identification is extremely important as that can adversely affect both quality and quantity of crops in agriculture. For identification of plant diseases, different approaches like image processing, machine learning, artificial neural networks, and deep learning are in use. This review focusses on an in-depth analysis on recently emerging deep learning-based methods starting from machine learning techniques. The paper highlights the crop diseases they focus on, the models employed, sources of data used and overall performance according to the performance metrics employed by each paper for plant disease identification. Review findings indicate that Deep Learning provides the highest accuracy, outperforming existing commonly used disease identification techniques and the main factors that affect the performance of deep learning-based tools. This paper is an attempt to document all such approaches for increasing performance accuracy and minimizing response time in the identification of plant diseases. The authors also present the attempts for disease diagnosis in Indian conditions using real dataset.

Key words: Image processing, Machine Learning, Plant disease identification

In the agriculture sector, plant diseases are responsible for major economic food losses across the globe. Food losses due to crop infections from pathogens such as bacteria, viruses, and fungi are persistent issues. The situation further gets complex by the fact that, nowadays, diseases are transferred globally more easily than ever before. In order to minimize the disease induced damage in crops during growth, prevention in crops are imperative. Traditionally, crop inspection and plant disorders were identified by farmers or experts with some training or experience. This manual method was expensive as it requires continuous monitoring and was not feasible for the larger fields. Due to complexity and variation in a large number of cultivated plant diseases, even experienced agronomists and plant pathologists fail to diagnose specific diseases accurately. It is also worth noting that many agricultural areas are too difficult to be properly monitored throughout (Barbedo 2013). Sankaran *et al.* (2010) recognized the need for developing a rapid, cost-effective, and reliable health monitoring disease detection technologies for advancements in agriculture. Automated image-based tools are needed for identification of plant diseases when human assessment is inappropriate, unreliable or unavailable (Mohanty *et al.* 2016, Yang and Guo 2017). They can be combined with different methods of image pre-processing in favour of better feature extraction (Hiary *et al.* 2011, Barbedo 2013, Kulkarni and Patil 2012). With this backdrop, the present study reviews the techniques for plant disease identification and classification. The main emphasis is on reviewing machine learning and deep learning techniques which has the potential to improve the accuracy of disease detection and diagnosis.

The bibliographic analysis in the domain involved two steps: (a) collection of relevant literature and (b) detailed review and analysis of the work. In the first step, a keyword-based search for conference papers or journal articles was performed from the scientific databases IEEE Xplore and Science Direct, and from the web scientific indexing services Google Scholar. The search criteria used for the purpose was - [“deep learning”] AND [“plant disease identification” OR [“plant disease classification “]. More than 20 papers had been initially identified. In the second step, the papers were analysed one-by-one, referring the related citations and considering the following research questions in mind:

1. Which was the crop disease problems they addressed?
2. Which machine learning or deep learning based models are employed?
3. What kind of data is used?
4. How is the pre-processing applied?
5. What is the performance level of the technique used in the selected papers?

Steps for disease identification

The basic steps for plant disease identification mainly

[1]Ph D Scholar (sapna.nigam1010@gmail.com), ICAR-Indian Agricultural Statistics Research Institute, New Delhi; [2]Principal Scientist (rajni.jain@icar.gov.in), ICAR-National Institute for Agricultural Economics and Policy Research, New Delhi.

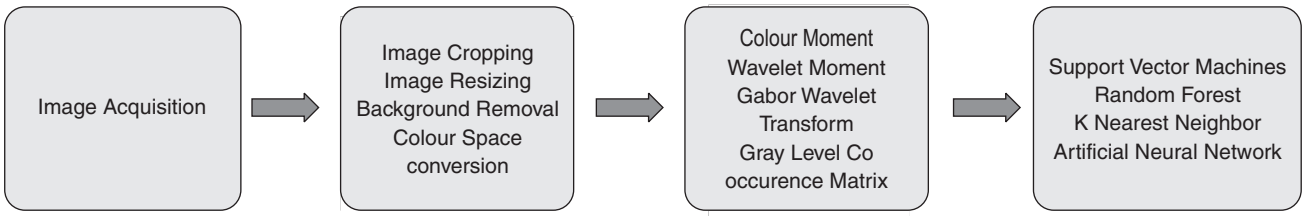

Fig 1 Four phases of plant disease identification using a leaf image.

involve four phases that are image acquisition, image pre-processing, feature extraction, and classification (Akhtar *et al.* 2013) as shown in Fig 1. The steps are further explained

Image acquisition: It is the process in which acquired images are converted to the desired output format for further processing. Images may be self-acquired by authors or may be any benchmarking dataset such as Plant Village database (Hughes and Salathe 2015).

Image pre-processing: The procedure of image pre-processing aims at highlighting the region of interest (disease infected area) in plant leaves (Hanson *et al.* 2017). Image pre-processing commonly involves image segmentation (Liu and Zhou 2009, Husin *et al.* 2012, Yao *et al.* 2009, Liu *et al.* 2017), image enhancement and colour space conversion. Firstly, digital image is enhanced by filter (Al Hiary *et al.* 2011, Mokhtar *et al.* 2015, Prasad *et al.* 2016, Sannakki *et al.* 2013, Semary *et al.* 2015). The leaf image is filtered from the background image and RGB colours are converted into colour space parameter (Dandawate and Kokare 2015, Le *et al.* 2015). Further image is segmented to a meaningful part which is easier to analyse. Unfortunately, removing background is quite difficult, and sometimes needs the intervention of the user, which decreases the automation of the system (Le *et al.* 2015).

