



Machine Learning Based Plant Disease Detection Using EfficientNet B7

Amit Kumar Bairwa , Sandeep Joshi ^(✉) , and Shikha Chaudhary 

Manipal University Jaipur, Jaipur, India
sjoshinew@yahoo.com

Abstract. Plant diseases have effects on the growth and production of the plant. Plant diseases can be figured out by using digital image processing, and nowadays, Deep learning has made a lot of progress in digital image processing to identify the disease efficiently. This paper finds the plant diseases using EfficientNet by focusing on three data steps: pre-processing, model selection, and detection network, using a canny edge detection algorithm. The model is trained and tested on plant disease data set. The model provides 97.2% accuracy in detecting the disease than existing CNN-based models.

Keywords: Plant diseases · Deep learning · Machine Learning · Classification · Segmentation · Detection

1 Introduction

Planting is one of the important ways for farmers to make money. Farmers can plant many kinds of crops, but diseases can make it hard for plants to grow. Plant diseases are a big reason agricultural products aren't as helpful or easy to sell as they could be. With the help of new technology, we can make enough food to feed more than 7 billion people. The safety of plants is still a risk because of things like climate change, fewer pollinators, plant diseases, etc. Farmers work hard to choose the best crop for production, but many diseases affect [18]. In agriculture, it's essential to figure out the issues with plants as soon as possible. This lets the reduction in damage to crops, lower production costs, and good quality of crops, which leads to profit for the farmers [11]. From the existing data, it has been analyzed that the disease affects and reduces crop production from 10% to 95%. There are several ways to get rid of plant diseases right now, such as removing infected plants by hand, planting them mechanically, or using pesticides [4].

A simple way to figure out what's wrong with a plant is to ask an expert in agriculture. But figuring out an infection manually takes a long time and is hard to implement. Pesticides can be used as a precaution and recovery from such diseases, but too much use can hurt crop yields, the environment, and people's health. Before using such things, the exact quantity must be calculated and used within a specific time window, and limit [16].

Plants and infectious diseases can be separated using digital imaging and machine learning algorithms for Timely Detection and recovery before spreading

the disease. Automatically diagnosing plant diseases is essential because it might be helpful for farmers to figure out and watch over large fields using state-of-the-art techniques using image processing and deep learning. Plants make up more than 80% of what people eat, and in many countries like India, the economy is majorly based upon farming. So, it is essential to find out the cause, detect the disease timely for instant recovery, and ensure that everyone has cheap enough, clean, healthy food to live a long, healthy life [7].

Traditional methods are less effective, necessitating automatic, quick, accurate, and cost-effective ways to identify plant diseases. Numerous agricultural applications utilize digital cameras to capture images of leaves, flowers, and fruits for disease identification. Through image processing and analytical techniques, valuable information is extracted for analysis. Precise farming data aids farmers in making optimal decisions for high agricultural productivity. This research explores various diagnostic approaches for plant diseases, leveraging image analysis and machine intelligence to achieve easy, automatic, and accurate identification of leaf diseases [5]. Early signs of disease detection, such as changes in leaf color and patches, enhance crop yields through automated disease identification.

Machine learning studies algorithms that change and improve on their own as more data is collected and used. Machine learning techniques are used to train the model on collected data (called “training data”) so that it can make predictions or decisions on its own without being told what to do [?]. Machine learning algorithms can be used in many fields, such as medicine, email filtering, speech recognition, and computer vision, where it is hard to make a single algorithm that can do all the tasks that need to be done. Machine learning algorithms can be categorized as supervised learning, unsupervised learning, and reinforcement learning [?]. Supervised learning is the process of using labeled data sets to teach models or algorithms predict outcomes. Classification is a well-known model in which the data can predict the label. It expects different kinds of responses, like whether an email is spam or a tumor is cancerous [1]. Support Vector Machine (SVM), Naïve Bayes, Random Forest, K-NN, discriminant analysis, etc. are some of the well-known methods of classification [12]. Regression techniques are trained using labels for what goes in and what comes out. Measures how one variable affects another to determine how they are connected using continuous values [?]. Some examples are predicting how much electricity will be used and trading based on algorithms. Linear regression, Ridge, LASSO, decision trees, neural network regression, KNN, and SVM are some of the well-known algorithms of regression algorithms [9].

