Archives of Computational Methods in Engineering (2022) 29:641–677
https://doi.org/10.1007/s11831-021-09588-5

REVIEW ARTICLE

Machine Learning and Deep Learning Based Computational Techniques in Automatic Agricultural Diseases Detection: Methodologies, Applications, and Challenges

Javaid Ahmad Wani[1] · Sparsh Sharma[1,2] · Malik Muzamil[1] · Suhaib Ahmed[4] · Surbhi Sharma[3] · Saurabh Singh[5]

Received: 20 November 2020 / Accepted: 8 April 2021 / Published online: 3 May 2021

Abstract
Plant disease detection is a critical issue that needs to be focused on for productive agriculture and economy. Detecting plant disease using traditional methods is a tedious job as it requires a tremendous amount of work, time, and expertise. Automatic plant disease detection is an important research area that has recently gained a lot of attention among the academicians, researchers, and practitioners. Machine Learning and Deep Learning can help identify the plant disease at the initial stage as soon as it appears on plant leaves. In this state-of-an-art review, a thorough investigation has been performed to evaluate the possibility of using Machine Learning models to identify plant diseases. In this study, diseases and infections of four types of crops, i.e., Tomato, Rice, Potato, and Apple, are considered. Initially, numerous possible infections and diseases on these four kinds of crops are studied along with their reason for the occurrence and possible symptoms for their detections. An in-depth study of the different steps involved in plant disease detection and classification using Machine Learning and Deep Learning is provided. Various datasets available online for plant disease detection have also been presented. Along with this, a detailed study on various existing Machine Learning and Deep Learning-based classification models proposed by different researchers across the world for four considered crops in terms of their performance evaluations, the dataset used, and the feature extraction method is discussed. At last, various challenges in the use of machine learning and deep learning for plant disease detection and future research directions are enumerated and presented.

Abbreviations

AI	Artificial intelligence
ANN	Artificial neural network
BNN	Binarized neural network
BPNN	Back propagation neural network
CNN	Convolutional neural network
DCNN	Deep convolutional neural network
DL	Deep learning
FT	Fourier transform
GA	Genetic algorithms
GDP	Gross DOMESTIC product
GLCM	Grey level co-occurrence matrix
HOG	Histogram of oriented gradients
HIS	Hue, saturation and intensity
HUE	Hue, saturation value
IoT	Internet of things
LDA	Linear discriminant analysis
ML	Machine learning
MSOFM	Modified self organizing feature maps
MSVM	Multiclass support vector machine
PCA	Principal component analysis
RBF	Radial based Function
ReLU	Rectified linear unit
RF	Random forest
RMS	Root mean square
SVM	Support vector machine

✉ Sparsh Sharma
sparsharma@outlook.com

Javaid Ahmad Wani
javaidcse14@gmail.com

Saurabh Singh
saurabh89@dongguk.edu

1 Department of Computer Science and Engineering, Baba Ghulam Shah Badshah University, Rajouri, India

2 Department of Computer Science & Engineering, NIT Srinagar, Srinagar, J&K, India

3 Department of Computer Science and Engineering, Shri Mata Vaishno Devi University, Katra, India

4 Department of Electronics & Communication Engineering, Baba Ghulam Shah Badshah University, Rajouri, India

5 Department of Industrial & Systems Engineering, Dongguk University, Seoul, South Korea

1 Introduction

The infections and diseases in plants and crops can have a drastic impact on both the quality as well as quantity of crops. This can make a negative impact on the countries economies, especially where the major source of income as well as the occupation is agriculture [1]. Henceforth, the detection, identification, and acknowledgement of the infection in the crops in the earlier stages is very crucial in order to prevent the damages in the crops as well as to improve the quality and the quantity of the yield. According to various existing studies, it has been found that almost 70% of the population of India depends undeviatingly or in a roundabout way on agribusiness and contributes around 17% to the overall GDP of the country [2, 3]. Around 3,50,000 hectares of India`s land is used for the tomato`s cultivation, which yields 5.3 million tonnes of tomatoes, and is ranked as the third-largest tomato yielder in the world [4]. India ranks second in potato production and produces about 48.60 million tonnes of potatoes [5]. With the production of about 2371.0 thousand million tonnes, India is the 6th largest apple producer in the world [6]. India is in the second position in rice production, with a yield of about 116.42 million metric tonnes [7].

Various factors, both environmental as well as adulteration, like the absence of quality land manures, choice of unsuitable crops, climatic changes, pests, weeds, and so forth, are responsible for the diseases and infections among the crops. Among all these factors, pests infections alone cause an annual loss of around 30-33% of the overall yield in India [8]. Fungus, viruses, and bacteria are the causes of infections in the plant. Because of these numerous infections and factors, farmers experience serious challenges in shifting from the adoption of one infection control strategy to another for limiting these infections, which in turn affects the overall yield as well as the quality of the crop.

Even in the technology-enriched era where technology has improved a lot, farmers are still following the orthodox and old way of identifying the diseases in the crops by personally and physically analyzing the crops through their eyes. This old technique of observing and analyzing the crops with the naked eyes by the experience of the farmer have various issues and hurdles. With this method, a farmer might be able to detect certain crop infections and disease which he is aware of, but this technique does not work well for the detection of new and those crop`s infections which are not known to him. This will lead the crop`s infection to get undetected or will lead to the use of the wrong control strategy, which can affect the crop`s yield and, in the worst case, can lead to deterioration of the whole crop. Some diseases do not have properly visible symptoms, and in such scenarios, it becomes difficult to decide when, what and how to act. In such situations, advanced and in-depth exploration becomes obligatory.

To prevent the infections from affecting the yield as well as the quality of the crop, it becomes vital for a farmer to be an expert and have deep knowledge of all the possible crop`s disease and their corresponding diagnose solution, which is not practically possible, especially in the current era where new infections are erupting due to current climatic disturbances. The convenient and exact finding of plant infirmities is one of the backbones of exactness agribusiness.

For dealing with the above-mentioned issues in current farming scenarios, the advancement in Computer Vision, Machine Learning, and Deep Learning technologies can be utilized for the accurate, quick, and earlier detection of crop diseases from a plethora of existing crop`s diseases. The benefits of using these technologies are that these computerized apprehensions using image processing methods give fast and accurate results. Advancements in computer vision and precision can be used for reducing the extra labor costs, time wastage, and for enhancing the quality as well as the overall yield of the crops. Early information on crop well-being and infection location can encourage the control of diseases through appropriate administration methodologies.

1.1 Motivation of this Study

Identification of diseases and choosing the appropriate remedies for controlling that crop`s diseases are the basic steps in farming which is being performed since ages by the farmers for preventing the crop`s loss as well as to maintain its quality. However, due to global climatic changes, changed lifestyles of people, pollution, and various other factors, the drastic rise in different crop`s disease have been seen. It becomes very important for the farmers to have the knowledge of these crop`s disease for their identification and control. However, due to numerous types of diseases, it is challenging for a farmer to have knowledge of all the diseases. Also, in large-scale cultivation, it is not feasible for a farmer, both economically as well as physically, to monitor such a large scale of crops. As a result, various crop`s diseases and infections usually go unnoticed and eventually affects the overall yield and quality of the crops. So, in order to deal with such issues, automatic detection and classification of crop`s disease is required and is the need of the hour. Machine Learning and Computer Vision are already being used by the farmers and researchers across the globe in agriculture for various purposes. Machine Learning-based techniques can be used for accurate and real-time detection as well as classification of diseases in crops, which will enhance the productivity and quality of the crops and will also reduce the labor, cost, and improve the accuracy of the farmer in crop cultivation.

1.2 Contribution of this Study

The Major Contribution of this study are as follows:

1. Various types of diseases and infections found in different crops like Tomato, Potato, Rice, Apples, etc. along with their symptoms for classification purpose are presented.
2. The Steps involved in the Automatic Detection and Classification of diseases in plants, along with the various possible techniques and algorithms in each step are presented.
3. Machine Learning and Computer Vision-based models presented by researchers for automatic detection and classification of plant diseases for different plant types in literature are thoroughly presented.
4. Challenges as well as the Issues in the deployment of these machine learning models for plant disease detection, enumerated from this comprehensive study, are also presented.

In Table 1, a comparison of this survey with different existing available related surveys is performed to elaborate on the innovation covered in this survey and to focus on how our study is unique and different from other existing surveys. It was observed that in most of the existing surveys on automatic plant disease detection using machine learning techniques, only a brief overview of the different aspects are covered. However, in our study, an in-depth focus has been laid on different aspects involved in automatic plant disease detection using machine learning techniques.

The rest of the paper is divided into six sections. In Sect. 2, various kinds of possible diseases found in different plants like tomato, rice, apple, etc. are discussed along with their symptoms that can be used for their detection and classification. Generic Steps, like Image Acquisition, Data Pre-Processing, Image Segmentation, Extraction of Images, and Classification of Disease, involved in the Automatic Plant Disease Detection and Classification using Machine Learning and Deep Learning along with the different algorithms and techniques commonly used in each step, are discussed in detail in Sect. 3. Afterwards, Various Existing Machine Learning based classification techniques proposed by different researchers in the literature for four crops: tomato, potato, apple, and rice crop are presented in Sect. 4. Based on the observation of various existing frameworks in the literature discussed in Sect. 4, various challenges and open issues in the deployment of these machine learning and computer vision models for the plant`s disease detection and classification along with the future research directions, are presented in Sect. 5. The exhibition of these frameworks relies significantly upon the size of the dataset and the classifiers used. Consequently, an endeavour to sum up the existing literature and challenges enumerated from this study, the concluding remarks of this paper, is given in Sect. 6.

The complete layout of this manuscript is shown graphically in Fig. 1.

Table 1 Comparison with other existing surveys in literature

Topics Covered	[9]	[10]	[11]	Our Paper	
Study of various possible infections and diseases in different crops like potato, tomato, rice and apple, along with their causes & probable symptoms	✓	#	x	✓	
Study of various dataset available in literature for plant disease detection	X	✓	x	✓	
In-depth study of various steps involved in automatic plant disease detection and classification using machine learning along with the numerous and most commonly used algorithms and techniques in each step	✓	#	#	✓	
In depth study of various existing machine learning and deep learning based classification techniques and frameworks in literature along with their performance evaluations, dataset used, and feature extraction method used	✓	✓	✓	✓	
Type of plant disease covered:	Classification of tomato crop`s diseases	#	*	#	✓
	Classification of potato crop`s diseases	#	#	x	✓
	Classification of rice crop`s diseases	✓	#	#	✓
	Classification of apple crop`s diseases	x	#	#	✓
Challenges and issues in automated plant disease detection and classification using machine learning and deep learning techniques	✓	x	x	✓	
Future research directions in automated plant disease detection and classification using machine learning & deep learning techniques	✓	x	x	✓	

✓ Covered, **x** Not covered, **#** Not covered in detail

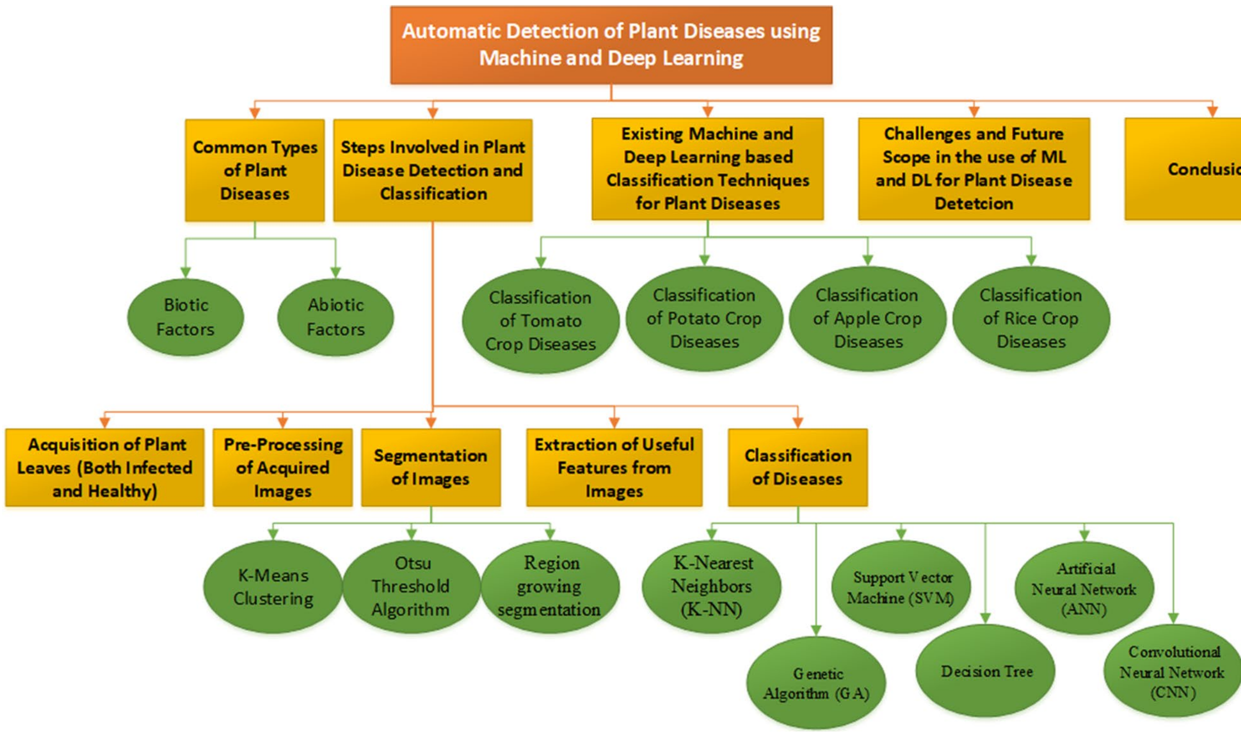


Fig. 1 Layout of the article

2 Common types of Plant Diseases

Mainly there are two kinds of plant diseases: (a) Biotic and (b) Abiotic. Biotic diseases are those disease which arises from dwelling organisms such as insects, bacteria, fungi, and viruses [12]. On the other hand, abiotic diseases in plants arise due to factors like extreme high temperature, extra moisture, bad mild, inadequate vitamins, bad soil pH, and greenhouses gasses. The factors causing the plant diseases are shown in Fig. 2. Identification and classification of such disease and infection in plants is a challenging task as these diseases are not always visible by just seeing them. Also, many plant`s disease and infections usually have symptoms that are common among many other plant diseases, which makes the recognition of the correct disease in plants a challenge. In some cases, the unusual signs of symptoms like abnormal leaf growth, color distortion of leaves, stunted growth, shrunk and degraded pods, etc. make the analysis and disease identification process hard. These indications, for instance, a blatant case of the influenced leaves, help to anticipate the disease. Early diagnosis and identification of these plant diseases, are very crucial for safeguarding the environment, crop`s yield, as well as the economy linked with the agriculture sector.

In this study, the main research focus is on the identification of leaf disease because, according to various existing studies, it is concluded that if leaves of the plant are healthy, then the immune system of the whole plant will be strong and capable enough to defeat any disease in other parts of the plant.

In crops like potatoes, pests and other diseases have a significant impact on their yield. Tomato`s crops are affected mainly by diseases like Anthracnose, Block mold, Early blight, Fusarium wilt, Leaf mold, Septoria leaf spot, Verticillium wilt, Bacterial speck, Late blight tomato, Tomato spotted wilt, and little leaf, which are explained in detail along with their causes and corresponding symptoms in Table 2 [13].

Similarly, Potato plant`s leaves are majorly infected by common diseases like Bacterial ring rot, Rhizoctonia canker, verticillium wilt, Potato leaf roll, Potato virus, Flea beetles, Late Blight, and Early Blight [13]. But the most frequently occurring two diseases of potato plants are Late Blight and Early Blight [14]. Each infection has some perceptible manifestations on the plant parts, which is explained in detail in Table 3, along with their causes and symptoms [13].