Feature extraction: Features are extracted from the image for constructing feature vectors. This extraction could be any of statistical, structural or signal processing. For example, colour moments are used to extract colour statistics (Semary *et al.* 2015) , in which Gabor Transform (GT) and Wavelet Transform (WT) are combined (GWT) for the extraction of multiscale features (Prasad *et al.* 2016). Gray Level Co-occurrence Matrix (GLCM) is used in many previous works (Mokhtar *et al.* 2015, Prasad *et al.* 2016, Semary *et al.* 2015, Xie and He 2016, Xie *et al.* 2015) to extract texture features. GLCM is a 256*256 matrix where each position in the matrix counts the co-occurrences of line colour and column colour in the analysed image (Dandawate and Kokare 2015). Scale Invariant Feature Transform (SIFT) is used to analyse the shape features of leaves. The advantage of deep learning is automatic feature extraction which ultimately holds a good contribution in higher accuracy as compared to other conventional techniques (Kamilaris and Boldú 2018, Amara *et al.* 2017, Sladojevic *et al.* 2016, Brahimi *et al.* 2017, Cruz *et al.* 2017, Ferentinos 2018, Liu *et al.* 2018, Mohanty *et al.* 2016).

Classification: The last phase identifies the plant disease existing in leaf using a classification model. The model should be trained using learning algorithms and examples with known disease images. Classification algorithms are explained in following section.

Techniques for disease identification: Techniques for disease identification can be broadly classified into two types image processing based techniques and machine learning techniques. For the purpose of disease identification, image processing techniques are necessarily followed by some machine learning methods which can perform on large datasets. On the other hand, machine learning methods can work on both image based dataset as well as textual attribute based data which does not require image. For disease identification using attribute based tables, one need not use image processing techniques but other data cleaning and pre-processing should be followed.

Image processing techniques: Image processing techniques were widely and successfully used for accurate detection and classification of the plant. Egmont *et al.* (2002) categorized various applications for image processing algorithms. The various pre-processing techniques such as image clipping, image smoothing, image enhancement is carried out for increasing the efficient detection of diseases (Khirade and Patil 2015). Image segmentation can be done using various methods like Otsu' method, k-means clustering, converting the RGB image into HIS model etc. A methodology for early and accurate plant diseases detection using diverse image processing techniques such as Gabor filter has been used for feature extraction and classification with identification accuracy up to 91% has been achieved through ANN based classifier (Kulkarni and Patil 2012). Husin *et al.* (2012) captured the chilli plant leaf image and processed to determine the health status of the chilli plant to ensure chemicals should be applied to the diseased chilli plant only. Similar techniques like Sobel and canny filter has been used to identify the edges that help in identifying disease spots (Revathi and Hemalatha 2012). They proposed a homogeneous pixel counting technique for cotton diseases detection algorithm and claims the accuracy of 98.1% over existing algorithms. Jaware *et al.* (2012) proposed a novel and improved k-means clustering technique to solve low-level image segmentation. Statistical texture features were extracted using Spatial grey-level dependence matrices method (Dhaygude and Kumbhar 2013).

Machine learning: Machine learning focuses on algorithms capable of learning on their own from a given set of input data according to the objective. Its high-performance computing creates new opportunities in the

agriculture domain.

Machine learning and statistical pattern recognition have been the subject of tremendous interest in the agriculture domain because they offer promise for improving the sensitivity of disease detection and diagnosis (Jain *et al.* 2005, Sajda 2006). To the best of the available literature, Jain *et al.* (2005) introduced the concept of machine learning techniques for disease diagnosis in agriculture domain for the first time (Jain *et al.* 2005). Later, Upadhyaya *et al.* (2006) explored the concept using clustering approach of machine learning for identification of soybean disease using bench-marking dataset available at UCI repository. The data set in this approach was not from India. Further, Jain *et al.* (2009) investigated and reported the potential of three machine learning models based on Decision Tree (DT) induction using C4.5, Rough Set (RS) and hybridized rough set based decision tree induction (RDT) in comparison to traditional logistic regression method for the dataset collected in India during real time condition. A DT is a classification scheme which generates a tree and a set of rules representing the model of different classes from a given dataset. A java implementation of C4.5, called CJP by the authors, was used for DT induction. A variant of RDT called RJP, combined merits of both RS and DT induction algorithms. Powdery Mildew of Mango (PWM) is a devastating disease and has assumed a serious threat to mango production in India resulting in yield losses of 22.3% to 90.4%. As a case study, prediction models for forewarning PWM disease using variables, viz. temperature and humidity were developed. The results obtained from machine learning techniques, viz. RS, CJP and RJP were compared with the traditional prediction model using LR technique. The authors were astonished to discover that machine learning models showed better performance over traditional approach. Many other machine learning studies for sugar beet diseases, brown spot and the leaf blast diseases in rice, wheat plant, banana, beans, jackfruit, lemon are summarised in Table 1.

