Journal of Research Administration
Society of Research Administrators International

DISEASE DETECTION AND RECOGNITION OF ROSE PIANTS USING IMAGE PROCESSING AND MACHINE LEARNING

Dr.M.P.Vani
Associate Professor. (sr), SITE (School of Information Technology and Engineering), Vellore Institute of Technology, Vellore (632014), Tamil Nadu, India

Abstract:
The main motivation behind doing this research is to help the agriculture sector, especially the farmers to predict, detect and prevent plant diseases. In this paper we will work on black spot disease in leaves of rose plant. The current methodology which is being used is outdated and highly dependent upon human resources, it is practically impossible for the plant pathologist to help every farmer and it is also not possible for farmers to afford plant pathologists. The main objective of this research is not only to detect plant diseases using image processing techniques but also to prevent diseases by studying the hidden pattern. We'll be using the following algorithms – Random Forest from machine learning and Convolutional Neural Networks from deep learning. This paper includes various phases of implementation namely data creation, feature extraction, training the classifier and classification.

Keywords: Diseased and Healthy leaf, Feature extraction, Training, Classification, CNN, Random Forest

Introduction:

In India, a lot of people are dependent upon agriculture as their source of income, and flowers constitute a major chunk of that. Most of the flowers farmers cultivate are rose plants, as it is known that roses are one of the most commonly used flowers and have a high market value. India is one of the highest rose exporters in the international market. In 2019, accounting for $353,503. Diseases like black spots and powdery mildew affect the growth of the plant, thereby hindering the revenue of the farmers. The traditional method of disease detection is highly dependent on human resources and is normally very expensive; hence, most farmers cannot afford expert help.

So, to tackle this problem, the paper proposes a new method of disease detection in the rose plant that will include the deep learning algorithm CNN convolutional neural network, which is used for image detection, recognition, and feature extraction from the given input. In addition to this, the model will be using the machine learning algorithm RF Random Forest, which is a classification algorithm used to classify the result. The selection of the proposed model is done because the general research shows that both the machine learning and deep learning algorithms have certain kinds of advantages and disadvantages depending upon the size of the data. Deep Learning tends to perform better when there is abundance of data in the dataset, it is designed to handle big data yet when the data in the dataset is less than the performance of the deep learning algorithm is not good. On the other hand, Machine learning does the exact opposite it excels when data in the data set is less or limited, the performance of machine learning is bad when a big dataset is used. So, the proposed solution is the best of both worlds, a combination of Deep Learning and Machine Learning.

Plant Diseases:

1) Downy mildew: It is caused by Peronospora sparsa, an obligate biotrophic oomycete. It is one of the deadliest diseases when it comes to roses. It attacks all types of roses - greenhouses, farms, nursery.

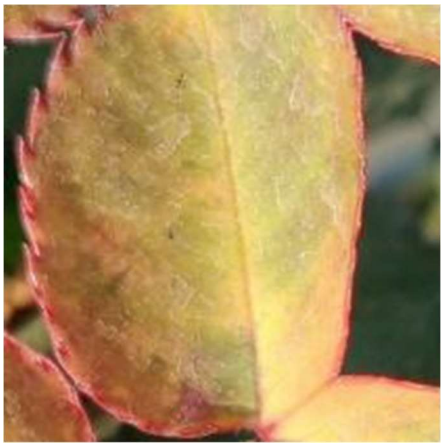


Figure 1

2) Black spot: It is caused by the fungus, Diplocarpon rosae. Black spots are one of the most common diseases in roses. It doesn't kill the plant immediately but rather takes more time and with time weakens the plant and later ends in the death of the plant. A cold and moist environment is the most suitable environment for this plant disease.

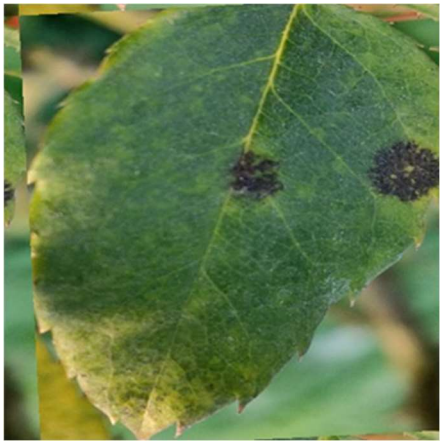


Figure 2

Literature Survey:

[1] O.A. Miloserdov proposed the following model, the dataset inserted into the model is manually pre-categorized into 5 groups. After the augmentation of the data, the model used CNN for image processing and classification.

