Journal of Plant Diseases and Protection (2022) 129:469–488
https://doi.org/10.1007/s41348-021-00500-8

REVIEW

Artificial intelligence in tomato leaf disease detection: a comprehensive review and discussion

Rajasekaran Thangaraj[1] · S. Anandamurugan[2] · P Pandiyan[3] · Vishnu Kumar Kaliappan[4]

Received: 10 February 2021 / Accepted: 6 July 2021 / Published online: 21 August 2021

Abstract
Accurate and fast tomato plant disease identification is significant to enhance its sustainable agricultural productivity. In the conventional technique, human experts in the field of agriculture have been accommodated to find out the anomalies in tomato plants caused by pests, diseases, climatic conditions, and nutritional deficiencies. Automatic tomato leaf disease identification is initially solved through conventional image processing and machine learning approaches which result in less accuracy. In order to produce greater prediction accuracy, deep learning-based classification is introduced. This paper provides an overall review of recent work performed in the field of tomato leaf disease identification using image processing, machine learning, and deep learning approaches. And also discuss both public and private datasets available to detect tomato leaf disease, methods employed, and adopted deep learning frameworks. Consequently, suggestions are provided to figure out the appropriate techniques in order to obtain the better prediction accuracy. Finally, the challenges encountered in implementing the machine learning and deep learning models are discussed.

Keywords Tomato leaf disease · Deep learning · Image processing · Convolution neural network · Artificial intelligence · Machine learning

Introduction

Agriculture is one of the foremost human activities which aids for nation development. Recently, major activities are taking place in the farming and food industry owing to population growth and satisfy their food requirements to facilitate their lives in a better way. The economy of the country mostly depends on agriculture which not only provides the raw material and food; however, it creates employment opportunities (Gebbers and Adamchuk 2010; Rajasekaran and Anandamurugan 2019). Globally, the food supplement decreases annually with an average value of 40% because of plant diseases and insect attacks (Shalaby et al. 2011). The agricultural fields face major issues such as loss in crop yield and production. Plant leaf diseases are the most significant issue reducing crop production which is caused by a variety of bacteria, fungi, insects, and viruses, etc.(Mishra et al. 2014). The reduction in the crop yield ultimately creates starvation in dry areas and insufficient food requirements (Oerke 2006). The plants that are affected by the diseases shows the symptoms such as leaf blight, leaf spots, fruit rots, root rots, fruit spots, dieback, decline, and wilt (Slavin 2016).

✉ S. Anandamurugan
dranandamurugan@gmail.com

Rajasekaran Thangaraj
rajasekaran30@gmail.com

P Pandiyan
pandyyan@gmail.com

Vishnu Kumar Kaliappan
vishnudms@gmail.com

1 Department of Computer Science and Engineering, KPR Institute of Engineering and Technology, Coimbatore, Tamilnadu, India

2 Department of Information Technology, Kongu Engineering College, Perundurai, Tamilnadu, India

3 Department of Electrical and Electronics Engineering, KPR Institute of Engineering and Technology, Coimbatore, Tamilnadu, India

4 Konkuk Aerospace Design Airworthiness Institute, Konkuk University, Seoul, Republic of Korea

In worldwide, tomato cultivation plays a vital role in the world's agricultural trade and production because of its rich nutrition content (Riley et al. 2002). Tomato usage occupies a significant contribution among vegetable crops globally. Tomato production is noticeably increased yearly according to the Food and Agriculture Organization of the United States statistical report (Strange and Scott 2005). In the case of tomato farming, leaf diseases are considered as one of the dominating factors for production loss which leads to significant losses in the agricultural economy (Ma et al. 2015a). For example, early blight disease is the most commonly occurring disease worldwide which drastically reduces the yield. Similarly, late blight disease represents severe damages to crops (Ma et al. 2015b). Consequently, protection is essential for tomato crops to prevent disease which enhances the quality and quantity of the crops. Early prediction of the disease directs to select proper treatment in order to prevent severe damages (Ma et al. 2017).

The foremost importance in agriculture is the early diagnosis of plant disease. Detection of plant disease through leaf is generally used technique to figure out the disease as it shows the change in its original structure for different diseases. Identifying the disease by the naked eye of an expert needs vast professional experience with extensive knowledge about the causes of disease on the crops (Ma et al. 2015b). Furthermore, the expert should have sufficient knowledge to inform the details related to signs and symptoms caused by disease. Even today, manual evaluation is done in remote villages, but it does not identify the exact disease and its variants. Manual assessment is a time-consuming process for larger farms and requires huge manpower. Moreover, cultivation is a continuous process, thus needs periodic monitoring of crops to find out the disease. As a result, the alternative method is needed to identify the diseases automatically utilizing leaf images (Barbedo, 2013).

Conventional image processing methods like Grey Level Co-occurrence Matrix (GLCM) (Jinzhu et al. 2013), Scale Invariant Feature Transform (SIFT) (Wang et al. 2013), Speeded Up Robust Features (SURF) (Bay et al. 2006) etc., contributes reasonable output for disease detection through leaf images. However, this method uses fewer datasets and provides theoretical-based results. Recently, advancements in Artificial Intelligent (AI) techniques and computer vision approach to detect and classify objects have been growing in interest (Barbedo 2017). Segmentation of disease-affected parts on leaves can be utilized as a key factor to acquire disease-related information using currently available computer vision technology (Ren et al. 2017).

Machine learning (ML) and deep learning (DL) has transformed computer vision technology, particularly in image-based detection and classification. Nowadays, deep learning models are a prominent tool to enhance this type of automated process to attain accurate results for real-time plant disease detection and classification. Convolution neural network (CNN)-based deep learning has proven technique for achieving the best results in image classification. CNN-based architecture such as AlexNet (Zhao and Jia 2016), LeNet (Xu et al. 2017), GoogLeNet (Sainath et al. 2015), VGGNet (Shelhamer et al. 2017), ResNet (Ribeiro et al. 2016), DenseNet (LeCun et al. 1998), Inception V3 (Krizhevsky et al. 2012) and Xception (Thangaraj et al. 2020) were employed to identify the tomato leaf disease with greater accuracy.

This paper aims to discuss the research work performed in tomato plant leaf disease identification and classification. Further, the investigations are carried out to identify the challenges faced in detecting and classifying the tomato disease using leaf image. The study has the following points:

- Discussion on ML-based tomato disease identification and classification.
- Investigation on DL models employed in tomato disease identification, classification, and improving the recognition accuracy by incorporating transfer learning concepts.
- Report the public and privately available tomato leaf disease datasets and deep learning frameworks used in disease identification.
- Evaluation metrics to assess the effectiveness of the ML/DL models.

Search strategy

In this study, search and selection approach for identifying the tomato leaf disease generally focused on electronic repositories namely IEEE Xplore, ScienceDirect, Google scholar and ACM library. These repositories are chosen due to their highest volume of important research studies on tomato plant leaf diseases detection with use of image processing techniques which includes ML and DL algorithms. In recent years, the tomato plant disease detection and classification through artificial intelligence (ML/DL) algorithm has inspired the scientific community. Therefore, publications related to this study are taken from 2015 to till date. The search process started with keyword-based search for journal and conference paper from the scientific repositories. The search keywords that were used for this research work as follows:

["Tomato plant leaf disease identification" OR "tomato plant leaf disease detection" OR "tomato plant leaf disease classification" OR "tomato plant disease"].

AND

["artificial Intelligence" OR "Image processing" OR "Machine learning" OR "Deep learning" OR "CNN"].

The procedure followed for drafting this research work is illustrated in Fig. 1. Initially, papers are downloaded from the electronic database relevant to tomato leaf disease detection/identification/recognition/classification using artificial intelligence technique. Next, read the paper and classify it according to the techniques such as traditional image processing technique, machine learning and deep learning. As a result, 79 papers have been found as a result of this search. The number of papers was reduced to 44 by proper search optimization and analysis provided excellent results by means of selective citations and precise conclusions.

The articles are reviewed individually by considering all relevant citations and the following research issues:

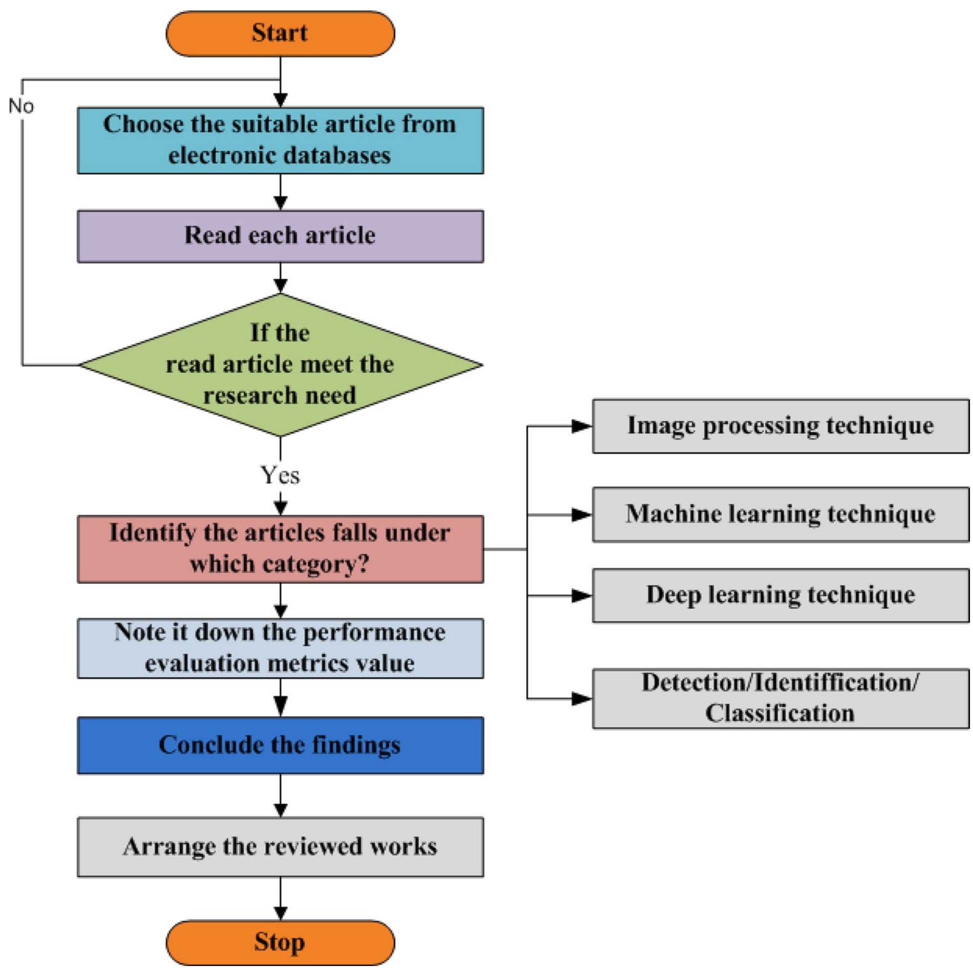
- Which plant disease issues are addressed?
- What type of artificial intelligence model is used in the research study?
- What type of dataset is used?
- What is the performance level of the ML and DL techniques employed in the chosen research work?

