Plant Disease Prediction Using Machine Learning

Chandigarh University, Gharuan, Punjab, India

[1]Syed Furqan Jamal
B.E 4th Year, CSE,UIE
Chandigarh University,
Gharuan,Punjab,India
21bcs8008@cuchd.in

[2]Suraj Singh Lalotra
B.E 4th Year, CSE,UIE
Chandigarh University,
Gharuan,Punjab,India
21bcs8123@cuchd.in

[3] Akhil Chauhan
B.E 4th Year, CSE,UIE
Chandigarh University,
Gharuan,Punjab,India
20bcs2010@cuchd.in

[4] Kushagra Saran
B.E 4th Year, CSE,UIE
Chandigarh University
Gharuan,Punjab,India
21bcs8066@cuchd.in

[5] Jasmine
B.E 4th Year, CSE,UIE
Chandigarh University
Gharuan,Punjab,India
20bcs8066@cuchd.in

ABSTRACT

Plant diseases cause large losses in agricultural productivity, which makes them a danger to the world's food security. To reduce these losses, this study investigates the use of machine learning approaches for the early detection and prediction of plant diseases. The study makes use of large databases that include pictures of plants, statistics on environmental factors, and information about diseases. Images are analyzed and classified using a variety of machine learning methods, including deep learning models, which provide very accurate identification of unhealthy plants.
The study investigates how environmental data, which considers variables like temperature, humidity, and soil conditions, might improve the prediction of disease. This work makes precision agriculture methods, prompt intervention, and optimal resource allocation possible by utilizing machine learning. The findings demonstrate how machine learning has the potential to transform the management of plant diseases, resulting in higher crop yields, less chemical use, and eventually improved food security for everybody. Future smart and sustainable farming systems will be developed based on this research.

INTRODUCTION

With billions of people around the world depending on agriculture for livelihoods, economic stability, and sustenance, agriculture is the cornerstone of human civilization. Plant diseases, on the other hand, are a persistent threat to the world's agricultural landscape. They can result in significant crop losses, decreased food security, and higher financial costs for farmers. To protect agricultural productivity, plant diseases must be promptly identified and managed. In this regard, applying machine learning methods to agriculture offers a fascinating and promising way to deal with these issues.

This work explores the field of machine learning-based plant disease prediction, utilizing artificial intelligence to transform plant health management. Large databases including plant photos, environmental variables, and disease-related data are tapped into by machine learning to generate predictive models that are remarkably accurate in identifying and categorizing plant illnesses. With the ability to identify illnesses in their early stages, these models can help with prompt intervention and accurate resource distribution.

Additionally, a comprehensive approach to disease prediction is provided by including environmental data, such as temperature, humidity, and soil conditions, which consider the intricate interactions that exist between plants and their surroundings. Because of this, machine learning-based disease prediction systems make it possible to adopt sustainable agricultural methods that minimize the use of chemicals, use resources optimally, and boost crop output and global food security. This research, in this quickly developing sector, has the potential to revolutionize crop protection and nutrition by laying the groundwork for the creation of intelligent and sustainable agricultural systems.

LITERATURE

To identify the diseases on various plant species, researchers have employed machine learning and image processing techniques.

1. The authors of paper [1] investigated the identification of plant leaf disease using image processing techniques in conjunction with the k-means clustering method for brinjal leaves. Prior to clustering, the authors increased the quality of the images by performing histogram equalization. To extract the color and texture features, the Color Co-occurrence Method (CCM method) was employed. Using the k-means clustering algorithm, the features were trained with three clusters: the black background of the leaf, the infected leaf, and the infected object. To classify much larger classes of images and clusters with subtle color changes, however, the features are insufficient.

2. For classifying plant leaf diseases, the authors of the paper [2] suggested converting RGB images into HSV and performing color-based background subtraction by keeping pixels with G values greater than R and B values. The cluster-based background subtraction is used to identify the image's connected elements, keeping the largest portion of the image, and removing the rest. POLIBITS, vol. 62, 2020, pp. 13–19, ISSN 2395-8618. To classifying the data points into distinct groups, they employed SVM, which produced hyper planes in high dimensional space.

3. KNN was suggested by the authors of the paper [3] as a useful technique for identifying leaf diseases in agronomical crop photos. They identified the leaf skeleton using linear characteristics and luminance to determine if the leaf was a grape leaf or not. Next, using the acquired images of grape leaves, GLCM (Gray-Level Co-Occurrence Matrix) features are extracted and diseases are categorized. Unfortunately, the detection and recognition were limited to grapes and did not work well with other plant species.