2 Literature Review

According to a suggestion by M.P. Vaishnave [10], one of the essential elements that bestow reduced yield is disease assault. The groundnut plant is susceptible to diseases caused by a fungus, viruses, and soil-borne organisms. The software determination to robotically classify and categorize groundnut leaf illnesses is shown to us in this paper. The output of the crops will increase as a result of using this strategy. Image capture, image pre-processing, segmentation, feature

extraction, and classifier with K Nearest Neighbor are some of the processes it consists of (KNN). The KNN classification is used instead of the SVM classifier, which improves the present algorithm's performance.

Agriculture is a significant part of the Indian economy, which Debasish Das [?] mentioned in his cite-singh2019comparative article. The primary objective of this research is to determine the various illnesses that can affect leaf tissue. Several feature extraction strategies have been tried and tested to improve the accuracy of the categorization. Statistical methods such as Support Vector Machine (SVM), Random Forest, and Logistic Regression have been utilized to categorize the various leaf diseases. When the outputs of the three classifiers are compared, the support vector machine comes out on top. The findings demonstrate that the model applies to situations that occur in real life.

In this review, Shruti [2] reviewed the comparative study on five different types of machine learning classification algorithms for recognizing plant disease. Compared to other classifiers, the SVM classifier is frequently utilized by writers for disease classification. The findings indicate that the CNN classifier is superior in precisely identifying a more significant number of diseases.

Training a convolutional neural network as a method for disease detection in plants was offered by Prasanna Mohanty [8] in a recent study. The CNN model has been trained to distinguish healthy plants from ill plants across 14 species. The model's accuracy was determined to be 99.35% based on the test set data. When applied to pictures obtained from reliable internet sources, the model achieves an accuracy of 31.4%. While this is superior to the accuracy achieved by a straightforward model based on random selection, the accuracy might be improved by utilizing a more varied training data collection.

Sharada P. Mohanty [2] described crop diseases and the methods for quickly identifying them. The rising prevalence of smartphone users globally and recent developments in computer vision made feasible by deep learning have paved the path for disease diagnosis that may be performed with a smartphone. We train a deep convolutional neural network to recognize 14 crop species and 26 diseases using a public data-set of 54,306 photos of damaged and healthy plant leaves taken under controlled settings (or absence thereof). This strategy is viable when the trained model achieves an accuracy of 99.35% on a held-out test set. Table 1 shows the accuracy comparison among all the reviewed techniques discussed in the section using various classification techniques.

3 Statement of the Problem

Plant disease detection using machine learning approach, applying various techniques/ algorithms, analyzing their efficiency, comparing them, and defining the best out of them.

4 Methodology

We have developed a plant disease detection model using EfficientNets [6], and it has been observed to perform well on the plant disease dataset [3]. The methodology of the model is described here in the subsections below.

Table 1. Accuracy Comparison

Classification Technique	Culture	No. of diseases	Accuracy
SVM	Citrus	2	95%
	Grape	2	88.9%
	Oil Palm	2	97%
	Potato	2	90%
	Tea	3	93%
	Soyabean	3	90%
KNN	Sugarcane	1	95%
	Cotton	1	88.5%
CNN	Peach, Cherry, Pear, Apple and Grapevine	13	96.3%
	14 crops	12	99.35%
	Soyabean	3	99.32%
	25 Plants	58	99.53%

KNN. K-means algorithm is an iterative algorithm that tries to partition the dataset into pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group. It is a centroid-based algorithm or a distance-based algorithm. In K-Means, each cluster is associated with a centroid.