There are more than a thousand kinds of apples grown around the world. Apples are subservient to a category of issues that can lead to various insignificant and superficial loss or, additionally, a high-quality catastrophe, including dwindled production and tree expulsion. The subsequent and common infections found in Apple`s crop are Apple scab, Black rot, Cedar apple rust, Powdery mold, Fire blight, and Spider mites. Table 4 lists out various possible infections

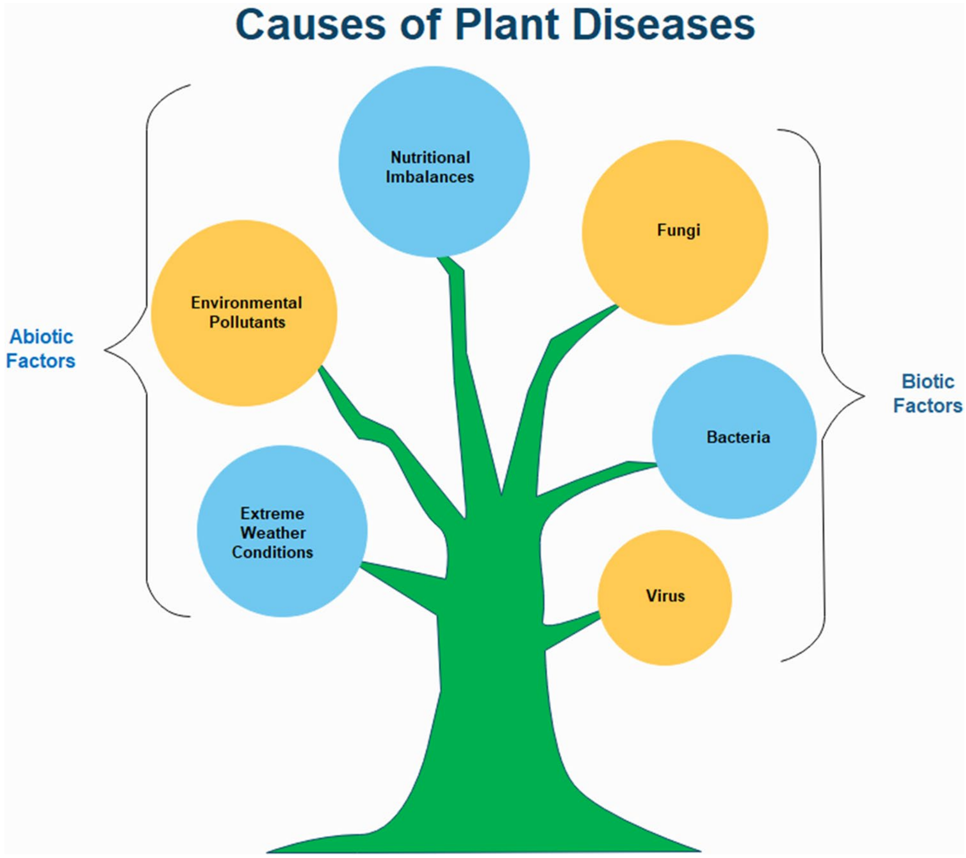


Fig. 2 Causes of plant diseases

found in Apple crop along with their causes and symptoms [13].

Rice, which is the most consumable food in India as well as across the globe, suffers from more than forty types of diseases that are genetically similar to each other, and the damage caused by these diseases is easy to spread [15]. The rice crop gets most affected by diseases like Bacterial leaf blight, Brown spot, Leaf smut, Rice blast, Tungr, Narrow leaf spot, and sheath blight [16]. The information on these diseases along with their cause and symptoms is presented in detail in Table 5 [13].

3 Steps Involved in Plant Disease Detection and Classification

Automatic Identification of Plant`s Diseases and Fig. 2 Infections is the need of the hour, and is required to minimize the crop`s infections and the losses that arise from these infections. The automated process of identification of plant`s diseases will also reduce the labor cost required in continuous and close monitoring of crops for possible infections.

These automated process of finding the possible diseases in crops and plants using Computer Vision and Machine Learning involves some generalized sequence of steps which ranges from acquiring and capturing of the images of the plants using different IoT Sensors deployed in the farm field, to the processing of the captured images for feeding to the machine learning models for classification of plants as healthy or infected. The detailed sequence of steps required for the identification of plant`s disease using Computer Vision and Machine Learning Models are shown in Fig. 3, and are as follow:

3.1 Acquisition of Plant Leaves (both Infected and Healthy)

Image acquisition is a process of capturing and acquiring the relevant images of the object, which is to be fed to the machine learning model for the learning and classification purpose. It is the foremost and crucial step in digital Image Processing employed to convert a visible picture into an array of binary records for further processing on a computer.

The images are captured in this step by making use of high-resolution digital cameras[9, 25]. Smartphones can also be brought into use to capture image samples in various supported formats like jpg, png, tiff, etc. [26]. After the successful acquisition of required images, images are sent

Table 2 Tomato leaf diseases

Disease	Picture	Cause	Symptoms
Anthracnose		Fungus	Disease causes the leaf to have distinctive sunken round lesions
Black mold		Fungus	The disease induces black or brown lesions to develop on the top of tomato leaves
Early blight		Fungus	Symptoms of early blight arise as circular shaped lesions across the lesion with a yellow chlorotic patch. On infected plants, concentric leaf lesions can be observed
Fusarium wilt		Fungus	Fusarium wilt symptoms can appear first as yellowing and wilting of the leaves on one surface of the leaf or of the plants
Leaf Mold		Fungus	On the upper surface, the older leaves show pale greenish to yellow spots

Table 2 (continued)

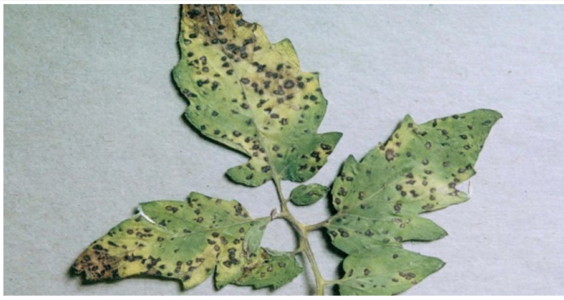
Disease	Picture	Cause	Symptoms
Septoria leaf spot		Fungus	On the underside of the older leaves, water-soaked patches or round grayish-white patches develop
Verticillium wilt		Fungus	The initial infection signs can be seen as yellow patches on the plant's lower leaves
Bacterial speck		Fungus	Black spots, often with a yellow halo, develop on the leaves
Late blight		Oomycete	Initial signs of the disease occur on leaves as water-soaked green to black areas that transform quickly to brown lesions
Tomato spotted wilt		Virus	Diseased plants show the upper sides of young leaves bronzing or purpling and establish necrotic patches

Table 2 (continued)

Little leaf		Physiological disorder	Symptoms of little leaves include interveinal chlorosis of new leaves, distortion and leaf inability to broaden along the midrib

to the image preprocessing stage, based upon the requirement. If the captured images are not suitable for processing, then the image enhancement strategies are performed on the images [27].

The Image Acquisition approach includes the following three steps [27]: -

1. Energy is accumulated using the Optical system.
2. The surface of the target object reflected energy.
3. Sensors measure the amount of energy.

This complete phase of Image Acquisition plays a vital role in the success of the Automated Plant Disease Classification as the correctness and accuracy of the whole framework lies on the quality and quantity of the captured images as these images are used for training of the Machine Learning models [9].

The images for the classification of a plant as healthy or infected can be acquired by deploying the IoT-based capturing devices in the farm field, which can be configured to capture the images of the crops periodically. The machine learning models for classification and identification of infected plant`s diseases can also be trained using the already existing database of plant images available in the literature. There are various famous datasets available in literature containing both infected and healthy images of various plants. Some of them are paid, but a maximum of them are freely available that can be used for training our model and for evaluating the performance of the proposed Machine Learning based classification model. Some of the available plant infections related datasets online are PlantVillage [17], New Plant Diseases [18], Rice Diseases Image [19], Plant Pathology 2020–FGVC7 [20], IPM Images [21], OSF [22], UCI [23], and APS Images[24] datasets.

There are also some research centers that give access to their datasets like IRRI, INIBAP, and BRRI [28–30]. Numerous amount of research has been done to collect images from the base fields, inside laboratories, and sampling box [31–34]. The image data can also be collected by a hyperspectral imaging system [35].

Some publicly accessible datasets for various plant diseases containing different types of infections found in a specific crop are listed in Table 6 along with the dataset size and their downloading link.

3.2 Pre-processing of Acquired Images

Image Pre-Processing is performed on the images acquired in the first phase of Images Acquisition, in order to bring the images to the same benchmark standard features onto which the various machine learning models can be applied. Various pre-processing tasks involved in this phase are Image Size Regularization, adjusting the image as per the required image color scale, etc. Pre-processing is usually required to be applied to the images to reduce the computational costs and standardize the image resolutions to a specific standard benchmark [36].

Image pre-processing strategies are basically the mathematical/statistical methods that enhance the apparent look as well as geometric characteristics of an image and transforms the image into a required common format. Images processed in this phase are fed to the next phase in which segmentation and feature extraction is applied. Any type of image`s noise and distortion is removed in this phase before passing the image for the next phases of segmentation and feature extraction, which, if not performed, may have an adverse effect on the performance of the system [37].

The goal of the pre-processing on the images is to alter the intensities of the images with the aim to highlight target areas [28]. The white base is used as a background of images to minimize complexity [38]. Widely used preprocessing methods are distortion removal, noise removal, color space conversion, image cropping, smoothing, and enhancement. Hue, saturation, and value (HSV) method is the most

Table 3 Potato leaf diseases

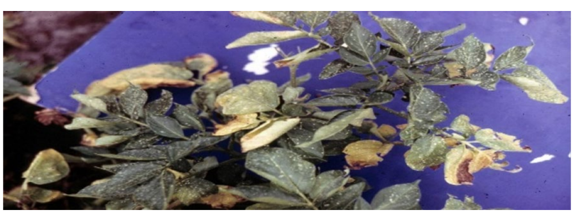
Disease	Picture	Cause	Symptoms
Potato early blight		Fungus	Dim sores with a yellowish fringe which may additionally grow concentric hoops and pessimistic membrane on the blades and stems
Potato late blight		Oomycete	Erratically growing brownish lesions on leaves with peculiar white fleecy sporulation at lesion boundaries on the underneath of the leaf in wet circumstances
Bacterial ring rot		Bacterium	Wilting stems and leaves
Rhizoctonia canker		Fungus	Lesions may surround the stem originating leaves to fold and turn yellowish
Verticillium wilt		Fungi	Pamphlets vanishing on particularly one aspect of the stalk or branching peduncle
Potato leaf roll		Virus	Young leaves rolled and yellow or pink
Potato virus		Virus	Manifestations range broadly from the slight mosaic of leaves to leaf mildew and plant loss of life relying on the diversity of potato and the stress of the infection

Table 3 (continued)

Disease	Picture	Cause	Symptoms
Flea beetles		Insects	Tiny openings or cavities in leaf that deliver the foliage a function shothole look

commonly applied pre-processing method as it mirrors the sensing capability of humans [39–41].

The masking and background removal process is used to speed up the processing as well as to improve the accuracy [42]. A colored image is transformed into famous HSI (Hue, Saturation, Intensity) color space representation since this resembles sensing features of humans. H component of HSI is mainly used for further analysis [1, 43, 44]. Using Low Pass high-frequency, components are attenuated. The high pass filter uses negative weighting coefficients for the nearby pixels that adequately enhance the regions of an image with more intensity gradient so that more precise information is highlighted [45]. Laplacian filter emphasizes contours in the image and is used to alter the distribution of gradients in an image [46]. Minimal and Maximal filters act as analytical filters that switch each pixel value by either the minimal or maximal value from the collection of most adjoining pixels [47]. The Fourier transform (FT) filter transforms the picture to the spatial frequency domain by making use of a Fast Fourier Transform algorithm [48]. A straightforward yet powerful commotion smoothing channel is used by the sigma likelihood of the Gaussian dispersion, and it smooths the picture clamor by averaging just those local pixels which include the powers inside a fixed sigma scope of the middle pixel [49].

The histogram equalization that circulates powers of the pictures is used on the images to upgrade the plant sickness pictures. Histogram modifications are specifically helpful to stress the shading varieties present in a scene or a chosen part of the image for segregating and to highlight the qualities, in case any gets unnoticed. The cumulative distribution function is employed to share intensity values [2]. In order to achieve the high accuracy in disease recognition, Segmentation of the diseased leaf Image is required to fetch the effect spot images before extracting the features [41].

3.3 Segmentation of Images

Segmentation is a process of subdividing an image into its constituent objects. The main aim of segmentation is to further analyze each object to extract some useful features. On the basis of extracted features, we can simply differentiate the healthy and infected segments [2, 9]. Acquired and pre-processed images are sent for performing segmentation to extract the various possible useful features required in the learning and classification of diseased leaves.

Various commonly used segmentation algorithms are K-Means Clustering, Otsu threshold [2], Region Growing [50]. Thresholds and locality-based segmentation techniques are believed to yield better results for plant disease detection systems. The Entropy and Otsu algorithms are a few famous thresholds primarily based on segmentation techniques. The K-Means clustering shows the best result in contrast with Sobel, prewit, Fuzzy C-Means, and conny based segmentation techniques [9]. The features inclusive of area, shape, statistical parameters, and texture are extracted for further analysis of images [50]. Segmentation can also be done based on the color and the luminosity into constituents of the image in the L*a*b* color space [51].

The most commonly used Segmentation methods in plant disease detection and classification process are:

3.3.1 K-Means Clustering

The K-Means clustering works by categorizing the objects into K-groups based on the Euclidean distance of the pixels of an image [52].

The following are the steps involved in K-Means clustering [2]:

- Initially, K-points are randomly selected.
- Each pixel of the image is then added to the cluster based on the Euclidean distance.
- The cluster center is calculated by equalizing all the pixels in the group, repeat the last two steps until the desired cluster is obtained.

K-Means clustering has numerous advantages, but it has some serious issues as well like the determination

Table 4 Apple leaf diseases

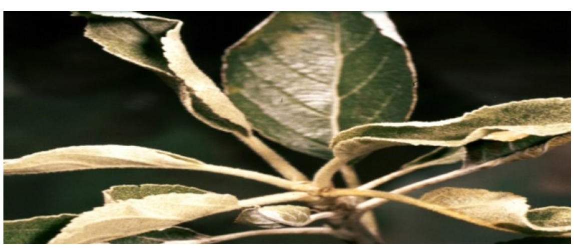
Disease	Picture	Cause	Symptoms
Apple scab		Fungus	Yellowish or chlorotic pinches at a leaf or dull drab-green pinches at leaf or perhaps a velvety maturity on bits on underneaths of the leaf
Black rot		Fungus	Purplish patches or round sores that are brownish in the midst and purple on the border
Cedar apple rust		Fungus	Fluorescent orange or yellowish patches on the uppermost surface of leaf encircled by using a crimson ring and tiny black pinches within the core
Powdery mildew		Fungus	Soft milky spots on base of the leaf or greensickness pinch on the uppermost surface of the leaf
Fire blight		Bacterium	The leaves and shoots seems as if it has been roasted through burning or washed exudate can be commenced on maladies regions

Table 4 (continued)

Disease	Picture	Cause	Symptoms
Spider mites		Arachnid	Leaf dappled by way of yellowish or leaf may additionally resemble bronzed or parasites that can be apparent as small flowing marks at the below of leaf

of the K values considering the constant cluster size, the assessment of cluster quality produced is a challenging job. Furthermore, the K-Means technique does not function efficiently with non-globural clusters, and a selection of preliminary partitions can result in a difference in concluding clusters.

3.3.2 Otsu Threshold Algorithm

This algorithm is used to obtain binary images from grey-level images using the threshold value. The pixel value under the threshold are treated as zero and above the threshold as one.

Steps of the Otsu Algorithm are as [53]:

- Set the threshold and cluster the pixels on the basis of threshold value.
- Calculate the median of every cluster.
- Squaring the difference among the medians.
- Product the number of pixels in a single cluster with the number in the other cluster.

3.3.3 Region Growing Segmentation

This approach works at the pixel level, and the foremost step of this approach is to select the seed points. The region starts growing based on certain criteria such as pixel intensity, grayscale, color, adjacency, and similarity.

Steps of region growing segmentation:

- Segment the entire image into pattern cells.
- Each patter cell is related to its adjoining cells and is added to the region based on the similarity of intensity values.
- Continue developing the region by examining all its adjacents until no combinable region remains.
- Repeat the above steps until all cells are added to the respective region or segment.

