PLANT HEALTH AND DISEASE DETECTION SYSTEM

Rishabh Singh Rana
Computer Science and Engineering
Graphic Era Hill University
Dehradun, India
t

Shrey Goel
Computer Science and Engineering
Graphic Era Hill University
Dehradun, India

Somesh Thapliyal
Computer Science and Engineering
Graphic Era Hill University
Dehradun, India

Preeti Chaudhary
Computer Science and Engineering
Graphic Era Hill University
Dehradun, India

***Abstract*—Disease identification in crops is a crucial and important task in agricultural industries it is also labour intensive. This process takes a lot of time and skilled labor. This study gives an intelligent approach for crop disease detection using computer vision and machine learning methods. This project is capable of identifying different distinct diseases 4 plants with an accuracy of 90%. Furthermore, the research the research gives historical and real time data of all the diseases and can identify the disease of the plants mentioned in deep research. Which gives an advantage in plant health management, fostering a dynamic and adaptive approach to plant health management. The results of the research gives the significance advancement to the AI technology in plants and agricultural field.**

***Index Terms*—Plant Disease Detection, Digital Image Processing, Machine Learning, Treatment Recommendation**

I. INTRODUCTION

About 70% of India's population depends on agriculture and plant management so to recognise plant diseases by air technology gives an advancement and is very beneficial for the population. It is it is very labour intensive and requires considerable hard work and expertise in the field of agriculture to identify diseases in plants in a profound agricultural facility. Does the use of AI machine learning and image processing can be used to detect diseases which can which makes it a very good bet. This research propose a method of detecting plant diseases from their leaves images by using image processing and machine learning.

The objective of machine learning is primarily to comprehend data and create models that are useful in making of decision and predicting outcomes with accuracy which are based on a lot of training data. In this project we have taken parameters like leaf color, textures, marks, etc. A lot of image parameters and features are checked to have optimal accuracy in identifying plant leaf diseases. Normally, plants or plant products were checked visually by experts at various stages of production and goes through a physical test using chemicals which might be harmful for the plant. This requires a constant observation of the plants which is only possible in a big farm or a plant, as even tiny errors could majorly affect the quality. Detection of diseases based on visible symptoms at plant leaves is also a cost-effective way to observe large crop fields. This project can work through this with accuracy. Using AI machine learning an image processing is very time effective and easy to learn process which can be implemented. Our project helps to detect diseases time effectively in cost effectively.

Identify applicable funding agency here. If none, delete this.

II. LITERATURE REVIEW

Earlier, detection methods for plant diseases depended upon visual checking and recognition of symptoms. These methods are still used, but they depend highly on experts which are humans and often misses. This weakness of that method is clear to researchers, and this has led many people to test more advanced and accurate detection techniques. This brings to the rapid developments in sensor technology, like hyperspectral imaging which is now widely used, so too is infrared thermography for plant disease detection. Such non-destructive methods without the use of harmful chemicals, makes it possible for crops to be quickly and remotely monitored, making the early detection of disease an objectively easier than before. Scientists have investigated the use of these sensors for early diagnosis, which gives detailed analysis of the affected area. Medha Wyawahare and Shiroop Madiwalar gave diverse

image processing methods for plant disease detection. They conducted experiments on a dataset of 110 RGB images and examined their colours and textures. Classification features included mean and standard deviation of RGB and YCbCr channels, (GLCM) features, and mean and standard deviation of the image coiled with a Gabor filter. Sharath D. M. et al. created a Bacterial Blight detection system for pomegranate plants, using features such as colour, mean, homogeneity, standard deviation, variance, etc. They successfully created a system which predicts he disease levels in pomegranate fruits. Garima Shrestha et al. gave a CNN for plant disease detection, achieved an 88.8% accuracy in classifying 12 plant diseases using a dataset of high-resolution RGB images. The network consisted of three blocks of convolution and pooling layers, making it very intensive for computers. However, the F1 score of the model was low at 0.12 because of a higher number of false negative predictions.

