INNSPUB

Journal of Biodiversity and Environmental Sciences (JBES)
ISSN: 2220-6663 (Print) 2222-3045 (Online)
Vol. 22, No. 4, p. 42-48, 2023
http://www.innspub.net

RESEARCH PAPER | OPEN ACCESS

Hybrid ResNet50-PCA based deep transfer learning approach for classification of tomato leaf diseases

Rubul Kumar Bania*, **Sumit Dey**, **Nilutpal Buragohain**

Department of Computer Application, North-Eastern Hill University, Tura Campus, Meghalaya, India

Article published on April 17, 2023

Key words: Tomato, Leaf diseases, Deep learning, PCA, Detection

Abstract

Tomato is one of the world's most indispensable and consumable vegetable items. In the Indian market, it has high commercial value, and it is produced in huge quantities. The crop sensitivity and climatic conditions have made diseases familiar in the tomato crop during all the stages of its growth. It is a difficult task to monitor plant diseases manually due to its complex nature and time-consuming process. Artificial intelligence (AI) based computational models can detect leaf diseases in their early stages. In this article, ResNet50 a deep transfer learning based Convolutional Neural Network (CNN) amalgamated with Principal component analysis (PCA) to classify tomato leaf diseases effectively. Subset of publicly available 'Plant Village' dataset is used in this study. The architecture has attained the highest accuracy of 98.18% for identifying tomato leaf diseases. The experimental results show that the computational model effectively identifies tomato leaf disease and could be generalized to other plant diseases.

***Corresponding Author:** Rubul Kumar Bania ✉ rubul.bania@nehu.ac.in

Introduction

Tomatoes are one of the most nutrient-dense crops on the planet, and their cultivation and production significantly impact the agricultural economy. Worldwide, approximately 160 million tons of tomatoes are consumed annually (Schreinemachers *et al.*, 2019). Tomato is a significant contributor to reducing poverty and can be seen as an income source for farm households. The tomato is nutrient-dense and possesses pharmacological properties that protect against diseases such as hypertension, hepatitis, and gingival bleeding (Stilwell, 2023). Tomato crop cultivation area in India spans around 350,000 approximately hectares, and the production quantities roughly sum up to 5,300 000 tons, making India the third largest tomato producer in the world. According to statistics, small farmers usually produce more than 80% of agricultural output, but about 40-50% of their crops are lost due to diseases and pests. In the Indian market, along with other vegetables, the tomato has high commercial value, and it is produced in huge quantities. The crop sensitivity and climatic conditions have made diseases familiar in the tomato crop during all the stages of its growth. Disease affected plants constitute 10-30% of the total crop loss (Basavaiah *et al.*, 2020). Identification of such conditions in the plant is essential in preventing any heavy losses in yield and the quantity of the agricultural product. It is difficult to manually monitor plant diseases due to their complex nature and time-consuming process. Therefore, there is a need to reduce the manual effort put into this task while making accurate predictions and ensuring that the farmers' lives are hassle-free (Basavaiah *et al.*, 2020, Hasan *et al.,* 2019). Various researchers have used cutting-edge technology, such as computer vision with image processing technologies, machine learning, and deep neural network architectures, to construct automated disease detection systems (Adhikari *et al.,* 2018).

Plant diseases pose a significant hazard to the global food source. The agricultural industry requires cutting-edge expertise for disease control, which is currently under process. The motivation of this study is to support farmers in identifying the tomato leaf diseases in early stages and informing them about these diseases. AI and computer vision-based techniques are advantageous for detecting and analysing plant infections. For farmers and ranchers, physically inspecting plants and detecting diseases is very laborious and tiring. There is a chance of error during the process of disease detection. Consequently, the use of these techniques is very advantageous because they are not particularly taxing, they do not require a great deal of labor, and they reduce the likelihood of misleading information. Nine different tomato leaf diseases such as: 'Bacterial spot', 'Early Blight', 'Late Blight', 'Leaf mold', 'Mosaic', 'Septoria spot','Spedier mites', 'Yellow curl', 'Septoria spot disease prediction', are considered in this study.