The Support Vector Machine (SVM) (Mokhtar *et al.* 2015, Semaryetal. 2015, Dandawate and Kokare 2015, Rumpf *et al.* 2010, Schikora and Schikora, 2014), k-nearest neighbors (KNN) (Prasad *et al.* 2016 and He 2016), and artificial neural network (ANN) (Bashish *et al.* 2010, Hiary *et al.* 2011, Sannakki *et al.* 2013, Xie *et al.* 2015, Schikora and Schikora 2014) represent the most commonly used learning algorithms in the literature reviewed. The SVM algorithm maximizes the margin between classes in linearly separable cases. The KNN algorithm classifies an image by voting between the K closest examples in the features space.

As the paper aims for review of an important machine learning technique called Deep Learning (DL), it is appropriate to discuss neural networks (an ancestor of DL) and DL machine learning techniques separately before the comparison is made among different disease identification techniques.

Neural networks: Neural networks, with their outstanding ability to derive meaning from complex data, can be applied for extracting patterns and detecting patterns that are too difficult to be observed by human brain or computer techniques. Other advantages of ANNs are adaptive learning, self-organization, real-time operations, and so forth. Table 1 shows the studies which used ANN for disease identification, respective crops and their performance.

Back propagation neural network (BPNN) has been used for efficient grape leaf colour extraction with a complex background for diagnostic system of grape leaf diseases (Menukaewjinda *et al.* 2008). They explored a modified self-organizing feature map (MSOFM) and genetic algorithm (GA) and found that these techniques provide automatic adjustment in parameters for grape leaf disease colour extraction. Similar method was used for classifying the healthy and diseased part of rice leaves (Liu and Zhou 2009).

The ANN is a model organized in layers, in which each layer is connected to the next one starting from the input to output. ANN represents the old version of deep learning algorithms used in this paper. In the present paper, we emphasise the use of deep learning and specifically Convolutional Neural Network (CNN) as an alternative approach for building a model of disease identification.

Deep learning: Deep learning means deep neural networks and learns hierarchical representations of a data with multiple levels of abstraction (Schmidhuber 2015). One of the most powerful and basic DL tools for modelling complex processes and performing pattern recognition is Convolutional Neural Networks (CNNs). CNN provides mapping between an input such as an image of a diseased plant to an output, i.e. crop disease (Mohanty *et al.* 2016).

A CNN is composed of three main layers namely convolution, pooling and fully connected layers (Kamilaris and Boldú 2018). The essential purpose of convolution is to extract features automatically from each input image (Fig 2). It consists of a set of learnable filters. Each filter is applied to the raw pixel values of the image in a sliding window fashion, computing the dot product between the filter pixel and the input pixel. This results into a two-dimensional activation map of the filter called the feature map. Hence, the network learns filters (i.e. edges, curves) that will activate when they find known features in the input. The CNN learns the values of these filters on its own during the training process. These convolution layers are followed by sub-sampling layers. Each sub-sampling layer reduces the size of the convolution maps, and introduces invariance to low rotations and translations that can appear in the input. The output of pooling layer is given by the maximum activation value in the input layer over sub windows within each feature map and hence reducing the dimensionality of the feature. At the end of the model, the fully connected layer is based on the SoftMax activation function for computing the classes scores. The input of the SoftMax classifier is a vector of features resulting from the learning process and the output is a probability that an image belongs to a predefined class. The different types of CNN architectures include LeNet (LeCun *et al.* 2015),

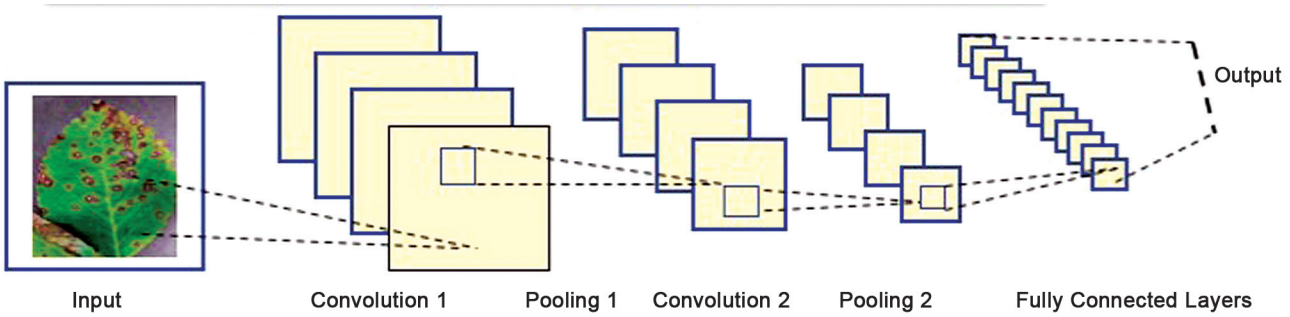

Fig 2 Convolution Neural Network Architecture.

AlexNet (Krizhevsky *et al.* 2012), GoogleNet (Szegedy *et al.* 2015), VGGNet (Simonyan and Zisserman 2014), Inception- ResNet (Szegedy *et al.* 2017).