SVM. Support Vector Machine (SVM) is a supervised machine learning algorithm for classification and regression. Though we say regression problems as well, it's best suited for classification. The objective of the SVM algorithm is to find a hyperplane in an N-dimensional space that distinctly classifies the data points.

4.1 Implemented Algorithms

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm that can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image, and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower than in other classification algorithms. While in primitive methods, filters are hand-engineered, with enough training, ConvNets can learn these filters/characteristics.

4.2 Data Pre-processing

Exploratory Data Analysis (EDA) is a process of performing initial investigations on data to discover patterns, spot anomalies, test hypotheses, and check assumptions with the assistance of summary statistics and graphical representations. This is a crucial step in the pre-processing phase. The leaf image is distributed in R-G-B channels separately, the green part or healthy area has shallow blue values, but by contrast, the brown parts have high blue values. This might suggest that the blue channel may be the key to detecting plant diseases.

The red channel values seem to roughly normal distribution, but with a slight rightward (positive skew). This indicates that the red channel tends to be more concentrated at lower values, at around 100 as in figure. There is large variation in average red values across images. The green channel of contains the high contrast region of image due to this Microaneurysms are clearly visible in figure. The blue channel has the most uniform distribution out of the three colour channels, with minimal skew (slight leftward skew). The blue channel shows great variation across images in the dataset as in figure.

4.3 Model Selection

The performance of the base network is heavily dependent on the model's measurements. Therefore, to further enhance performance, we are also developing a new primary network by doing neural architecture searches using the AutoML MNAS framework, which improves accuracy and efficiency. These searches are being done to improve performance (FLOPS) further. The produced structures use mobile inverted bottleneck convolution (MBCnv), which is com-Machine, learning-based plant diseases detection, using EfficientNet B7 parable to MobileNetV2 and MnasNet. However, these structures are significantly more significant due to an enhanced FLOP budget. After that, we broaden the scope of the initial network to discover a family of models known as EfficientNets. EfficientNets was tested on eight different databases, too, for transfer learning. EfficientNets attained modern accuracy in five out of eight data sets, such as CIFAR-100 (91.7%) and Flowers (98.8%), with only a few parameter settings (up to 21 times in the parameter reduction), which suggests that EfficientNets also transmits well. We anticipate that EfficientNets will provide considerable gains in model performance and, as a result, will serve as a new foundation for future operations involving computer vision.

The performance of our EfficientNets is compared to that of other CNNs hosted on ImageNet. EfficientNet models can attain higher accuracy and better performance than currently available CNNs. This is accomplished by reducing the parameter size and FLOPS by size system. For instance, in a high-precision system, our EfficientNet-B7 achieves 84.4% with a maximum accuracy of 1/97.1% a maximum of 5 in ImageNet, while at the same time being 8.4 times more compact and 6.1 times in CPU precision than Gpipe was previously. Our EfficientNet-B4 employs the same number of FLOPS as the widely used ResNet-50, but it improves the maximum accuracy of 1 from 76.3% of ResNet-50 to 82.6% (+6.3%).

4.4 Image Processing Using Canny Edge Detection Algorithm

Canny edge detection algorithm is used to detect edges of a leaf and the detected region/ edges will be used to fed in the proposed model. A common edge detection method that recognizes picture edges based on the multistep algorithm.

1. Noise reduction: Edge detection is sensitive to picture noise, it is eliminated using a 5×5 Gaussian filter.

2. Finding Intensity Gradient of the Image: The smoothed picture is then filtered using a Sobel kernel horizontally and vertically to produce the first derivative (G_x , G_y). These two photos show each pixel's gradient and orientation.
3. Rounding: Always perpendicular to edges. It's rounded to a vertical, horizontal, or diagonal angle.
4. Non-maximum suppression: After collecting the gradient magnitude and direction, a thorough scan of the picture is done to eliminate any unnecessary pixels. Every pixel is checked for a local maximum in the gradient's direction.
5. Hysteresis Thresholding: This step determines edges. minVal and maxVal are needed. Any edges with an intensity gradient over maxVal are deemed edges, while those below minVal are removed. Based on their neighborhood, those between these two thresholds are edges or non-edges. They're regarded as edges if they're near "sure-edge" pixels; otherwise, they're ignored. Five stages provide a two-dimensional binary map (0 or 255) of picture edges. Leaves show Canny edge.