Many solutions have been proposed with the advent of technology (Adhikari *et al.,* 2018, Hasan *et al.,* 2019). But the existing models or methods consumed several amounts of time for producing results. Moreover, to train a computational model enough image features are required. This paper provides a solution having less computational cost to recognize tomato leaf diseases. This work is mainly focused on investigating an optimal solution for tomato plant leaf disease detection. In this paper, we present a hybrid technique to classify crop leaf diseases by extracting the image features by ResNet50 deep transfer learning model and then reducing the feature dimension by employing PCA. The model outperformed and achieved 98.18% classification accuracy. The model was trained and tested on independent data with different disease classes. The organization for the rest of the paper is as follows. Section 2 reviews previous study related to plant leaf detection then illustrates the proposed approach. Section 3 reports the experimental results and analyses, and finally Section 4 presents the conclusions and future work of this research work.

Materials and methods

Related Study

Traditional machine learning algorithms or models are highly applied in various fields, but feature engineering remains the main problem. With the emergence of deep neural networks, promising results are available for plant pathology without laborious feature engineering. Deep neural networks significantly increase image classification accuracy. This section provides various deep-learning techniques used by researchers in plant disease identification.

Hasan and Tanawala *(*Hasan *et al.,* 2019) have used a dataset contains 2100 images of tomato leaves from internet and 500 images from local farms. Inception v3 Transfer learning model was for identifying diseases in tomato leaves. Leaves are classified in 3 broad categories as good, average, and bad for pesticide intensity. In another work, utilized the CNN-based classifiers and tested on a subset of images related to tomato plant leaf diseases. The dataset consists of 3 leaf diseases of the tomato plant, including Gray spot (113 samples), Late Blight (121 pieces), and Bacterial Canker (111 samples). Healthy tomato leaf images are also used and added to the used dataset (Adhikari *et al.,* 2018*)*.

Authors have used the Decision tree classification model and segmentation techniques are used to detect and classify six different types of tomato leaf disease with a dataset of 300 images (*Sabrol and Satish*, 2016). Another machine learning based technique has been proposed to detect and classify plant leaf disease with an accuracy of 93.75% (Salih *et al.,* 2020*)*. The image processing technology and classification algorithm detect and classify plant leaf disease with better quality (Wu *et al.,* 2021).

A simple CNN model with eight hidden layers has been used to identify the conditions of a tomato plant. The proposed techniques show optimal results compared to other classical models (Suryanarayana *et al.,* 2021). Authors have applied CNN model and ResNet 50 transfer learning model to the tomato leaf dataset. The CNN shows the better accuracy than the ResNet50 model (Sharma *et al.,* 2022). Alex Net, ANN and CNN models are used to identify tomato leaf diseases and got an accuracy of 95.75%, 92.94% and 98.12% accuracy respectively (Agarwala *et al.,* 2019).

Methodology

In this section, the pipeline of the methodology followed in this research is elaborated. It is basically divided into three phases; in the first phase data collection and data pre-processing is done; in the second phase by applying one deep ResNet50 model various features are extracted from the input images to train the model but as the number of features are very high therefore to reduce the dimension PCA is used; finally, in the third phase with the help of fully connected layer classification task is performed. The block diagram with architecture for the classification of tomato leaf diseases using the deep transfer learning-based model is shown Fig. 1.

The accuracy of a deep CNN model always dependent on the quality of the dataset. Therefore, after collecting the data, data cleaning process is performed to eliminate some of the faulty images discovered in the dataset. The images are resized into a fixed size of 224 × 224, which helps to reduce the load on the machine while training and to provide optimum results. The images are properly labelled with a one-hot encoding system. The array of images is then transformed into a NumPy array for quicker computation.

The ResNet (residual neural) was proposed in 2015 by researchers at Microsoft Research. The architecture of the model is shown in Fig. 2. ResNet has many variants that run on the concept of CNN but have different numbers of layers. Resnet50 denotes the variant that can work with 50 neural network layers (Weiss, 2016). A common problem in deep learning is associated with that called vanishing/exploding gradient. This causes the gradient to become 0 or too large. Thus, the training and test error rate increases when we increase the number of layers. After analyzing more on error rate, the authors concluded that it is caused by vanishing/exploding gradient. To solve the problem of the vanishing/exploding gradient Residual network is used. In this network skip connections are used. This skip connection basically skips training from a few layers and connects directly to the output. The approach behind this network is instead of layers learn the underlying mapping; it allows the network to fit the residual mapping. So, instead of say $H(x)$, initial mapping, the network will fit, $F(x) := H(x) - x$ which gives $H(x) := F(x) + x$.