III. ALGORITHMS USED

A. BPNN:

Its full form is Back Propagation Neural Network, It is a type of artificial network that used important learning Technique called backpropagation for training. It is a descent optimization algorithm that is used to minimize. The error between the predicted output and the actual target output . It reduces the error by adjusting weights of connections between neurons. It is basically used for task like regression and classification, like in image and speech recognition. It may require a significant length of data.

B. Random Forest:

This algorithm basically combines the result of lots of decision tree to make a more reliable overall decision. This help because it have many perspective and ways to proceed things in different styles so that can get different outcomes or better outcomes It is a ensemble learning method that basically make multiple decision trees while training. It improves accuracy

C. Decision Trees:

As name is self-explanatory , this algorithm makes a tree like structure to make decisions, It is used for classification and regression purposes, it is very famous algorithm in ML Where decision is represented by internal node, branch tells the outcome of decision, and final output is represented by leaf.. It is used to classify objects and patterns of a image

D. CNN:

CNN are those algorithms that are mostly used for work like recognizing for example image recognition, Because this algorithm automatically learns features like edges and textures and then it is able to process the more complex features like objects and patterns. CNN is also very good at doing segmentation,image classification, object detection. All the data Like images are preprocessed to maintain and increase the features. It includes steps like resizing, cropping, normalizing and augmenting so that we can increase the diversity. It is used to find patterns and similarities in the images.

E. K-Nearest Neighbours (KNN):

(KNN) K-Nearest Neighbours is a simple algorithm used in image classification. It is used to Classify an object based on the majority class of its k-nearest neighbours. When we apply this algorithm first matching images will be selected from images and these Pictures include some features in it like texture , colours.

IV. METHODOLGY PROPOSED

Disease Detection System comprises of the following steps:

A. Data Collection:

We used the Plant village dataset. The PlantVillage dataset contains 87000 RGB images having both healthy as well as the diseased plant leaves.For our project we used 15 classes form the 38 classes from this datset.

B. Classes are:

- Apple: Healthy, Scab, Black rot, Cedar apple rust
- Corn: Healthy, Cercospora leaf spot, Common rust, Northern Leaf Blight
- Grapes: Healthy, Black rot, Esca (Black Measles) Diseased: Leaf blight (Isariopsis)
- Potato: Healthy, Early blight Diseased: Late blight

C. Pre-Processing:

In this step we basically filter out the redundant data that may be in the raw form it's because there could be many data that may be incomplete and may cause problem in our algorithm. We will simply discard the redundant data and keep the consistent and reliable data so that our algorithm works perfectly and give the most accurate result.

D. Feature Extraction:

This stage is dedicated to the identification and extraction of the most pertinent attributes from the dataset. Through the elimination of irrelevant and redundant information, the dataset is refined to a form that is highly conducive for the subsequent application of classifiers. This process not only streamlines the dataset for improved computational efficiency but also enhances the discriminatory power of the selected attributes, thereby optimizing the overall performance of the classifier in subsequent stages of analysis.

E. Methodology

Getting a computer to understand pictures is tricky but important. Before we can teach it anything, we need to clean up the pictures. Imagine it's like turning a colorful photo into a simpler black and white version and then smoothing out any rough parts. Once we've cleaned up the main part of the picture, we highlight it in color again. We then look closely at the shape, texture, and color of what we're interested in, like a leaf. To do this, we use lines to measure how big and curvy the leaf is. After that, we look at the colors in the picture. We figure out how much green there is by comparing it to the total colors in one part. Anything that's not green is also noted. This careful cleaning and checking help us make sure the computer

understands the pictures better and gives us more accurate information. The GLCM (Gray-Level Co-occurrence Matrix) captures the spatial relationships among pixels in an image. Extracting texture features from GLCM is a traditional method widely utilized in computer vision. Following the extraction of the features (contrast, energy, homogeneity, dissimilarity) from all the images in the dataset, a subsequent task involves feature selection.