5 Results

The model is trained and tested on plant disease dataset [3]. The dataset contains 71.7% unhealthy leaves having multiple diseases, rust, and scab, whereas 28.3% are healthy leaves available in the dataset. The 80% of data from the dataset is used to train the designed model, and 20% data is used for testing. In the Fig. 1, we can see that the healthy leaves are entirely green and do not have any brown/yellow spots or scars. Healthy leaves do not have scabs or rust.

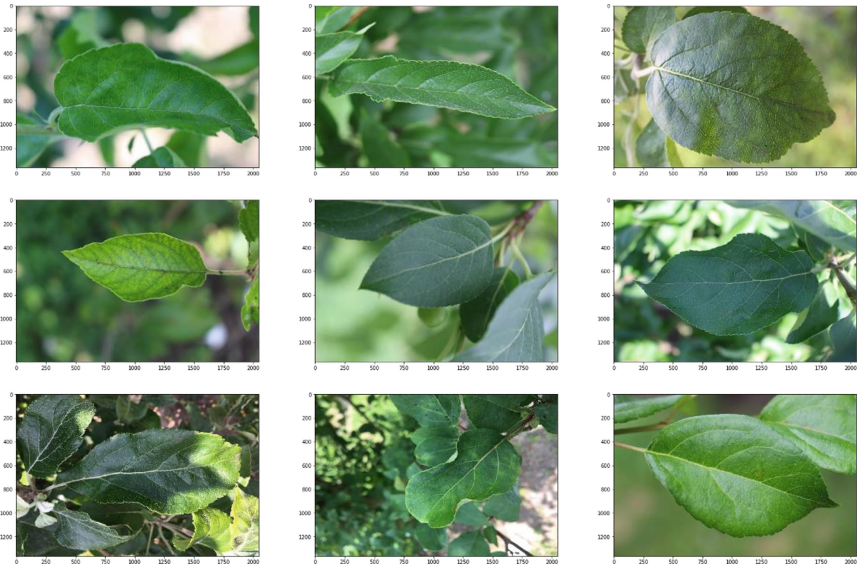


Fig. 1. Sample of Healthy Dataset (Color figure online)

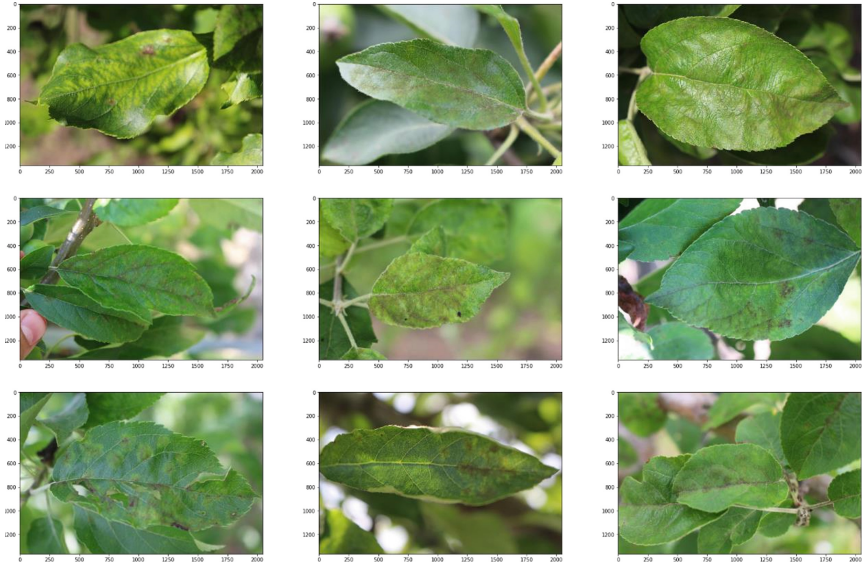


Fig. 2. Sample of rusty Dataset (Color figure online)