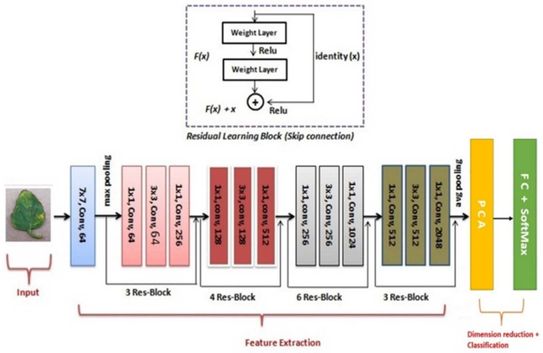


Fig. 1. Architecture of ResNet-50 model with PCA.

The benefit of features extracted using the ResNet50 deep learning algorithms is that the network learns plant image features automatically layer-by-layer (Bania, 2023). Generally, the last layer of any deep learning network such as ResNet50 produces an output class predication using softmax classification. The high-level features are extracted before the FC layers. As the dimension of the extracted features are very high so to reduce an transformed the features PCA is applied.

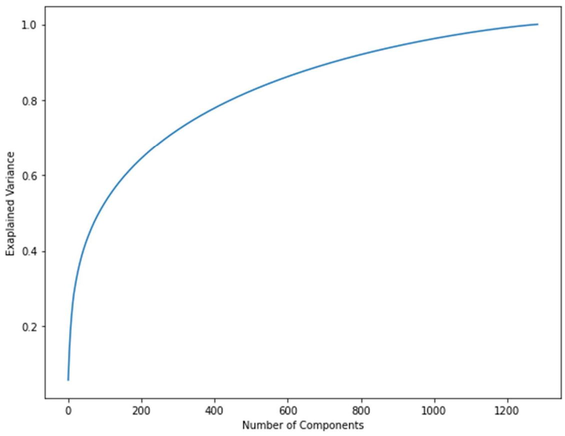


Fig. 2. PCA analysis.

PCA uses orthogonal transformations to identify the correlated variables and convert them into uncorrelated ones. For efficient classification, the dimension of the fused feature (FF) vectors is reduced by employing PCA. We can observe from Fig. 2 that the first 1200 components can capture above 90% of the variability in the original image data. That is enough to retain the quality of the original images. So, we selected the first 1200 components and apply PCA with that number of selected components.

Empirical study

After discussing the methodology, this section will describe this study's experimental details and findings.

Experimental Setup

Various models were developed to make a comparative analysis along with the proposed model. All the methods are implemented in Python 3.9. Programs are simulated in a machine with Processor: Xenon(R) CPU- E5-1630, 3.70 GHz clock speed and random-access memory of 32 GB having Windows 10 environment. The deep learning models are implemented using Tensorflow, Keras functional API, which provides a flexible way to design neural networks with non-linear topology and shared layers.

Evaluation metrics

The experimental results are validated by applying various measures (Bania, 2023) such as (i) accuracy, (ii) precision, (iii) recall and (iv) F-score measure.

(i) *Accuracy:* Overall effectiveness of the classifier can be calculated as follows.

$$\frac{TP + TN}{TP + FN + FP + TN}$$

Here, TP = True Positive, TN = True Negative, FN= False Negative and FP = False Positive.