Machine learning-based tomato leaf disease classification

Conventional ML algorithms related to tomato disease identification is discussed in this section. In general, the crop leaves are considered as the first source of tomato plant disease identification as well as symptoms of major diseases that may occur on plant leaves. The identification of tomato disease through leaf images in the current research field attracts the researcher globally for the prospective advantages and favorable outcomes. Plenty of works have been proposed for identifying the tomato leaf disease in which each reported variety of models, methods, and features. Thus, a literature review was performed to conclude the study that has previously been completed in this area. The flow diagram for applying a ML model for identification and classification of plant disease is illustrated in Fig. 2.

Sabrol et al. (Sabrol and Satish 2016) experimented to detect the tomato disease using leaf image by utilizing Otsu's segmentation algorithm with a decision tree approach. In this work, shape, texture, and color features

Fig. 1 Search strategy flowchart



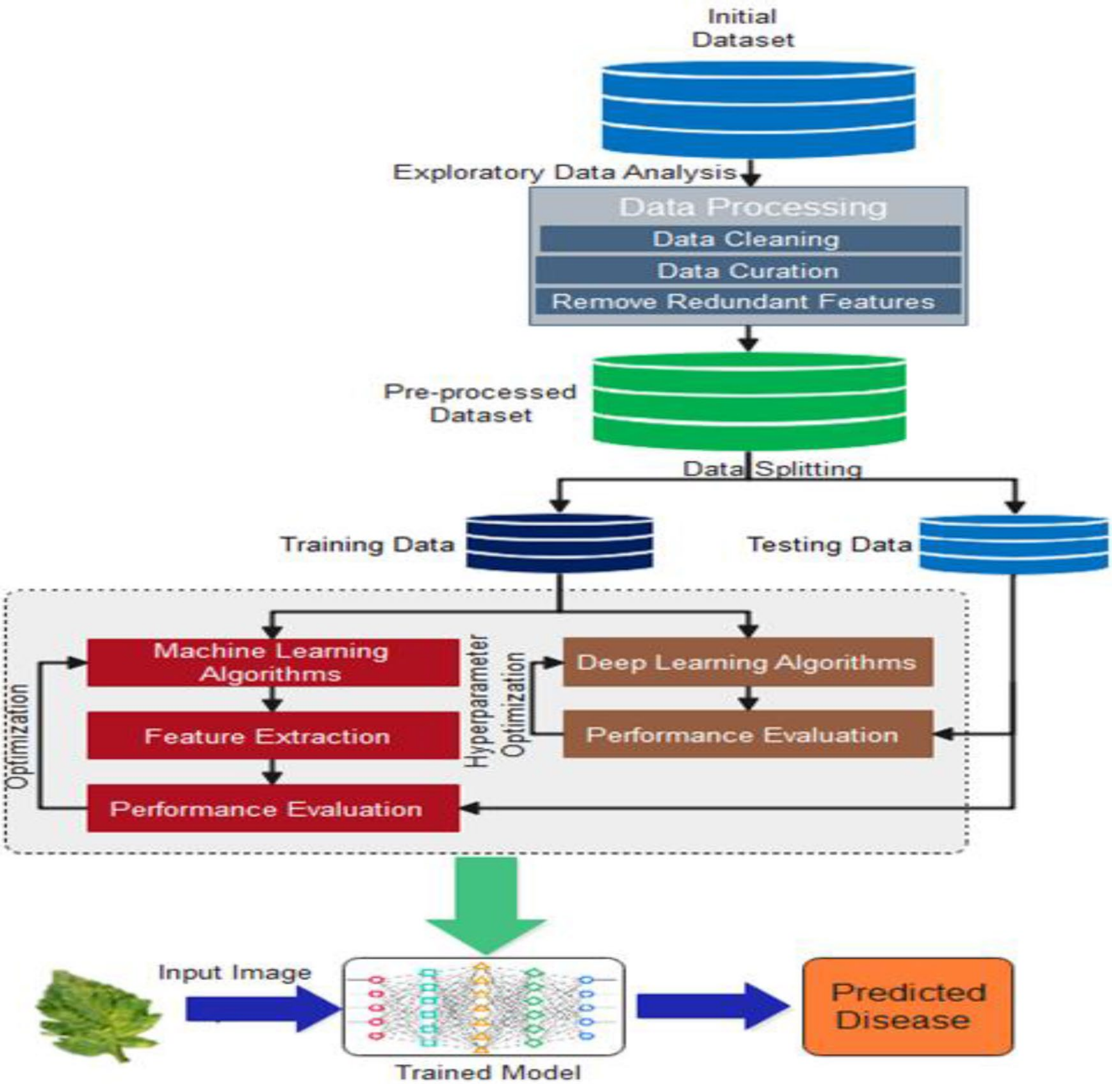


Fig. 2 Conceptual machine learning/ deep learning model flow diagram

are used for learning the leaf disease characteristics. The proposed method achieved a classification accuracy of 97.30%. The research work proposed by Hlaing et al. (Hlaing et al. 2017) used statistical model to detect six types of tomato leaf diseases. In order to decrease the SIFT feature vector dimensions, the generalized extreme value (GEV) distribution was adopted which reduces algorithm computational time and attained an accuracy value of 84.7%. Xie et al. used a hyperspectral imaging technique to identify the occurrence of the late blight and early blight on tomato leaf utilizing images. The 310 hyperspectral images are employed and spectral analysis is performed to figure out the disease-affected images for choosing the exact wavelength with SPA-ELM (Successive Projection Algorithm-Extreme Learning machine model) and textural features for detection. The classification accuracy is obtained from the experiment ranging between 97.10% and 100% (Xie et al. 2015).

Mokhtar et al. developed a tomato disease detection model based on support vector machine (SVM). This method is used to identify whether the leaves of tomato are infected with the disease such as early blight or powdery mildew. The method employs the Gabor wavelet transform technique to obtain the features of the tomato leaves and SVM performs the disease classification. Furthermore, three different types of the kernel such as Invmult Kernel, Cauchy kernel, and Laplacian Kernel have used to access the disease detection. The experimental result confirms that SVM using Cauchy kernel achieved the highest classification accuracy of 100% followed by Laplacian kernel which is 98% and Invmult Kernel which is 78% (Mokhtar et al. 2015a). Similarly, the same authors developed SVM-based classification model to identify tomato leaf infected with yellow leaf curl virus. The model involves geometrics and histogram features for the classification of diseases. This model is evaluated using different types of kernel functions such as radial basis function (RBF), linear, quadratic (QP), multi-layer perceptron (MLP), and polynomial. The experimental result shows that the model using quadratic kernel achieved the highest accuracy of 91.5% tested using the N-fold cross-validation technique (Mokhtar et al. 2015b).

Hassanien et al. introduced the Moth-Flame Optimization (MFO) and Moth-Flame Optimization Rough Set (MFORSFS) approach to identify a couple of tomato leaf diseases such as early blight and powdery mildew. The proposed approach is evaluated by comparing with the genetic algorithms (GA) and particle swarm optimization (PSO) with rough sets algorithms. The experimental results confirm that the proposed approach shows better performance metrics namely accuracy 86%, specificity 86%, F-score 85.7%, and recall 86% (Hassanien et al. 2017).

Sabrol and Kumar (Sabrol and Kumar 2016a) used 360 leaf color images consisting of six different classes in which five belong to the disease category and the remaining one is in the healthy category. Conventional image processing techniques were used to transform RGB images into CIE XYZ color space model with three different types of classifiers such as FIS (Fuzzy Inference system), MLBPNN (MultiLayer Feed Forward Back Propagation Neural Network) and ANFIS (Adaptive Neuro-Fuzzy Inference System). The accuracy reported for the MLBPNN algorithm is of 87.20% which is highest compared with the other two algorithms.

Lu et al. (Lu et al. 2018) identify the yellow leaf curl disease through 166 images collected from the hyperspectral camera which includes healthy and disease-infected leaves and carried out spectral dimension analysis and band selection. Zhang et al. (Tm et al. 2018) investigated the work on finding out the late blight-infected tomato crop using hyperspectral images of leaves in order to prevent severe damage. Lu et al. employ a hyperspectral imaging technique to detect the infection of the yellow leaf curl virus on the tomato leaf. These leaf features are extracted utilizing GLCM (Grey Level Co-occurrence Matrix), and its performance is evaluated through ROC (Receiver Operator Characteristic) curve. The result indicates that the proposed approach achieved accuracy between the range 87.20%– 92.30% (Lu et al. 2013).

In (Mokhtar et al. 2015c), Mokhtar et al. differentiate healthy and disease-infected tomato leaves through the support vector machine (SVM) algorithm employing different kernel functions namely linear, radial basis function (RBF), polynomial, and multilayer perceptron (MLP). Totally, 400 images are used for training and 800 images for testing and attain greater recognition accuracy of 99.83% for SVM with a linear kernel function. Sabrol and Kumar (Sabrol and Kumar 2016b) carried out experimental work on the classification of healthy and unhealthy tomato leaves using a decision tree algorithm. The results confirm the classification accuracy of 78% in identifying the infection caused due to bacterial canker, fungal late blight, leaf curl diseases, and bacterial leaf spot.

Annabel and Muthulakshmi reports detection of three types of tomato leaf disease including bacterial spot, tomato mosaic virus, late blight, and healthy leaf through a random forest (RF) algorithm. The algorithm achieved an accuracy of 94.10% which is highest compared to SVM which is of 82.60% and MDC 87.60% obtained on the same dataset (Annabel et al. 2019). Muthukannan et al. (Muthukannan and Latha 2015) developed the fuzzy rule to identify the disease-affected tomato leaf regions. Feature extraction is performed with the color feature of the leaf images and fuzzy logic is employed for classifying healthy, less affected, and more affected areas of the leaf. The experimental result demonstrates that the proposed method attains a classification accuracy of 95%. Hlaing et al. (Hlaing et al. 2018) implemented quadratic SVM to detect the disease-affected tomato crop through plantvillage leaf image dataset which consists of seven different types of classes. In this work, the images are pre-processed initially to fill the regions, assign suitable channel values, remove salt and pepper noise, and so on. The extraction of statistical texture features is performed employing the Johnson SB distribution system and attains a classification accuracy of 85.10%. Das et al. developed a system to detect seven different types of tomato leaf disease using SVM, logistic regression (LR), and RF. The texture features of the leaves are extracted employing the Haralick algorithm, and these features are classified using SVM, LR, and RF. The result confirms that SVM outperforms with an accuracy of 87.60% followed by RF 70.05% and LR 67.30% (Das et al. 2020). Basavaiah and Anthony presented multiple features fusion method to identify four main leaf disease of tomato which includes bacterial spot, septoria spot, yellow curl virus, and mosaic virus. The features such as Hu moments, color histogram, local binary pattern, and haralick are extracted from the leaves. Subsequently extracted features are used by the random forest and decision tree algorithms to perform classification. The experimental result confirms that random forest shows the highest detection accuracy of 94% whereas the decision tree is 90% (Basavaiah and Anthony 2020).