The application of deep learning to plant pathology and specifically on leaf image classification and plant disease identification has started to gain momentum in recent years. In this approach, features are extracted automatically and learned from the data in the training phase. To confirm the superiority of deep models against state-of-the-art methods, we have reviewed both small and large dataset studies containing 500 to 87848 images (Amara *et al.* 2017, Fuentes *et al.* 2017, Ferentinos 2018, Brahimi *et al.* 2017, Lu *et al.* 2017, Zhang *et al.* 2017, Liu *et al.* 2017, DeChant *et al.* 2017, Oppenheim and Shani, 2017, Arivazhagan *et al.* 2018, Cruz *et al.* 2017). Literature review shows that deep learning provides high accuracy and outperforms commonly used image processing techniques (Kamilaris and Boldú 2018). Other methods based on conventional machine learning techniques and image-processing techniques are successful under limited and constrained setups only. Pawara *et al.* (2017) compared the performance of some conventional pattern recognition techniques with that of CNN models, in plants identification, using three different databases concluding that CNNs outperforms conventional methods.

Nigam *et al.* (2019) developed a Convolutional neural network model to perform plant disease identification using wheat crop images of healthy and yellow rust infected leaves, through deep learning. Training of the models was performed with the use of own created database of 2000 images clicked at the experimental field of ICAR-Indian Agricultural Research Institute. CNN architecture based model achieved the 97.37% testing accuracy for classification of yellow rust infected leaf and healthy leaves. The application to the wheat disease identification shows that the proposed CNN model can correctly and effectively recognize yellow rust infected plants. The model was trained on MacBook air having operating system macOS Mojave with a 4 GB RAM and 1.6ghz Intel core i5 processor. The total parameters trained during the model training were 1212513 parameters. The significantly high success rate makes the model suitable for early warning tool in image based disease identification. Different Experiments were carried out to observe the effect of hyperparameters on the accuracy such as variation in number of epochs, batch size, optimization algorithm and others. The authors identified the potentials of deep learning methods such as high performance and accuracy, automatic feature extraction, faster processing of test data, and its suitability for mobile app development. They also observed its limitations namely: (i) larger datasets requirement, (ii) comprehensive training data covering diverse features for better computational accuracy, (iii) time consuming for training model, (iv) real time images must be similar in shape, size and location of the disease to images used for model building and (iv) high-end hardware requirement.

Usually, hardware requirement is quite a challenge in deep learning. It needs high computational machines for faster and training of model. Deep Learning algorithms were mostly implemented on the GPU of an NVIDIA® GTX1080 card, using the CUDA® parallel programming platform, in Linux environment (Ubuntu operating system). Caffe, Keras, Tensorflow, Theano and Deeplearning4j are the popular software for disease identification using CNN architecture models. But, setting of parameters and hyper parameters is crucial and important step in using deep learning. The process of setting the hyper-parameters requires expertise and extensive trial and error. There are no simple and easy ways to set hyperparameters specifically, learning rate, batch size, momentum, and weight decay.

Comparison of deep learning studies on disease identification

In agriculture, plant diseases cause substantial economic and environmental losses, thus careful and expert monitoring is must for early detection followed by consequent application of control measures in order to improve the quality and quantity of crop yield. The use of deep learning includes improvements in performance and high computational accuracy as per the reviewed work shown in Table 1 and Table 2. Some factors impacting the performance of CNNs for plant disease recognition are limited annotated datasets, symptom representation, covariate shift, image background, image capture conditions, symptom segmentation, multiple simultaneous disorders, symptom variations, disorders with similar symptoms and others (Barbedo 2016, 2018).

In real-life conditions, the systems developed through

deep learning has high levels of performance and precision in detecting and diagnosing specific diseases. It could even be operated through an appropriate, easy-to-use, and user-friendly mobile application for detection of multiple diseases in plants (Johannes *et al.* 2017, Picon *et al.* 2018). DL techniques are able to achieve accuracy between 90-99% (Table 1-2). MATLAB is mostly commonly used software for image pre-processing techniques. Some comparative features with reference to crops, accuracy, historical evolution and software for disease identification are presented below.

Crops: Most of the studies (50%) are based on the plant disease identification for horticultural crops while 27.3% for cereals and 22.7% for other crops. This may be due to availability of the open public datasets like PlantVillage Dataset having maximum images related to horticultural crops. This further shows the need for independent development of image repository for other crop categories to promote the machine learning or deep learning based research for disease identification in these crops. Reason for a very few studies dealing cereal crops is probably due to the long-time requirement (may be months) for image capturing of diseases in cereal crops. Thus, to meet the food security of the growing population, for a country like India, more research is needed for disease identification in cereals crops (Nigam *et al.* 2019).

Performance: Comparison of performance of deep learning algorithms with other algorithms reveal that deep learning algorithms have capability to achieve accuracy up to 99.8% while other algorithms show comparatively lower accuracy for disease detection in plants (Table 2). It is observed that SVM and other related algorithms show accuracy ranging between 79.5 to 97%. However, time requirement is higher for model building in deep learning algorithms (Nigam *et al.* 2019).