3.4 Extraction of Useful Features from the Image

The Conversion of the input information into a set of features is called feature extraction. The three global feature descriptors that are extracted from the input images are shape, color, and texture. Feature extraction plays a vital role in various classifications problems, including plant disease detection. This phase of plant disease detection is used for the extraction of various unique properties from the images dataset, which is fed as input to this phase, for the purpose of classification and identification of plants as healthy and infected. Feature extraction is also known as feature engineering in the machine learning discipline.

The basic features of the image are color, texture, morphology, etc. The morphologic features show better results than other features for the disclosure of the infected area of leaves [2, 37]. The color moments and Gabor texture are commonly used color features that can be acquired by numerous techniques like color histogram [54], color correlogram [55], color R moment [56], etc. The attributes like contrast, homogeneity, variance, and entropy can be added to the texture. The texture feature gives better results for plant disease detection problems[9]. The texture properties viz energy, entropy, contrast, homogeneity, moment of inertia, etc., of the desired portion can be computed using Gray Level Co-occurrence Matrix (GLCM) [51, 57]. The texture features can also be extracted using other techniques like FT, difference operator, discrete cosine transform (DCT), and wavelet packet decomposition methods [9].

The other features like Speed-up robust feature (SURF), Histogram of oriented gradients (HOG), Scale-invariant feature transform (SIFT), and Pyramid histogram of visual words (PHOW) also show better performance in the plant disease detection [9]. The Back-Propagation Neural Networks (BPNN), Modified Self-Organizing Feature Maps (MSOFM), and Genetic Algorithm (GA) can also produce

Table 5 Rice leaf diseases

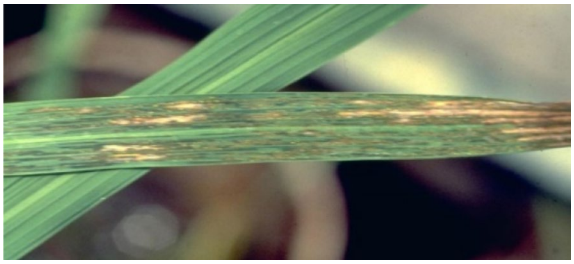
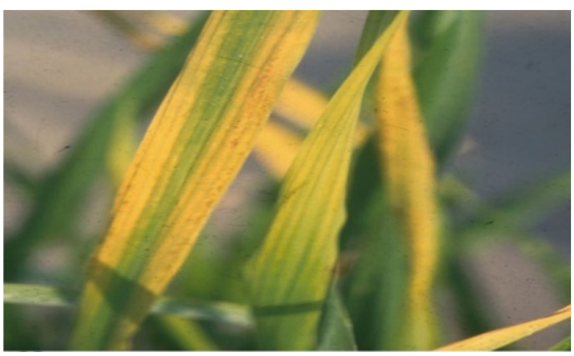
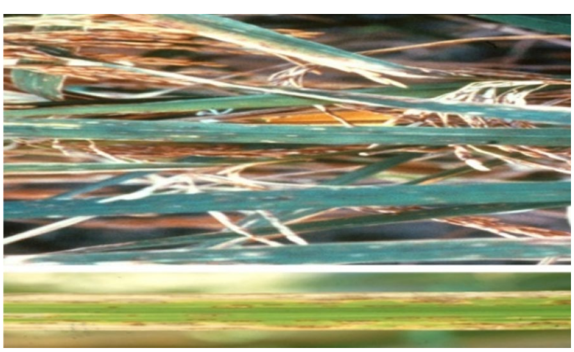
Disease	Picture	Cause	Symptoms
Bacterial leaf streak		Bacterium	Water-soaked strips leaf that's initially dark inexperienced and then turns translucent.
Rice bacterial blight		Bacterium	The color of leaf turn to yellow or grayish
Rice blast		Fungus	White to greenish or grizzled diamond-formed tumors with darkish grassy edges
Tungr		Virus	Shoots are stunted with a yellow-orange blemish or plants may have a diminished fraction of tillers and rust-colored pinches on leaves
Narrow leaf spot		Fungus	Tiny, cryptic or elongated brownish sores on rice leaf

Table 5 (continued)

Disease	Picture	Cause	Symptoms
Brown spot		Fungus	Oval-shaped brown spots on rice leaf
Leaf smut		Fungus	Black spots on both sides of the rice leaf

the desired, and promising results in disease classification of the plant leaves on the basis of the color feature [58]. Different mathematical formulae involved in the extraction of various features in the feature extraction process are listed in Table 7.

3.5 Classification of Diseases/Infections in Plant Leaves

Classification is a process of identification and categorization of input data into different possible known classes. In this phase, actual detection and separation of input data into different classes is performed. This is the primary task in computer vision. In this phase of classification of plant`s leaves, the plant images are classified and organized into different categories based on identified diseases. Leaves are classified as healthy and infected in this phase. The leaves which are classified as infected are further classified into the specific possible disease class which the leaf may be suffering from. Disease detection is a bit simpler than disease classification, so selecting appropriate classifiers based upon the problem is crucial in the case of Disease Classification.

Broadly, machine learning techniques fall into two types: Supervised and Unsupervised techniques. Mostly used classifier techniques are K-means, Support Vector Machine (SVM), Decision tree, Random forest (RF), K-NN, Genetic algorithm (GA), Fuzzy logic, Fuzzy C-means, Artificial Neural Network (CNN), BPNN, and Convolutional Neural Network (CNN). However, ANN, BPNN, and CNN are considered majorly for various classification problems, including plant disease detections, as these techniques show good performance in the classification of plant leaves with automatic features extraction. Few Classification techniques are discussed below.

3.5.1 K-Nearest Neighbors (K-NN)

K-NN is the oldest non-parametric classification algorithm i.e. it does not make any underlying assumptions about the distribution of data. In addition to classifications problems, it can solve regression problems too. K-NN technique works by storing each single accessible case and characterizing new cases based on a measure of their comparability. K-NN classifier is used to identify an object by considering majority vote of its neighbor on the basis of k most related vectors present in feature space, where k is a positive integer and generally small [59]. It has applications in statistical estimation and pattern recognition too. The distance function in KNN can be calculated using the Euclidean equation.

$$E(i,j) = \sqrt{\sum_{i=1}^{n} \left(i_k - j_k\right)^2}$$

where E (i, j) is distance measure between vector I and vector j and n, k is the participation of significant worth, and n is the quantity of highlights in vectors.

3.5.2 Genetic Algorithm (GA)

GA [60] is a loose optimization approach and is based on the impressions of natural selection and genetics. GA is typically

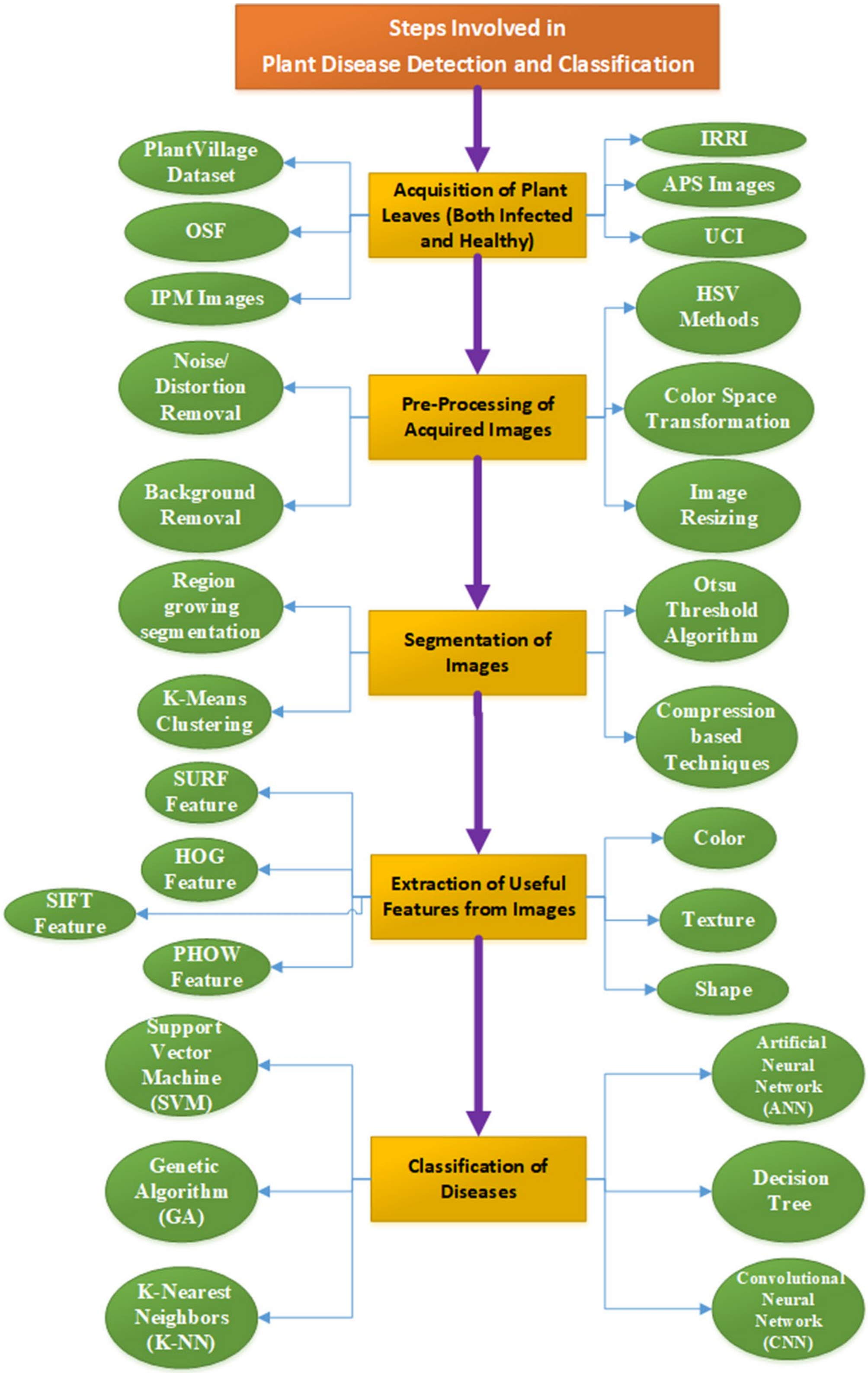


Fig. 3 Steps involved in automatic plant diseases detection using ML and DL

Table 6 Dataset for plant disease

Reference	dataset name	Crops	Datase Dataset set size	Link
[17]	PlantVillage	Pepper_Bell Tomato Potato	Pepper_Bell Infected:997 Healthy:1478 Tomato Healthy:1591 Infected:14,525 Potato Healthy:152 Infected:2000	https://www.kaggle.com/emmarex/plant disease
[18]	New Plant diseases dataset	Apple Blueberry Cherry Corn Grape Orange Peach Pepper_Bell Potato Raspberry Soybean Squash Strawberry Tomato	Apple Healthy:2510 Infected:7204 Blueberry Healthy:2270 Infected:2247 Cherry Healthy:2247 Infected: 2133 Corn Healthy:2339 Infected:6821 Healthy:2115 Grape Infected:6912 Orange Infected:2513 Peach Healthy:2160 Infected:2341 Pepper_Bell Healthy:2485 Infected:2416 Potato Healthy:2260 Infected:4848 Raspberry Healthy:2226 Soybean Healthy:2527 Squash Infected:2170 Strawberry Healthy:2280 Infected:2218 Tomato Infected:4547	https://www.kaggle.com/vipoooool/new-plant-diseases-dataset
[19]	Rice diseases image dataset	Rice	Rice BrownSpot:513 Healthy:1488 Hispa:565 LeafBlast:779	https://www.kaggle.com/minhhuy2810/rice-diseases-image-dataset?

Table 6 (continued)

Reference	DataSet Name	Crops	Dataset Size	Link
[20]	Plant pathology 2020—FGVC7	Apple	Total: 3642 Images Dataset for apple scab, cedar apple rust, and healthy leaves	https://www.kaggle.com/c/plant-pathology-2020-fgvc7/data?select=images
[21]	IPM Images	Tomato Potato Pepper Pea Soya Beans Turnip rice Carrot Onion Garliccorn	Tomato:1957 Potato: 3745 Pepper:771 Pea: 7516 Soya Beans:2678 Turnip:110 Rice: 430 Carrot:1916 Onion:1827 Garlic: 847 Corn:464	https://www.ipmimages.org/browse/Areathumb.cfm?area=63
[22]	OSF	Maize	Maize:123,957 This dataset consists of 18,222 real image of maize and 105,735 annotated by human expert	https://osf.io/p67rz/
[23]	UCI	Soybean Rice Leaf	Soybean: 307 The dataset is classified into 19 classes. There are 35 attributes, some are nominal and some ordered Rice Leaf:120	https://archive.ics.uci.edu/ml/datasets/Soybean+(Large)
[24]	APS Images	Tomato Potato Pepper pea	Tomato: 366 Potato:501 Pepper:18 Pea:671 Soya Beans:325	https://imagedatabase.apsnet.org/

Table 7 Mathematical formulation of features

Features	Mathematical Formulation
Contrast	$\sum_i \sum_j (i-j)^2 g_{ij}$
Correlation	$\frac{\sum_i \sum_j (ij) g_{ij} - \mu_x \mu_y}{\sigma_x \sigma_y}$
Energy	$\sum_i \sum_j g_{ij}^2$
Homogeneity	$\sum_i \sum_j \frac{1}{1+(i-j)^2} g_{ij}$
Mean	$\sum_{i=0}^{L-1} g(i) P(g(i))$
Standard deviation	$\sqrt{\sum_{i=0}^{L-1} (g(i)-M)^2 P(g(i))}$
Skewness	$\frac{1}{S^3} \sum_{i=0}^{L-1} (g(i)-M)^3 P(g(i))$
Entropy	$\sum_{i=0}^{L-1} P(g(i)) \log_2 P(g(i))$
Kurtosis	$\frac{1}{S^k} \sum_{i=o}^{L-1} (g(i)-M)^k P(g(i))$
RMS	$\sqrt{\frac{1}{L*L} \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (g(i,j)-I)^2}$
Cluster shade	$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i+j-\mu_z-\mu_y)^3 \times P(i,j)$
Cluster-prominence	$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i+j-\mu_x-\mu_y)^4 \times P(i,j)$
Sum of squares	$\sum_{i=0}^{G-1} \times \sum_{j=0}^{G-1} (1-\mu)^2 P(i,j)$

employed to deliver extremely good solutions for optimization issues and search problems. GA works on the logic of genetic structure and function of the chromosome of the population and uses the principle of survival of the fittest. The population is passed through successive generations in a sense that at each generation, suitable results are permitted to continue, to exist and reproduce. In contrast, the unfit ones are forced to flash out. Each result, additionally known as a chromosome, symbolizes a possible feasible solution to the problem being solved. The degree of correctness and quality of a particular solution in solving the problem is known as fitness. From a random set of solutions, an initial population is generated. Different types of probability distribution, such as uniform distribution or a random selection from a population, can be used for selection so that the best candidate has the highest probability of being picked. Crossover is usually applied on to the chosen sets of chromosomes with a probability corresponding to a predefined crossover rate. As the generation progress, the average fitness is expected to improve, and the best individual solutions among the generation are chosen as the solution to the problem.

GA can also be used for feature selection which enhances the classification performance and reduces the computations

[61]. GA transmits only the best part of chromosomes from ancestor's genetic code to the next generations, and thus, the best combinations of features result in improving the fitness of their ancestor's chromosomes.

3.5.3 Support Vector Machine (SVM)

SVM is a non-probabilistic algorithm, formally defined by separating hyperplane [62]. SVM can provide a feasible solution for both Classification as well as Regression problems but is mostly used as a classifier. The use of SVM has been seen in numerous applications ranging from particle identification, plant disease identification, text categorization to engine knock detection, bioinformatics, and is also referred to as the Kernel method. There are several well-known kernel functions like Radial Based Function (RBF), Linear, nonlinear, polynomial, and Gaussian kernel that can be applied for different classification problems [63]. The N-fold cross-validation method is used to measure performance.

SVM is of two kinds, Linear SVM and Non-Linear SVM. Linear SVM is used for linearly separable data, i.e. separates the given data into two groups. Here, the hyperplane is a straight line, as shown in Fig. 4 This variation of SVM is used when there are only two classes as output, like a healthy and unhealthy leaf.