The comparison of the adopted machine learning algorithm to detect the tomato leaf disease is shown in Table 1. Table 1 includes data source, number of images, number of class, types of diseases, name of the method, and accuracy obtained. Furthermore, the number of conference and journal papers published related to the tomato leaf disease classification with respect to years are depicted in Fig. 3.

Table 1 Comparison table showing the performance of machine learning algorithm applied to the detection of tomato plant disease through leaf images

Author	Data source	Number of images	Number of class	Types of disease	Method	Accuracy (%)
Muthukannan et al.(Muthukannan and Latha 2015)	Private	120	2	Healthy, Infected	Fuzzy Logic	95
Lu et al.(Lu et al. 2013)	Private	116	1	Yellow leaf curl virus	Grey Level Co-occurrence Matrix (GLCM), Receiver Operator Characteristic (ROC)	87.20%–92.30%
Mokhtar et al. (Mokhtar et al. 2015c)	Private	800	2	Healthy,Infected	SVM-Linear	99.83%
Xie et al.(Xie et al. 2015)	Private	310	3	Late blight, Early blight, Healthy	Extreme Learning Machine(ELM)	97.10% to 100%
Mokhtar et al.(Mokhtar et al. 2015a)	Private(Collected from farms in bani seef city)	200	2	Powdery mildew, Early blight	SVM Cauchy kernel	100
					SVM-Invmult Kernel	98
					SVM Laplacian Kernel	78
Mokhtar et al.(Mokhtar et al. 2015b)	Private	200	2	Tomato Yellow Leaf Curl Virus, Tomato Spotted Wilt Virus	SVM-Quadratic	91.5
Sabrol et al. (Sabrol and Satish 2016)	Private	383	6	Septoria spot, Late blight,Bacterial canker, Bacterial spot, Leaf curl,Healthy	Decision Tree	97.30
Sabrol and Kumar(Sabrol and Kumar 2016a)	Private	180	6	Tomato leaf curl virus, Septoria leaf spot, Bacterial canker, Bacterial leaf spot, Late blight, Healthy	MLBPNN (MultiLayer Feed Forward Back Propagation Neural Network	87.20
Sabrol and Kumar(Sabrol and Kumar 2016b)	Private	598	6	Bacterial Leaf Spot,Septoria Leaf Spot, Bacterial Canker, Fungal Late Blight, Leaf Curl, Healthy	Decision Tree	78
Hlaing et al. (Hlaing et al. 2017)	Plantvillage	3474	6	Leaf Mold, Septoria Leaf Spot, Two Spotted Spider Mite, Late Blight, Bacterial Spot, Target spot	Statistical-model	84.70
Hassanien et al.(Hassanien et al. 2017)	Private(Real Time Images)	200	2	Powdery mildew, Early blight	Moth-Flame Optimization (MFO) and rough set (MFORSFS)	86
Lu et al. (Lu et al. 2018)	Tomato leaves(College of Agriculture and Biotechnology, Zhejiang University)	166	1	Yellow leaf curl	Hyperspectral Imaging	100
Zhang et al.(Tm et al. 2018)	Tomato leaves(Salinas Valley, California)	66	2	Healthy,Late blight	Hyperspectral Remote Sensing	–
Hlaing et al. (Hlaing et al. 2018)	Plantvillage	3535	7	Bacterial Spot, Late Blight, Septoria Leaf Spot, Leaf Mold, Mite, Target Spot, Two Spotted Spider, Healthy	SVM-Quadratic	85.10

Table 1 (continued)

Author	Data source	Number of images	Number of class	Types of disease	Method	Accuracy (%)
Annabel and Muthulakshmi(Annabel et al. 2019)	Private	2175	3	Bacterial spot late blight, Tomato mosaic, Healthy	SVM Random Forest	82.60
					MDC	87.60
					Random Forest	94.10
Das et al.(Das et al. 2020)	Kaggle(Public)	14,000	7	Bacterial Spot, Leaf Mold, Septoria Leaf Spot,Target Spot, Two Spotted Spider Mite, Yellow Mosaic virus, Leaf Curl Virus	SVM	87.60
					LR	70.05
					RF	67.30
Basavaiah and Anthony(Basavaiah and Anthony 2020)	Plantvillage	500	4	Bacterial spot, Septoria spot, Mosaic virus, Yellow curl virus	Random Forest	94
					Decision Tree	90

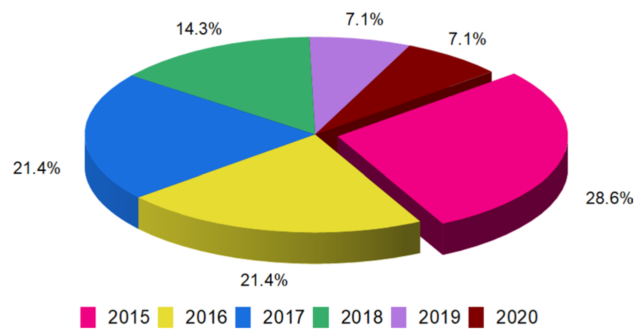


Fig. 3 Year-wise published research papers using ML for tomato leaf disease classification

Deep learning-based tomato leaf disease classification

Deep learning is a type of automated learning architecture evolved from artificial neural networks with multilayer architecture (Guo et al. 2016). This architecture is used to devise the information from simple to complex in order to segregate it into unsupervised, supervised, and reinforcement learning. Many applications involving deep learning models have found solutions for numerous image recognition problems and gave greater results in research areas viz medical diagnosis, natural language, and automatic plant disease identification. Usage of deep learning to plant species and leaf disease identification is a new approach.

The flow diagram for applying a deep learning algorithm for the detection and classification of tomato leaf disease is illustrated in Fig. 2. This section discusses work carried out employing deep learning models in the identification of tomato diseases through leaf images.

Arvind et al. (Rangarajan et al. 2018) attempted the pre-trained deep learning model adopting transfer learning (TL) concepts namely AlexNet and VGG16 to extract the features from tomato images for classification of healthy and disease-affected classes. This work produces a classification accuracy of 97.49% and 97.23% for AlexNet and VGG16, respectively. Keke Zhang et al. (Zhang et al. 2018) employed TL-based CNN model to predict the tomato plant disease through leaves images. The AlexNet, ResNet, and GoogLeNet are the pre-trained models used which utilize stochastic gradient descent (SGD) and adaptive moment estimation (Adam) optimizer for classifying the tomato diseases automatically. The ResNet model with SGD optimizer produces greater accuracy of 97.28% compared with other pre-trained models such as GoogLeNet and AlexNet.

Fuentes et al., 2017 presented different types of meta-architectures namely single shot multibox detector (SSD), faster region-based CNN (Faster R-CNN), and region-based fully CNN (R-FCN) to identify the disease-affected parts in leaf images. These architectures are utilized by CNN models such as VGG16 and ResNet. The experimental result shows that ResNet with R-FCN outperforms with an accuracy of 85.98%. Durmuş et al., 2017 employed SqueezNet and AlexNet to detect the tomato disease by means of leaf images acquired from the plantvillage database. The highest accuracy of 95.65% is obtained by AlexNet whereas SqueezNet produces 94.30%.

Brahimi et al., 2017 did the experimental work to detect tomato plant disease using GoogLeNet and AlexNet models with leaf images. The result confirms that GoogLeNet provides better performance with an accuracy of 99.18% compared with AlexNet which is 98.60%. Similarly, Suryawati et al., 2018 evaluated the performance of the deep CNN model to identify the tomato diseases. In this work, VGGNet, AlexNet, GoogleNet, and baseline CNN models are proposed to find out the disease-affected tomato plants, among which VGGNet model has several layers compared to other models and attains a better accuracy of 95.24%. The experimental results confirm that deeper architecture provides the best accuracy.

Ruedeeniraman et al. (Foysal et al. 2020a) developed embedded-based VegeCare tool using deep CNN model which identifies six varieties of tomato leaf diseases. Sardogan et al., 2018 proposed a learning vector quantization (LVQ)-based deep CNN model to classify four types of tomato disease and a healthy class. The proposed method achieves a classification accuracy of 86%. Elhassouny and Smarandache 2019 developed a deep CNN model based on MobileNet for identifying the ten varieties of tomato diseases taken from plantvillage dataset which run on a mobile platform and achieved a greater prediction accuracy of 90.3%. Rangarajan et al. (Ma et al. 2015c) proposed the VGG16 and AlexNet adopting transfer learning approach to classify the tomato disease employing plantvillage datasets and produces the recognition accuracy of 97.29% for AlexNet and 97.49% for VGG16.

Brahimi et al., 2017 demonstrated two approaches to identify ten varieties of tomato leaf disease. The first technique is to train the GoogLeNet- and AlexNet-based CNN models from the base level. In the second technique, the application of transfer learning to random forest (RF) and support vector machine (SVM) optimizer. From the experimental results, the pre-trained CNN model gives better performance compared to RF and SVM classifiers. Likewise, Manpreet Kaur and Rekha Bhatia used a pre-trained CNN model, ResNet101 to identify the six types of tomato diseases using leaf images. The tomato leaf images from the plantvillage dataset are used for the experiment. The result signifies that the ResNet101 model achieved a classification accuracy of 98.8%. (Kaur and Bhatia 2019). Kumar and Vani presented CNN models such as LeNet, VGG16, ResNet50, and Xception for tomato leaf disease classification. The models are trained using 14,903 images belonging to nine diseased and one healthy class from the plantvillage dataset. The experimental result shows that VGG16 provides better performance with an accuracy of 99.25% compared with LeNet 96.27%, ResNet50 98.65%, and Xception 98.13%. Furthermore, the models are evaluated on tomato-segmented leaf images and VGG16 achieved the highest accuracy of 99.11% compared with other models (Kumar and Vani 2019).

Tm et al. reported the LeNet model to find out the ten different classes of tomato leaf classes in which 18,160 tomato leaf images are resized to 60×60 pixels resolutions. From the experimental result, it shows that the model proposed to achieve a classification accuracy of 94.85% with 30 epochs (Tm et al. 2018). Foysal et al. reported a novel CNN model that consists of 15 layers to identify the occurrence of five different tomato diseases using leaf images. The experimental result shows that the proposed model achieved an accuracy of about 76% on the test dataset which consists of 600 images (Foysal et al. 2020b).

Jiang et al. introduced an improved ResNet50 model to classify the three different types of tomato leaf diseases such as spot blight, yellow leaf curl, and late blight. The ResNet50 model uses a leaky ReLU activation function and to enhance the accuracy the filter size is modified to 11×11 in each convolution layer. The result shows that ResNet50 obtained an accuracy of 98% on the test dataset (Jiang et al. 2020). Prabhakar et al. developed an intelligent system using foldscope and ResNet101 to classify the severity level like mild, moderate, and severe of the early blight disease in the tomato leaf. The ResNet101 achieves an accuracy of 94.6% in severity assessment of disease which is highest as compared which the other CNN models such as VGG16, VGG19, AlexNet, GoogleNet, and ResNet101(Prabhakar et al. 2020).