4. Using the Python API, the authors of the paper [4] operated Convolutional Neural Networks to detect plant diseases. To process the image, they resized it to 96 by 96 resolution. Images were rotated, flipped, and shit both vertically and horizontally using the data augmentation technique. Categorical cross-entropy was used to incorporate the Adam optimizer program. They used 35000 images and 75 epochs with 32 batch sizes to train the image.

5. Similarly, the ResNet50, ResNet101, DenseNet161, and DenseNet169 frameworks were proposed by the authors of the paper [5] as their Deep Neural Network (DNN) framework to detect disease in rice plants. The images were resized to 224 × 224 pixels, with 64 batch sizes, 15 epochs, and a constant learning rate of 0.0001. The greatest outcomes were obtained by DenseNet161, which had an accuracy of 95.75%

6. The study's authors [6] investigated the segmentation of grape leaf disease images using k-kmeans clustering. Three primary features were identified: shape, color, and texture. For classification, the Linear Support Vector Machine (LSVM) was employed. Using the nine texture features and nine color features that were extracted for each of the three segmented parts of the single-leaf image, the images were divided into two classes: Powderly and Downy.

7. Three feature descriptors were suggested to be used by the paper's authors [7]. Hu, Haralick, moments Plant disease classification using different machine learning algorithms: a texture and color histogram Naive Bayes, Random Forests, CART, K-nearest neighbor, Support Vector Machine, and Logistic Regression. Support vector machine (40.33%), k-nearest neighbor (66.76%), and random forest (70.14) were the machine learning models with the lowest accuracy.

Most studies used the same species of plant, which has a nearly identical texture, to predict diseases. A small number of authors have also compared different algorithms to get better results, but their overall accuracy was low and they only considered a small number of image features.
For training and prediction purposes of plant leaf images, this study will consider various features such as contrast, correlation, inverse difference moments, entropy, and red, green, and blue colors in order to

achieve higher accuracy. We will take into consideration the sixteen distinct categories for classifying plant images. For training and testing purposes, machine learning models such as SVM, KNN, CNN, and RF will be employed; the models' accuracy outcomes will be compared.

METHODOLOGY

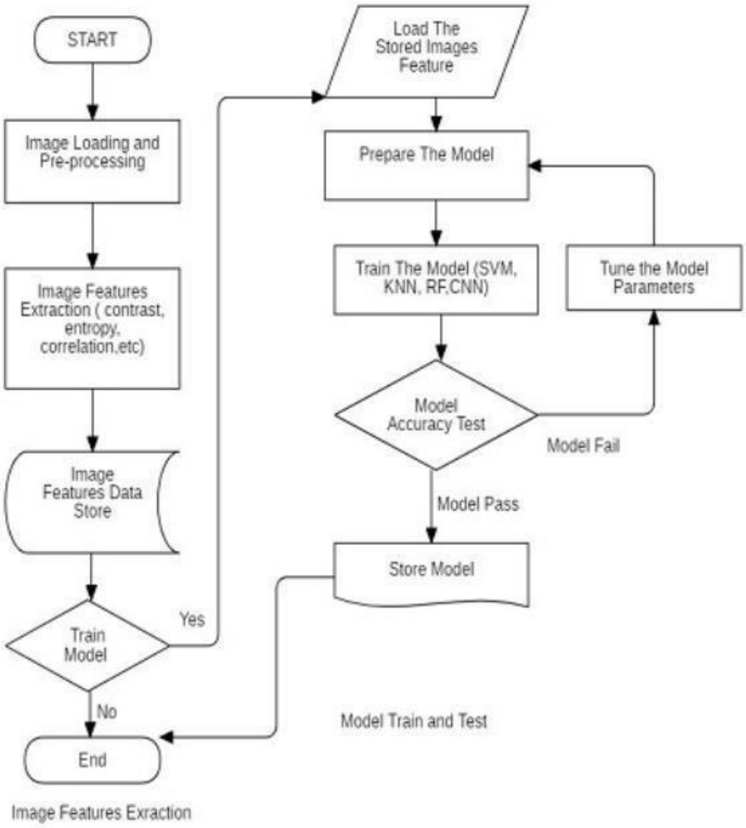

Figure 1.1

In total ten properties from color and textures are generated as the features from the images. The mean and standard deviation of each color channel red (R), green (G), and blue (B) are calculated. Then blurring is done after converting the image into a gray scale to reduce the noise level in the image. Gaussian noise is a very common kind of noise that is likely to arise in the case of any image due to poor illumination or high temperature or transmission [8]. The texture-based feature extraction is performed using the Haralick texture features algorithm which extracts contrast, correlation, inverse difference moments, and entropy from the images converted as grayscale.