Non-Learning SVM is used for Non-linear separable data for classification. Here given data cannot be categorized using a single straight line, as shown in Fig. 5. However, the data can be converted to linearly separable data in a higher dimension. A multi-class SVM classifier is used to classify more than two class data [57].

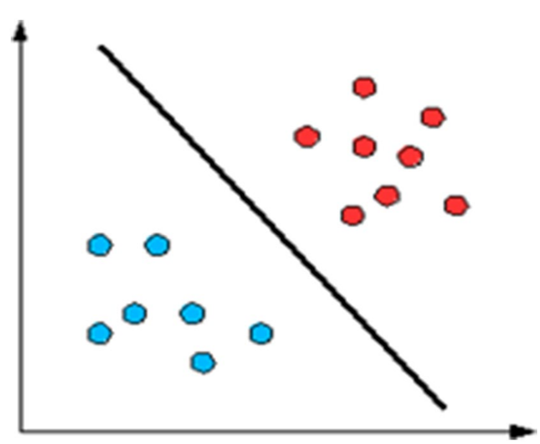


Fig. 4 Linear SVM

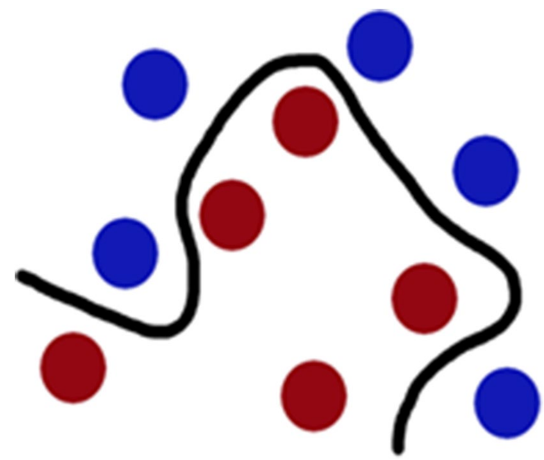


Fig. 5 NON-Linear SVM

3.5.4 Decision Tree

A decision tree is another robust and successful algorithm for classification and regression tasks. It is mostly used for applications like medical diagnosis, speech recognition, character recognition, etc. [50, 90]. It is capable enough to handle the high dimensional data. The random forest is made up of multiple decision trees. It summarizes the outcome at the training time of all the decision trees and outputs the class in case of classification problem and prediction in case of a regression problem. It surmounts the decision tree algorithm's overfitting problem.

A tree can be said to be learned by dividing the source into subsets based on a value-test attribute. This process is repeated recursively on each derived subset, which is referred to as recursive partitioning. The recursion is completed if the subset at a node has the same variable value or if its split no longer affects the value of the predictions.

By combining many decision trees, a Random forest can be constructed by actually summing up the outcomes of all at the time of training, which could further be used to solve classifications as well as regression problems [64].

3.5.5 Artificial Neural Network (ANN)

ANN is biologically inspired by the structure and function of the human brain and also known as parallel distributed processing systems [65]. ANN consist of multiple number of nodes, also called as neurons, which are interlinked with each other and allows the communication between them. These neurons operate in parallel and acts as simple processors. ANN is feedforward networks and uses learning algorithms to adjust the biases and weights parameters with a purpose to acquire the desired network output. An activation function is used in a neural network which is basically a special function that decides when and which neuron to fire an output. The ANN architecture is constituted of 3 layers, (a) Input layer, (b) Hidden layer, and (c) Output layer, as shown in Fig. 6.

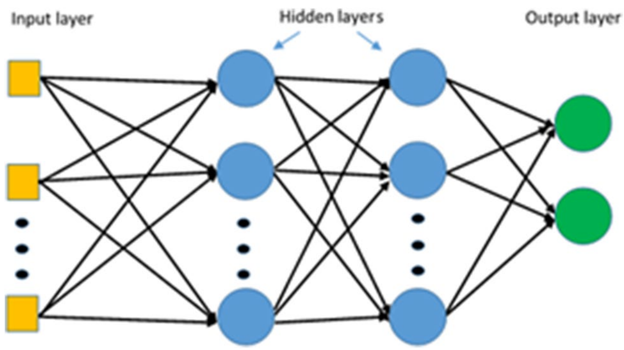


Fig. 6 Artificial neural networks

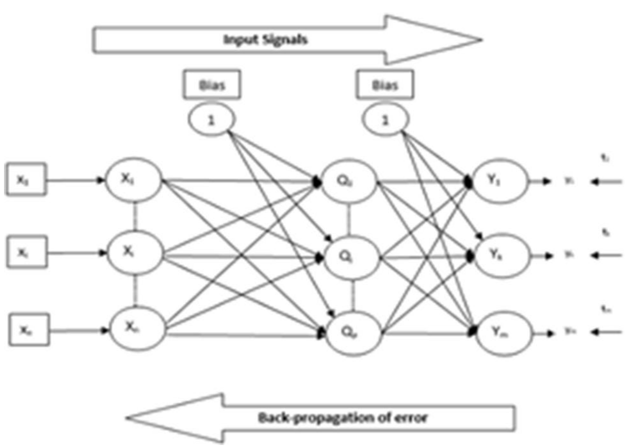


Fig. 7 Backpropagation ANN

A single layer is composed of a number of nodes connected with nodes of other layers, but have no connection within a layer. Input data is fed through the input layer, which is then forwarded to the hidden layer. The Input Layer is exposed to the outside world. The hidden layer consists of a single or multiple layers that are accountable for information processing and acts as the heart of the model. The last layer is the output layer which receives information from the hidden layer and sends the output to the outside world.

BPNN has the same architecture as that of ANN, but the major distinction is that in the case of BPNNs, the information can propagate backwards also in addition to the forward direction. The architecture of BPNN is shown in Fig. 7. Two processes take place in BPNN. One involves moving of data in the forward direction, while the other involves the back-propagation of errors to minimize the error and to enhance the learning rate [66, 67]. Backpropagation algorithms are often used for training the ANN models. ANN and BPNN can be used to solve classification, clustering, regression, pattern recognition, dimension reduction, structured prediction, machine translation, anomaly detection, decision making, visualization, and computer vision problems.

3.5.6 Convolutional Neural Network (CNN)

Deep Learning (DL) is a sub-field of machine learning that makes uses of ANN. DL decreases the task of generating a new feature extractor for every new problem. ANN is the core of deep learning methodologies. CNN based classifiers can be directly trained using raw images without the intervention of humans in feature extraction.

Recent advancements in Computer Vision, Hardware technologies and DL have enabled the researchers to considerably enhance the accuracy of disease detection and other classifications category models. CNN has shown outstanding performance in agriculture, automation, and robotics, artificial intelligence (AI), audio, video recognition, and image recognition.

CNNs are feedforward neural networks wherein data moves from the input layer to the output layer. CNN's architectures consist of input, hidden, and fully-connected (output) layers, as displayed in Fig. 8. The hidden layers are Convolutional, ReLU (rectified linear unit), and Pooling layers which are assembled to form a single network.

The pixel values of the image are fed in the form of an array into the Input layer [68]. The feature extraction work is performed automatically by the hidden layers on its own. The convolutional layers play a role of a feature extractor. The convolution operation is performed between the image matrix and filters.

ReLU introduces non-linearity into the network by setting all negative pixels to zero with the help of ReLU activation function. In the Pooling layer, different filters are used to reduce the dimensionality of features. Before forwarding the feature map to the fully connected output layer, flattening of high dimension feature map into a single linear vector is performed. It is then fed to the next layer.

Finally, the classification of images into different predefined classes is done using fully connected layer. CNNs depend upon numerous hyperparameters, including the number of epochs, batch size and learning rate. The training performance and learning pace of a model can be intensified with the help of optimizer. Adam optimizer is the most widely used optimizer.

In recent decades, there is a remarkable advancement in CNNs architectures, supervision components, regularization

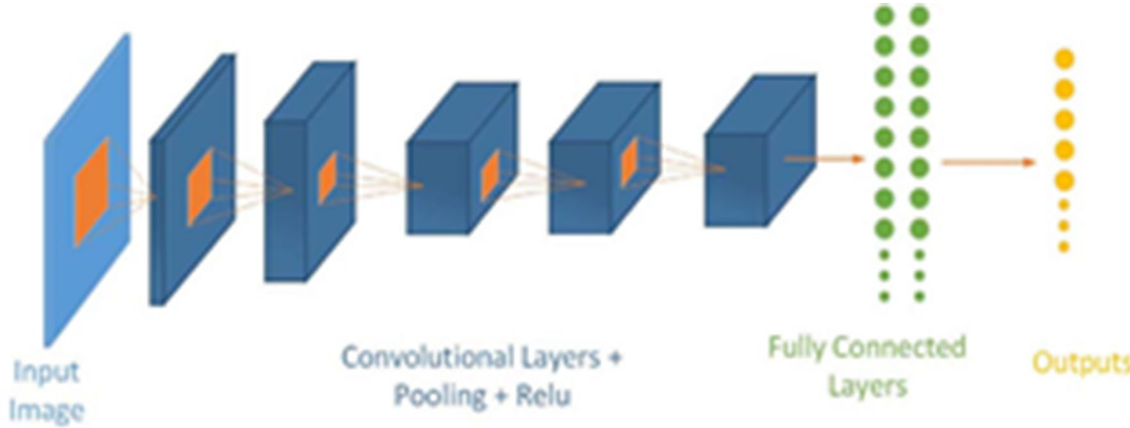


Fig. 8 Convolutional neural network

mechanisms, optimization techniques, and computations for feature extraction and transformation as well as for pattern analysis and image classification [68].

CNN sometimes can suffer from issues like overfitting, which arises due to fewer available data items to train the model and can be overcomed by the concept of data augmentation in which multiple images are generated from the single available image by cropping, using different angle of the available image.

Making and preparing a CNN model without any preparation is a dull procedure when contrasted with the utilization of existing profound learning models for different applications to accomplish the greatest exactness. So, relying upon the application, different models can be utilized or retrained, and this concept is known as transfer learning. Fine-tuning means change in terms of quantity of convolutional layers, variety of filters, stride window length, filter size, max-pooling, and including dropout among layers. There are a number of pre-trained models for DNN available in [69] Keras, which are used for prediction, feature extraction, and fine-tuning. The performance of CNNs models highly relies upon their architectures. Some well-known CNN architectures available, which shows better performance in image classification are AlexNet [70], VGG-16 [71], GoogleNet [72], ResNet [73], DenseNet [74], Genetic CNN [75], SqueezeNet [76], LeNet [77], Inception [77], MobileNet [78], and Xception [79].

4 Existing Machine Learning based Classification Techniques for Plant Diseases

In this section, various plant disease detection and classification frameworks using Machine Learning, Deep Learning, Computer Vision and Transfer Learning based techniques, available in the literature, are presented.

4.1 Classification of Tomato Crop Diseases

Tomato is one of the widely consumed crops, especially in India. Its significant constituents are vitamin E, vitamin C, and beta carotene. They are additionally plentiful in potassium, a significant mineral for good wellbeing. It is a very popular vegetable and is rich in nutrients. However, tomato crop frequently gets infected by different diseases because of bacteria, microorganisms, and pests.

In order to categorize and classify the infected and healthy tomato leaves, authors have proposed an SVM based model in [4]. The model was trained with 400 healthy and 400 infected tomato leaf images. 36 training features were extracted for classification by SVM, and for getting the enhanced performance, exceptional kernel functions like Linear kernel, RBF kernel, MLP kernel, and Polynomial kernel have been used. The evaluation of the proposed SVM based model was carried out with N-fold cross-validation. Through the usage of a linear kernel, the model was able to achieve high accuracy of 99.83%.

A DCNN based Neural Network based model for the identification of 10 different infections of tomato leaves has been proposed by authors in [57]. AlexNet and SqueezeNet based transfer learning architectures were used to run on a real-time robot for automatic detection of plant diseases in fields or greenhouses. AlexNet and SqueezeNet networks were trained and validated on the Nvidia Jetson TX1. The Caffe framework was used to train Convolutional neural networks, and implementation was done using Python. PlantVillage dataset was used to train the model, in which 10 classes, i.e., bacterial spot, early blight, late blight, leaf mold septoria leaf spot, spider mites, target spot, mosaic virus, yellow leaf curl virus and healthy, were considered for classification. AlexNet was able to give an accuracy of 97.22 %, while, SqueezeNet shows an accuracy of 94.3 % as the size of AlexNet is much larger and has a higher number of hidden layers than SqueezeNet.

Another system aimed to disclose nitrogen and potassium scarcity in tomato plants was proposed by authors in [34], which can diagnose the disease even about one week earlier than experts can identify the signs and symptoms. The image sample of tomato leaves was taken daily from the tomato field to record the changes until symptoms of the disease become visible to the naked eye. The system was trained using color and texture features. Fuzzy K-NN is used as a classifier, and GA is used for feature selection. The dataset is composed of three classes: normal, nitrogen, and potassium deficient. From each class, 40 images were used for training, and 40 images for testing the model. The color features and percent histogram were used to extract color features, and yellow-pixels in leaves images indicate nitrogen deficiency. Similarly, the intensity features were used to differentiate the healthy and potassium deficient leaves. The accuracy in classification of normal, nitrogen-deficient, and potassium-deficient leaves of the model were 92.5%, 85%, and 82.5%, respectively.

CNN based classifiers can be directly trained using raw images without the intervention of humans in feature extraction. Another system was developed using deep-learning for real-time tomato diseases and pests recognition by a group of researchers in [80]. The model can detect the following diseases and pests: Gray mold, Canker, leaf mold, Plague, Leaf miner whitefly, and powdery mildew, and also detects low temperature and nutritional deficiency of the tomato plant. Dataset consists of 5000 image samples that were collected directly from the tomato crop field. The heart of the proposed system was the deep learning architecture on which different feature extractors were adapted to perform said tasks. The deep architectures R-FCN with ResNet-50 feature extractors generates the performance of 85.98% mAP (mean average accuracy), and Faster R-CNN with VGG-16 yields 83% mAp.

The authors in a work in [81] have proposed a CNN based model that identifies whether a leaf of tomato and apple is healthy or infected. The model consists of 4 convolutional layers, each followed with the aid of pooling layers and two completely connected dense layers and a sigmoid function. The model was trained and tested with 3663 image samples taken from the PlantVillage dataset. The system shows good accuracy of 88.7%.

The authors in the article [4] have proposed an approach to identify and detect disease in tomato leaves. The authors made use of three pre-trained deep learning architectures: AlexNet, GoogleNet, and LeNet architecture, and compared their corresponding performance. After fine-tuning these architectures, the model shows the best performance in LeNet architecture. The dataset that was used contains 18160 images of samples collected from PlantVillage. The model was trained using 13360 images and tested using 4800 images. The model was successfully able to classify 10 tomato leaf diseases. The model implementation was carried out using Keras framework and achieves 94 % accuracy.

Similarly, authors in [82] presented a study on different ML techniques used to recognize and categorize healthy and diseases leaves of the tomato plant. A dataset composed of 8 classes of a total of 140 images collected from PlantVillage was used. The three global feature descriptors used in work were texture, color and shape, extracted using Haralick Texture, Color Histogram, and Hu Moments methods, respectively. In this work, authors have used different ML methods such as K Nearest Neighbors, Logistic Regression, Decision Tree, RF, SVM etc., to train the model. The following tomato diseases: Early Blight, Late Blight, Septoria Leaf spot, Spider mite, Mosaic Virus, Yellow leaf curl virus, and Target spot were detected and classified with the help of the proposed model. RF achieved an accuracy of 95.2 %, among other ML algorithms.

Authors in their work in [83] presented a comprehensive study of KNN and PNN to detect and classify the tomato plant leaf disease. The dataset used consists of 600 image samples of both healthy and infected tomato leaves collected directly from the field. The model was able to classify following tomato disease: Verticillium wilt, Powdery mildew, Leaf miners, Septoria leaf spot, and Spider mites. Firstly, the KNN classifier was trained using the GLCM feature to classify healthy and infected leaf images, then both KNN and PNN were used for further classification of diseases. PNN classification performance was 91.88 % which was found to be better than the KNN.

Similarly, in other work in [84], the K-means method was applied for feature extraction, and classification of infected leaves was done using the BPNN. The model successfully classified Septoria, leaf spot, and leaf mold infections of tomato leaves. Seven features were extracted viz, contrast, correlation, energy, homogeneity mean, standard deviation, and variance to train the BPNN. The model achieved a fascinating classification of 100%.