Chen developed a framework combining Artificial Bee Colony algorithm (ABCK), Binary Wavelet Transform combined with Retinex (BWTR), and Both-channel Residual Attention Network model (B-ARNet) to recognize five different types of tomato leaf diseases such as early blight, late blight, citrinitas leaf curl, leaf mold, and bacterial leaf spot. The experiment is conducted using 8616 images of tomato leaves obtained from the Hunan Vegetable Institute. The proposed method based on the combination of ABCK-BWTR and B-ARNet achieved a classification accuracy of 89% (Chen et al. 2020). Gadekallu et al. proposed a deep learning model using principal component analysis (PCA)-whale optimization algorithm (WOA) and deep neural network (DNN) to classify tomato leaf diseases. The tomato leaf images used in this work are collected from the plantvillage database which consists of nine types of diseased class and a healthy class. The approach PCA-WOA is used to identify the significant features of the tomato leaf and it is fed to the

DNN which performs classification. The proposed model achieves a test accuracy of 94% (Gadekallu et al. 2020).

Kaushik et al. implemented pre-trained ResNet50 applying transfer learning concept for detection of five different types of tomato leaf disease class and a healthy class. The proposed model achieved a detection accuracy of 97.01% on the test dataset (Kaushik et al. 2020).

Agarwal et al. developed a CNN model for disease identification of tomato plants using leaf images. The proposed model uses three convolution layer, three max-pooling layers, and two fully connected layers to classify nine diseased class and a healthy class. The experimental output signifies that the proposed model achieves the highest test accuracy of 91.20% compared with pre-trained models such as VGG16-77.20%, MobileNet-63.75%, and Inception V3-63.40% (Agarwal et al. 2020). Thangaraj et al. proposed the transfer learning-based modified Xception model for automatic identification of tomato leaf diseases. The tomato dataset extracted from the plantvillage database comprises nine diseased class and one healthy class. The proposed model is tested using three different types of optimizers such as adaptive moment estimation (Adam), root mean square propagation (RMSprop), and stochastic gradient descent (SGD). The experimental result proves that the proposed model with Adam optimizer outperforms with an accuracy of 99.55% compared with SGD and RMSprop (Thangaraj et al. 2020).

Shijie et al. proposed the hybrid model with a combination of VGG16 and SVM to classify tomato leaf diseases. VGG16 acts as the feature extractor and SVM performs the classification of disease. The transfer learning (TL) approach is introduced to fine-tune the pre-trained model to enhance model efficiency. This hybrid model achieved an accuracy of 89% in the classification of diseases (Shijie et al. 2017).

Karthik et al. (Karthik et al. 2020) developed an attention-embedded deep residual network model to identify the type of disease infected in the tomato leaves. The plantvillage dataset is used to conduct the experiment which comprises of three different types of disease such as early blight, late blight, and leaf mold. The proposed model utilizes the features learned by the CNN at different processing levels. The experiment results show that the proposed model achieved a detection accuracy of 98% on the validation dataset.

Fuentes et al. (Fuentes et al. 2018) proposed deep neural network-based framework that performs real-time detection of diseases and pests of tomato crop. The framework comprises of three major units: First, A key diagnosis unit (bounding box generator) which obtains bounding box where it denotes the location and type of diseases and pests; next, secondary unit (CNN filter bank) which trains each CNN classifier independently to filter the misclassified samples; finally, integration unit which merges the information of key diagnosis and auxiliary unit by allowing true positives and removing false positives. The experimental result shows that the proposed method achieved 96% of accuracy for recognition of tomato disease and pest.

Liu and Wang build improved YOLO V3 model to identify the tomato diseases and pests in the real-time environment. The proposed model employs image pyramid to perform multiscale feature detection in order to enhance the detection accuracy and speed of the model. Subsequently, the model can accurately detect the region and type of the diseases and pests spotted in the tomato plant. Based on the experimental result, improved YOLO V3 model achieved the highest average recognition accuracy of 92.36% compared with models such as SSD, Faster R-CNN, and original YOLO V3 models (Liu and Wang 2020a).

Liu and Wang (Wang and Liu 2021) proposed novel YOLO-Dense model to detect the tomato anomalies under complex background environment. The model utilizes the multiscale training strategy to improve the accuracy of anomalies detection. Based on the experimental result, YOLO-Dense model achieved the recognition accuracy of 96.41% which is highest compared with SSD, Faster R-CNN, and YOLO V3 employed in the same task.

In another study, Liu and Wang et al. (Liu and Wang 2020b) developed YOLO V3-MobileNetV2 model to recognize the tomato Gray leaf spot in the early stage. The experimental result shows that the proposed model provides the high recognition accuracy for early disease prediction compared with the Faster-RCNN and SSD models.

Fuentes et al. (Fuentes et al. 2019) developed a diagnostic system which automatically detects the location of anomalies in the tomato plant image and also provides the detailed information about the disease symptoms. The system composed on two main units: (1) detector is trained employing the set of features that contains the region of anomalies in the tomato plants using deep neural network; (2) long-short term memory (LSTM) is used to obtain the symptoms description based on the input features from the detector. This approach uses the newly generated dataset named tomato plant anomalies description dataset and achieved an average recognition accuracy of 92.50%.

Rubanga et al. (Rubanga et al. 2020) detect the infestation of the tuta absoluta virus in the tomato leaves using pre-trained CNN architecture such as Inception V3, VGG16, VGG19, and ResNet. The result obtained from the experiment shows that Inception V3 produces the highest detection accuracy of 87.20% than VGG16, VGG19 and ResNet.

Nandhini and Ashokkumar proposed CNN-based approach to identify the four different types of tomato leaf diseases such as bacterial spot, Septoria leaf spot, late blight, and tomato mosaic virus. Two different CNN architectures such as VGG16 and Inception V3 are employed in this work and its parameters are optimized using an improved crossover-based monarch butterfly optimization (ICRMBO) algorithm. The experiment is conducted using the tomato leaf disease images obtained from the plantvillage dataset. The test result confirmed that optimized VGG16 and Inception V3 achieves the classification accuracy of 99.98% and 99.94%, respectively (Nandhini and Ashokkumar 2021).

The comparison chart describing the adopted deep learning model to detect the tomato plant disease is shown in Table 2 in terms of the data source used, number of images used, number of the class used, and types of disease identified, employed CNN model and achieved accuracy. Yearwise published research papers using DL for tomato leaf disease classification is illustrated in Fig. 4

Figure 5 provides the details about the number of papers published to identify the tomato plant disease using ML and DL approaches. DL models with the highest accuracy achieved for tomato leaf disease identification are illustrated in Fig. 6, and specific number of research papers published using various DL models for tomato leaf disease detection is illustrated in Fig. 7.

From the observation, the deep learning technique is the best way to identify the diseased plant through raw image datasets. Generally, CNN-based classifiers are used to classify the changes in the provided input image datasets depending upon the optimal parameters such as object shape, texture, and color. These optimal parameters can be trained and tested easily when the input images are linear.

A lot of challenges are still available when employing a more complex CNN model. All image-based problems cannot be dealt with image processing and computer vision technology. The CNN model-based work proposed till now depends on data acquisition background. This type of situation leads to capture a variety of characteristics which can analyze the image datasets difficult in tomato disease identification and classification. In near future, some deep learning models are anticipated to produce greater accuracy, and also researchers required to train and test the proposing and developed architecture on real-time image datasets. Recently, many deep learning architectures are proposed by the researchers using plantvillage image datasets to evaluate the performance of their model. This dataset comprises images with a plain and simple background. Even though, real environmental conditions should be included in order to get greater accuracy when the model to be implemented in the real-time automated plant disease identification and classification.

An overall review is needed to know the exact factors which affect the identification of tomato plant diseases such as size and classes of image datasets, illumination, and learning rate, etc. This section concludes that so many deep learning models are implemented for tomato leaf disease identification; however, there is a requirement for better identification of tomato leaf disease in a variety of scenarios like changes in illumination conditions and real environmental conditions.

Open access dataset for tomato leaf images

The major issue faced by deep learning researchers is collecting datasets with sufficient quantity in order to produce the best accuracy. This section discusses the open-access datasets holding healthy and diseased leaf images of different plants.

The plantvillage (Hughes and Salathé 2015) datasets which is the largest open-access crop image repository. This dataset comprises 54,306 leaf images including infected and healthy leaf images of 14 different types of plants. All leaf images in this dataset are taken with the plain background in a controlled environment and leaves labeled by the experts.

PlantDoc (Singh et al. 2020) dataset consists of 2598 real-time leaf images of 13 different plant species with 27 classes including 17 diseased and 10 healthy classes. The images presented in this dataset are captured under non-controlled environment. The plant varieties include bell pepper, apple, blueberry, corn, cherry, grape, potato, peach, squash, strawberry, raspberry, tomato, and soybean.

PlantDisease dataset employed in Arsenovic et al. (2019) is the extended version of the plantvillage dataset which consists of 79,265 images holding 12 different plant species and 42 classes including healthy and diseased class. This dataset includes the new images of both healthy and diseased leaves captured under different environmental backgrounds. The plant species present in the dataset are tomato, sugar beet, strawberry, plum, potato, peach, onion, wheat, grape, cherry, bell pepper, and apple.

The tomato disease datasets used in Tian et al. (2019) are generated at South China Agricultural University which consists of 1000 leaf disease images among which 200 images are with white backgrounds and 800 images have a natural background. In this dataset, 200 images with the white color background are exclusively used for experimental purpose and remaining 800 images with the real background is used for testing purpose. Table 3 lists the summary of tomato image datasets.