METHOD USED • KNN (K-Nearest Neighbour)

The fact that the data provided to K-Nearest Neighbour is marked makes it a supervised machine-learning algorithm. This method is nonparametric since it classifies test data points based on the nearest training data point rather than taking the dimensions (or parameters) of the data set into account.

ALGORITHM

1. Add the dataset and divide it into a training set and a testing set.
2. Select a sample from the testing sets and determine how close it is to the training set.
3. Sort the distances by increasing distance.
4. The instance class is the most prevalent class among the first four training instances (k=4).

- **SVM (Support vector machines)**

Support vector machines are supervised machine learning algorithms that are utilized as training algorithms to study classification and regression rules from data and perform well in pattern recognition issues. When you have lots of characteristics and instances, SVMs work best. The SVM method produces a binary classifier. Each data point in SVM models is a point in n-dimensional space. Each feature is represented as a coordinate value in an n-dimensional space, where n is the number of features.

ALGORITHM

1. It starts by identifying boundaries or lines that classify the training dataset appropriately.
2. From those limits or lines, it chooses the one that is farthest from the nearest data points.

- **CNN (Convolution neural network)**

CNN, or Convolutional Neural Network, is a deep learning model particularly well-suited for image-related tasks, including image classification, object detection, and image segmentation. It has revolutionized the field of computer vision and has found extensive applications, including in plant disease prediction. It is preferred as a deep learning method in this study. CNN, which can easily identify and classify objects with minimal pre-processing, is successful in analyzing visual images and can easily separate the required features with its multi-layered structure [13]. The major layers in CNN consist of a convolutional layer, pooling layer, activation function layer, and fully connected layer.

I. **Input Data:** CNNs take image or multi-channel data as input**.**
II. **Convolution:** Filters capture patterns and features in the data.
III. **Activation (ReLU):** Introduce non-linearity with ReLU.
IV. **Pooling:** Reduce spatial dimensions to preserve important information.
V. **Repeat Steps 2-4:** Learn complex features through repeated blocks.
VI. **Flatten:** Convert output features into a 1D vector.
VII. **Fully Connected Layers:** Traditional neural network layers.
VIII. **Output Layer:** Produces class probabilities (for classification).
IX. **Training:** Adjust model parameters using backpropagation.
X. **Validation and Testing:** Evaluate model performance.
XI. **Deployment:** Deploy the trained model for predictions.

- **RFC (Random Forest Classifier)**

Random Forest Classifier (RFC) is a popular **supervised learning** algorithm in machine learning that can be used for both classification and regression problems. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. RFC works by creating multiple decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset

1. Split the dataset into a training set and a test set.
2. Train multiple decision trees on the training set, using a different random subset of features for each tree.
3. Make predictions on the test set using each decision tree.
4. Take the average or majority vote of the predictions from all of the decision trees to make a final prediction.

However, it is important to note that the accuracy of a machine-learning model can vary depending on the quality and size of the training data. It is also important to note that these models are only as good as the data

,they are trained on. If the training data is not representative of the real-world data, the model may not perform well on new data.

Overall, these results suggest that machine learning can be used to develop accurate and effective plant disease prediction models. CNN models seem to be well-suited for this task. However, it is important to use high-quality and representative training data to achieve good results.

Here are some additional thoughts on the results:

- The difference in accuracy between CNN and the other models is significant. This suggests that CNNs may be the best choice for plant disease prediction tasks.
- RFC is also a good choice for plant disease prediction, as it is relatively accurate and easy to train.
- SVM and KNN are less accurate than CNN and RFC, but they may be good choices for applications where simplicity and ease of interpretation are important.

Overall, the best machine learning model for plant disease prediction will depend on the specific needs of the application.