Evolution: Although disease detection through machine learning has begun in the beginning of 21st century, yet deep learning based disease detection studies are observed after 2015 (Jain *et al.* 2005, Jain *et al.* 2009, Upadhyaya 2006). Further, this type of work gained momentum since 2016 (Table 1).

Software: With increasing focus on deep learning studies, many software are prevailing in the industry for automatic disease identification. Caffe, Tensorflow, frameworks are used along with popular python libraries such as Keras. MATLAB is used for the efficient pre-processing of images. Most recently, deep learning libraries are introduced in R software too. Caffe is the most popular among deep learning researchers probably because of its ease of use.

Dataset: Most of the studies (68%) are based on the data collected on their own by authors (Table 1). The authors who have done image acquisition on their own in field or experimental setup clicked pictures by the digital cameras or phone camera has a large variation and that leads to better accuracy. However, 32% of the studies use PlantVillage dataset available in public domain. The size of the dataset varies from 107 images to 87848 images in the reviewed studies. It is observed that authors have selected the images from the PlantVillage dataset based on their requirements of the crops. Hence, the size of the PlantVillage dataset is different in different studies.

Location of the studies: Most of the studies for crop disease identification are located outside India. As per meta-analysis, out of 10% crops involved in the study of disease identification are from Indian agriculture (Hanson *et al.* 2017) on the other hand more than 40% are from China and other developed countries (Lu *et al.* 2017, Zhang *et al.* 2017, Liu *et al.* 2017, Arivazhagan and Ligi 2018, DeChant *et al.* 2017, Oppenheim and Shani 2017, Mohanty *et al.* 2016, Amara *et al.* 2017). This also accounts for the need for using Artificial Intelligence for disease identification of crops in India with the main emphasis on cereal crops. Thus, the status of research on disease identification in India is low and the main reason for such less application in India could be lack of expertise in DL applications. Efforts are needed to harness the potential of this promising technology for Indian agriculture.

Architecture: As discussed above, CNN is a special kind of multi-layer neural networks, designed to recognize visual patterns directly from pixel images with minimal pre-processing. The ImageNet project is a large visual database designed for use in visual object recognition software research. As a result, many competitors have developed software to implement CNN architectures. Literature review suggests that studies for disease detection by deep learning differs in using one or the other in network. At surface level, there are tons of new architectures that are different from each other. Upon closer inspection, most of them are reapplying well established principles. Universal principles seem to be having shorter sub paths through the networks. Any of these architectures may be used for disease detection by using either Tensorflow or Caffe platform. Error rates are declining as more and more advancement are taking place in CNN architectures.

Parameters: Deep learning comprises of parameters and hyperparameters. Parameters are the configuration variables whose value can be estimated or learned from the data whereas hyperparameters are the variables which determine the network structure. They are decided before the training and determine how the network is trained. Hyperparameters related to a network structure are described below.

Batch and batch size: Total number of training examples present in a single batch. Batch size is the number of sub samples given to the network after which parameter update happens. A good default for batch size might be 32**.**

Number of epochs: Number of epochs is the number of times the whole training data is shown to the network while training. The number of epochs are increased until the validation accuracy starts decreasing even though the training accuracy is increasing (overfitting).

Number of hidden layers and units: Hidden layers are the layers between input layer and output layer. Layers can be added till the test error improve.

Table 1 Studies employing deep learning for plant disease

Crops	Algorithm, software and accuracy	Dataset with number of images	Study
Wheat	Quadratic discriminating model, Spectral Reflectance; (96%)	Own (120)	Moshou *et al.* (2003)
Soybean	Rough set based clustering machine	Benchmarking data (307)	Upadhyaya *et al.* (2006)
Grape	Self-organizing feature map, SVM; (97.80%)	Own (1478)	Meunkaewjinda *et al.* (2008)
Rice	Back Propagation Neural Network (90%)	Own (400)	Liu and Zhou (2009)
Mango	Decision tree, rough sets, hybridised decision tree	Own (time series data of 15 years)	Jain *et al.* (2009)
Rice	Image processing techniques and SVM (97.20%)	Own (216)	Yao *et al.* (2009)
Sugar beet	SVMand spectral vegetation indices; (92.46%)	Own	Rumpf *et al.* (2010)
Apple, blueberry, grape, corn, peach. soybean and others.	Back Propagation Algorithm; MATLAB; (93%)	Own	Bashish *et al.* (2010)
Rice	Bayes' and SVM Classifier (79.5% and 68.1%)	Own (>2000)	Phadikar *et al.* (2012)
Chilli	Image processing techniques; MATLAB	Own (107)	Husin *et al.* (2012)
Apple, cherry, corn, grape, orange, peach, bell pepper, potato, squash, strawberry, tomato	AlexNet, GoogleNet; Caffe; (99.35%)	Plant Village (54,306)	Mohanty *et al.* (2016)
Pear, cherry, peach, apple, pair, grapevine	Caffenet; Caffe; (96.30%)	Own (4483)	Sladojevic *et al.* (2016)
Banana	LeNet architecture; deep learning4j; 96%	PlantVillage (3700)	Amara *et al.* (2017)
Tomato	Deep CNN, Faster R-CNN; VGG-16 (83%)	Own (5000)	Fuentes *et al.* (2017)
Tomato	CNN; DIGITS; (99.18%)	PlantVillage (14,828)	Brahimi *et al.* (2017)
Rice	CNN; MATLAB; (95.48%.)	Own (500)	Lu *et al.* (2017)
Maize	Cifar10 and GoogLeNet; Caffe; (98.8% and 98.9%)	PlantVillage and Google (500)	Zhang *et al.* (2017)
Apple	AlexNet; Caffe; (97.62%)	Own (13,689)	Liu *et al.* (2017)
Maize	CNN; Keras and Theano; (96.7%)	Own (1796)	DeChant *et al.* (2017)
Potato	VGG; Matlab; (95.85%)	Own (400)	Oppenheim and Shani (2017)
Olive	Modified LeNet; MATLAB; (99%)	Own (299)	Cruz *et al.* (2017)
Cassava	CNN; Tensorflow; (95%-98%)	Own (11,670)	Ramcharan *et al.* (2017)
Different crops	CNN; (96.3%)	Internet (30880 images)	Hanson *et al.* (2017)
Mango	CNN; (96.67%)	Own (1200)	Arivazhagan *et al.* (2018)
Apple, banana, blueberry, cabbage, cassava, gourd, cucumber, eggplant etc.	CNN Models; Torch7; 99.53% through VGG	PlantVillage (87,848)	Ferentinos (2018)
Tomato	CNN; (99.84%)	PlantVillage (9000)	Ashqar and Naser (2019)
Wheat	Deep learning (98%)	Own dataset (2000)	Nikam *et al.* (2019)