Deep learning framework

Recent advancements in deep learning contribute to several open-source deep learning frameworks that are used to implement the DL models. These frameworks offer building blocks to design, train and validate the deep CNNs with high-level programming interfaces. This section provides a brief overview of DL frameworks used in the development of DL models. TensorFlow (Abadi et al. 2016) is one of the popular deep learning framework

Table 2 Comparison table showing performance of deep learning algorithm applied to the detection of tomato plant disease through leaf images

Author	Data source	Number of images	Number of Class	Types of disease	CNN model	Accuracy (%)
Brahimi et al.(Brahimi et al. 2017)	Plantvillage	14,828	9	Late blight, Early blight, Bacterial Spot, Leaf Mold, Septoria spot, Target Spot, Spider mites, Yellow Leaf Curl Virus, mosaic virus	AlexNet GoogLeNet	98.60 99.18
Durmuş et al. (Durmuş et al. 2017)	Plantvillage	19,742	10	Late blight, Early blight, Bacterial Spot, Leaf Mold, Septoria spot, Target Spot, Spider mites, Yellow Leaf Curl Virus, Healthy, mosaic virus	AlexNet SqeezeNet	95.65 94.3
Shijie et al. (Shijie et al. 2017)	Plantvillage	7040	11	Late blight, Early blight, Bacterial Spot, Septoria spot, Target Spot, Leaf Mold, Tomato gray spot, Spider mites, mosaic virus, Yellow Leaf Curl Virus, Healthy,	VGG16 + SVM	89
Zhang et al. (Zhang et al. 2018)	Plantvillage	5,550	9	Late blight, Early blight, Corynespora leaf spot disease, Leaf Mold, Septoria spot, Two-spotted Spider mites, Yellow Leaf Curl Virus, Mosaic virus, Healthy	AlexNet GoogLeNet ResNet	95.83 95.66 97.28
Arvind et al. (Rangarajan et al. 2018)	Plantvillage	13,262	7	Leaf Mold, Two-spotted Spider mites, Late blight, Mosaic virus, Target spot, Yellow Leaf Curl Virus, Healthy	AlexNet VGG16	97.49 97.29
Sardogan et al.(Sardogan et al. 2018)	Plantvillage	500	5	Late blight, Bacterial spot, Yellow leaf curl virus, Septoria leaf spot	CNN	86
Suryawati et al. (Suryawati et al. 2018)	Plantvillage	18,160	10	Early blight, Late blight Bacterial Spot, Septoria spot, Leaf Mold, Target Spot, Spider mites, Yellow Leaf Curl Virus, mosaic virus, Healthy	VGGNet AlexNet GoogleNet Baseline	95.24 91.52 89.68 84.58
Tm et al.(Tm et al. 2018)	Plantvillage	18,160	10	Early blight, Late blight Bacterial Spot, Leaf Mold, Septoria spot, Target Spot, Spider mites, Yellow Leaf Curl Virus, Mosaic virus, Healthy	LeNet	94.85
Elhassouny et al.(Elhassouny and Smarandache 2019)	Plantvillage	7176	10	Late blight, Early blight, Bacterial Spot, Leaf Mold, Septoria spot, Target Spot, mosaic virus, Spider mites, Yellow Leaf Curl Virus, Healthy	CNN	90.30
Kaur and Bhatia (Kaur and Bhatia 2019)	Plantvillage	6888	7	Late blight, Bacterial spot, Septoria leaf spot. Yellow Leaf Curl Virus, Mosaic virus, Leaf Moldmanual, Healthy	ResNet101	98.8
Kumar and Vani (Kumar and Vani 2019)	Plantvillage	14,903	10	Late blight, Early blight, Bacterial Spot, Leaf Mold, Septoria spot, Spider mites, Target Spot, mosaic virus, Yellow Leaf Curl Virus, Healthy	LeNet ResNet50 Xception VGG16	96.27 98.65 98.13 99.25

Table 2 (continued)

Author	Data source	Number of images	Number of Class	Types of disease	CNN model	Accuracy (%)
Foysal et al.(Foysal et al. 2020b)	Plantvillage	3000	6	Late bright, Bacterial spot, Septoria leaf spot. Spider mites, Tomato leaf curl virus, Healthy	CNN	76
Jiang Et al (Jiang et al. 2020)	Plantvillage	3000	3	Spot blight, Yellow leaf curl, and Late blight	ResNet50	98
Prabhakar et al.(Prabhakar et al. 2020)	Plantvillage	2388	1	Early blight	VGG16	85.10
					VGG19	86.60
					GoogleNet	86.40
					AlexNet	89.50
					ResNet50	90
Chen (Chen et al. 2020)	Private(Hunan Vegetable Institute)	8616	5	Late blight, Early blight, Citrinitas leaf cur, Leaf mold, Bacterial leaf spot	ABCK-BWTR and B-ARNet	89
Gadekallu et al.(Gadekallu et al. 2020)	Plantvillage	18,159	10	Early blight, Late blight Bacterial Spot, Leaf Mold, Septoria spot, Target Spot, Spider mites, mosaic virus, Yellow Leaf Curl Virus, Healthy	Deep Neural Network	94
Kaushik et al.(Kaushik et al. 2020)	Plantvillage	12,206	6	Bacterial spot, Early blight, Septoria leaf spot, Tomato yellow leaf curl virus, Tomato mosaic virus	ResNet50	97.01
Agarwal et al.(Agarwal et al. 2020)	Plantvillage	10,000	10	Late blight, Early blight, Bacterial Spot, Leaf Mold, Target Spot, Spider mites, Septoria spot, Yellow Leaf Curl Virus, mosaic virus, Healthy	VGG16	77.20
					MobileNet	63.75
					Inception V3	63.40
					CNN	91.20
Thangaraj(Thangaraj et al. 2020)	Plantvillage	16,578	10	Late blight, Early blight, Bacterial Spot, Septoria spot, Leaf Mold, Target Spot, mosaic virus, Spider mites, Yellow Leaf Curl Virus, Healthy	Xception	99.55
Karthik et al.(Karthik et al. 2020)	PlantVillage	95,999	4	Healthy, Early blight, Late blight, Leaf mold	CNN	98
Fuentes et al.(Fuentes et al. 2018)	Private	8927	11	Leaf mold, Gray mold, Canker, Plague, Miner, Powdery mildew, Whitefly, Yellow leaf curl, low temperature, Nutritional excess, Background	Faster R-CNN + CNN	96
Liu and Wang(Liu and Wang 2020a)	Private	15,000	12	Early blight, Late blight, Yellow leaf curl virus,Brown spot,Coal pollution, Gray mold, Leaf mold,Navel rot, Leaf curl disease, Mosaic virus, Leaf miner,Greenhouse whitefly	SSD	84.32
					Faster R-CNN	90.67
					YOLO V3	88.31
					Improved YOLO V3	92.39

Table 2 (continued)

Author	Data source	Number of images	Number of Class	Types of disease	CNN model	Accuracy (%)
Liu and Wang(Wang and Liu 2021)	Private	15,000	12	Early blight, Late blight, Yellow leaf curl virus, Brown spot, Coal pollution, Gray mold, Leaf mold,Navel rot, Leaf curl disease, Mosaic virus, Leaf miner,Greenhouse whitefly	SSD	84.32
					Faster R-CNN	90.67
					YOLO V3	88.31
					YOLO-Dense	96.41
Liu and Wang(Liu and Wang 2020b)	Private	2385	1	Tomato Gray Leaf Spot	YOLO V3-MobileNet V2	90.29
Fuentes et al.(Fuentes et al. 2019)	Private	8927	11	Leaf mold, Gray mold, Canker, Plague, Miner, Powdery mildew, Whitefly, Yellow leaf curl, low temperature, nutritional excess, background	Deep Neural Network + Long-Short Term Memory (LSTM)	92.50
Rubanga et al.(Rubanga et al. 2020)	Private	3460	3	No tuta, Low tuta, High tuta	VGG16	87.10
					VGG19	78.30
					ResNet50	83.70
					Inception V3	87.20
Nandhini and Ashokkumar(Nandhini and Ashokkumar 2021)	PlantVillage	6218	5	Healthy, Late Blight, Bacterial Spot, Septoria Leaf Spot, Mosaic Virus	VGG16	99.98
					Inception V3	99.94

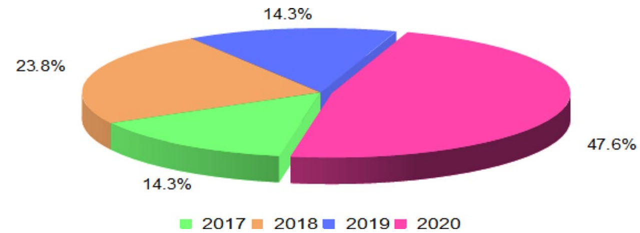


Fig. 4 Year-wise published research articles using DL for tomato leaf disease classification

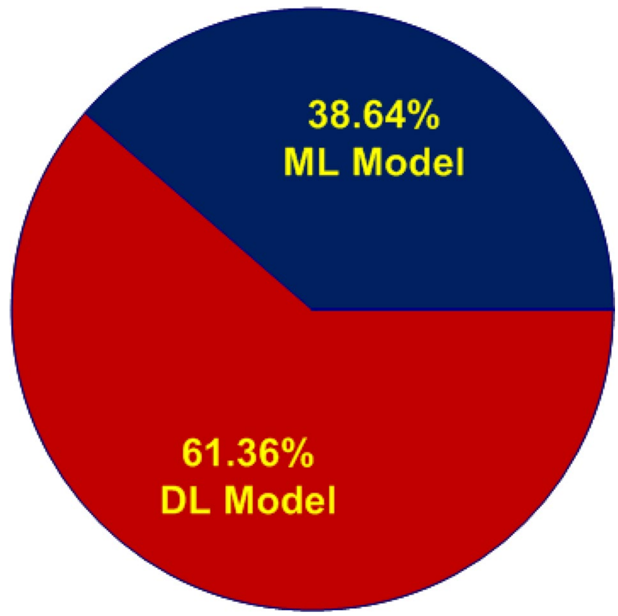


Fig. 5 Comparison of the number of machine learning and deep learning article published for plant disease classification

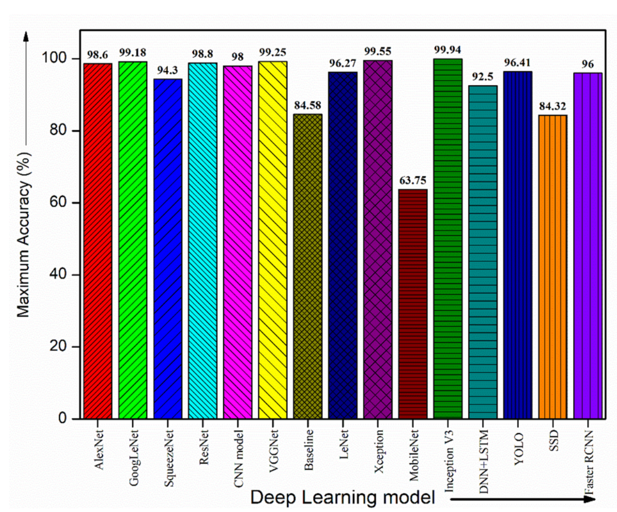


Fig. 6 Accuracy comparison of DL models for tomato leaf disease identification

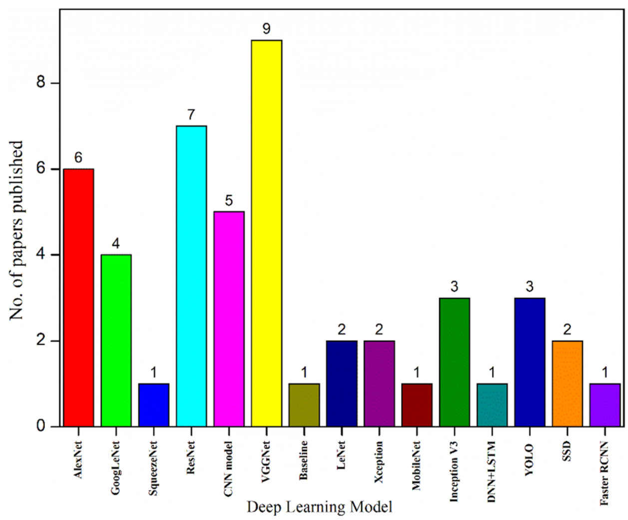


Fig. 7 Number of research papers published using DL models for tomato leaf disease detection

released by the Google Brain team. Following, Keras (Chollet 2015) is developed using python and it runs on top of TensorFlow. Subsequently, PyTorch (Paszke et al. 2017) developed by Facebook which is one of the basic software tools for DL framework after TensorFlow. This is the port to the torch deep learning framework which can be utilized to build DCNNs and performing tensor computations. Caffe (Vedaldi et al. 2014) is the one more widely used open-source deep learning framework developed by Yangqing Jia at the University of California, Berkeley to solve the image processing-based problems. Sonnet (Sonnet, 2019) is another deep learning framework developed by DeepMind which provides a platform to design complex neural network architecture. The next deep learning framework employed recently is MXNet (Chen et al. 2015) which is highly scalable framework used to build DL models. Other than above-mentioned deep learning framework, there are few frameworks employed in the DL model development are Gluon, Chainer, Swift, Theano, DeepLearning4J, Microsoft Cognitive Toolkit, ONNX, and PaddlePaddle.