[2] Sahadat Hossain Khan proposed that the model the data is manually augmented and the algorithm ConvNet of CNN is used for training the model, it has used various libraries and tools like Scikit-Learn, Pandas, Matplotlib and NumPy as well. The modified CNN model has four layers consisting of a 2- D Convolutional layer and max-pooling layers available in each layer.

The features of the image are usually extracted using these layers. The model produces an accuracy of 96%.

[3] The paper proposed a "Rose Plant Leaves: Disease Detection and Pesticide Management using CNN" model, the paper is focused on the comparative study conducted on the usage of 5 different machine learning classification techniques like SVM, ANN, KNN, Fuzzy, used for recognition of plant disease. CNN is used for feature extraction with the help of RESNET for skipping one or more layers without hindering the performance of the model. Lastly, a suitable classification technique was selected.

[4] Sakshi Sharma proposed a model which can identify two diseases of potato leaves – early blight and late blight. The author proposed a model that included several machine learning approaches, namely SVM, KNN, Nave Bayes, and Decision Tree. K means clustering is used for obtaining the ROI for diseased leaf. Then classification is done using the SVM because of the highest accuracy which is 92.9%.

[5] S.Arun Kumar proposed a model, which identifies whether the tomato leaf is diseased or healthy. Data was pre-processed, then HIS space is used for image segmentation for the diseased portion of the leaf. Then GLCM and LBP for feature extraction based on texture. Features extracted from both GLCM and LBP are distinct. Hence, feature fusion was employed to form a single feature and CCA is used to examine the mutual relation between 2 features before fusion. Then SVM is used for classification. The model has an accuracy of 96.93% for the binary class and 85.27% for the multiclass system.

[6] Shima Ramesh proposed a method to check whether the plant is diseased or not. Dataset was collected from a public database. The data was pre-processed, HOG was applied for feature extraction and RF was used for classification because of the best accuracy score. The model scored an accuracy of 70.14%.

[7] Dr. V. Brindha Devi proposed a method to detect disease in crops using RF. After data augmentation HOG was applied for feature collected, after that multiple methods were implemented which are Logistic Regression, SVM, KNN, CART, RF, and Naïve Bayes. Then based on accuracy RF was selected and the model was trained and tested. The accuracy of the model was 97.2%.

Materials and Methods

Dataset Collection:

The required dataset of rose leaves is collected from. The dataset is divided into two types and even in those two types there are three different segregations, the dataset is divided into Train and Test and there both are having three categorizations i.e.

1. Black Spot
2. Downy Mildew
3. Fresh Leaves

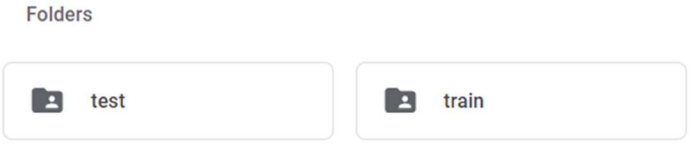


Figure 4

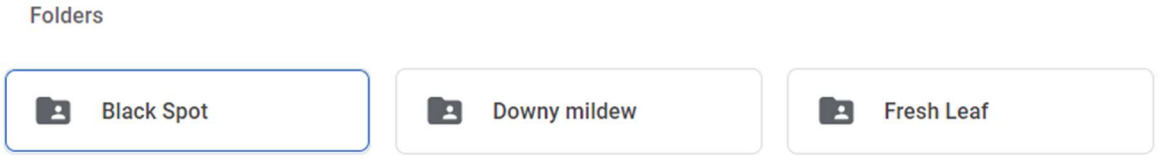


Figure 3

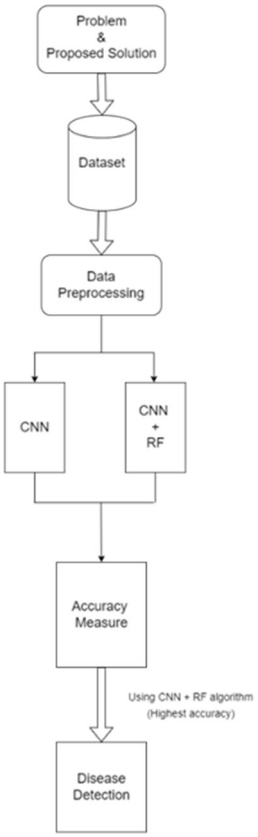


Figure 3

Experimental Procedure:

The problem stated above has been thoroughly analysed and we have produced a solution which is better than the existing ones, there is nothing new in building an image recognition model using Deep Learning Algorithm CNN (Convolutional Neural Network). But in the proposed model we are trying to implement a combination of a Deep Learning algorithm and Machine Learning algorithm to identify and classify the diseased leaves and the healthy leaves of the rose plant.

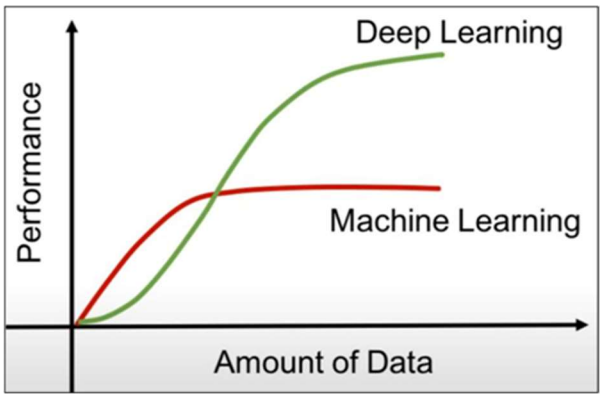
The combination of the below two is being used:

1. Deep Learning – Convolutional Neural Network (CNN)
2. Machine Learning – Random Forest (RF)

The proposed model will use the layers of CNN for feature extraction which are conv2D and Maxpooling layers then Batch-normalization will be performed (re-centring and re-scaling). After that model will use flatten layer to flatten the processed data into a single column because Random Forest does not accept 2D or multidimensional inputs.

The output of the feature Extractor will be the input for the RF classifier then the model will train. Random Forest is a Machine Learning algorithm widely used for classification and regression problems. RF builds decision trees with different samples and takes majority votes for classification. Now the test dataset features will be extracted, and the model will predict results for test data that was not the part of training data.

The need of this approach:



The figure (6) represents the performance of Deep Learning and the performance of Machine Learning based on the amount of data used in the training of the respective models.

The graph here shows that the Machine Learning approach works best when the data in the dataset is in a small quantity, as the graph depicts there is a stunning growth in the ML curve, but when the data in the dataset increases the Machine Learning curve tends to remain the same which shows stagnation.

The other curve which shows the performance of Deep Learning gradually increases when there is less data in the dataset and its performance peaks when the amount of data in the dataset is maximum.

Hence, to tackle both problems the proposed model suggests using a combination of both to obtain the optimum result as depicted in the curve when both ML and DL curve intersects each other.

Results and Comparison:

1) Training Accuracy:

a) CNN (Convolutional Neural network)

First, we used the CNN technique alone to recognize and classify images and the result of the test is given below:

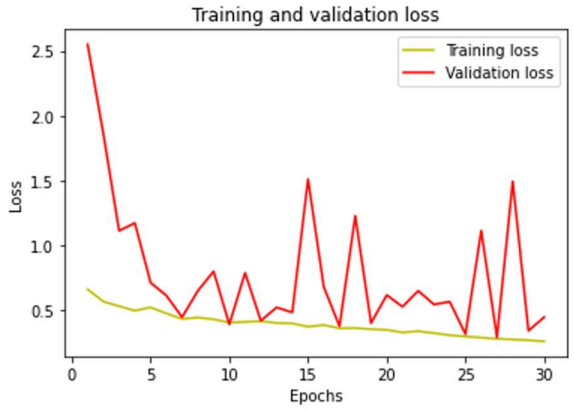


Figure 7

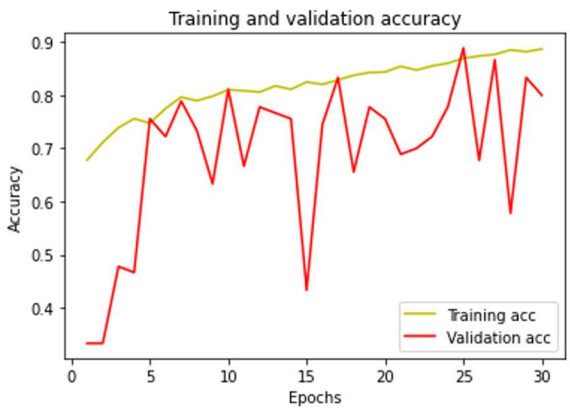


Figure 8

- Training accuracy: 88.6874
- Validation Accuracy: 78.1124
- Validation Loss: 0.2611
- Training Loss: 0.1662

b) CNN+ RF (Proposed Model)