In Fig. 2, leaves with “scab” have significant brown marks and stains across the leaf. Scab is “any of various plant diseases caused by fungi or bacteria resulting in crust like spots on fruits, leaves, or roots”. The brown marks across the leaf are a sign of these bacterial/fungal infections. Once diagnosed, scabs can be treated using chemical or non-chemical methods.

In the Fig. 3, we can see that leaves with “rust” have several brownish-yellow spots across the leaf. Rust is “a disease, especially of cereals and other grasses, characterized by rust-colored pustules of spores on the affected leaf blades and sheaths and caused by any of several rust fungi”. The yellow spots are a sign of infection by a particular type of fungi called “rust fungi”. Rust can also be treated with several chemical and non-chemical methods once diagnosed.

In Fig. 4, we can see that the leaves show symptoms of several diseases, including brown marks and yellow spots. These plants have more than one of the above-described diseases.

Table 2. Accuracy comparison of Proposed model using various classification techniques

Classification Technique	Accuracy(Training Data)	Accuracy(Testing Data)
SVM	88.9%	%
K-Means	85%	%
CNN	96.3%	%
EfficientNet-B7	97.2%	90%

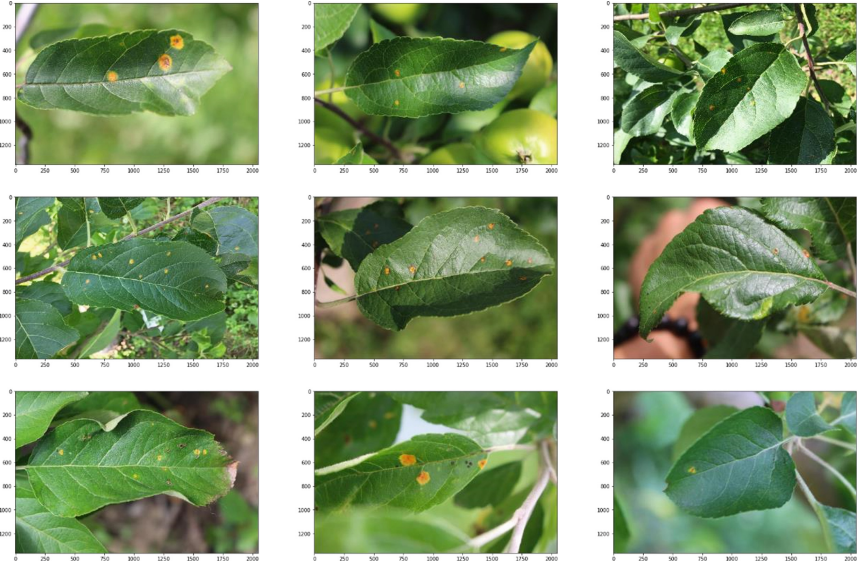


Fig. 3. Sample with “scab” Dataset (Color figure online)

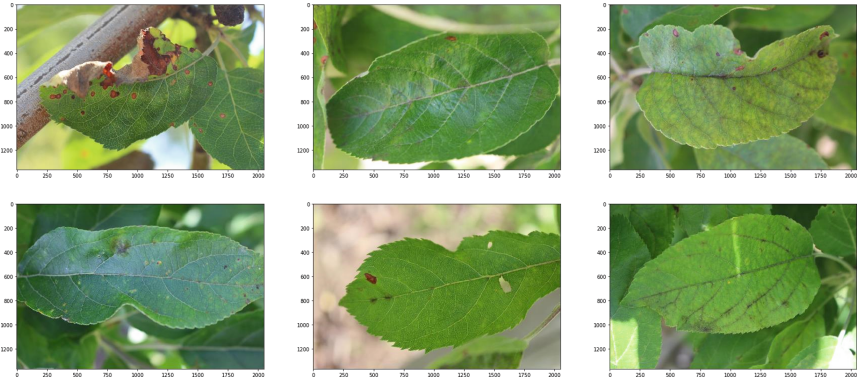


Fig. 4. Sample of Diseased Dataset (Color figure online)