Evaluation metrics

The quantitative performance analysis of deep learning models in plant disease classification is performed through statistical evaluation measurements. The four statistical measurement data used for performance analysis of deep learning models are TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative). TP represents the number of true positive images which are exactly predicted as infected one. TN denotes perfectly predicted as healthy images whereas actual value is also healthy. A number of sample images that are wrongly identified as defective correspond to FP. FN represents a number of image samples that are incorrectly predicted as non-infected.

$$Sensitivity/Recall = \frac{TP}{TP + FN} \tag{1}$$

$$Specificity = \frac{TN}{TN + FP} \tag{2}$$

$$Accuracy = \frac{TP + TN}{TN + TP + FN + FP} \tag{3}$$

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$F1 - score = 2 \times \frac{(Sensitivity \times \Pr ecision)}{(Sensitivity + \Pr ecision)} \tag{5}$$

Factors that affect the AI-based classifiers

The main factors which affect the machine learning/deep learning-based classifiers are discussed in this section.

Factors that affect the ML-based classifiers

a) The ML approach will not produce the expected results if the training data comprise of more irrelevant features and insufficient relevant features.

Table 3 Summary of tomato image datasets

Datasource	Total number of images	Number of plant species	Number of classes	Number of tomato leaf images	Number of tomato classes
Plantvillage	54,306	14	38	16,578	10
PlantDoc	2,598	13	27	741	9
PlantDisease	79,265	12	42	7832	3
South China Agricultural University(Tomato)	–	–	–	1000	2

b) For extracting significant features from data, ML is not a good option.
c) ML utilizes hand-crafted features as input such as a gradient histogram, local binary patterns etc. in order to produce better classification based on images.
d) A person must define and manually code the implemented features in ML systems based on the data type
e) Most implemented features in ML must be defined by professional and then manually coded according to the domain and data type.

Factors that affect the DL-based classifiers

a) Shallow DL models are preferred for few image datasets
b) The effect of diversity of target dataset and selecting the best model opt for the target class is very much important compared quantity of images available in the datasets
c) No standardized computer vision technologies for automatic classification of tomato leaf disease.
d) Geographical and environmental-related information has a prominent effect on gathering the input image datasets and also have an impact on analyzing the disease identification
e) There are no defined disease symptoms
f) More challenging to discrete the healthy leaf images and diseased region
g) The similarity of different disease symptoms makes the researchers depend on the existing methods to discriminate
h) Methods are used to detect the disease affected on tomato plant but it is failed to inform about the severity of the disease and how to rectify it.
i) CNN/DCNN models with smaller datasets produce greater prediction accuracy; however, that's not reliable and trustworthy results
j) The higher computational cost for running the CNN/DCNN models in CPU's compared to GPU's
k) Most of the images in the datasets are taken in perfect lighting conditions but in real-time conditions slightly differ and yield different output.
l) The CNN models does not provide the best data classification by incorporating multiple convolutions.
m) The lack of large datasets is the obstacle for applying deep learning approach in the area of plant leaf disease detection. Even though, plantvillage is an open-source database which has a massive database with thousands of images. This database don't have actual field images.
n) Another issue faced by DL researchers is annotating self-collected data with the aid of an agriculture field expert.
o) Early detection of plant diseases is critical. Farmers can take cost-effective corrective action if they detect an infected plant at the early stage. For this reason, hyperspectral imaging has been used, but the area captured on the ground using thermal sensors and light reflector sensors is very large, making detection of a disease or contaminated area difficult.
p) There are no specified shapes in the leaves for the mild symptoms of tomato leaf diseases and lesion spots.

Discussion

For a long time, traditional image processing and machine learning approaches have been employed to identify tomato leaf diseases. It is very difficult to identify the significant features of the various diseases in the tomato leaves employing image processing and machine learning techniques since it uses the hand-crafted method to extract the features. Therefore, the feature selection has to be performed automatically and optimal set of features to be learned for accurate disease classification. The study (26–35,37–41,61) employing image processing and ML approaches explores low accuracy in recognition of tomato disease using leaves images. Moreover, the studies (26–35,37–41,61) presented in the (see Machine learning-based tomato leaf disease classification section) uses small dataset and small number of class which achieves significant performance in disease identification. Furthermore, image processing and ML methods using large dataset and increase in number of classes results a reduction in recognition accuracy of tomato leaf diseases. From the results, image processing and machine learning provides better performance in terms of minimum dataset with a smaller number of classes.

Subsequently reviewing various papers which have employed DL in identifying tomato plant leaf diseases, it noted that DL-based classification resulted in greater prediction accuracy. In all comparisons that has been made among DL and image processing ML methods, it is observed that deep learning always outperformed. Several well-known deep learning models such as ResNet(Zhang et al. 2018; Kaur and Bhatia 2019; Kumar and Vani 2019; Jiang et al. 2020; Prabhakar et al. 2020; Tian et al. 2019; Rubanga et al. 2020), VGGNet(Rangarajan et al. 2018; Suryawati et al. 2018; Kumar and Vani 2019; Prabhakar et al. 2020; Agarwal et al. 2020; Shijie et al. 2017; Rubanga et al. 2020; Nandhini and Ashokkumar 2021), GoogleNet (Zhang et al. 2018; Brahimi et al. 2017; Suryawati et al. 2018; Prabhakar et al. 2020), AlexNet(Rangarajan et al. 2018; Zhang et al. 2018; Durmuş et al. 2017; Brahimi et al.2017; Suryawati et al. 2018; Prabhakar et al. 2020), SqeezeNet(Durmuş et al. 2017), LeNet(Tm et al. 2018; Kumar and Vani 2019), Xception(Thangaraj et al. 2020; Kumar and Vani 2019), MobileNet(Agarwal et al. 2020), InceptionV3(Agarwal et al. 2020; Rubanga et al. 2020; Nandhini and Ashokkumar 2021) and customized CNN model(Sardogan et al. 2018; Elhassouny and Smarandache 2019; Foysal et al. 2020b; Gadekallu et al. 2020; Agarwal et al. 2020; Karthik et al. 2020; Fuentes et al. 2018),Faster-RCNN(Fuentes et al. 2018; Liu and Wang 2020a; Wang and Liu 2021),SSD(Liu and Wang 2020a; Wang and Liu 2021),YOLO(Liu and Wang 2020a, 2020b; Wang and Liu 2021) and DNN-LSTM(Fuentes et al. 2019) have been published in the literature for handling the tomato plant diseases employing leaf images. Many researchers have been motivated by the performance of these models to use pre-trained models in leaf disease identification task. This review paper has found that pre-trained models with transfer learning approaches provided higher prediction accuracy compared to other approaches such as custom CNN model and the model trained from scratch. This study demonstrates that 85% (Thangaraj et al. 2020; Tm et al. 2018; Rangarajan et al. 2018; Zhang et al. 2018; Durmuş et al. 2017; Brahimi et al. 2017; Suryawati et al. 2018; Kaur and Bhatia 2019; Kumar and Vani 2019; Jiang et al. 2020; Prabhakar et al. 2020; Kaushik et al. 2020; Shijie et al. 2017; Rubanga et al. 2020; Nandhini and Ashokkumar 2021) of deep learning models rely on transfer learning and hyperparameter tuning concepts to improve the prediction accuracy of finding tomato leaf diseases. The majority of the study uses plantvillage (Thangaraj et al. 2020; Hlaing et al. 2017, 2018; Tm et al. 2018; Rangarajan et al. 2018; Zhang et al. 2018; Durmuş et al. 2017; Brahimi et al. 2017; Suryawati et al. 2018; Sardogan et al. 2018; Elhassouny and Smarandache 2019; Kaur and Bhatia 2019; Kumar and Vani 2019; Foysal et al. 2020b; Jiang et al. 2020; Prabhakar et al. 2020; Gadekallu et al. 2020; Basavaiah and Anthony 2020; Kaushik et al. 2020; Agarwal et al. 2020; Shijie et al. 2017; Karthik et al. 2020; Nandhini and Ashokkumar 2021) dataset, which is freely available and it contains more than 50,000 images, which is sufficient to train any type of CNN model. This open-source database leverages the researchers to use machine learning and deep learning techniques in plant disease identification. As far tomato dataset is concerned, plantvillage database consists of ten tomato classes of leaf images including nine diseased class and one healthy class. However, the number of images in each class is unbalanced. The datasets including actual field images are still lacking and it is not adequate to train the deep learning models. To enhance the size of real-time and plantvillage datasets, data augmentation is employed. The number of images are relied on the data augmentation to obtain an adequate number of images for training a deep learning model. Augmentation techniques have been employed almost universally to artificially extend the dataset in order to increase the dataset's output capabilities. Cropping, rotation, grayscale conversion, and adding noise are some of the augmentation techniques. This review found that 64% (Thangaraj et al. 2020; Hlaing et al. 2017, 2018; Tm et al. 2018; Rangarajan et al. 2018; Zhang et al. 2018; Durmuş et al. 2017; Brahimi et al. 2017; Suryawati et al. 2018; Sardogan et al. 2018; Elhassouny and Smarandache 2019; Kaur and Bhatia 2019; Kumar and Vani 2019; Foysal et al. 2020b; Jiang et al. 2020; Prabhakar et al. 2020; Gadekallu et al. 2020; Basavaiah and Anthony 2020; Kaushik et al. 2020; Agarwal et al. 2020; Shijie et al. 2017; Karthik et al. 2020; Nandhini and Ashokkumar 2021) of research article utilized plantvillage dataset and 36%(Sabrol and Satish 2016; Xie et al. 2015; Mokhtar et al. 2015a, 2015b, 2015c; Hassanien et al. 2017; Sabrol and Kumar 2016a, 2016b; Lu et al. 2018, 2013; Tm et al. 2018; Annabel et al. 2019; Muthukannan and Latha 2015; Chen et al. 2020; Fuentes et al. 2018, 2019; Liu and Wang 2020a, 2020b; Wang and Liu 2021; Rubanga et al. 2020) of research article employs the real-time dataset collected from various agricultural field. Figure 8 depicts the percentage distribution of data sources like plantvillage and actual field to acquire tomato leaf images. The research studies discussed in (see Machine learning-based tomato leaf disease classification section) and (see Deep learning-based tomato leaf disease classification section) explores low accuracy in recognition and classification of images obtained from the actual field environment. It is noted that the models trained specifically with the datasets generated in the controlled environment provides low detection accuracy once tested with the actual field images acquired under the different conditions such as variation in illumination, different background, size and resolution. Comparing the results of the traditional image processing and ML methods, 52% of the methods used in the