The proposed method was used to recognize and classify images and the result of the test is given below:

```
RF_model = RandomForestClassifier(n_estimators = 25, random_state = 49)
RF_model.fit(X_for_RF, y_train)
print('Training Accuracy : ',
      metrics.accuracy_score(y_train,
                             RF_model.predict(X_for_RF))*100)
print('Validation Accuracy : ',
      metrics.accuracy_score(y_test,
                             RF_model.predict(x_test_RF))*100)

Training Accuracy :  99.90270007297495
Validation Accuracy :  84.44444444444444
```

Figure 4

- Training Accuracy: 99.9027
- Validation Accuracy: 84.4444

The results of both the techniques used show different results, both took different times to train, and in all aspects, the proposed method performs better than CNN. The training time in CNN was approximately 1 hour and 30 minutes and showed a Training accuracy of 88.86% (Approx.) whereas the purposed methods perform exceedingly well with a training time of less than 1 minute and it showed a Training accuracy of 99.90% (Approx.) which is better and faster than CNN.

2) Test Results:

a) CNN (Convolutional Neural network)

Type of Leaf Model Prediction Result

Downy Mildew

Downy Mildew

(Correct)
Black Spot
Black Spot
(Correct)
Fresh Leaf
Black Spot
(Incorrect)

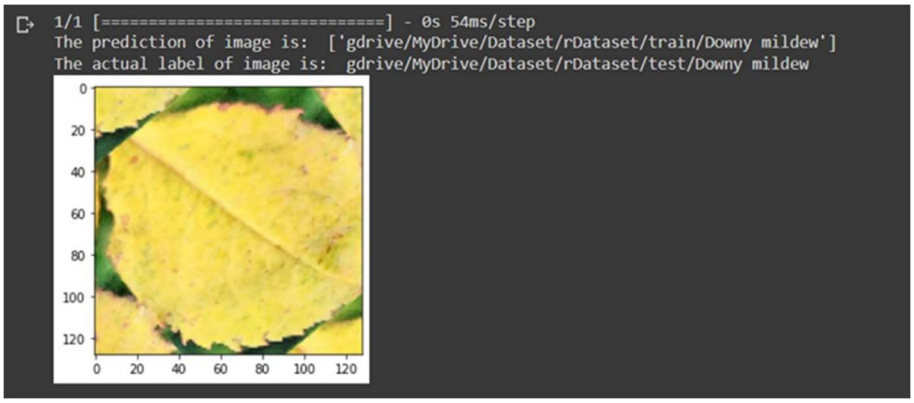


Figure 5

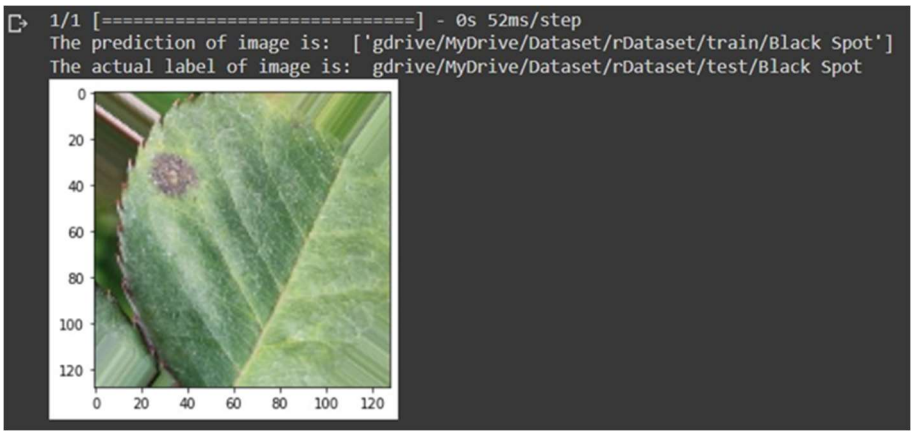


Figure 6

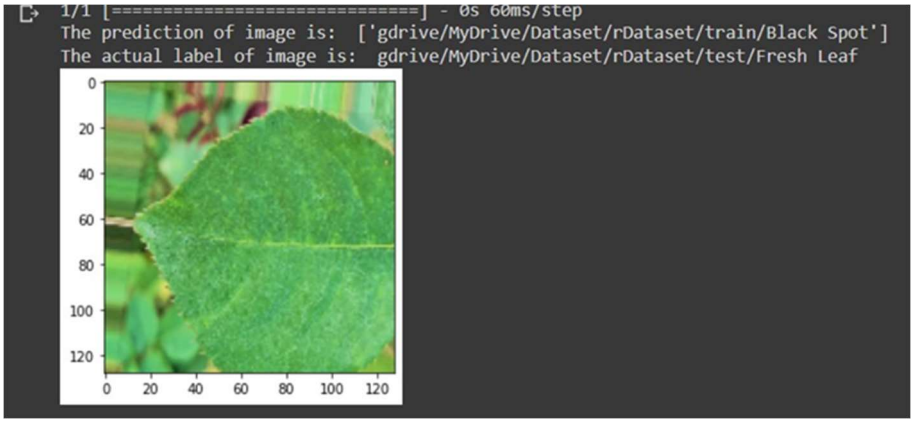


Figure 7

b) CNN+RF (Proposed Method)

Type of Leaf Model Prediction Result

Downy Mildew
Downy Mildew
(Correct)

Black Spot
Black Spot
(Correct)

Fresh Leaf

Fresh Leaf
(Correct)

Conclusion and Future Work:

In future, this model may be utilised in website development or to create a mobile application for people. More disease categories can be added to this model as an addition. More datasets can be added to this model to improve the outcome. Machine learning and deep learning techniques are now widely employed in everyday applications.