(ii) *Precision:* It calculates the r*atio of the correctly classified positive image data samples to a total number of classified positive image samples.*

$$\frac{TP}{TP + FP}$$

(iii) *Recall:* It calculates the ratio between the numbers of positive *image data samples* correctly classified as Positive to the total number of *positive image samples.*

$$\frac{TP}{TP + FN}$$

(iv) *F-score:* By combining the Precision and Recall calculations F-score can be calculated. Basically, Harmonic mean of the two fractions is computed.

$$\frac{2*(Precision*Recall)}{(Precision+Recall)}$$

Dataset in used

Images of tomato disease have been taken from Plant Village dataset. 2219 numbers of images are randomly selected from that dataset. Dataset was divided into train-test set (80%-20%). There were ten unique classes of disease in the sample.

Tomato leaves of nine types were infected, and one class was resistant or healthy. Reference photos and disease names are used to identify images from the dataset and those are shown in Fig. 3.

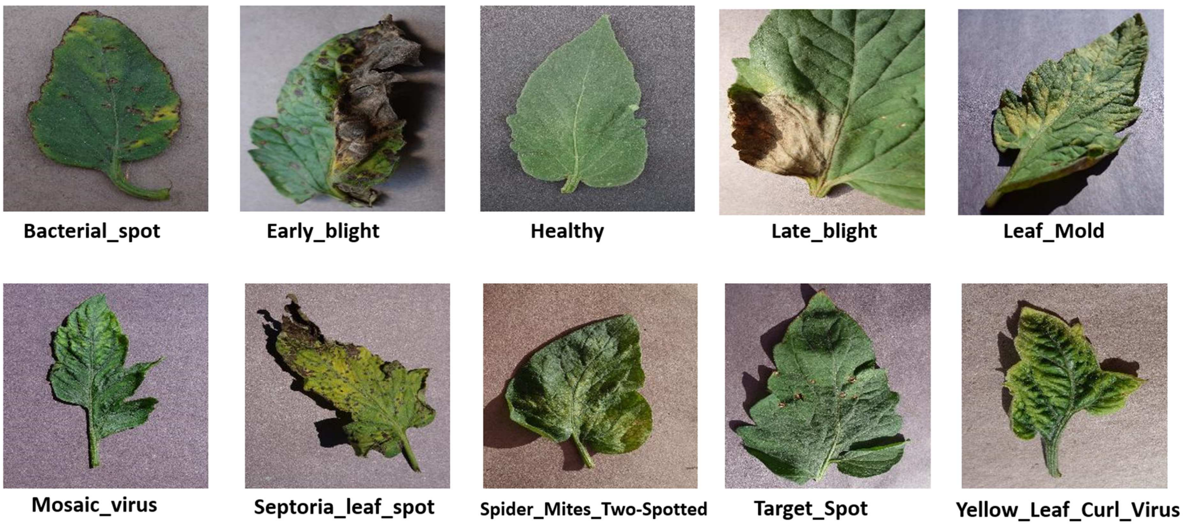


Fig. 3. Ten different samples of the dataset.

Results and discussions

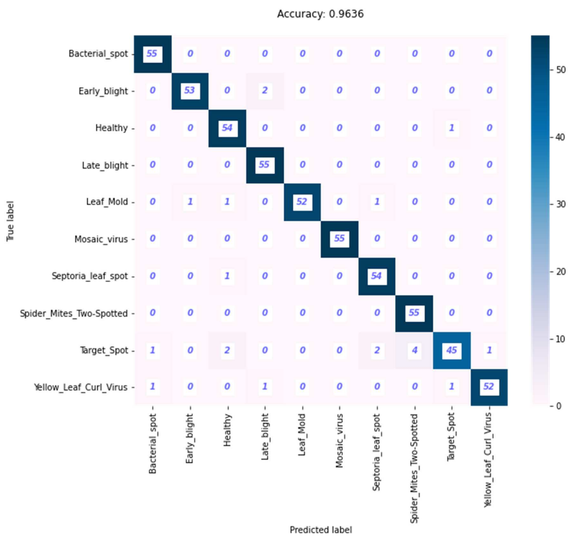
In the literature, several existing methods are available so it is beyond the scope of the paper to compare the proposed model with several existing methods. Here in this work, by considering the popularity and latest state-of-the-art work, proposed model is compared with other five models namely model proposed by Sharma (Sharma *et al.*, 2022), model proposed by Agarwal (Agarwal *et al.*, 2019) VGG-16, ResNet50 and Inception50 models.