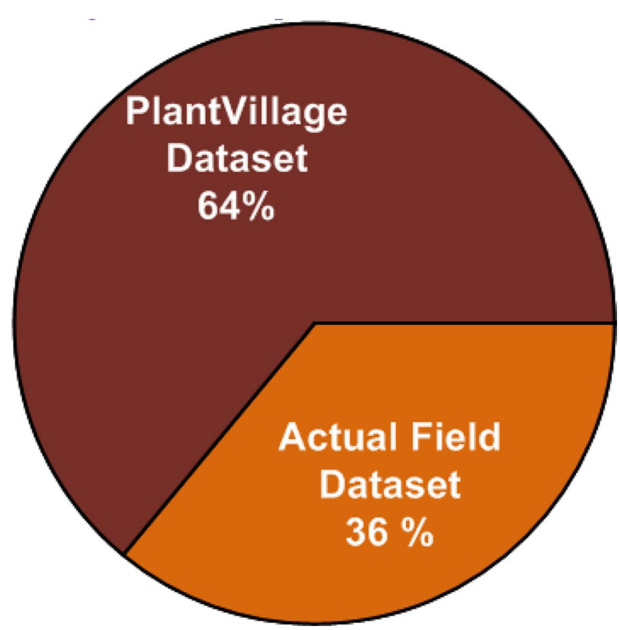


Fig. 8 Percentage distribution of data sources utilized to acquire tomato leaf images

research article listed in (see Machine learning-based tomato leaf disease classification section) achieves an accuracy more than 90% (Sabrol and Satish 2016; Xie et al. 2015; Mokhtar et al. 2015a, 2015b, 2015c; Lu et al. 2018, 2013; Annabel et al. 2019; Muthukannan and Latha 2015; Basavaiah and Anthony 2020) and remaining 48% of the article produce classification accuracy below 90%(Hlaing et al. 2017, 2018; Hassanien et al. 2017; Sabrol and Kumar 2016a, 2016b; Das et al. 2020). The graphical representation of percentage distribution is depicted in Fig. 9. As observed in (see Machine learning-based tomato leaf disease classification section), traditional methods achieve good results only with small amount of data. Similarly, DL models are concerned, 68% of models employed in the research article listed in (see Machine learning-based tomato leaf disease classification section) achieved accuracy above 90% (Thangaraj et al. 2020; Tm et al. 2018; Rangarajan et al. 2018; Zhang et al. 2018; Durmuş et al. 2017; Brahimi et al. 2017; Suryawati et al. 2018; Elhassouny and Smarandache 2019; Kaur and Bhatia 2019; Kumar and Vani 2019; Jiang et al. 2020; Gadekallu et al. 2020; Kaushik et al. 2020; Agarwal et al. 2020; Karthik et al. 2020; Fuentes et al. 2018, 2019; Liu and Wang 2020a, 2020b; Wang and Liu 2021; Nandhini and Ashokkumar 2021) in identification of tomato leaf diseases and below 90%(Suryawati et al. 2018; Sardogan et al. 2018; Foysal et al. 2020b; Prabhakar et al. 2020; Chen et al. 2020; Agarwal et al. 2020; Shijie et al. 2017; Liu and Wang 2020a; Wang and Liu 2021; Rubanga et al. 2020) is achieved by the remaining 32% article which is presented in Fig. 10.

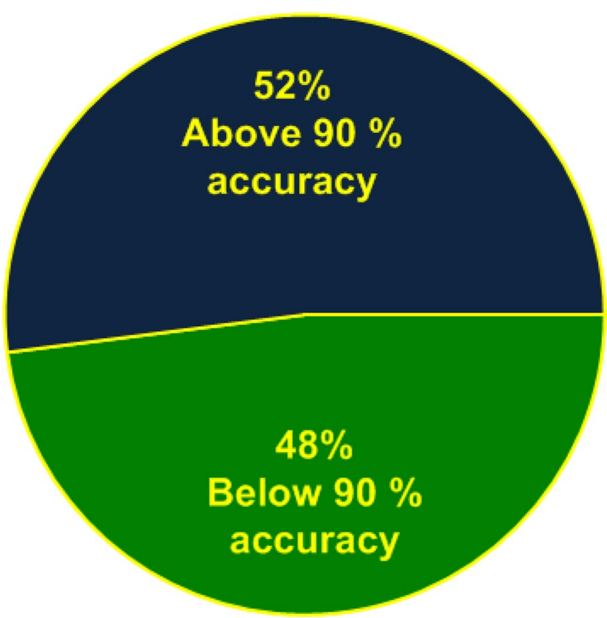


Fig. 9 Percentage distribution representing the efficiency of the ML models in tomato leaf disease identification

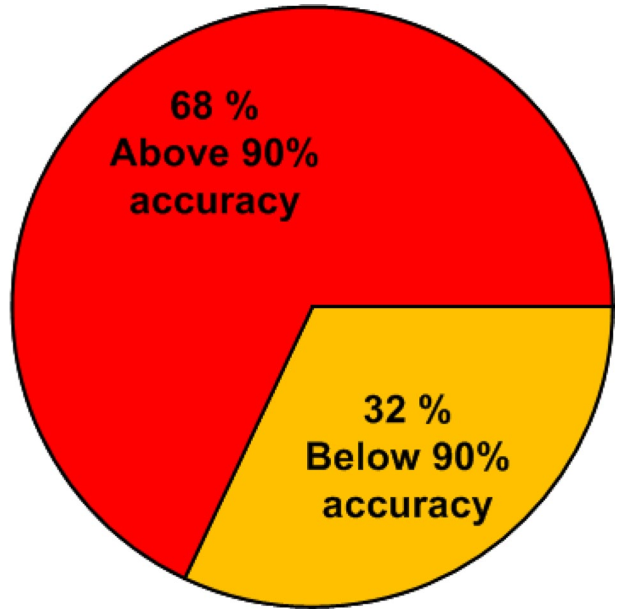


Fig. 10 Percentage distribution representing the efficiency of the DL models in tomato leaf disease identification

The accuracy of the models varies for each class of diseases, and this study confirms that models have minimal variations in the prediction of class accuracies. The variation happened because individual models are not developed for different disease classes. All deep networks surveyed for this study outperformed traditional ML-based feature extraction and classification techniques. Additionally, some DL-based models were found to be more frequently used than others in recent publications on tomato leaf disease identification as shown in Fig. 7. From Fig. 7, VGGNet model is the most widely used model in recent literature on tomato leaf disease detection. Following that, ResNet and AlexNet are all well-known and have been in the spotlight for the past five years. The comparisons Tables 1 and 2 demonstrate that the classification accuracy for detecting tomato leaf disease is significantly higher when using different types of deep learning models compared with traditional machine learning approach.

Conclusion

In recent years, the agricultural sector has faced numerous challenges. This paper provides an up-to-date analysis of current research in this field of tomato leaf disease identification based on artificial intelligence technique. The key objective of this work is to analyze different machine learning and deep learning techniques broadly used to classify the tomato leaf diseases. In this work, 44 related research work have read and their works are analyzed based on dataset, pre-processing techniques employed, models used, and overall prediction accuracy. We focused on analyzing data source (public and private), highest recognition accuracy and methods. From the literature study, the deep leaning model compared to other conventional methods like image processing, machine learning, and neural networks outperforms for tomato disease identification using leaf images. The early identification of tomato plant disease reduces the costs by skipping the unnecessary application of pesticide to the plants. The use of deep learning with hyperspectral imaging is a current emerging technique that is recommended for the early detection of tomato leaf disease. The severity of tomato plant diseases exposed to nearby plants with increasing time, so that the customized deep learning models can be employed to identify and classify the tomato leaf diseases for the duration of its entire cycle of occurrence. In order to reduce the convergence time and also enhance the prediction accuracy, fusion of low- and high-level features of the CNN models can be used. In future, incorporating agriculture robots and drones for classifying disease-affected plant by capturing leaf images automatically.

Declarations

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

References

Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Zheng X (2016) Tensorflow: A system for large-scale machine learning. In *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)* (pp. 265–283).

Agarwal M, Singh A, Arjaria S, Sinha A, Gupta S (2020) ToLeD: tomato leaf disease detection using convolution neural network. Procedia Computer Science 167:293–301

Annabel LSP, Muthulakshmi V (2019) AI-Powered Image-Based Tomato Leaf Disease Detection. In *2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)* (pp. 506–511). IEEE.

Arsenovic M, Karanovic M, Sladojevic S, Anderla A, Stefanovic D (2019) Solving current limitations of deep learning based approaches for plant disease detection. Symmetry 11(7):939

Barbedo JGA (2013) Digital image processing techniques for detecting, quantifying and classifying plant diseases. Springerplus 2(1):660

Barbedo JGA (2017) A review of the main challenges in automatic plant disease identification based on visible range images. Biosyst Eng 144:52–60

Basavaiah J, Anthony AA (2020) Tomato leaf disease classification using multiple feature extraction techniques. Wireless Pers Commun 115(1):633–651

Bay H, Tuytelaars T, Van Gool L (2006) SURF: speeded up robust features. Springer, Berlin, Heidelberg, pp 404–417

Brahimi M, Boukhalfa K, Moussaoui A (2017) Deep learning for tomato diseases: classification and symptoms visualization. Appl Artif Intell 31(4):299–315

Chen T, Li M, Li Y, Lin M, Wang N, Wang M, Zhang Z (2015) Mxnet: A flexible and efficient machine learning library for heterogeneous distributed systems. *arXiv preprint* arXiv:1512.01274.

Chen X, Zhou G, Chen A, Yi J, Zhang W, Hu Y (2020) Identification of tomato leaf diseases based on combination of ABCK-BWTR and B-ARNet. Comput Electron Agric 178:105730

Chollet F (2015) keras.

Das D, Singh M, Mohanty SS, Chakravarty S (2020) Leaf Disease Detection using Support Vector Machine. In *2020 International Conference on Communication and Signal Processing (ICCSP)* (pp. 1036–1040). IEEE.

Durmuş H, Güneş EO, Kırcı M (2017) Disease detection on the leaves of the tomato plants by using deep learning. In *2017 6th International Conference on Agro-Geoinformatics* (pp. 1–5). IEEE.

Elhassouny A, Smarandache F (2019) Smart mobile application to recognize tomato leaf diseases using Convolutional Neural Networks. In *2019 International Conference of Computer Science and Renewable Energies (ICCSRE)* (pp. 1–4). IEEE.

Foysal MFA, Islam MS, Abujar S, Hossain SA (2020a) A novel approach for tomato diseases classification based. In proceedings of international joint conference on deep convolutional neural networks computational intelligence. Springer, Singapore

Foysal MFA, Islam MS, Abujar S, Hossain SA (2020) A novel approach for tomato diseases classification based on deep convolutional neural networks. In Proceedings of International Joint Conference on Computational Intelligence (pp. 583–591). Springer, Singapore.