Table 2 Summary of performance of different plant disease identification approaches

Algorithms	Number of research papers	Literature reference (from Table 1)	Range	Average
SVM and Others	3	11, 10, 8	(79.5-97.2) %	88.35%
Back propagation neural networks	2	7, 9	(90-93) %	91.5%
Image Processing	3	13, 12, 9	(90-97.20) %	93.6%
CNN	16	1, 2, 3, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26	(95-99.84) %	97.42%
Deep learning	2	1,27	(96-98)%	97%

Activation function: Activation functions are used to introduce non-linearity to models, which allows deep learning models to learn nonlinear prediction boundaries. SoftMax is used in the output layer while making multi-class predictions.

Learning rate: The learning rate defines how quickly a network updates its parameters. Usually a decaying learning rate is most preferred.

Momentum: It helps to know the direction of the next step with the knowledge of the previous steps. It helps to prevent oscillations. A typical choice of momentum is between 0.5 to 0.9.

The present paper reviews some of the non-invasive techniques which have been used for plant disease identification. We have identified relevant papers and examined the diseases, crops models employed, sources of data and overall performance according to the performance metrics employed. Findings from the review indicates that the Deep Learning outperforms other popular image processing techniques among various approaches of disease identification and classification such as machine learning, neural networks, image processing techniques. DL techniques shows higher accuracy and use images directly as input for automatic feature extraction in disease identification. This paper reviewed machine learning and deep learning-based research efforts applied for plant disease identification and classification. It is observed that applications of deep learning for disease identification or in general for other agricultural uses are still in nascent stage but offers a huge potential. Lack of benchmarking images in agricultural domain is a major constraint in boosting the research in this domain. Hence, agricultural experts should come forward in developing such standardized image repository of plant diseases. Most of the studies either use plant disease dataset available at (https://www.kaggle.com/emmarex/plantdisease) or they develop their own set of images for disease identification. However, most of these data are not available in public domain. Thus, there is a need for common repository of images in agriculture domain just like ImageNet database for other domains.

CNN is identified as the suitable and powerful architecture for image-based disease identification. Various implementations like GoogleNet, AlexNet, LeNet, ResNet etc. are available in public domain for this architecture. However, efforts to harvest these implementations are very limited in public institutions because of lack of awareness and training facilities. Thus, there is a need for creating awareness and expertise in this important technique at student as well as faculty level. Further, inclusion of this important technique in curriculum of computer application courses at various levels should be initiated. The authors are hopeful that this survey would motivate more researchers to experiment with deep learning, applying it for solving various agricultural problems involving classification or prediction as the overall benefits of deep learning are encouraging for its further use towards smarter, sustainable farming and more secure food production.

REFERENCES

Akhtar A, Khanum A, Khan S A and Shaukat A. 2013. Automated Plant Disease Analysis: Performance comparison of machine learning techniques, (*In*) *Eleventh International Conference on Frontiers of Information Technology*, IEEE, pp 60–65.

Amara J, Bouaziz B and Algergawy A. 2017. A deep learning-based approach for banana leaf diseases classification. (In) Proceedings of Datenbanksysteme für Business, Technologie und Web (BTW 2017) – Workshop Bonn: Gesellschaft für Informatik.

Arivazhagan S, Shebia RN, Ananthi S and Varthini S V. 2013. Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. *Agricultural Engineering International: CIGR Journal* **15**(1): 211–17.

Arivazhagan S and Ligi S V. 2018. Mango Leaf Diseases Identification Using Convolutional Neural Network. *International Journal of Pure and Applied Mathematics* **120**(6): 11067–79.

Ashqar B A and Abu-Naser S S. 2019. Image-based tomato leaves diseases detection using deep learning. *International Journal of Academic Engineering Research* **2**(12): 10–16.