All the models mentioned in the proposed research were implemented with Tensorflow in python. Further, Kaggle was used to train the models mentioned, with the following specs - GPU Tesla P100-PCI-E-16GB computes capability: 6.0 and 16 GB GPU RAM.

EfficientNet predicts leaf diseases with great accuracy as describe in the Fig. 5, 6, 7, 8 and 9. No red bars are seen. The probabilities are very polarized (one very high and the rest shallow), indicating that the model is making these predictions with great confidence. The semi-supervised weights seem to set this model apart from EfficientNet. Once again, the red and blue bars are more

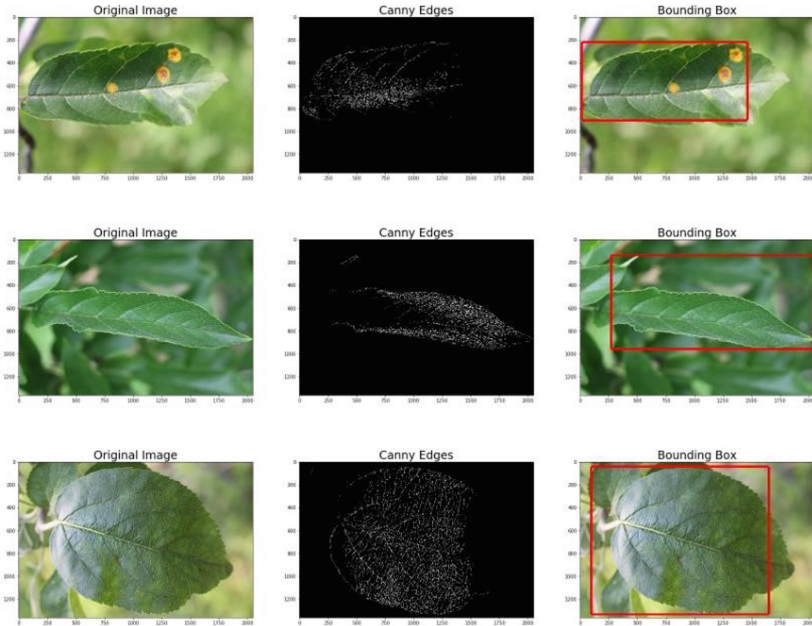


Fig. 5. Canny Edge Detection (Color figure online)

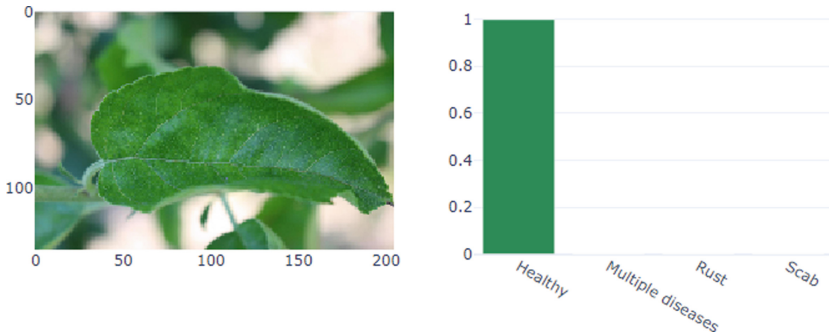


Fig. 6. Result of Healthy leaf (Color figure online) .

prominent in the last (fourth) leaf labeled “multiple_diseases”. This is probably because leaves with multiple diseases may also show symptoms of rust and scab, thus slightly confusing the model.