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

Fuentes AF, Yoon S, Lee J, Park DS (2018) High-performance deep neural network-based tomato plant diseases and pests diagnosis system with refinement filter bank. Front Plant Sci 9:1162

Fuentes A, Yoon S, Park DS (2019) Deep learning-based phenotyping system with glocal description of plant anomalies and symptoms. Front Plant Sci 10:1321

Gadekallu TR, Rajput DS, Reddy MPK, Lakshmanna K, Bhattacharya S, Singh S, Alazab M (2020) A novel PCA–whale optimization-based deep neural network model for classification of tomato plant diseases using GPU. J Real-Time Image Process. 1–14.

Gebbers R, Adamchuk VI (2010) Precision agriculture and food security. Science 327(5967):828–831 (**PMID:20150492**)

Guo Y, Liu Y, Oerlemans A, Lao S, Wu S, Lew MS (2016) Deep learning for visual understanding: a review. Neurocomputing 187:27–48

Hassanien AE, Gaber T, Mokhtar U, Hefny H (2017) An improved moth flame optimization algorithm based on rough sets for tomato diseases detection. Comput Electron Agric 136:86–96

Hlaing CS, Zaw SMM (2017) Model-based statistical features for mobile phone image of tomato plant disease classification. In *2017 18th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT)* (pp. 223–229). IEEE.

Hlaing CS, Zaw SMM (2018) Tomato plant diseases classification using statistical texture feature and color feature. In *2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS)* (pp. 439–444). IEEE.

Hughes D, Salathé M (2015) An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv preprint* arXiv:1511.08060.

Jiang D, Li F, Yang Y, Yu S (2020) A tomato leaf diseases classification method based on deep learning. In 2020 chinese control and decision conference (CCDC) (pp. 1446–1450). IEEE.

Jinzhu LU, Di CUI, Jiang H (2013) Discrimination of tomato yellow leaf curl disease using hyperspectral imaging. In 2013 Kansas

City, Missouri, July 21-July 24, 2013. American Society of Agricultural and Biological Engineers, United States

Karthik R, Hariharan M, Anand S, Mathikshara P, Johnson A, Menaka R (2020) Attention embedded residual CNN for disease detection in tomato leaves. Appl Soft Comput 86:105933

Kaur M, Bhatia R (2019) Development Of An Improved Tomato Leaf Disease Detection And Classification Method. In 2019 IEEE Conference on Information and Communication Technology (pp. 1–5). IEEE.

Kaushik M, Prakash P, Ajay R, Veni S (2020) Tomato leaf disease detection using convolutional neural network with data augmentation. In 2020 5th International Conference on Communication and Electronics Systems (ICCES) (pp. 1125–1132). IEEE.

Krizhevsky A, Sutskever I, Geoffrey E. Hinton (2012) Imagenet classification with deep convolutional neural networks.In Proceedings of the 25th International Conference on Neural Information Processing Systems NIPS'12. 1: 1097–1105, USA. Curran Associates Inc.

Kumar A, Vani M (2019) Image Based Tomato Leaf Disease Detection. In 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1–6). IEEE.

LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324

Liu J, Wang X (2020a) Tomato diseases and pests detection based on improved Yolo V3 convolutional neural network. Front Plant Sci 11:898

Liu J, Wang X (2020b) Early recognition of tomato gray leaf spot disease based on MobileNetv2-YOLOv3 model. Plant Methods 16:1–16

Lu J, Cui D, Jiang H (2013) Discrimination of tomato yellow leaf curl disease using hyperspectral imaging. American Society of Agricultural and Biological Engineers, Kansas City, Missouri

Lu J, Zhou M, Gao Y, Jiang H (2018) Using hyperspectral imaging to discriminate yellow leaf curl disease in tomato leaves. Precision Agric 19(3):379–394

Ma J, Li X, Wen H (2015a) A keyframe extraction method for processing greenhouse vegetables production monitoring video. Comput Electron Agric 111:92–102

Ma J, Li X, Zhang L (2015b) Monitoring video capture system for identification of greenhouse vegetable diseases. Trans Chin Soc Agric Mach 46(3):282–287

Ma J, Li X, Wen H, Chen YY, Fu ZT, Zhang L (2015c) Monitoring video capture system for identification of greenhouse vegetable diseases. Trans Chin Soc Agric Mach 46(3):282–287

Ma J, Du K, Zhang L, Zheng F, Chu J, Sun Z (2017) A segmentation method for greenhouse vegetable foliar disease spots images using color information and region growing. Comput Electron Agric 142:110–117

Mishra RK, Jaiswal RK, Kumar D, Saabale PR, Singh A (2014) Management of major diseases and insect pests of onion and garlic: A comprehensive review. J Plant Breed Crop Sci 6(11):160–170

Mokhtar U, Ali MA, Hassenian AE, Hefny H (2015) Tomato leaves diseases detection approach based on support vector machines. In *2015 11th International Computer Engineering Conference (ICENCO)* (pp. 246–250). IEEE.

Mokhtar U, Ali MA, Hassanien AE, Hefny H (2015b) Identifying two of tomatoes leaf viruses using support vector machine. In information systems design and intelligent applications. Springer, New Delhi

Mokhtar U, El Bendary N, Hassenian AE, Emary E, Mahmoud MA, Hefny H, Tolba MF (2015c) SVM-based detection of tomato leaves diseases. intelligent systems. Springer International Publishing, Berlin, pp 641–652

Muthukannan K, Latha P (2015) Fuzzy inference system based unhealthy region classification in plant leaf image. Int J Comput Info Eng 8(11):2103–2107

Nandhini S, Ashokkumar K (2021) Improved crossover based monarch butterfly optimization for tomato leaf disease classification using convolutional neural network. *Multimedia Tools and Applications*. 1–28.

Oerke EC (2006) Crop losses to pests. J Agric Sci 144:31

Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, Lerer A (2017) Automatic differentiation in pytorch.

Prabhakar M, Purushothaman R, Awasthi DP (2020) Deep learning based assessment of disease severity for early blight in tomato crop. Multimed Tools Appl 79(39):28773–28784

Rajasekaran T, Anandamurugan S (2019) Challenges and applications of wireless sensor networks in smart farming—a survey In Advances in big data and cloud computing. Springer, Singapore, pp 353–361

Rangarajan AK, Purushothaman R, Ramesh A (2018) Tomato crop disease classification using pre-trained deep learning algorithm. Procedia Computer Science 133:1040–1047

Ren S, He K, Girshick R, Sun J (2017) Faster R-CNN: towards realtime object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 39(6):1137–1149

Ribeiro E, Uhl A, Wimmer G, Ha¨fner M (2016) Exploring deep learning and transfer learning for colonic polyp classification. Comput Math Methods Med. 1–16

Riley MB, Williamson MR, Maloy O (2002) Plant disease diagnosis. Plant Health Instr

Rubanga DP, Loyani LK, Richard M, Shimada S (2020) A Deep Learning Approach for Determining Effects of Tuta Absoluta in Tomato Plants. *arXiv preprint* arXiv:2004.04023.

Sabrol H, Kumar S (2016a) Fuzzy and neural network based tomato plant disease classification using natural outdoor images. Indian J Sci Technol 9(44):1–8

Sabrol H, Kumar S (2016b) Intensity based feature extraction for tomato plant disease recognition by classification using decision tree. Int J Comput Sci Inf Secur 14(9):622–626

Sabrol H, Satish K (2016) Tomato plant disease classification in digital images using classification tree. In *2016 International Conference on Communication and Signal Processing (ICCSP)* (pp. 1242–1246). IEEE.

Sainath TN, Kingsbury B, Saon G et al (2015) Deep convolutional neural networks for large-scale speech tasks. Neural Netw 64:39–48

Sardogan M, Tuncer A, Ozen Y (2018) Plant leaf disease detection and classification based on CNN with LVQ algorithm. In *2018 3rd International Conference on Computer Science and Engineering (UBMK)* (pp. 382–385). IEEE.

Shalaby MY, Al-Zahrani KH, Baig MB, Straquadine GS, Aldosari F (2011) Threats and challenges to sustainable agriculture and rural development in Egypt: implications for agricultural extension. J Anim Plant Sci 21(3):581–588

Shelhamer E, Long J, Darrell T (2017) Fully convolutional networks for semantic segmentation. IEEE Trans Pattern Anal Mach Intell 39(4):640–651

Shijie J, Peiyi J, Siping H (2017) Automatic detection of tomato diseases and pests based on leaf images. In 2017 Chinese Automation Congress (CAC) (pp. 2537–2510). IEEE.

Singh D, Jain N, Jain P, Kayal P, Kumawat S, Batra, N (2020) PlantDoc: a dataset for visual plant disease detection. In *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD* (pp. 249–253).

Slavin P (2016) Climate and famines: a historical reassessment. Wiley Interdiscip Rev Clim Change 7(3):433–447. https://doi.org/10.1002/wcc.395

Sonnet, https://sonnet.dev/. Accessed Sept. 11, 2019.

Strange RN, Scott PR (2005) Plant disease: a threat to global food security. Annu Rev Phytopathol 43(1):83–116

Suryawati E, Sustika R, Yuwana RS, Subekti A, Pardede HF (2018) Deep structured convolutional neural network for tomato diseases detection. In *2018 International Conference on Advanced Computer Science and Information Systems (ICACSIS)* (pp. 385–390). IEEE.

Thangaraj R, Anandamurugan S, Kaliappan VK (2020) Automated tomato leaf disease classification using transfer learning-based deep convolution neural network. J Plant Dis Prot. 1–14.

Tian K, Li J, Zeng J, Evans A, Zhang L (2019) Segmentation of tomato leaf images based on adaptive clustering number of K-means algorithm. Comput Electron Agric 165:104962

Tm P, Pranathi A, SaiAshritha K, Chittaragi NB, Koolagudi SG (2018) Tomato leaf disease detection using convolutional neural networks. In *2018 Eleventh International Conference on Contemporary Computing (IC3)* (pp. 1–5). IEEE.

Vedaldi A, Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Darrell T (2014) Convolutional architecture for fast feature embedding. *Cornell University, arXiv: 1408.5093 v12014*.

Wang X, Liu J (2021) Tomato anomalies detection in greenhouse scenarios based on YOLO-Dense. Front Plant Sci 12:533

Wang YY, Li ZM, Wang L, Wang M (2013) A scale invariant feature transform based method. J Inf Hiding Multimed Signal Process 4(2):73–89

Xie C, Shao Y, Li X, He Y (2015) Detection of early blight and late blight diseases on tomato leaves using hyperspectral imaging. Sci Rep 5:16564

Xu Y, Yu G, Wang Y, Wu X, Ma Y (2017) Car detection from low altitude UAV imagery with the faster R-CNN. J Adv Transport 11:1–11

Zhang K, Wu Q, Liu A, Meng X (2018) Can deep learning identify tomato leaf disease?. *Advances in Multimedia*.

Zhao L, Jia K (2016) Multiscale CNN's for brain tumor segmentation and diagnosis. Comput Math Methods Med 2016:1–8