Barbedo J G A. 2013. Digital image processing techniques for detecting, quantifying and classifying plant diseases. *Springer Plus* **2**(1): 660.

Barbedo J G A. 2016. A review on the main challenges in automatic plant disease identification based on visible range images. *Biosystems Engineering***144**: 52–60.

Barbedo J G A. 2018. Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. *Computers and Electronics in Agriculture* **153**: 46–53.

Bashish A D, Braik M, and Bani-Ahmad S. 2010. A framework for detection and classification of plant leaf and stem diseases. (*In*) *Signal and Image Processing International Conference*, IEEE.

Brahimi M, Boukhalfa K and Moussaoui A. 2017. Deep learning for tomato diseases: classification and symptoms visualization. *Applied Artificial Intelligence* **31**(4): 299–315.

Cruz A C, Luvisi A, De Bellis L and Ampatzidis Y. 2017. X-FIDO: An effective application for detecting olive quick decline syndrome with deep learning and data fusion. *Frontiers in plant science* **8**: 1741.

Dandawate Y and Kokare R. 2015. An automated approach for classification of plant diseases towards development of futuristic Decision Support System in Indian perspective. (*In*) *International conference on advances in computing, communications and informatics (ICACCI),* IEEE.

DeChant C, Wiesner-Hanks T, Chen S, Stewart E L, Yosinski J, Gore M A and Lipson H. 2017. Automated identification of northern leaf blight-infected maize plants from field imagery using deep learning. *Phytopathology* **107**(11): 1426–32.

Dhaygude S B and Kumbhar N P. 2013. Agricultural plant leaf disease detection using image processing. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering* **2**(1): 599–602.

Egmont-Petersen M, de Ridder D and Handels H. 2002. Image processing with neural networks - a review. *Pattern recognition* **35**(10): 2279–2301.

Ferentinos K P. 2018. Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture* **145**: 311–8.

Fuentes A, Yoon S, Kim S C and Park D S. 2017. A robust deep-learning-based detector for real-time tomato plant diseases and

pests recognition. *Sensors* **17**(9): 2022.

Hanson A M G J, Joel M G, Joy A and Francis J. 2017. Plant leaf disease detection using deep learning and convolutional neural network. *International Journal of Engineering Science* 5324.

Hiary A H, Bani-Ahmad S, Reyalat M, Braik M and Alrahamneh Z. 2011. Fast and accurate detection and classification of plant diseases. *Machine Learning* **14**(5): 58–60.

Hughes D and Salathé M. 2015. An open access repository of images on plant health to enable the development of mobile disease diagnostics. Retrieved from arXiv preprint arXiv:1511.08060.

Husin Z B, Shakaff A Y B M, Aziz A H B A and Farook R B S M. 2012. Feasibility study on plant chili disease detection using image processing techniques. (*In*) *Third International Conference on Intelligent Systems, Modelling and Simulation*, IEEE.

Jain R, Minz S and Subramanian R V. 2009. Machine learning for forewarning crop diseases. *Journal of the Indian society of Agricultural Statistics* 63(1): 97-107

Jain R, Minz S and Subramanian R V. 2005. Performance of machine learning techniques vis-à-vis logistic regression in forewarning incidence of crop diseases. (*In*) *Proceedings of 59th Annual conference of the Indian Society of Agricultural Statistics*, Sher-e- Kashmir University of Agricultural Sciences and Technology, Jammu, 11-13 November.

Jaware T H, Badgujar R D andPatil P G. 2012. Crop disease detection using image segmentation. *World Journal of Science and Technology* **2**(4): 190–4.

Johannes A, Picon A, Alvarez-Gila A, Echazarra J, Rodriguez-Vaamonde S, Navajas A D and Ortiz-Barredo A. 2017. Automatic plant disease diagnosis using mobile capture devices, applied on a wheat use case. *Computers and Electronics in Agriculture* **138**: 200–9.

Kamilaris A and Boldú P F X. 2018. Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture* **147**: 70-90.

Khirade S D and Patil A B. 2015. Plant disease detection using image processing. (*In*) *International Conference on Computing Communication Control and Automation*, IEEE.

Krizhevsky A, Sutskever I and Hinton G E. 2012. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems* **(84)**: 1097–1105.

Kulkarni A H and Patil A. 2012. Applying image processing technique to detect plant diseases. *International Journal of Modern Engineering Research* **2**(5): 3661–4.

Le T L, Duong N D, NguyenV T, Vu H, Hoang V N and Nguyen T T N. 2015. Complex background leaf-based plant identification method based on interactive segmentation and kernel descriptor. (In) ACM Proceedings of the 2nd International Workshop on Environmental Multimedia Retrieval.

LeCun Y, Bengio Y and Hinton G. 2015. Deep learning. *Natur*e **521**(7553): 436.

Liu B, Zhang Y He D and Li Y. 2017. Identification of apple leaf diseases based on deep convolutional neural networks. *Symmetry* **10**(1): 11.

Liu L and Zhou G. 2009. Extraction of the rice leaf disease image based on BP neural network. (*In*) *International Conference on Computational Intelligence and Software Engineering*, IEEE.