From the Fig. 10, we can see that the accuracy of the model achieved in training data is 97.2% and using the testing data is 90%. The training metrics settle down very fast (after 1 or 2 epochs), whereas the validation metrics have

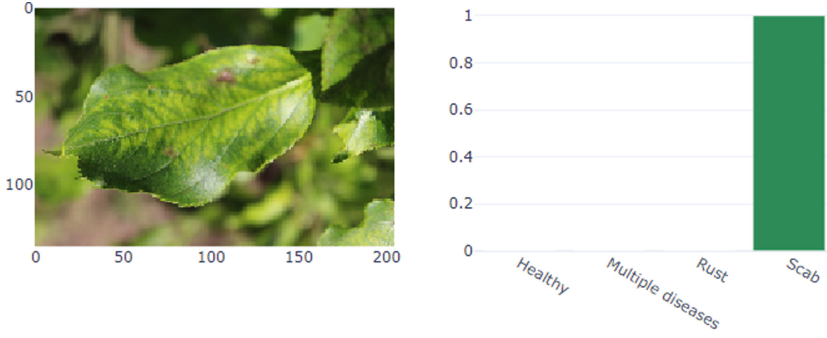


Fig. 7. Result of Scab Leaf (Color figure online)

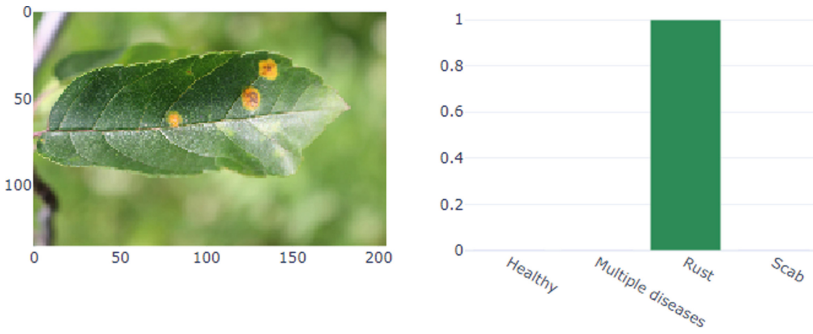


Fig. 8. Scenario 1: Result of Rust Leaf (Color figure online)

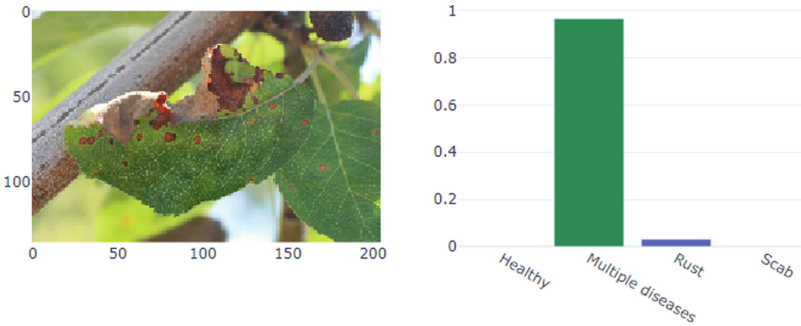


Fig. 9. Scenario 2: Result of Rust Leaf (Color figure online)

much greater volatility and start to settle down only after 12–13 epochs (similar to DenseNet). This is expected because validation data is unseen and more challenging to predict than training data.