In another work by authors in [85], Early Blight disease identification in tomato leaves was performed using the Deep Learning. The considered dataset consisted of 4281 image samples and was collected from Plant Village. The authors proposed model aims to classify healthy and Early Blight tomato leaf. Two architectures were fine-tuned: ResNet and Xception, which are popular image classifiers, and achieved an accuracy of 99.952% in classification of healthy and Early Blight of tomato plant leaves.

Authors in [86] have used and proposed a deep learning-based model for automatic detection of viruses and pests on tomato leaf. The model was trained with 880 images and tested with 580 image samples. The model was able to classify healthy and ten diseases of tomato leaf, i.e., bacterial spot, Early Blight, Late Blight, leaf mold, Septoria Leaf spot, Spider mite, target spot, Mosaic Virus, Yellow leaf curl virus, and gray spot. In this work, the VGG16 along with the SVM, was applied to detect diseases from an image of tomato leaf, in which VGG16 was used to extract features, after which categorization was performed using SVM. The second Fine-tuning original VGG16 based classification model was used for the classification of infections in tomato leaves, and their average classification accuracy was found to be 88% and 89%, respectively.

Another work by group of authors in which various fine-tuning pre-trained models for plant disease classification and symptoms visualization was presented in [87]. The system was trained on a large dataset of PlantVillage, containing a total of 14,828 images of tomato leaves infected with nine diseases. In this work, authors fine-tuned and compared the performance of pre-trained deep learning architectures, AlexNet, and GoogleNet with a shallow based model, SVM, and Random Forest. The model classifies nine tomato diseases easily, i.e., Bacterial Spot, Early Blight, Late Blight, Leaf Mold, Septoria Spot, Spider mite, Target Spot Tomato, mosaic virus, and Tomato Yellow Leaf Curl Virus, with high accuracy and also visualize its symptoms. The Result of shallow models shows 95.476% accuracy, while deep models show very high accuracy of 99.918% in classifying the tomato disease.

Authors in [88] have presented a study of different fine-tuned CNN architectures for recognizing diseases in tomato leaves. The PlantVillage dataset consists of 14,903 images of diseased, and healthy tomato leaves were used in this work. Ten class of tomato diseases in the dataset, i.e., Yellow Leaf Curl Virus, Target Spot, Spider mites, Septoria leaf spot, Mosaic virus, Leaf Mold, Late blight, Healthy, Early blight, and Bacterial Spot, were identified in this research work. The authors carried experiments on four famous transfer learning-based architectures, i.e., LeNet, Xception, ResNet50, and VGG16, and their corresponding obtained classification accuracies to identify the disease of the tomato leaves were 96.27 %, 98.13 %, 98.65 %, and 99.25 %, respectively. Among all, VGG16 was found to be showing better performance.

4.1.1 Discussions

To identify and classify the disease from a leaf, two tasks are required to be performed: Firstly, Image Processing techniques like pre-processing, segmentation, etc. are applied on the considered image samples, and then machine learning techniques are applied to the images. For nutrients deficiency and the potassium deficiency of tomato leaves, Fuzzy K-NN was found to be quite useful and accurate. On the other hand, SVM and CNN show better performance in disease detection and classification of tomato leaves followed by K-NN and random forest algorithms. The CNN achieve maximum accuracy of 99.25% in detecting 10 diseases of tomato leaves, Bacterial Spot, Early Blight, Late Blight, Leaf Mold, Septoria Spot, Spider mite, healthy, Target Spot Tomato mosaic virus, and Tomato Yellow Leaf Curl Virus.

Also, in most of the existing research work in the literature related to the classification of disease and infected tomato leaves, the PlantVillage dataset was the most used dataset for the training and validation of the proposed machine learning models. Description of the above literature concerning tomato leaf disease detection is also presented in Table 8.

4.2 Classification of Potato Crop Diseases

Potato is a fourth-largest crop cultivated throughout the world and is also one of the most affected crops in agriculture by a variety of diseases. In this sub-section, various research work in literature having their focus on the diagnosis of Potato crop`s disease is presented.

A comparative analysis of various Machine Learning techniques including RF, SVM and one deep learning technique, i.e. ANN, for the detection of Blight disease in Potato Leaf Images was performed by authors in [89]. The dataset for the training as well as testing of the abovementioned techniques was collected from PlantVillage dataset and Plant Pathology Department, University of Agricultural Sciences, Dharwad, India. The dataset consisted of 892 image samples of healthy, early and late blight of potato leaves. Fuzzy c-mean clustering algorithm was applied to each image to cluster into healthy and diseased clusters. The various texture features were extracted using GLCM from the defected regions. These extracted texture features were used to train

Table 8 Summary of tomato crop

Reference	Year	Disease or Deficiency	Features	Dataset	Classifier	Performance Measure Accuracy
[85]	2020	Early blight	YOLO V3 was used for object detection	4281 Tomato Leaves extracted from PlantVillage Dataset. 4141 images were used for training while 100 images were used for testing	CNN based Transfer Learning Models like RestNet, Xception	RestNet: 99.735% Xception: 99.952%
[82]	2019	Yellow Leaf Curl Virus, Late Blight, Spider Mite, Mosaic Virus, Early Blight, Septoria Leaf Spot and Target Spot	texture, color and shape	1120 Images of tomato leaves collected from PlantVillage Dataset	SVM and random forest	SVM: 22% RF: 95.2%
[83]	2019	Healthy, Leaf Miners, Verticillum Wilt, Spider mites, Powdery Mildew, and Septoria Leaf Spot	Textural Features	Self-created dataset of 600 images of Tomato leaves	KNN and PNN	91.88%
[84]	2019	Septoria leaf spot and leaf mold	Feature Extraction by K-Mean Clustering and 6 features were collected	PlantVillage Dataset	Back propagation neural network	100%
[81]	2019	Healthy and infected	Auto Feature Selection by CNN	3663 image dataset	CNN	88.7%
[88]	2019	Tomato Target Spot, Mosaic Virus, Septoria Leaf Spot, Yellow Leaf Curl Virus, Spider Mites, Late Blight, Early Blight, Bacterial Spot, Leaf Mold	Auto Feature selection by CNN	14,903 images from PlantVillage Dataset	CNN	99.25
[4]	2018	10 Disease categories of tomato leaves	Auto Feature Selection by CNN	18,160 images from plant village dataset	CNN	94%
[80]	2017	Leaf Mold, White Fly, Gray Mold, Plague Leaf Minor, Powerdery Mildew Low Temperature, and Nutritional deficiency	Auto Feature Selection by CNN	5000 images were capture from areas of Korean Peninsula from which 43,398 image samples were generated after image augmentation	Faster Region based CNN (FR-CNN), Region based Fully CNN (RFCNN), Single Shot Multibox Detector (SSD)	FR-CNN: 83.06% RFCNN: 85.59% SSD: 85.23%
[86]	2017	10 Diseases i.e. Early Blight, Late Blight, Leaf Mold, Target Spot, Mosaic Virus, Yellow Leaf Curl Virus, Gray Spot, Spider Mite, Septoria Spot,	VGG16 was used for Feature Extraction and Selection	7040 images captured from fields of China	SVM and transfer learning methods like VGG16	89%
[87]	2017	9 Diseases found in Tomato leaves	Auto Features extraction for deep models while manual feature extraction for shallow models	14,828 images of tomato leaves collected from PlantVillage Dataset	AlexNet, GoogleNet, SVM, Random forest	AlexNet: 98.66% GoogleNet: 99.18% SVM: 94.53% RF: 95.46%
[57]	2015	Healthy and infected	GLCM based Features	800 images of Tomato Leaves	SVM	99.83%
[34]	2011	Nitrogen Deficient and Potassium Deficient	Genetic Algorithm was used for extraction of Textural and multi-angle color features	240 images of tomato leaves were self captured	FUZZY K-NN	82.5%

the considered classifiers, i.e., SVM, RF and ANN. Based upon the simulation results, it was observed that the classification accuracy of ANN in disease detection was the highest among the other considered machine learning techniques with the accuracy of 92%, followed by SVM with 84% accuracy and RF with 79% accuracy.

A supervised learning-based efficient automated disease management tool for diagnosing potato infections like late blight and early blight diseases from potato leaf images was presented by a group of authors in [51]. In this work, multi-class SVM (MSVM) classifier was trained using color and texture features. GLCM was used for extracting the statistical texture features and histograms for color features. The image samples were taken from the online PlantVillage dataset and consisted of three potato leaves classification classes, i.e., healthy, late blight, and early blight, where each class contain 100 images. For training and testing the model, 60 % of images were used for training the models while 40 % for testing the model`s performance. Based on the simulation results, the author`s proposed MSVM based model was able to generate excellent accuracy of 96%.

A model based on Computer Vision was developed to detect defects and diseases in unwashed potatoes using Genetic Algorithm (GA) as feature selector to classify good, rotten, and green potatoes, by authors in [61]. The histogram and co-occurrence matrix was used to extract texture features from each segment.

Authors in [90] have proposed a robust model to recognize Potato Late Blight disease using Neural Network and image processing techniques. The dataset used for the training and testing was small in size and consists of only 27 image samples, captured directly from the farm to recognize Late Blight disease. The FCM algorithm was used for the clustering the image of potato leaf, and texture features were extracted from each cluster using GLCM, which was then used to train the BNN classifier. The proposed model achieves an accuracy of 93% in detecting Late Blight disease.

In other work, a machine learning-based automatic system proposed to classify potato leaf diseases using distinct color features combined with texture features, with the help maximum-minimum color difference method was presented by authors in [91]. The segmentation of image samples was carried out with the k-means clustering algorithm, and classification was done using Euclidean distance. 300 potato leaf images were obtained from PlantVillage dataset. The author`s considered system for the classification of infected potato leaves was the combination of two methods MCD and TTF (three texture features), and was capable of achieving an extreme category accuracy of 91.67% in classifying late blight, early blight, and healthy Potato leaf images.

Authors in [92] introduced the feasibility of application of CNN and AlexNet architectures on potato and mango leave disease detection. The PlantVillage dataset containing a total of 4004 potato images was used for the model training and testing. 3523 images were used for training & validation, while, 481 images were used for testing the model. Data Augmentation step was performed on the images in order to get good accuracy in classification. Based upon the models simulation and analysis, it was observed that the AlexNet architecture takes more time in training the model since it has an abundance of layers in contrast to the CNN Architecture. However, AlexNet architecture was able to achieve higher accuracy of 98.33%, as compared to CNN 90.85%.

Similarly, in another work, a DCNN based approach to detect the common diseases affecting the potato crops, namely late blight and early blight from the healthy potato leaves was presented by authors in [14]. The dataset for training and testing the model consists of 2250 image samples of potato leaves, in which 1000 image samples were of late blight and early blight, while 250 images of healthy potato leaves. The DCNN based architecture for the diagnosis of infected potato leaves was developed from scratch and consisted of 11 layers, i.e. 2 fully connected dense layers, 5 convolution layers, 1 dropout layer and 3 MaxPooling layers. The prediction performance of the model yields high accuracy of 98.33%.

4.2.1 Discussions

It can easily be observed that late blight and early blight disease in potato leaves is explored in almost all the papers in literature. Pretrained Transfer Learning techniques like AlexNet architecture can be the easy choice for not only achieving high detection accuracy, but it also saves both costs as well as coding efforts. The dataset mostly used by all authors is PlantVillage [17] which consists of image samples of potato disease infections like healthy, late blight and early blight images. Amongst the results from available classifiers, CNN presents the very best accuracy of 98.33% followed by BPNN, MSVM, ANN, SVM and RF methods.

Table 9 manifests the review of several classifiers implemented for classification of potato disease detetction in which crucial parameters like, traning features, dataset used, and their corresponding accuracies, are presented.

4.3 Classification of Apple Crop Diseases

Apples are rich in minerals and grown throughout the world. Currently, there are more than 3,000 varieties of apples. Apple is a popular commercial crop and has high nutritional and medicinal value. However, apples often get infected by different diseases which impact its yield and production.

The most common diseases of apple fruit are apple blotch, apple rot, and apple scab, which can easily and accurately be detected using various machine learning techniques. The

Table 9 Summary of potato crop

Ref	Year	Disease or Deficiency	Features	Dataset	Classifier	Performance Measure Accuracy
[91]	2019	Late blight, Early blight, and Healthy potato leaves	Texture and Color based Features extracted by Mean of Maximum-Minimum Color difference and Euclidian Distance method, respectively	300 images of potato leaves collected from PlantVillage Dataset	Classification using Maximum-Minimum Color Difference and Euclidian Distance	91.67%
[92]	2019	Healthy and infected with early blight	Automatic Feature Selection and Extraction done by CNN Model	4004 images collected from PlantVillage dataset	CNN and transfer learning based AlexNet model	CNN: 90.85% AlexNet: 98.33%
[14]	2019	Late blight, Early blight, and healthy	Automatic Feature Selection and Extraction done by CNN Model	2250 images of potato leaves collected from PlantVillage Dataset	DCNN	98.33%
[89]	2017	Late blight, early blight, & healthy	Textural features were used	892 potato leaves images of which few were collected from PlantVillage dataset and few from Pathology Dept., University of Agricultural Sciences, Dharwad, INDIA	Random forest (RF), SVM, & ANN	RF: 79% SVM: 84% ANN: 92%
[51]	2017	Late blight, early blight, & healthy	Textural and color-based Features	300 potato leaves collected from Plant-Village Dataset	Multi-Class SVM	95%
[90]	2014	Late blight	Texture	27 images, which after applying Fuzzy C-Means (FCM) generated 340 Region of Interest (ROI)	Back propagation neural network	93%
[61]	2009	Healthy, Rotten, and Green potato	Textural features were used	5.040 pictures from 47 bags of potato were captured	Nearest neighbor with GA as optimization	86%

scab is a more dangerous and common disease in apple crop and has a drastic effect on both quality and quantity of apple`s crop.

Authors in [93] have used MSVM for detection of infected apple crops using the leaves of apple. The author`s proposed machine learning model was trained for the detection of four types of apple`s diseases, viz. (i) Apple Blotch, Apple Rot, and Apple Scab. The dataset used by authors was collected from Google Images in which images were collected against four classification classes, i.e. Apple Blotch, Normal Apple, Apple Rot, and Apple Scab. The dataset contains a total of 320 images of apple fruits in which each class contains 80 image samples. 70 images from each class were used to train the model and rest were used for testing the classification model. For classification of infected and healthy apples, steps like background removal from the images and region of interest-based segmentation were performed using the K-means method. The Global Color Histogram and color coherence vector were used to extract color features, while the Local Binary Pattern and Completed Local Binary Pattern methods were used to extract texture features. For the extraction of shape features, Zernike moments was used. These mentioned extracted features were used to train the MSVM based disease prediction model. The system was showing better performance when all the features were integrated with a classification accuracy of 95.94%.

Similarly, in another work, MSVM was used as a set of binary classifiers for the automatic detection and classification of apple fruit disease by authors in [94]. The author`s proposed model was successfully able to classify apple rot and apple scrub disease. The dataset used for the training as well as evaluating the performance of the proposed machine learning-based MSVM model was comprised of 300 image samples of apple fruits. In this procedure, 100 image samples were used for each training and testing per class, i.e. normal apples, apple rot, and apple scrub. The features like color, shape, and texture were used for training the MSVM classifier. The different kernel function like quadratic kernel, polynomial kernel, linear kernel, and Radial kernel basis function was used for classification by the MSVDM, and their corresponding performance was found to be 69%, 72%, 79%, and 98%, respectively.

Authors in their work in [95] have proposed a Hybrid Approach for apple fruit disease detection and classification using RF algorithm. The dataset used for the training and testing consists of 320 images containing healthy and 3 classes of diseased apples, i.e. apple scab, apple rot, and apple blotch images, where each class had 80 images. The model was implemented in three phases. In the first phase, image Segmentation using K-means was performed to find the infected regions in an image. In the second phase, Feature Extraction was performed to extract features like GCH, CCV, and texture-based features like Local Binary Pattern (LBP), Completed Modelling of Local Binary Pattern (CLBP), and Gabor. The whole RF was trained by fusing features like Gabor, LTP, and CLBP to obtain better results. Finally, the classification of the diseases was carried out by RF, and an accuracy of 80% was obtained.