A combined Machine Learning and Deep Learning approach is used in our paper to predict the intended outcome. Our primary goal was to develop a model that may aid in the efficient and accurate identification of illness in rose leaves. Our approach is more efficient, faster, and produces more accurate results when compared to other models. This model can also be helpful in the area of research.

References:

[1] Title: Use of a deep convolutional neural network to diagnose disease in the rose by means of a photographic image
Authors: O.A. Miloserdov, N.S. Ovcharenko, AndreyV. Makarenko
Published in: 2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP)
[2] Title: Rose Leaf Disease Detection Using CNN
Authors: Sahadat Hossain Khan, Md. Nayan Hossain, Sithi Mozumder, Shayla Sharmin, Nazmun Nessa Moon
Published in: 2021 IEEE International Conference on Robotics, Automation, Artificial-Intelligence and Internet-of-Things (RAAICON)
[3] Title: Rose Plant Leaves: Disease Detection and Pesticide Management using CNN
Authors: Mohammad Ibrahim Khaleel, Paineni Gowtham Sai, Adabala Uday Raghavendra Kumar, Purushothamman Raja, Vinh Truong Hoang
Published in: Proceedings of the Sixth International Conference on Trends in Electronics and Informatics (ICOEI 2022)
[4] Title: Classification of Diseased Potato Leaves Using Machine Learning
Authors: Sakshi Sharma, Vatsala Anand, Swati Singh
Published in: 10th IEEE International Conference on communication systems and network technologies (CSNT) in September/2021
[5] Title: Disease Detection in Tomato Leaves using Machine Learning and Statistical Feature Fusion
Authors: S.Arun Kumar, S. Sasikala
Published in: International Conference on advancements in electrical, Electronics, Communication, Computing and Automation (ICAECA) in March 2021
[6] Title: Plant Disease Detection Using Machine Learning

Authors: Shima Ramesh, Mr. Ramachandra Hebbar, Niveditha M, Pooja R, Prasad Bhat N, Shashank N, Mr. P V Vinod
Published in: International Conference on Design Innovations for 3Cs Compute Communicate Control in 2018
[7] Title: An Efficient and Robust Random Forest Algorithm for Crop Disease Detection
Authors: An Efficient and Robust Random Forest Algorithm for Crop Disease Detection
Published in: International Conference on Communication, Computing and Internet of Things (IC3IoT) in March 2022