Lu Y, Yi S, Zeng N, Liu Y and Zhang Y. 2017. Identification of rice diseases using deep convolutional neural networks. *Neurocomputing* **267**: 378–84.

Meunkaewjinda A, Kumsawat P, Attakitmongcol K and Srikaew A. 2008. Grape leaf disease detection from colour imagery using hybrid intelligent system. *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology* (**1**): 513–6.

Mokhtar U, El Bendary N, Hassenian A E, Emary E, Mahmoud M A, Hefny H andTolba M F. 2015. SVM-based detection of tomato leaves diseases. *Intelligent Systems* **20**(14): 641-652.

Nigam S, Jain R, Marwaha S, Arora A, Singh K V. 2019. Deep learning for plant disease identification. (*In*) *Proceeding of International Conference on Agricultural Statistics*, New Delhi, November 18-21.

Oppenheim D and Shani G. 2017. Potato disease classification using convolution neural networks. *Advances in Animal Biosciences* **8**(2): 244–9.

Pawara P, Okafor E, Surinta O, Schomaker L and Wiering M. 2017. Comparing local descriptors and bags of visual words to deep convolutional neural networks for plant recognition. (*In*) *International Conference on Pattern Recognition Applications and Methods (ICPRAM),* pp 479–86.

Phadikar S, Sil J and Das A K. 2012. Classification of rice leaf diseases based on morphological changes. *International Journal of Information and Electronics Engineering* **2**(3): 460.

Picon A, Alvarez-Gila A, Seitz M, Ortiz-Barredo A, Echazarra J and Johannes A. 2018. Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild. *Computers and Electronics in Agriculture* **161**: 280–90 .

Prasad S, Peddoju S K and Ghosh D. 2016. Multi-resolution mobile vision system for plant leaf disease diagnosis. *Signal, Image and Video Processing* **10**(2): 379–88.

Ramcharan A, Baranowski K, McCloskey P, Ahmed B, Legg J and Hughes D P. 2017. Deep learning for image-based cassava disease detection. *Frontiers in Plant Science* **8**: 1852.

Revathi P and Hemalatha M. 2012. Classification of cotton leaf spot diseases using image processing edge detection techniques. (*In*) *International Conference on Emerging Trends in Science Engineering and Technology (INCOSET),* IEEE.

Rumpf T, Mahlein A K, Steiner U, Oerke E C, Dehne H W and Plümer L. 2010. Early detection and classification of plant diseases with Support Vector Machines based on hyperspectral reflectance. *Computers and Electronics in Agriculture* **74**(1): 91–9.

Sajda P. 2006. Machine learning for detection and diagnosis of disease. *Annual Review of Biomedicine Engineering* **8:** 537-565.

Sannakki S S, Rajpurohit V S, Nargund V B and Kulkarni P. 2013. Diagnosis and classification of grape leaf diseases using neural networks. (*In*) *Fourth International Conference on Computing, Communications and Networking Technologies*, IEEE.

Sankaran S, Mishra A, Ehsani R and Davis C. 2010. A review of advanced techniques for detecting plant diseases. *Computers and Electronics in Agriculture* **72**(1): 1-13.

Schikora M and Schikora A. 2014. Image-based analysis to study plant infection with human pathogens. *Computational and Structural Biotechnology Journal* **12**(20-21): 1-6.

Schmidhuber J. 2015. Deep learning in neural networks: An overview. *Neural networks*, **61**: 85–117.

Semary N A, Tharwat A, Elhariri E and Hassanien A E. 2015. Fruit-based tomato grading system using features fusion and support vector machine. *Intelligent Systems* **20**(14): 401–10.

Simonyan K and Zisserman A. 2014. Very deep convolutional networks for large-scale image recognition. Retrieved from *arXiv preprint arXiv.*

Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov Dand

Rabinovich A. 2015. Going deeper with convolutions. (In) *Proceedings of conference on computer vision and pattern recognition*, IEEE.

Szegedy C, Ioffe S, Vanhoucke V and Alemi A A. 2017. Inception-v4, inception-resnet and the impact of residual connections on learning. (In) *Thirty-First AAAI Conference on Artificial Intelligence*.

Upadhyaya S, Arora A and Jain R. 2006. Rough set based cluster analysis for soybean disease diagnosis. (*In*) *Proceeding of International Conference on Statistics and Informatics in Agricultural Research*, New Delhi, December.

Xie C and He Y. 2016. Spectrum and image texture features analysis for early blight disease detection on eggplant leaves. *Sensors* **16**(5): 676.

Xie C, Shao Y, Li X and He Y. 2015. Detection of early blight and late blight diseases on tomato leaves using hyperspectral imaging. *Scientific Reports* **5**: 165–64.

Yang X and Guo T. 2017. Machine learning in plant disease research. *European Journal of BioMedical Research* **3**(1): 6–9.

Yao Q, Guan Z, Zhou Y, Tang J, Hu Y and Yang B. 2009. Application of support vector machine for detecting rice diseases using shape and colour texture features. *(In*) *International Conference on Engineering Computation*, IEEE.

Zhang X, Qiao Y, Meng F, Fan C and Zhang M. 2018. Identification of maize leaf diseases using improved deep convolutional neural networks. *IEEE Access* **6**: 30370–7.