Similarly, authors in their work in [96] have proposed a system to identify apple leaf disease using GA and correlation as a feature selection (CFS) method and SVM as classifier. The dataset consists of 3 classes, and each contains 90 image samples. The infected portion of the image was segmented by region growing algorithm. 38 features of color, texture, and shape were derived from each portion of the image. In order to improve the accuracy of the model, the most valuable features were selected by GA and CFS. The system was able to identify three apple leaf diseases, powdery mildew, mosaic, and rust with a high accuracy of 90%.

Similarly, in another work, SVM based classifier was used for plant disease classification [97]. Otsu method and k-means clustering were used for the segmentation of the region of interest in the considered image. Crucial features like color, texture, and morphological features were extracted for disease detection and based upon the results, it was observed that morphological features were giving better results than the other features. Finally, after the extraction of the features, the classification of the plant leaf image was carried by the SVM classifier.

A deep learning-based ANN model was proposed by authors in [98] to detect diseases in grapes, apples, and Pomegranates using features like Color, morphology, texture, structure of hole, etc. The K-Means clustering method was used for the segmentation of images. Grape diseases like Black rot, powdery mildew, and downy mildew were considered, whereas, apple diseases like Apple Scab, Apple Rot, and Apple Blotch were considered in this work by authors. The model was able to attain an accuracy of 90%. On the classification and identification of the disease from which the crop is suffering, the model also has the support for a treatment recommendation.

The use of CNNs in a model proposed by authors in [99] provides a feasible real-time solution for early diagnosis and detection of apple leaf diseases automatically with high accuracy. This system helps to detect five types of apple leaf diseases, viz. Alternaria leaf spot, Brown spot, Mosaic, Grey spot, and Rust, each having the affect on the qualitative and quantitative aspect of apple production. In this work, the CNN-based INAR-SSD model was trained using 75% of an apple leaf disease dataset (ALDD) consisting of 26,377 images of diseased apple leave, while, remaining 25% was used for the testing purpose. Image annotation and data augmentation are the two steps that were employed on each image to increase the performance of the whole model. In this work, the concept of transfer learning was also used in which re-training and testing of various existing pretrained different deep convolution networks like AlexNet, GoogLeNet, InceptionV3, ResNet-101, ResNet-50, ResNet-34, ResNet-18, VGGNet-16 were performed. The recognition performance of these existing transfer learning-based models was compared with that of the proposed INAR-SSD model on ALDD dataset. The detection performance and speed of INAR-SSD was found to be 78.80% mAP, and 23.13FPS (Frames Per Second), respectively.

In another work, CNN based on a robust model was used to detect black rot of apple leaf disease [36]. The dataset consists of a total of 35,000 images of healthy as well as infected plant leaves collected from PlantVillage. The model was able to detect and recognize 32 different plant varieties and plant diseases using CNN. The model achieved a high accuracy of 96.5% to detect and recognize plant diseases.

A comparative analysis of the performance of CNN, InceptionV3, VGG16 and traditional ML algorithms for apple Leaf Disease identification was carried out by authors in [100]. The Plant Village dataset was used for the training as well as testing of the models, which consisted of 2486 image samples. The author`s proposed CNN architecture consists of three Convolution layers, three Max-Pooling layers, and 2 dense layers. In this work, three apple diseases viz. apple scab, black rot, and Apple Cedar were classified from the healthy apple image samples. The image augmentation to increase the dataset size was carried to get rid of overfitting. ML algorithms like SVM, Decision Tree, Logistic Regression, K-NN, LDA, Naive Bayes, and RF were also applied on the same dataset and were accuracies of 85%, 80%, 86%, 96%, 93%, 70%, and 95% respectively. The accuracy of the author`s proposed CNN model was 99%, and of pre-trained models, VGG16 has obtained the accuracy of 97%, while, the accuracy of 89.5 % was obtained in case of Inception V3.

4.3.1 Discussions

From the literature, it is evident that the most studied apple disease is Apple Scab, followed by Apple Rot and Apple Blotch. Deep learning models like DCNN showed the fascinating performance while machine learning methods like MSVM, ANN, SVM, and RF have also shown impressive performance in disease identification and their classification. DCNN can be seen as the preferred classifier and may help in improving the effectiveness of a system on large datasets.

Table 10 summarizes all the studies considered for apple disease detection and classification.

4.4 Classification of Rice Crop Diseases

Rice is considered a staple food in many countries and is one of the significant sources of food in India. Henceforth, the disease-free cultivation of rice is extremely crucial to guarantee the financial development of the nation as well as of the farmers. Paddy Plants are prone to viral, bacterial, and parasitic infections. Analysis, as well as detection of rice diseases, is essential as it can aid in the early detection of the diseases in rice crops followed by their treatment, which will considerably improve the quality and quantitative production of rice.

Authors in [101] have offered a machine learning-based automated method for recognition of fungal diseases in rice, viz. rice blast and brown spot diseases of paddy leaves using KNN as a classifier. The dataset that was used for training and testing of the proposed model contains 330 images captured directly from the paddy field. Otsu method was used in the segmentation of image samples to extract diseased parts of paddy leaves. Using Geometrical features, kNN was trained on 60% of images, and the remaining 40% of images were used for testing. The proposed model by authors was able to achieve the testing accuracy of 76.59% in the detection of diseased paddy leaves.

In another work, a machine learning-based classifier was proposed by authors in [102] to classify the paddy leaf diseases using Gaussian Naïve Bayes. Each pixel of the infected part of the paddy leaf was represented in terms of Red, Green, and Blue value. The blue pixels were discarded to ease the computations. The percentage of RGB value of the diseased portion of paddy leaf as a parameter was used in this model to train Gaussian Naïve Bayes method and ANN. The dataset used in this model consists of a total of 60 image samples of rice leaves. The Gaussian Naïve Bayes classifier was trained with 75% of image samples, while the rest 25% were used for testing. In this proposed disease detection model, the processing speed was found to be higher than other existing techniques in literature as in this model, only the affected part of the paddy leaf was taken under consideration. This model was used for the detection and identification of rice blast, rice blight, and rice brown spot diseases. The classifying accuracy of the author`s proposed model was found to be higher than 90%.

Authors in [16] have come with another automatic disease detection and classification model of the paddy plant using SVM as a classifier. A total of 120 images of three different rice diseases was collected directly from the rice field for training the proposed paddy leaves disease detection. For the extraction of the diseased portion of the leaf image, segmentation was carried by making use of K-means clustering with feeding centroid values. In this model, 88 distinct and useful features were extracted from the input training images to train the model. The model was found to be successful in recognizing rice diseases like brown spot, bacterial leaf blight, and leaf smut. For the multiclass classification, the Radial Basis kernel function was used. The model was able to attain training accuracy of 93.33% and 73.33% as testing accuracy in classifying these diseases of paddy plants.

Table 10 Summary of apple crop

Ref	Year	Disease or Deficiency	Features	Dataset	Classifier	Accuracy
[36]	2019	Black rot	Automatic Feature Extraction by CNN	Plant Village dataset was used with 35,000 total images	CNN	96.5%
[99]	2019	Brown Spot, Alternaria Leaf Spot, Mosaic, Grey Spot, and Rust	Automatic feature Extraction by DCNN	2029 images of 5 classes of apples disease were collected to form Apple Leaf Disease Dataset (ADNN). 26,377 images generated using image augmentation techniques	DCNN based transfer learning model, VGGNet and SSD	97.14%
[100]	2019	Apple Scab, Black rot, Apple cedar	Automatic feature Extraction by CNN	2486 Apple Images from Apple Dataset of PlantVillage. After applying image augmentation, 4000 images were generated	CNN	99%
[96]	2017	Powdery Mildew, Mosaic and Apple Rust	Color, Shape, Texture based Features were extracted. GA & Correlation Feature Selection (CFS) used for Feature Selection	90 images for each disease class was collected	SVM	94.22%
[95]	2016	Apple Rot, Apple Scab, Apple Blotch, and Normal Apples	Color and texture based features were extracted	320 images with 80 images of each class (Source not mentioned)	Random forest	80%
[93]	2016	Apple rot, Apple Blotch, Apple Scab and Normal Apples	Color, Shape and Texture based Features were extracted	320 images with 80 images of each class extracted from google images	Multi-Class SVM	95.94%
[98]	2015	Apple Rot, Apple Scab, Apple Blotch	Texture, Color, Morphology, and Structure of Hole	Not Mentioned	ANN	90%
[94]	2015	Apple Scrub and apple rot	Color, Shape, Texture	Not Mentioned	Multi-Class SVM with different kernels	Multi-Class SVM with radial basis Kernel: 98%

Similarly, in other work, a proposed automatic model to identify different diseases of the rice plant using SVM was proposed by authors in [103]. The dataset considered for evaluating the performance of the proposed SVM based model consists of a total of 145 images samples, in which 30 images were of the healthy leaf, while infected rice leaves consists of 46 images of bacterial leaf blight, 44 images of leaf smut, and 25 images of brown spot collected directly from field and segmentation were performed on the image set using K-means clustering. The SVM with Gaussian kernel function was trained using about 50 features of image samples to classify the rice disease.

In similar to another work, few authors have used SVM as a classifier to classify the healthy and blast rice leaves [15]. The dataset used for the training and testing consists of 84 healthy and 83 rice blast image samples of paddy leaves. The SVM was trained using color features and implemented with RBF kernel. The classification model was able to generate an accuracy of around 90%.

A Deep Learning based ANN model was proposed by authors in [104] to detect rice blast disease in the paddy leaves. For training and evaluating the performance of the proposed ANN-based model in classifying the infected as well as healthy rice leaves, a dataset containing 300 image samples of normal and diseased leaves of paddy was used, in which 60 % of images were used for training and the rest of them for testing purposes. K-Means Clustering method was used for Image Segmentation of image samples. The features, mean value, standard deviation, and GLCM were calculated for the healthy and infected leaves in order to train the model. The proposed ANN-based algorithm consists of five layers in which three were hidden layers that were interconnected to each other, one input layer and one output layer. The model was able to produce 99% and 100% classification accuracy for blast and healthy leaf, respectively, during training and 90% and 86% accuracy testing, respectively.

Similarly, in other work, for the automatic disease detection and classification of paddy leaves, authors in [105] made use of two machine learning methods, i.e., SVM and ANN, along with the texture and color descriptors features. The dataset considered for the training of the model was extracted directly from Rice Research Center Karjat, Maharashtra, India, and consists of 3 classes of rice leaves, namely, rice blast, brown spot, and healthy paddy leaves, where each class consists of 20 images samples. The textural descriptors using GLCM, and color moments were extracted from the infected area of the paddy leaf, which in turn results in a 21-D feature vector. Segmentation of image samples was carried with the help of the K-means algorithm. GA was used for the selection of relevant features and to discard redundant features, which reduces feature vector from 21D to 14D in order to lower the computations. The proposed model shows the classification accuracy of 92.5% using SVM and 87.5% using ANN.

Authors in their work in [67] have proposed a BPNN based deep learning model to detect the rice brown spot diseases of paddy leaves. The dataset considered for model learning and testing consists of 400 image samples of rice leaves directly collected from the field. BPNN was used because errors are back-propagated into the algorithm to increase the learning rate. The architecture of the model is composed of one input layer, three hidden layers, and an output layer. The R, G, L (red, green, and color brightness) color feature values were used to train the BPNN model. The model was able to show 90% accuracy in identifying rice brown spot and was capable of identifying other diseases as well.

Authors in [106] have performed the comparative simulation analysis on the use of different machine learning algorithms carried for automatic classification of diseases in rice plants using leaf images. Various distinct image processing algorithms were used by authors to detect the leaves as well as likely disease-induced lesions on the leaves. The model identifies and classifies leaf blight, brown spot, and leaf smut diseases, which are induced by bacterias. 144 features were extracted based on morphology and color of leaves and lesion to train the model. The dataset for evaluating the performance was collected from open-source, International Rice Research Institute. The 19 classifiers were used in which ZERoR and OneR served as baselines, while three rule-based methods: (i) PART, (ii) Decision Table, and (iii) Jrip were also used. Apart from that five tree-based methods: (i) J48, (ii) RF, (iii) LMT, (iv) Random Tree, and (v) Rep Tree, three Function-based methods: (i) SMO, (ii) SimpleLogistic and (iii) Logistic, three Bayesian methods: (i) Naïve Bayes, (ii) Bayes Net with limited parents and iii) Bayes Net with unlimited parents and three lazy methods: (i) IBL, (ii) Kstar, and (iii) LWL were evaluated, and their performance in terms of prediction accuracy was compared. Based upon the simulation results, it was found that they were able to generate classifying accuracy of 35%, 60%, 78%, 76%, 76%, 79%, 92%, 92%, 80%, 75%, 92%, 92%, 77%, 71%, 77% 82%,92%, 65%, and 70%, respectively. It is quite evident that algorithms like RF, LTM, SMO, Slog, and IBK were showing better classification performance.

An exemplified hybrid model by combining DCNN and SVM to classify diseases on paddy leaves was proposed by a group of authors in [107]. The dataset used was collected directly from the rice research field and some online resources like IRRI[108], BRKB [109], BRRI, and Plantix, and was composed of 9 classes of infected paddy leaves, viz. bacterial leaf blight, rice blast, rice brown spot, false smut, leaf smut, red stripe, leaf scald, sheath blight, and tungro. 1080 image samples of the dataset were used for

Table 11 Summary of rice crop

Ref	Year	Disease	Features	Dataset	Classifier Used	Accuracy
[106]	2020	Brown Spot, Leaf Smut, Bacterial Leaf Blight	17 out of total 144 features	Not Mentioned	Decision Table JRip ZeroR OneRJ48, Random Forest, LMT, Random Tree, REP TreeSMO, Simple LogisticNaive Bayes, IBk, KStar	92%
[110]	2020	bacterial blight, blast, brown spot and tungro	Features extracted using pre-trained CNN models	5932 Rice leaf Image Samples collected from fields of Odisha, INDIA	SVM & Transfer Learning Models	SVM: 98.38% RestNet50: 98.38%
[111]	2020	bacterial blight, brown spot, sheath rot and blast diseases	Color features like Standard Deviation and Mean, while Texture Features like GLCM	650 images of rice leaf captured from rice fields of Tamil Nadu India	DNN with Jaya Optimized Algorithm	98.9%
[112]	2020	rice sheath blight, brown spot, rice stem borer	Automatic selection using DNN	1800 image samples collected from rice field of China	DCNN	90.9%
[113]	2020	Stem Rot, False Smut, Sheath Blight, Leaf Blast	Texture, Color, Structure and Morphology	150 images	Minimum Distance Classifier (MDC), Bayes	MDC: 69.53% Bayes: 81.06%
[107]	2019	Rice Blast, Flase Smut, Bacterial Leaf Blight, Leaf Smut, Sheath Blight, Brown Spot, Tungro, Leaf Scald, and Red Stripe	DCNN was used for feature extraction & 23.5 million trainable parameters were used	1350 rice images extracted from field while some from existing online rice datasets like IRRI, BRKB, BRRI and Plantix	SVM	97.5%
[114]	2019	Rice Blast	Automatic selection using CNN	5267 Samples obtained using IoT	CNN	89.4%
[115]	2019	Brown Spot, Rice Blast	Color, Texture, Shape, Hue Saturation	300 images	RBFNN	95%
[102]	2018	Rice brown spot, bacterial blight, rice blast	RGB Value	60 rice images were collected from rice field and internet	Gaussian Naive Bayes	90%
[15]	2018	Healthy and Rice blast	Color i.e. value of the pixel	167 Image Samples	SVM	90%
[104]	2018	Healthy, rice blust	GLCM, Mean and Standard deviation	300 rice samples from field of Tamil Nadu	ANN	90%
[116]	2018	Brown Spot, Bacterial Blight, Leaf Scald, Leaf Blast	GLCM	16 Image Samples were considered	SVM	86.35%
[101]	2017	Rice blust and Rice brown spot	Geometrical Feature like Area, Perimeter, Major axis, Minor axis	330 image samples from paddy field of Karnataka, INDIA	KNN	76.59%
[16]	2017	Brown spot, bacterial leaf blight, leaf smut	88 Features based on texture, color and shape	120 image samples from rice field of Gujrat, INDIA	SVM	73.33%
[105]	2017	Healthy, blust, brown spot	GLCM, color, area	60 image samples from rice field of rice research centre Maharashtra, INDIA	SVM ANN	92.5%

Table 11 (continued)

Ref	Year	Disease	Features	Dataset	Classifier Used	Accuracy
[117]	2017	rice blast, rice false smut, rice brown spot, rice bakanae disease, rice sheath blight, rice sheath rot, rice bacterial leaf blight, rice bacterial sheath rot, rice seeding blight, rice bacterial wilt	PCA and whitening were performed to get the features	500 image samples collected from rice field of China	DCNN	95.48%
[103]	2016	Healthy, bacterial leaf bight, leaf smut, brown spot	50 features including color, texture, GLCM	145 image samples from paddy field of Gujrat, INDIA	SVM	Not Mentioned
[67]	2009	Brown spot	Color Space	400 images were taken from rice field of Ningxia Hui Autonomous Region	BPNN	90%

the purpose, out of which 270 images were used for testing the model and rest for the training of the model. The DCNN Inception-V3 architecture was fine-tuned for extracting more invariant, high-level, and relevant features of rice image samples and was then flattened by a CNN pooling layer named 'global average-pooling' to get better results. SVM was trained with all extracted features using RBF kernel. The model was able to achieve high whooping accuracy of 97.5% in the identification and classification of nine rice diseases.