Also, model is compared with by hybridizing VGG16-PCA and Inception50-PCA. The summary of the classification experimental results achieved by the different models in terms four validity measures namely percentage accuracy, precision, recall, and F1-score are reported in Table 1. It can be seen in the table that transfer learning-based models such as VGG16, ResNet50 and Inception50 have attained an accuracy below 85%. But with the amalgamation of PCA all the transfer learning-based models have achieved higher accuracy.

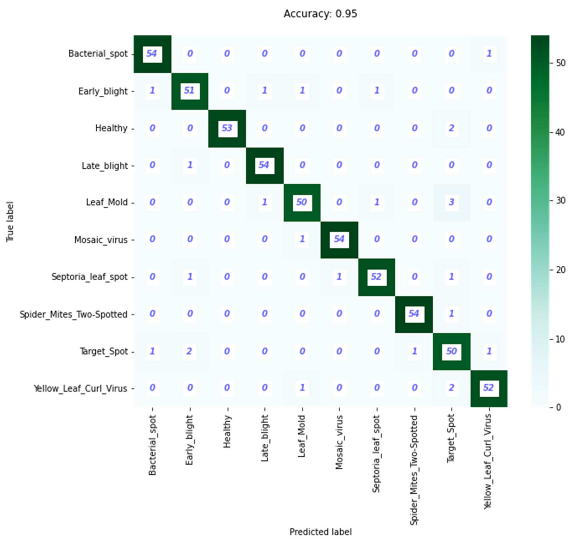
The proposed architecture i.e., ResNet50-PCA has abled to achieved highest accuracy. In Fig. 4 confusion matrices achieved by VGG16-PCA, Inception50-PCA and ResNet50-PCA are shown. In Fig. 4 (c) among the 768 testing image samples (20% of dataset), 10 images were misclassified by the proposed architecture (ResNet50-PCA).

Table 1. Summary of experimental results achieved by various models.

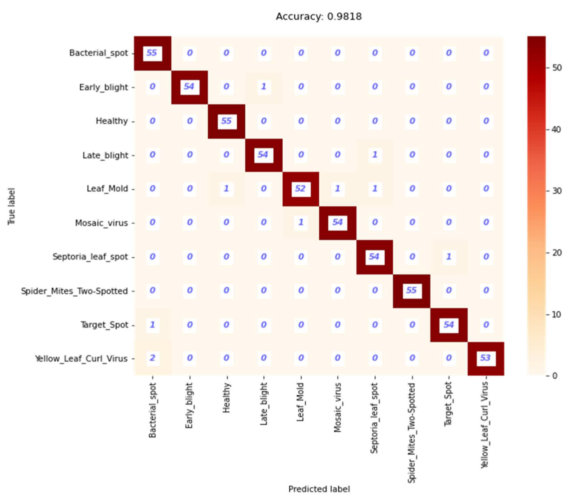
Methods	Precision	Recall	F-score	Accuracy
Sharma *et al.*, 2022	0.918	0.912	0.915	92.10%
Agarwal *et al.*, 2019	0.901	0.905	0.910	91.20%
VGG16	0.829	0.829	0.826	83.23%
ResNet50	0.851	0.851	0.849	85.02%
Inception50	0.844	0.843	0.844	84.66%
VGG16-PCA	0.960	0.960	0.960	96.36%
InceptionV3-PCA	0.951	0.950	0.951	95.22%
Proposed Model (ResNet50-PCA)	0.981	0.980	0.981	98.18%



(a)



(b)



(c)

Fig. 4. Confusion matrix generated by various models.

From the experimental result highlights of the article's findings, which are comparable to the current state of knowledge in their respective fields of study. In addition, when the proposed ResNet50-PCA model was trained with a greater number of parameters, overall performance significantly improved. The use of PCA has decreased the model training time by reducing the feature dimensions and it improves the performance. This study mainly concentrates on the 'Plant Village' dataset which consists of images captured in a variety of environments; however, it was collected in a specific location and contained images of specific tomato leaf disease varieties. Our trained model may allow for the automated and early detection of tomato plant diseases. Professionals require years of training and experience to diagnose an illness through a visual examination. Still, anyone can utilize our methodology, regardless of their experience or expertise. If there are any new users, the network will operate in the background, receiving input from the user and immediately notifying them of the result so they can take the appropriate action. As a result, anticipatory procedures may be taken sooner rather than later.