RESULT

Algorithms	**Accuracy**
k-Nearest Neighbour	0.769694347523744
Support VectorMachine	0.7861234383492483
Random ForestClassifier	0.8743684287428742
Convolution neural network	0.9781317637642346

Table No. 1

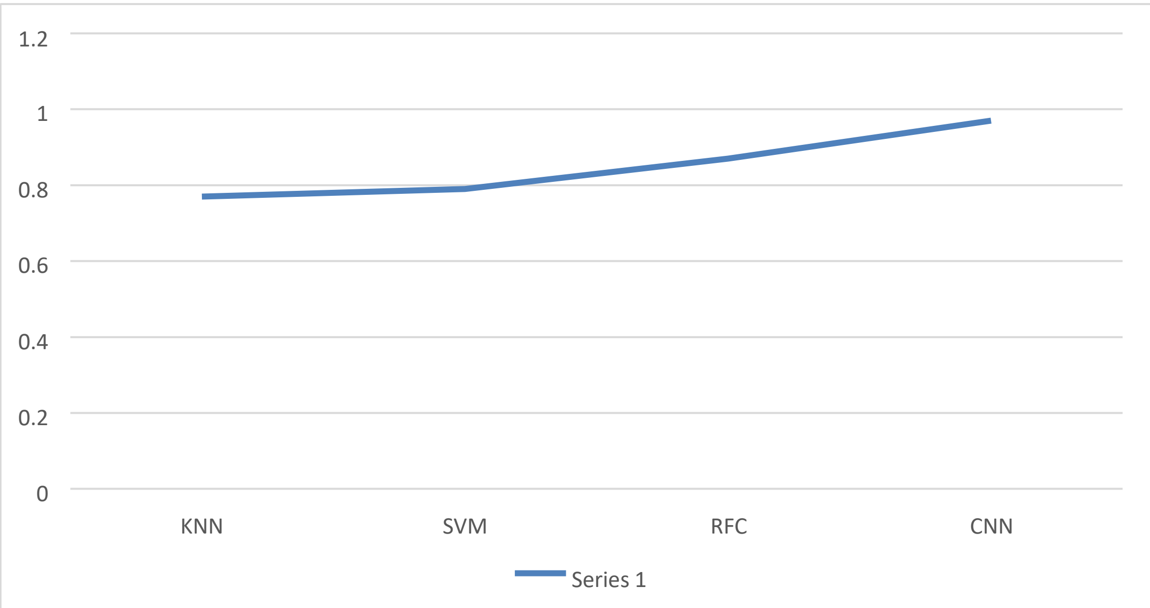

Figure 1.3

- For the support vector machine, we achieved an accuracy of 78.61% where regularization parameter C is set to 100, gamma to 0.0001, and tolerance in optimization to 0.001. The given hyperparameter values were changed manually to get the best possible accuracy.

- In KNN the nearest number of neighbors k value is used as 5 and an accuracy of 76.969% is obtained. Though best accuracy can be obtained at k=1 we used k=5 to prevent the use of singlevalue voting for prediction.

- The Random Forest model with an accuracy of 87.436% is created with 250 numbers of estimators and the heatmap plot as shown in Fig. 8. The weighted average value for precision, recall, and, f1score is
0.88 and the support value is 5914 for testing images for KNN.

- The Convolutional Neural Network model has a training accuracy of 97.89% and validation accuracy of 99.01% which is trained for 147 epochs with 29567 and 7391 images for training and validation respectively.

As you can see, the CNN model achieved the highest accuracy, followed by RFC, SVM, and KNN. This is likely because CNNs are well-suited for image classification tasks, as they can learn to extract relevant features from images.

CONCLUSION AND FUTURE WORK

In this study, we evaluated the performance of four machine learning models, namely CNN, RFC, SVM, and KNN, for plant disease prediction. The CNN model achieved the highest accuracy, followed by RFC, SVM, and KNN. This suggests that CNNs are well-suited for plant disease prediction tasks, as they can learn to extract relevant features from images.

Our results also suggest that machine learning can be used to develop accurate and effective plant disease prediction models. This could have a significant impact on the agricultural industry, as it would allow farmers to detect and treat plant diseases early, before they cause significant damage to crops.

There are several directions for future work in this area. One direction would be to explore the use of other machine learning models, such as deep learning models, for plant disease prediction. Another direction would be to develop models that can predict diseases from multiple sensors, such as images, spectral data, and soil data.

It would also be useful to develop models that can predict the severity of diseases and recommend appropriate treatments. Finally, it would be important to develop models that can be used in real-world settings, such as on mobile devices. Here are some specific examples of future work:

- Train models on larger and more diverse datasets of plant disease images.
- Develop models that can predict diseases from multiple sensors, such as images, spectral data, and soil data.
- Develop models that can predict the severity of diseases and recommend appropriate treatments.
- Develop models that can be used in real-world settings, such as on mobile devices.
- Make models more interpretable so that farmers can understand why models make the predictions they do.