F. Feature Selection:

Feature selection constitutes a crucial phase in addressing machine learning challenges. Within this project, our approach involves selecting features based on their correlation with the target variable. The correlation matrix illustrates the relationships among variables in the apple dataset. Notably, the correlation between the features ”green part of leaf (F1)” and ”green part of leaf (F2)” is exceptionally high (1), signifying their interdependence. Consequently, one of these variables (F2) has been omitted.
For predicting apple diseases, certain features exhibit lower correlations, such as ”green channel mean,” ”red channel standard deviation,” ”blue channel standard deviation,” ”dissimilarity (f5),” and ”correlation (f8).” These variables are deemed less impactful for model development, and therefore, they have been removed. Following the feature selection process, the refined dataset is subsequently fed into machine learning classifiers to discern patterns within the data.

G. KNN Classifier:

The K-Nearest Neighbours (KNN) classifier is applicable to both classification and regression problems, although its primary strength lies in addressing classification challenges. KNN operates as a non-parametric algorithm, specifically functioning as a diagnostic of a distribution tree. When there is a lack of predefined assumptions about the distribution of data, it is categorized as non-parametric. In this context, the algorithm constructs a structure based on the attributes of the dataset. KNN is particularly useful for prediction tasks in scenarios where datasets do not conform to theoretical mathematical models. Notably, KNN does not necessitate any training data for subsequent processing, earning it the designation of a Slow Learning (SL) algorithm.

H. Random forest Classifier:

Random Forests are best known for their high acuuracy when there is a large amount of dataset present. The random forest classifier is used in many sector such as health , finance and task like image classification.Its versatile that’s why it’s so popular among various domains and is used widely. This is a useful tool for sorting the things into different groups. We make use of decision tree and each tree acts like a decision maker where each person looks at the different information. This team work reduces mistakes and makes prediction more accurate.. We use a method called K-fold cross-validation to make sure our predictions are accurate and fair.We use this information to train our team and the rest is used for checking whether they are doing good job or not. After we are done with the training of our team we check that how well they are doing by checking certain things such as precision recall , F1 score and acuuracy that how much accurate it is. We also try to find the mistakes that where we could be making an error by looking at the graphs and charts

RESULT

The table below shows us the result or the performance metrics for each model created for individual plants. It can be seen that the F1 scores of each plant closely aligns with the accuracy score which shows that our model is has accuracy of 93

Performance metric for all the models.

TABLE I
TABLE DATA

Plants	Accuracy	F1
Apple	0.90	0.95
Corn	0.96	0.97
Grapes	0.95	0.93
Potato	0.92	0.92

I. Treatment Recommendation System:

The Treatment Recommendation System is the important part of this application. It’s because when the disease is identified it gives us the suggestions for the treatment of the disease. It depends on various factors such as environment conditions , the identified disease and all of the historical data stored.

J. Disease Pattern Analysis:

When the disease is detected successfully then the system analyses the identified disease and analyze a pattern. This involves a detailed examination of the disease and the main goal is to extract the meaningful insights by the identified disease.

K. Historical Data Utilization:

The systems uses the historical data of the disease pattern by using this data the systems basically refines it’s treatment when the disease is identified . So basically this feature helps us to improve our treatment recommendation system as well as identifying the diseases.

L. Machine Learning Algorithms:

ML algorithms like reinforcement algorithms and decision trees are used to model the relationship between characteristics of disease, environmental factors, and outcomes of the treatment. These algorithms learn from data which is already available and from the feedback of users, updating the knowledge base of the system. This makes the Treatment Recommendation System more accurate and adaptive as it gets new cases and learns from their outcomes.

M. Personalized and Adaptive Recommendations:

The output of this application will be the solutions that would be given to the farmers so that they can save their crops early. This solution may include certain pesticides or insecticides that the farmer could use after using the application and finding the disease that the plant could have.