4.4.1 Discussions

Based upon analyzing various existing works of researchers in literature, it was observed that most of the authors had captured images for their work from various rice fields. The authors in [107] have used some standard database such as IRRI, BRRI, BRKB, and plantix. Various dataset of rice is openly available on Kaggle and UCI database [19, 23]. CNN based models have achieved maximum accuracy in detecting and classifying the diseases in paddy leaves.

Table 11 summarizes different work for rice disease identification.

5 Various Challenges in the Use of Machine Learning and Deep Learning in Plant Disease Detection

5.1 Challenges

In the previous section, vast applicability of Machine Learning in the field of Agriculture for Plant and Crops Disease detection was presented in which various existing Machine Learning frameworks and models for disease detection in different varieties of crops proposed in the literature by researchers across the globe were discussed.

From the study conducted in the previous section, it is evident that machine learning models have great potential in the field of agriculture, especially for plant and crop disease detection. However, there exist few challenges also which requires attention and consideration that are restricting the performance of these models for the crop`s disease detection. The challenges that mostly affects the performance of the automatic plant disease detection and classification are presented in this sub-section and are as follow:

5.1.1 Unnecessary Noise and Background in Crop`s Image

Segmentation is the technique that is used to extract the infected segment from an image. However, this extraction of the infected leaf segment from an image becomes difficult when the image background contains color other than white and black, or it contains other elements like plants, leaves, soil, grass, etc. [118]. If farmers want to identify the diseases of crops in real-time from the fields, the photo may contain many elements in the background. Therefore, the system must be capable of removing all the unnecessary elements from the picture to get only the desired segment.

5.1.2 Image Capture Conditions

From literature, it is evident that the images in all available datasets are captured in a controlled environment in laboratories, and even in some work, images were created using the animation techniques. However, suppose a farmer or a capturing device deployed in the field tries to capture the image of the same object at different times of a day. In that case, it is difficult to get a similar image because of various varying factors like different light intensity, wetness, and other environmental factors. So, it is required to capture the images of the same leaf from different angles and at different intervals of times and under different environmental conditions [118]. Inevitably this will cause the size of the dataset to increase, which eventually will increase the performance and accuracy of the prediction model but at the cost of high computations and storage. The image samples in automatic disease detection and classification system are as important as blood in our body. The types of equipment used for image capturing also have a significant influence on the performance of the system. So, factors like the type of sample capturing device, light intensity, time of day when capturing, wetness, etc. can impact the prediction accuracy of the machine learning-based plant disease detection. Hence, training, as well as the real-time deployment of the automatic disease prediction model, is required to consider these aspects

5.1.3 Symptom segmentation

The symptom segmentation is an intrinsic factor in digital image processing. Segmentation becomes difficult when the number of diseases shows the same symptoms and are superimposed on each other [118]. Some symptoms do not show clear edges, and some diseases change the symptoms with time and with the change in environmental conditions. Therefore, to extract the desired segment and to correctly identify the disease with which the crop is suffering while showing overlapping symptoms of multiple diseases is also a challenge. So, alternative approaches to get the correct symptom segment is required to be adopted in addition to the existing model. The accuracy of symptom segmentation is considered as per the context in which segmentation is carried.

5.1.4 Size of Dataset

In literature, all most all the authors used only a few thousands of images to train their model. One solution provided in deep learning to increase the size of the learning dataset is data augmentation. To train the model using large datasets indeed will increase the accuracy of the disease prediction model but at the cost of the computational penalty. The situation covariate shift arises from the difference between distributions of the training data to learn the model and data on which model is applied.

5.1.5 Symptom Variations

Generally, the characteristics of disease symptoms are more prone to changes like variation in the color, shape, and size of symptoms, and only specific diseases show the same symptoms [118]. The diseases show different symptoms at different stages, which makes quantification difficult. The plant genotype, healthy tissue color variation, and environmental factors like humidity, exposure to sunlight, temperature, wind, and other meteorological phenomena have a significant influence on the characteristics of the disease symptoms. So, it becomes very pertaining to capture the images at each stage of disease symptoms and store in a database for accurately training the prediction model and for increasing its accuracy.

5.1.6 Non-Uniformity and IN-balanced Data Distribution for Different Diseases

Another challenge that can hamper the performance, as well as the accuracy of the automatic machine learning-based Plant disease detection model, is the presence of imbalanced and non-uniformity in the training dataset. It is required that all the diseases have almost equal size of infected as well as healthy crop samples for a particular disease in order to prevent the biasing of the machine learning algorithms and models.

5.1.7 Multiple Simultaneous Diseases

Commonly many methods assume that only one disease is present per image, which is not good to say as there can more than one disease, other nutritional deficiencies, and pests present simultaneously [118]. Thus, in these scenarios, it becomes hard to trace a particular disease. Another reason is that characteristics of symptoms may vary with the geographic position also.

5.1.8 Different diseases with the same Symptoms and Patterns

Various diseases have the same symptoms and patterns, which makes it very challenging in the identification of diseases [118]. Generally, the visible spectrum is used only. However, there are other spectral bands like infrared to differentiate different diseases that can be used along with the other new sensors, which is also a complex, costly, and challenging task. These techniques are still prone to errors. However, these new techniques can minimize the use of large datasets.

The leaf is the index of almost all plant diseases. Through leaf images, nutrients deficiencies in the plant and crops can also be detected. Similarly, water scarcity in plants can also be detected from their leaves by considering the factors like change in shape, color, and collapsing of leaves.

There is a dire need for sophisticated, hybrid, automatic systems that can overcome all the abovementioned challenges.

5.2 Future Work

Despite the unique and plethora of advantages offered by various image processing and machine learning techniques for plant disease detection and their classification, there still exist several limitations among these techniques.

When it comes to the use of image processing techniques, they have the ability to detect and isolate the infected segment from an image. However, new methods and techniques are needed to be developed to address various existing issues like handling of noise, presence of irrelevant background objects, etc. In this manuscript, well-acknowledgement of a number of computer vision methods and techniques which have become a new hotspot of research in this domain has been provided.

The forecasting of disease in early stages, so that appropriate precautions are taken to minimize the crop loss is imperative. There are only a few mobile-based applications and websites available for the same purpose to common masses. There exist various highly efficient and accurate models in the literature. However, these models are required to be tested and deployed in real-time mobile applications and web services so that farmers can also make use of these advanced automatic disease identification techniques in their farming by directly taking a photo of suspected leaves of plants [119]. From literature, it is quite evident that there is limited to no real-time machine learning-based system available to identify the diseases.

The recommendation of chemicals and their ratio to control the further spread of diseases on the different parts of plants after the proper identification of diseases is also an important research topic. By spraying irrelevant and inappropriate proportion of chemicals have an adverse effect on both the quality and quantity of crops and also lead to soil, water, and air pollution. It is deplorable that in various instances, farmers combine the chemicals without checking their composition, which could lead to a chemical reaction, which is more dangerous for plants. This problem can be solved by using expert systems and various machine learning-based recommendation systems.

6 Conclusion

The correct recognition and categorization of Plant diseases are essential for the prosperous farming of crops. Manual detection of diseases in crops involves various obstacles as considerable efforts in terms of cost, labor, and expertise is required to identify the crop`s disease correctly. Also, considering the factors like the scale of farming, different disease types and similar symptoms for different diseases, it becomes a challenging task for a farmer to accurately and timely detect these diseases, and this also has the adverse impact on both the yield as well as the quality of the crops. So, considering the abovementioned issues, automatic detection of Plant`s disease using various machine learning and deep learning models can be brought into use. There exist various efficient and robust machine learning techniques that can be used for the early prediction of diseases by analyzing the plant`s leaf. However, for synthesizing maximum from the capability of these machine learning models for the detection of plant diseases, it is very important to identify various trending machine learning models best suitable for these applications and also the various basic steps involved in the automatic detection of infection from which the plant is suffering using the machine learning is also needed to be identified.

In this review, a thorough study has been conducted to identify various machine learning and deep learning models that can perform better in different challenging real-time farming scenarios. The manuscript also attempts to investigate the numerous studies, and reliable models for plant disease detection, their classification, and various possible infections found in different types of plants. Various challenges in the pathway of using machine learning models that has the affect on the performance of these automated plant disease detection systems are also identified.

Based upon the study conducted in this article, it is observed that there exist various ways to boost the accuracy as well as the overall performance of these machine learning-based plant disease detection systems. One such way to enhance the accuracy of such models is to make use of large datasets so that the machine can upgrade their learning rate. The training images to use must contain different samples keeping different environments as well as lightening parameters in mind. However, an increase in the size of the dataset also costs the storage as well as the computational penalty.

Also, it is observed that the Transfer Learning technique is used widely in various existing studies for the detection of plant diseases, as transfer learning techniques allow the improvement of the performance of existing CNN based model without developing the model from scratch.

Among traditional machine learning algorithms, SVM has shown extraordinary performance in the classification as well as identification of disease from which the plant is suffering. Among deep learning techniques, CNN has shown remarkable performance.

From the literature, it is clear that a lot of work has already been done in forecasting the different diseases in plants using their images. However, improvement in the accuracy, as well as real-time testing and deployment, are the few aspects that need to be focused on. Chemical and pesticide recommendation based on the identified disease is also a good research topic.

The review introduced here would be of great assistance not particularly to researchers and specialists in this area but also to the pathologists and farmers for forecasting the plant diseases.

Acknowledgements We would like to thanks World Bank and National Project Implementation Unit (NPIU), MHRD, Government of India for assisting and supporting in the completion of this study.

Declarations

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

References

1. D Al Bashish, M Braik, S Bani-Ahmad (2010) "A framework for detection and classification of plant leaf and stem diseases," In 2010 international conference on signal and image processing, IEEE, 113–118
2. SD Khirade, A Patil (2015) "Plant disease detection using image processing, " In 2015 International conference on computing communication control and automation, pp. 768–771: IEEE.
3. Agarwal A, Sarkar A, Dubey AK (2019) Computer vision-based fruit disease detection and classification. In: Smart innovations in communication and computational sciences. Springer, Singapore, pp 105–115
4. P Tm, A Pranathi, K SaiAshritha, NB Chittaragi,SG Koolagudi (2018) "Tomato leaf disease detection using convolutional neural networks," In 2018 Eleventh international conference on contemporary computing (IC3), IEEE, 1–5
5. Thapa S, Thapa S (2019) Scope of Value-addition in Potato. Int J Horticult Agri Food Sci. https://doi.org/10.22161/ijhaf.3.3.4
6. Bhat HA (2019) Trends and growth in area, production and productivity of apples in India from 2001–02 to 2017–18. Res Ambition 4(1):13–23
7. Dwivedi M, Saxena S, Ray S (2019) Assessment of rice biomass production and yield using semi-physical approach and remotely sensed data. Int Arch Photogramm Remote Sens Spatial Infor Sci XLII-3/W6:217–222
8. A Kumar, S Sarkar, C Pradhan (2019) "Recommendation system for crop identification and pest control technique in agriculture," In 2019 International conference on communication and signal processing (ICCSP), IEEE, 0185–0189
9. Kaur S, Pandey S, Goel S (2019) Plants disease identification and classification through leaf images: a survey. Arch Comput Method Eng 26(2):507–530
10. Gokulnath B, Usha Devi G (2017) A survey on plant disease prediction using machine learning and deep learning techniques. Intel Artif 22(63):1–19
11. Loey M, ElSawy A, Afify M (2020) Deep learning in plant diseases detection for agricultural crops: a survey. Int J Service Sci, Manage, Eng, Technol (IJSSMET) 11(2):41–58
12. ZB Husin, AYBM Shakaff, AHBA Aziz, RBSM Farook (2012) "Feasibility study on plant chili disease detection using image processing techniques," In 2012 Third International Conference on Intelligent Systems Modelling and Simulation, IEEE, 291–296
13. Potatol Diseases and Pests, Description, Uses, Propogation (n.d.) https://www.plantvillage.psu.edu/topics/potato/infos. Accessed 02 June 2020
14. M Al-Amin, TA Bushra, MN Hoq, "Prediction of Potato Disease from Leaves using Deep Convolution Neural Network towards a Digital Agricultural System," In 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), IEEE, 1–5
15. T Kodama, Y Hata (2018) "Development of Classification System of Rice Disease Using Artificial Intelligence," In 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE, 3699–3702
16. Prajapati HB, Shah JP, Dabhi VK (2017) Detection and classification of rice plant diseases. Intell Decision Technolog 11(3):357–373
17. PlantVillage Dataset: Dataset of diseased plant leaf images and corresponding labels (2019) https://www.kaggle.com/emmarex/plantdisease. Accessed 2 Oct 2020
18. New Plant Diseases Dataset: Image dataset containing different healthy and unhealthy crop leaves (2019)https://www.kaggle.com/vipoooool/new-plant-diseases-dataset. Accessed 19 July 2020
19. Rice Diseases Image Dataset (2020) https://www.kaggle.com/minhhuy2810/ricediseases-image-dataset. Accessed 16 July 2020
20. Plant Pathology 2020 - FGVC7 (2020) https://www.kaggle.com/c/plantpathology-2020-fgvc7/data. Accessed 28 June 2020
21. IPM Images- Vegetable (n.d.) https://www.ipmimages.org/browse/bareas.cfm?domain=12. Accessed 23 March 2020
22. Wiesner-Hanks Tyr, Stewart EL, Kaczmar N, DeChant C, Wu H, Nelson RJ, Lipson H, Gore MA (2018) Image set for deep learning: field images of maize annotated with disease symptoms. BMC Res Notes 11(1):1–3
23. Soybean (Large) Data Set (1988) https://archive.ics.uci.edu/ml/datasets/Soybean+(Large). Accessed 13 March 2020
24. APS Image Database (n.d.) https://www.imagedatabase.apsnet.org/. Accessed 10 Jan 2020
25. Sugimura D, Mikami T, Yamashita H, Hamamoto T (2015) Enhancing color images of extremely low light scenes based on RGB/NIR images acquisition with different exposure times. IEEE Trans Image Process 24(11):3586–3597
26. Shrivastava S, Hooda DS (2014) Automatic brown spot and frog eye detection from the image captured in the field. Am J Intell Sys 4(4):131–134