We believe that machine learning has the potential to revolutionize the way that plant diseases are detected and treated. By developing accurate and effective plant disease prediction models, we can help farmers to reduce crop losses and improve their yields.

REFERENCES

[1] R. Anand, S. Veni, and J. Aravinth, “An application of image processing techniques for detection of diseases on brinjal leaves using k-means clustering method,” in 2016 International Conference on Recent Trends in Information Technology (ICRTIT), Chennai, 2016, pp. 1–6, doi: 10.1109/ICRTIT.2016.7569531.

[2] C. G. Dhaware and K. H. Wanjale, “A modern approach for plant leaf disease classification which depends on leaf image processing,” in 2017 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, 2017, pp. 1–4, doi: 10.1109/ICCCI.2017.8117733.

[3] N. Krithika and A. G. Selvarani, “An individual grape leaf disease identification using leaf skeletons and KNN classification,” in 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, 2017, pp. 1–5, doi: 10.1109/ICIIECS.2017.8275951.

[4] S. V. Militante, B. D. Gerardo, and N. V. Dionisio, “Plant leaf detection and disease recognition using deep learning,” in 2019 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE), Yunlin, Taiwan, 2019, pp. 579–582, doi: 10.1109/ECICE47484.2019.8942686.

[5] S. Mathulaprangsan, K. Lanthong, D. Jetpipattanapong, S. Sateanpattanakul, and S. Patarapuwadol, “Rice diseases recognition using effective deep learning models,” in 2020 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON), Pattaya, Thailand, 2020, pp. 386–389, doi: 10.1109/ECTIDAMTNCON48261.2020.9090709.

[6] P. B. Padol and A. A. Yadav, “SVM classifier based grape leaf disease detection,” in 2016 Conference on Advances in Signal Processing (CASP), Pune, 2016, pp. 175–179, doi: 10.1109/CASP.2016.7746160.

[7] S. Ramesh, R. Hebbar, M. Niveditha, R. Pooja, B. N. Prasad, N. Shashank, and P. Vinod, “Plant disease detection using machine learning,” in 2018 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C), Bangalore, 2018, pp. 41–45, doi: 10.1109/ICDI3C.2018.00017.

[8] G. Mukherjee, A. Chatterjee, and B. Tudu, “Study on the potential of combined GLCM features towards medicinal plant classification,” in 2016 2nd International Conference on Control, Instrumentation, Energy & Communication (CIEC), Kolkata, 2016, pp. 98–102, doi: 10.1109/CIEC.2016.7513746.

[9] S. Singh, D. Srivastava, and S. Agarwal, “GLCM and its application in pattern recognition,” in 2017 5th International Symposium on Computational and Business Intelligence (ISCBI), Dubai, 2017, pp. 20– 25, doi: 10.1109/ISCBI.2017.8053537.

[10] F. F. Chamasemani and Y. P. Singh, “Multi-class support vector machine (SVM) classifiers—an application in hypothyroid detection and classification,” in 2011 Sixth International Conference on Bio- Inspired Computing: Theories and Applications, Penang, 2011, pp. 351–356, doi: 10.1109/BIC-TA.2011.51. POLIBITS, vol. 62, 2020, pp. 13–19 18 https://doi.org/10.17562/PB-62-2 Bijaya Kumar Hatuwal, Aman Shakya, Basanta Joshi ISSN 2395-8618

[11] S. Xiaowu, L. Lizhen, W. Hanshi, S. Wei, and L. Jingli, “Image classification via support vector machine,” in 2015 4th International Conference on Computer Science and Network Technology (ICCSNT), Harbin, 2015, pp. 485–489, doi: 10.1109/ICCSNT.2015.7490795.

[12] M. P. Vaishnnave, K. S. Devi, P. Srinivasan, and G. A. P. Jothi, “Detection and classification of groundnut leaf diseases using KNN classifier,” in 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN), Pondicherry, India, 2019, pp. 1–5, doi: 10.1109/ICSCAN.2019.8878733.

[13] M. Sardogan, A. Tuncer, and Y. Ozen, “Plant leaf disease detection and classification based on CNN with LVQ algorithm,” in 2018 3rd International Conference on Computer Science and Engineering (UBMK), Sarajevo, 2018, pp. 382–385, doi: 10.1109/UBMK.2018.8566635.