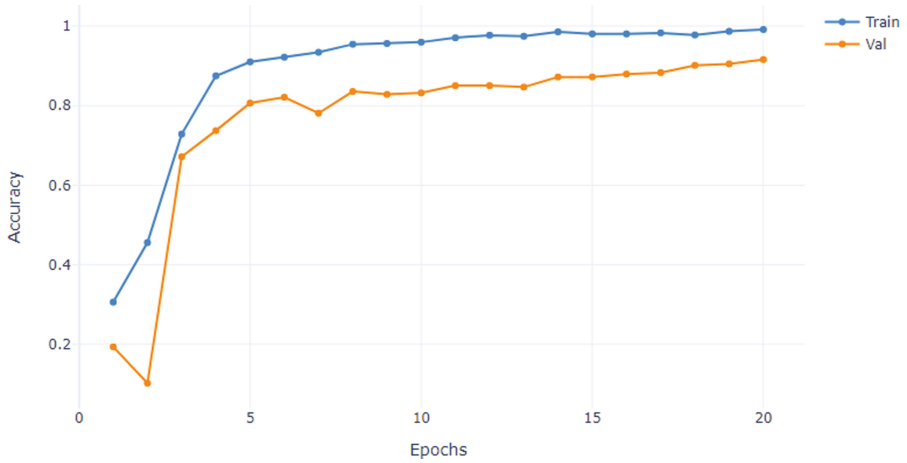


Fig. 10. Performance of EfficientNet-B7 (Color figure online)

6 Future Work

This paper develops a disease detection model using the convolutional neural network EfficientNet-B7 is developed and compared with other models by applying K-means and SVM techniques. The accuracy of the proposed model came out best from cross-validation, given its high scores. The model predicts the diseased plants with high precision so that unnecessary expenses in treatment can be avoided. It has been observed from the investigation that the proposed model provides the accuracy of highest 90% using EfficientNet-B7 on testing data-set and 97.2% accuracy on training dataset to detect the leaves more than other proposed and existing models.

References

1. Ai, Y., Sun, C., Tie, J., Cai, X.: Research on recognition model of crop diseases and insect pests based on deep learning in harsh environments. *IEEE Access* **8**, 171686–171693 (2020)
2. Alotaibi, F.S.: Implementation of machine learning model to predict heart failure disease. *Int. J. Adv. Comput. Sci. Appl.* **10**(6), 261–268 (2019)
3. AlSuwaidi, A., Grieve, B., Yin, H.: Feature-ensemble-based novelty detection for analyzing plant hyperspectral datasets. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **11**(4), 1041–1055 (2018)
4. Ashourloo, D., Aghighi, H., Matkan, A.A., Mobasheri, M.R., Rad, A.M.: An investigation into machine learning regression techniques for the leaf rust disease detection using hyperspectral measurement. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **9**(9), 4344–4351 (2016)
5. Azimi, S., Wadhawan, R., Gandhi, T.K.: Intelligent monitoring of stress induced by water deficiency in plants using deep learning. *IEEE Trans. Instrum. Meas.* **70**, 1–13 (2021)

6. Sunil, C.K., Jaidhar, C.D., Patil, N.: Cardamom plant disease detection approach using EfficientNetV2. *IEEE Access* **10**, 789–804 (2022)
7. Huang, S., Zhou, G., He, M., Chen, A., Zhang, W., Hu, Y.: Detection of peach disease image based on asymptotic non-local means and PCNN-IPELM. *IEEE Access* **8**, 136421–136433 (2020)
8. Jadon, S.: SSM-net for plants disease identification in low data regime. In: 2020 IEEE/ITU International Conference on Artificial Intelligence for Good (AI4G), pp. 158–163 (2020)
9. Joshi, P., Das, D., Udutalapally, V., Pradhan, M.K., Misra, S.: Ricebios: identification of biotic stress in rice crops using edge-as-a-service. *IEEE Sens. J.* **22**(5), 4616–4624 (2022)
10. Khosla, A., Cao, Y., Lin, C.C.-Y., Chiu, H.-K., Hu, J., Lee, H.: An integrated machine learning approach to stroke prediction. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 183–192 (2010)
11. Kumar, M., Kumar, A., Palaparthi, V.S.: Soil sensors-based prediction system for plant diseases using exploratory data analysis and machine learning. *IEEE Sens. J.* **21**(16), 17455–17468 (2021)
12. Liu, L., et al.: Deep learning based automatic multiclass wild pest monitoring approach using hybrid global and local activated features. *IEEE Trans. Ind. Inf.* **17**(11), 7589–7598 (2021)