27. A Akhtar, A Khanum, SA Khan, A Shaukat (2013) “Automated plant disease analysis (APDA): performance comparison of machine learning techniques,” In 2013 11th International conference on frontiers of information technology, IEEE, 60–65
28. Camargo A, Smith J (2009) An image-processing based algorithm to automatically identify plant disease visual symptoms. Biosys Eng 102(1):9–21
29. P Sanyal, U Bhattacharya, SK Parui, SK Bandyopadhyay, S Patel (2007) “Color texture analysis of rice leaves diagnosing deficiency in the balance of mineral levels towards improvement of crop productivity,” In 10th International Conference on Information Technology (ICIT 2007), IEEE, 85–90
30. Bangladesh Rice Research Institute (BRRI) (n.d.) http://www.brri.gov.bd/. Accessed 16 Feb 2020
31. G Anthonys, N Wickramarachchi (2009) “An image recognition system for crop disease identification of paddy fields in Sri Lanka," In *2009 International conference on industrial and information systems (ICIIS)*, IEEE, 403–407
32. Pydipati R, Burks T, Lee W (2006) Identification of citrus disease using color texture features and discriminant analysis. Comput Electron Agric 52(1–2):49–59
33. Bauer SD, Korč F, Förstner W (2011) The potential of automatic methods of classification to identify leaf diseases from multispectral images. Precision Agric 12(3):361–377
34. Xu G, Zhang F, Shah SG, Ye Y, Mao H (2011) Use of leaf color images to identify nitrogen and potassium deficient tomatoes. Pattern Recogn Lett 32(11):1584–1590
35. L Jinzhu, C Di, H Jiang (2013) “Discrimination of tomato yellow leaf curl disease using hyperspectral imaging," In 2013 Kansas City, Missouri, July 21-July 24, American Society of Agricultural and Biological Engineers, 1
36. SV Militante, BD Gerardo, NV Dionisio (2019) “Plant leaf detection and disease recognition using deep learning,” In 2019 IEEE Eurasia conference on IOT, communication and engineering (ECICE), IEEE, 579–582
37. Q Yao, Z Guan, Y Zhou, J Tang, Y Hu, B Yang (2009) “Application of support vector machine for detecting rice diseases using shape and color texture features,” In 2009 international conference on engineering computation, IEEE, 79–83
38. Jadhav SB, Patil SB (2015) Grading of soybean leaf disease based on segmented image using k-means clustering. Int J Adv Res Electron Commun Eng 4(6):1816–1822
39. Al-Hiary H, Bani-Ahmad S, Reyalat M, Braik M, Alrahamneh Z (2011) Fast and accurate detection and classification of plant diseases. Int J Computer App 17(1):31–38
40. L Wang, F Dong, Q Guo, C Nie, S Sun (2014) “Improved rotation kernel transformation directional feature for recognition of wheat stripe rust and powdery mildew,” In 2014 7th International congress on image and signal processing, IEEE, 286–291
41. Zhang S, Shang Y, Wang L (2015) Plant disease recognition based on plant leaf image. J Animal Plant Sci 25(3):42–45
42. SS Sannakki, VS Rajpurohit, V Nargund, P Kulkarni (2013) “Diagnosis and classification of grape leaf diseases using neural networks,” In 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), IEEE, 1–5
43. Arivazhagan S, Shebiah RN, Ananthi S, Varthini SV (2013) Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. Agric Eng Int CIGR J 15(1):211–217
44. Wiwart M, Fordoński G, Żuk-Gołaszewska K, Suchowilska E (2009) Early diagnostics of macronutrient deficiencies in three legume species by color image analysis. Comput Electron Agric 65(1):125–132
45. Makandar A, Halalli B (2015) Image enhancement techniques using highpass and lowpass filters. Int J Computer App 109(14):12–15
46. Aubry M, Paris S, Hasinoff SW, Kautz J, Durand F (2014) Fast local laplacian filters: theory and applications. ACM Trans Graphic 33(5):1–14
47. Verbeek P, Vrooman H, Van Vliet L (1988) Low-level image processing by max-min filters. Signal Process 15(3):249–258
48. Ell TA, Sangwine SJ (2006) Hypercomplex Fourier transforms of color images. IEEE Trans Image Process 16(1):22–35
49. Lee J-S (1983) Digital image smoothing and the sigma filter. Computer Vision, Graphics, and Image Process 24(2):255–269
50. Bins LS, Fonseca LG, Erthal GJ, Ii FM (1996) Satellite imagery segmentation: a region growing approach. Simpósio Brasileiro de Sensoriamento Remoto 8(1996):677–680
51. M Islam, A Dinh, K Wahid, P Bhowmik (2017) “Detection of potato diseases using image segmentation and multiclass support vector machine,” In 2017 IEEE 30th canadian conference on electrical and computer engineering (CCECE), IEEE, 1–4
52. Auliya YA, Mahmudy WF, Sudarto S (2019) Improve hybrid particle swarm optimization and K-means for clustering. J Infor Technol Computer Sci 4(1):42–56
53. Badnakhe MR, Deshmukh PR (2012) Infected leaf analysis and comparison by Otsu threshold and k-means clustering. Int J Adv Res Computer Sci Softw Eng 2(3):449–452
54. Singh SM, Hemachandran K (2012) Content-based image retrieval using color moment and gabor texture feature. Int J Computer Sci Issue 9(5):299
55. J Huang, SR Kumar, M Mitra, WJ Zhu, R Zabih (1997) “Image indexing using color correlograms,” In Proceedings of IEEE computer society conference on Computer Vision and Pattern Recognition, IEEE, 762–768
56. JL Shih, LH Chen (2002) “Color image retrieval based on primitives of color moments,” In VISUAL, 88–94
57. Mokhtar U, El Bendary N, Hassenian AE, Emary E, Mahmoud MA, Hefny H, Tolba MF (2015) SVM-based detection of tomato leaves diseases. In: Intelligent Systems’ 2014. Springer, Cham, pp 641–652
58. A Meunkaewjinda, P Kumsawat, K Attakitmongcol, A Srikaew (2008) “Grape leaf disease detection from color imagery using hybrid intelligent system,” In 2008 5th International conference on electrical engineering/electronics, computer, telecommunications and information technology, IEEE, vol. 1, 513–516
59. Harefa J, Alexander A, Pratiwi M (2017) Comparison classifier: support vector machine (SVM) and K-nearest neighbor (K-NN) in digital mammogram images. Jurnal Informatika dan Sistem Informasi 2(2):35–40
60. Goldberg DE, Holland JH (1988) Genetic algorithms and machine learning. Machine Learn 3(2):95–99
61. A Dacal-Nieto, E Vázquez-Fernández, A Formella, F Martin, S Torres-Guijarro, H González-Jorge, “A genetic algorithm approach for feature selection in potatoes classification by computer vision,” In 2009 35th Annual Conference of IEEE Industrial Electronics, IEEE, 1955–1960
62. JS Prakash, KA Vignesh, C Ashok, R Adithyan “Multi class support vector machines classifier for machine vision application,” In 2012 International conference on machine vision and image processing (MVIP), IEEE, 197–199
63. Yekkehkhany B, Safari A, Homayouni S, Hasanlou M (2014) A comparison study of different kernel functions for SVM-based classification of multi-temporal polarimetry SAR data. Int Arch Photogramm, Remote Sens Spatial Infor Sci 40(2):281
64. Biau G, Scornet E (2016) A random forest guided tour. TEST 25(2):197–227

65. Techopedia. (n.d.) *Artificial Neural Network (ANN)*. https://www.techopedia.com/definition/5967/artificial-neural-network-ann. Accessed 20 Jan 2020
66. W Jin, ZJ Li, LS Wei ,H Zhen (2000) "The improvements of BP neural network learning algorithm," In WCC 2000-ICSP 2000. 2000 5th international conference on signal processing proceedings. 16th world computer congress, IEEE, vol. 3, 1647–1649
67. L Liu, G Zhou (2009) "Extraction of the rice leaf disease image based on BP neural network," In 2009 International Conference on Computational Intelligence and Software Engineering, IEEE, 1–3
68. Rawat W, Wang Z (2017) Deep convolutional neural networks for image classification: A comprehensive review. Neural Comput 29(9):2352–2449
69. Gulli A, Pal S (2017) Deep learning with Keras. Packt Publishing Ltd, Birmingham
70. A Krizhevsky, I Sutskever, GE Hinton (2012) "Imagenet classification with deep convolutional neural networks," In Advances in neural information processing systems, 1097–1105
71. K Simonyan, A Zisserman (2014) "Very deep convolutional networks for large-scale image recognition," *arXiv preprint* arXiv:1409.1556,
72. C Szegedy et al.,(2015) "Going deeper with convolutions," In Proceedings of the IEEE conference on computer vision and pattern recognition, 1–9
73. K He, X Zhang, S Ren, J Sun (2016) "Deep residual learning for image recognition," In Proceedings of the IEEE conference on computer vision and pattern recognition, 770–778
74. G Huang, Z Liu, L Van Der Maaten, KQ Weinberger(2017) "Densely connected convolutional networks," In Proceedings of the IEEE conference on computer vision and pattern recognition, 4700–4708
75. L Xie, A Yuille (2017) "Genetic cnn," In Proceedings of the IEEE international conference on computer vision, 1379–1388
76. FN Iandola, S Han, MW Moskewicz, K Ashraf, WJ Dally, KJ Keutzer (2016) "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size", arXiv preprint arXiv:1602.07360
77. LeCun Y, Haffner P, Bottou L, Bengio Y (1999) Object recognition with gradient-based learning. In: Shape, contour and grouping in computer vision. Springer, Berlin, Heidelberg, pp 319–345
78. AG Howard et al. (2017) "Mobilenets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:.04861
79. F Chollet, "Xception: Deep learning with depthwise separable convolutions," In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1251–1258
80. Fuentes A, Yoon S, Kim SC, Park DS (2017) A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. Sensors 17(9):2022
81. M Francis ,C Deisy (2019) "Disease detection and classification in agricultural plants using convolutional neural networks—a visual understanding," In 2019 6th International conference on signal processing and integrated networks (SPIN), IEEE, 1063–1068
82. M Govardhan, M Veena (2019) "Diagnosis of tomato plant diseases using random forest," In 2019 Global Conference for Advancement in Technology (GCAT), IEEE, 1–5
83. Balakrishna K, Rao M (2019) Tomato plant leaves disease classification using KNN and PNN. Int J Computer Vision Image Process 9(1):51–63
84. CU Kumari, SJ Prasad, G Mounika (2019) "Leaf disease detection: feature extraction with K-means clustering and classification with ANN," In 2019 3rd International conference on computing methodologies and communication (ICCMC), IEEE, 1095–1098
85. AS Chakravarthy, S Raman (2020) "Early blight identification in tomato leaves using deep learning," In 2020 International Conference on Contemporary Computing and Applications (IC3A), IEEE, 154–158
86. J Shijie, J Peiyi, H Siping (2017) "Automatic detection of tomato diseases and pests based on leaf images," In 2017 Chinese automation congress (CAC), IEEE, 2537–2510
87. Brahimi M, Boukhalfa K, Moussaoui A (2017) Deep learning for tomato diseases: classification and symptoms visualization. Appl Artif Intell 31(4):299–315
88. A Kumar, M Vani (2019) "Image based tomato leaf disease detection," In 2019 10th International conference on computing, communication and networking technologies (ICCCNT), IEEEL, 1–6
89. P Patil, N Yaligar, S Meena (2017) "Comparision of performance of classifiers-SVM, RF and ANN in potato blight disease detection using leaf images," In 2017 IEEE International conference on computational intelligence and computing research (ICCIC), IEEE, 1–5
90. S Biswas, B Jagyasi, BP Singh, M Lal "Severity identification of potato late blight disease from crop images captured under uncontrolled environment," In 2014 IEEE Canada international humanitarian technology conference-(IHTC), IEEE, 1–5
91. U Suttapakti, A Bunpeng (2019) "Potato leaf disease classification based on distinct color and texture feature extraction," In 2019 19th International Symposium on Communications and Information Technologies (ISCIT), IEEE, 82–85
92. S Arya, R Singh (2019) "A comparative study of CNN and AlexNet for detection of disease in potato and mango leaf," In 2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), IEEE, vol. 1, 1–6
93. Dubey SR, Jalal AS (2016) Apple disease classification using color, texture and shape features from images. SIViP 10(5):819–826
94. Dewliya S, Singh P (2015) Detection and classification for apple fruit diseases using support vector machine and chain code. Inter Res J Eng Technol (IRJET) 2.04
95. Samajpati BJ, Degadwala SD (2016) "Hybrid approach for apple fruit diseases detection and classification using random forest classifier," In 2016 International conference on communication and signal processing (ICCSP), IEEE, 1015–1019
96. Chuanlei Z, Shanwen Z, Jucheng Y, Yancui S, Jia C (2017) Apple leaf disease identification using genetic algorithm and correlation based feature selection method. Int J Agricul Biolog Eng 10(2):74–83
97. Elangovan K, Nalini S (2017) Plant disease classification using image segmentation and SVM techniques. Int J Comput Intell Res 13(7):1821–1828
98. A Awate, D Deshmankar, G Amrutkar, U Bagul, S Sonavane (2015) "Fruit disease detection using color, texture analysis and ANN," In 2015 International Conference on Green Computing and Internet of Things (ICGCIoT), IEEE, 970–975
99. Jiang P, Chen Y, Liu B, He D, Liang C (2019) Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks. IEEE Access 7:59069–59080
100. M Agarwal, RK Kaliyar, G Singal,SK Gupta (2019) "FCNN-LDA: a faster convolution neural network model for leaf disease identification on apple's leaf dataset," In 2019 12th International Conference on Information & Communication Technology and System (ICTS), IEEE, 246–251
101. M Suresha, K Shreekanth, B Thirumalesh (2017) "Recognition of diseases in paddy leaves using knn classifier," In 2017 2nd

International conference for convergence in technology (I2CT), IEEE, 663–666
102. T Islam, M Sah, S Baral, R RoyChoudhury (2018) “A faster technique on rice disease detectionusing image processing of affected area in agro-field,” In 2018 second international conference on inventive communication and Computational technologies (ICICCT), IEEE, 62–66
103. JP Shah, HB Prajapati, VK Dabhi (2016) “A survey on detection and classification of rice plant diseases,” In 2016 IEEE International conference on current trends in advanced computing (ICCTAC), IEEE, 1–8
104. S Ramesh (2018) “Rice blast disease detection and classification using machine learning algorithm,” In 2018 2nd International Conference on Micro-Electronics and Telecommunication Engineering (ICMETE), IEEE, 255–259
105. BS Ghyar, GK Birajdar (2017) “Computer vision based approach to detect rice leaf diseases using texture and color descriptors,” In 2017 International Conference on Inventive Computing and Informatics (ICICI), IEEE, 1074–1078
106. SS Chawathe, “Rice disease detection by image analysis,” In 2020 10th Annual Computing and Communication Workshop and Conference (CCWC), IEEE, 0524–0530
107. MJ Hasan, S Mahbub, MS Alom, MA Nasim (2019) “Rice disease identification and classification by integrating support vector machine with deep convolutional neural network,” In 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), IEEE, 1–6
108. IRRI: International Rice Research Institute (n.d.) https://www.irri.org/. Accessed 22 Apr 2020
109. BRKB: Bangladesh Rice Knowledge Bank (n.d.) http://knowledgebank-brri.org/. Accessed 12 Apr 2020
110. Sethy PK, Barpanda NK, Rath AK, Behera SK (2020) Deep feature based rice leaf disease identification using support vector machine. Comput Electron Agric 175:105527
111. Ramesh S, Vydeki D (2020) Recognition and classification of paddy leaf diseases using optimized deep neural network with jaya algorithm. Infor Process Agri 7(2):249–260
112. Li D et al (2020) A recognition method for rice plant diseases and pests video detection based on deep convolutional neural network. Sensors 20(3):578
113. Sharma V, Mir AA, Sarwr A (2020) Detection of rice disease using bayes’ classifier and minimum distance classifier. J Multimed Infor Sys 7(1):17–24
114. Chen W-L, Lin Y-B, Ng F-L, Liu C-Y, Lin Y-W (2019) RiceTalk: rice blast detection using internet of things and artificial intelligence technologies. IEEE Internet Thing J 7(2):1001–1010
115. Rath A, Meher J (2019) Disease detection in infected plant leaf by computational method. Arch Phytopathol Plant Protect 52(19–20):1348–1358
116. Sethy PK, Negi B, Barpanda NK, Behera SK, Rath AK (2018) Measurement of disease severity of rice crop using machine learning and computational intelligence. IN: Cognitive science and artificial intelligence. Springer, Singapore, pp 1–11
117. Lu Y, Yi S, Zeng N, Liu Y, Zhang Y (2017) Identification of rice diseases using deep convolutional neural networks. Neurocomputing 267:378–384
118. Barbedo JG (2016) A review on the main challenges in automatic plant disease identification based on visible range images. Biosys Eng 144:52–60
119. S Anas, I Badhusha, Z Ot, K Faseela, M Shelly (2017) “Cloud based automated irrigation and plant leaf disease detection system using an android application,” In 2017 International conference of electronics, communication and aerospace technology (ICECA), IEEE, vol. 2, 211–214