Conclusion and future directions

It is a difficult task to monitor plant diseases manually due to its complex nature and time-consuming process. In this article, we have attempted to propose a deep neural network model by amalgamating ResNet50 with PCA for detecting and classifying tomato plant leaf diseases into predefined categories. The numbers of features extracted by the ResNet50 model is very high so to reduce the computational complexity with help of PCA, 1200 components are selected. One subset of publicly available 'Plant Village' dataset with 9 disease classes is used in this study. The proposed architecture has attained the highest accuracy of 98.18% for identifying tomato leaf diseases. In future, we will expand the model to include certain abiotic diseases due to the deficiency of nutrient values in the crop leaf. Also, our long-term objective is to apply the proposed model in intelligent drone cameras, advanced mobile phones, or may be in robotics. Also, we are working towards the unique data collection and accumulate a vast amount of data on several diseases of plants.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-from any profit sectors.

Conflict of interests

None declared

References

Adhikari S, Shrestha B, Baiju B, Kumar S. 2018. Tomato plant diseases detection system using image processing. In Proceedings of the 1st KEC Conference on Engineering and Technology **1(2),** 81-86, Laliitpur, Nepal, 27 September.

Agarwal M, Singh A, Arjaria S, Sinha A, Gupta S. 2019 ToLeD: Tomato Leaf Disease Detection using Convolution Neural Network, In the proceedings of International Conference on Computational Intelligence and Data Science (ICCIDS 2019), Procedia Computer Science **167(2020),** 293-301.

Bania KR. 2023. Ensemble of deep transfer learning models for real-time automatic detection of face mask, Multimedia Tools and Applications 1-23, https://doi.org/10.1007/s11042-023-14408-y.

Basavaiah J, Anthony AA. 2020. Tomato Leaf Disease Classification using Multiple Feature Extraction Techniques. Wirel. Pers. Commun **115(3),** 633-651.

Hasan M, Tanawala B, Patel KJ. 2019. Deep learning precision farming: Tomato leaf disease detection by transfer learning. In Proceeding of the 2nd International Conference on Advanced Computing and Software Engineering (ICACSE), Sultanpur, India, 8-9 February.

Sabrol H, Satish K. 2016. Tomato plant disease classification in digital images using classification tree. In Proceedings ofthe International Conference on Communication and Signal Processing (ICCSP), 1242-1246, Melmaruvathur, India, 6-8 April.

Salih TA, Ali A, Ahmed M. 2020. Deep Learning Convolution Neural Network to Detect and Classify Tomato Plant Leaf Diseases. Open Access Library Journal **7(12),** 1-12.

Schreinemachers P, Simmons EB, Wopereis MC. 2018. Tapping the economic and nutritional power of vegetables. Glob. Food Secur **16(1),** 36-45.

Sharma R, Panigrahi A, Garanayak M. 2022. Tomato Leaf Disease Detection Using Machine Learning, ACI'22: Workshop on Advances in Computation Intelligence, its Concepts & Applications at ISIC 2022, May 17-19, Savannah, United States 294-299.
https:// www.kaggle.com /kaustubhb999/tomatoleaf.

Stilwell M. 2023 The Global Tomato Online News Processing in 2018. Available online:
https:// www. tomatonews.com.

Suryanarayana G, Chandran K, Khalaf OI, Alotaibi Y, Alsufyani A, Alghamdi SA. 2021. Accurate Magnetic Resonance Image Super-Resolution Using Deep Networks and Gaussian Filtering in the Stationary Wavelet Domain. IEEE Access **9,** 71406-71417.

Weiss K. 2016. A survey of transfer learning, Journal of Big data **3(9),** 1-40.

Wu Y, Xu L, Goodman ED. 2021. Tomato Leaf Disease Identification and Detection Based on Deep Convolutional Neural Network. Intelligent. Automatic Soft Computing **28,** 561-576.