N. Weather-Based Crop Recommendation:

The application also has a crop recommendation system where the application gives a crop recommendation according to the weather. So that it's easy for the farmer to carry out the the farming of certain crops.

O. Real-Time Weather Data Integration:

The system stores the real time data such as the temperature , humidity , precipitation wind speed and many more and store these all into the database so that it can be used in the future references.

P. Seasonal and Climatic Considerations:

This feature basically takes the seasons and historic data of weather into consideration and then gives recommendations of crops. The systems basically identifies the relation between weather conditions and crop yield success.

Q. Crop Suitability Models:

The relationship between weather pattern and sustainability of various crops and whether they would be success or not. Machine learning models are used to establish this relationship. The models learn from the historical data such as the rainfall , sunlight , precipitation etc.

R. Personalized Recommendations:

This feature helps the farmer to get a personalized crop recommendation according the farmer's land , the type of soil and the region and what is the historical climate of that place as well the current climate. This all helps in giving a personalized recommendation of crops to the farmers.

S. Dynamic Adaptation:

When the weather or any conditions will be changed in the future the application provides information to the farmer so that the farmers can use certain precautions and farmers can adapt their crop choices for better farming.

In summary this features provides farmers with right knowledge of recommended plants according to real-time weather and recommends them with practices which are helpful for healthy plants. This features complements our project with other features.

T. Result And Discussion

The project provides a most valuable asset to the farmers by using our feature rich application it would make farming more easy for the farmers as we are adding a catalyst to the information that the farmers already know so farmers need not to know anything extra for using our application.

V. Conclusion:

We as a team have made this project to help farmers by identifying the diseases. So that farmers can early on catch the disease and can work on it. We have made the project such that it gives 97% of correct result. This would help the farmers in many ways. The UI is also made simple so that farmers can easily use it.

References

[1] S. D. Khirade and A. B. Patil explored "Plant Disease Detection Using Image Processing" at the 2015 International Conference on Computing Communication Control and Automation. [Link to the paper](https://doi.org/10.1109/ICCUBEA.2015.153).

[2] S. C. Madiwalar and M. V. Wyawahare conducted a "Plant disease identification: A comparative study" at the 2017 International Conference on Data Management, Analytics, and Innovation. [Link to the paper](https://doi.org/10.1109/ICDMAI.2017.8073478).

[3] P. Moghadam, D. Ward, E. Goan, S. Jayawardena, P. Sikka, and E. Hernandez presented "Plant Disease Detection Using Hyperspectral Imaging" at the 2017 International Conference on Digital Image Computing: Techniques and Applications. [Link to the paper](https://doi.org/10.1109/DICTA.2017.8227476).

[4] S. D.M., Akhilesh, S. A. Kumar, R. M.G., and P. C. shared insights on "Image based Plant Disease Detection in Pomegranate Plant for Bacterial Blight" at the 2019 International Conference on Communication and Signal Processing. [Link to the paper](https://doi.org/10.1109/ICCSP.2019.8698007).

[5] G. Shrestha, Deepsikha, M. Das, and N. Dey discussed "Plant Disease Detection Using CNN" at the 2020 IEEE Applied Signal Processing Conference. [Link to the paper](https://doi.org/10.1109/ASPCON49795.2020.9276722).

[6] Mohanty SP, Hughes DP, and Salathe´ M investigated "Using Deep Learning for Image-Based Plant Disease Detection" in their 2016 Frontiers in Plant Science publication. [Link to the paper](https://doi.org/10.3389/fpls.2016.01419).

[7] R. M. Haralick, K. Shanmugam, and I. Dinstein explored "Textural Features for Image Classification" in their 1973 IEEE Transactions on Systems, Man, and Cybernetics paper. [Link to the paper](https://doi.org/10.1109/TSMC.1973.4309314).

[8] L. Breiman introduced "Random Forests" in his 2001 Machine Learning paper. [Link to the paper](https://doi.org/10.1023/A:1010933404324).