ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 11 Issue 7 July 2024

Plant Disease Detection Technique Using Machine Learning

[1] Metharani. N, [2] Manoranjitham. T

[1] [2] SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India
Corresponding Author Email: [1] metharani53@gmail.com, [2] manorant@srmist.edu.in

Abstract*— Plant diseases pose a significant threat to agriculture, causing substantial crop and financial losses. Modern technologies enable precise monitoring of plant health and early disease identification. Employing image processing, particularly Convolutional Neural Network (CNN) techniques, allows accurate prediction of plant diseases. The aim is to provide an automated, reliable disease detection system, aiding professionals and farmers in timely action to prevent infections and reduce crop losses. Integrating cutting-edge technologies in agriculture holds vast potential to enhance profitability and production. The primary focus lies in developing an automated system proficient in analyzing plant images to detect disease symptoms and classify plants as healthy or disease affected. The system aims to simplify plant disease diagnostics for farmers, providing essential information about leaf name, integrity, and life span. The method aims to empower farmers by enabling easy identification of plant diseases, providing essential details like disease name, accuracy level, and life span. The CNN model accurately gauges the systems accuracy level. It further streamlines the process by offering a unified solution through a user-friendly web application, eliminating the need for separate interventions for affected leaves. The system saves farmers time by delivering crucial information directly. The Proposed web application proves to be a comprehensive solution, eliminating the need for farmers to search for separate interventions for affected leaves. The machine learning model exhibits a noteworthy accuracy of 96.67%, emphasizing its proficiency in making correct predictions for the given task.

Index Terms*— Agriculture, Plant diseases detection, Image processing, Machine Learning.

I. INTRODUCTION

Plant sickness detection is a vital element of agriculture that plays a vital position in making sure crop fitness, minimizing losses, and keeping meals security. Traditional strategies of disorder detection regularly rely upon manual inspection and visible commentary, which may be time-consuming, subjective, and liable to human errors. With the advancements in machine gaining knowledge of and synthetic intelligence, there's an opportunity to increase computerized structures that could appropriately come across and classify plant illnesses.

A. Objective

This paper aims to explore the application of machine learning techniques for the accurate sensing and classification of plant diseases. The primary focus is to develop an automated system capable of analysing plant images, detecting disease symptoms, and classifying plants as healthy or disease affected.

- Plant disease diagnostics helps in identifying the integrity and life span of the plant leaves without difficulty for the farmers to know the name, integrity and life span.
- It provides insights into the benefits and applications of developed plant disease detection systems, including early disease detection, precision agricultural practices, and effective disease management strategies.

B. Literature Survey

Nabobi et al. [5] (2022) had examined on the detection of plant leaf diseases, highlighting the limitations of manual visual inspection. They addressed these challenges by utilizing image processing and artificial intelligence algorithms. Their research was organized into three parts: the first part explored various image processing and AI techniques, the second part evaluated different frameworks and their accuracy, and the third part provided a detailed explanation of the disease detection and classification performance. Through their work, they contributed valuable insights into automated approaches for disease recognition, promising to advance agricultural practices and crop management.

Pranesh et al. [6] (2021) had presented an intelligent and effective method for crop disease detection, employing computer vision and machine learning techniques. Their system demonstrates an impressive 93% accuracy in detecting 20 diseases across 5 common plants. To accomplish this, the researchers utilized the Plant Village dataset, a publicly available collection of 87,000 RGB images of healthy and diseased plant leaves. The algorithm was rigorously tested across 25 different classes to validate its performance.

Rahul et al. [7] (2022) had described the significance of India's agriculture industry. Their research study demonstrated methods for identifying plant illnesses using image processing in leaves in an effort to provide a solution to the query of whether the grains and crops are chemical-free and healthy.

Kelothu et al. [8] (2023) had utilized a dataset comprising of 9127 images to perform a study on the detection of plant diseases. The study employed VGG16, VGG19 and CNN

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 11 Issue 7 July 2024

models. The outcome of the study favours VGG16 model with an accuracy of 0.96. Their work stated both the advantages and limitations of deep learning like automatic feature extraction and scalability, overfitting, need for sophisticated computational resources, and high-quality annotated data.

Bharath et al. [9] (2020) had carried out smart farming systems to improve agricultural production quality and quantity by addressing plant leaf disease, which is an important threat to food safety. They employed deep learning and computer vision for disease diagnosis in plant leaves. CNNs have been successful in classifying various plant leaf diseases using neuron-wise and layer-wise visualization methods.

II. PROCEDURE FOR ALGORITHMS

A. *Patel et al. [1] (2020)* had proposed a solution for identifying and categorizing leaf diseases in which deep learning and image processing methods are frequently utilized. These methods include image preprocessing, CNN, MRCNN, and FRCNN

B. Monigari et al. [2] (2021) had presented an innovative solution for plant disease detection using leaf photographs. Their approach involved utilizing image processing techniques like acquisition, filtering, segmentation, feature extraction, and classification.

C. *Figures* Sai et al. *[3] (2021)* had developed a novel disease detection system was introduced, utilizing color and

Texture-based analysis with deep learning techniques. The system employs Dense Net for image classification and 1D-CNN for semantic segmentation to distinguish healthy and defective leaf pixels.

III. FACTORS RESPONSIBLE FOR PLANT DISEASES

A wide range of agricultural diseases can arise at various stages of plant development and harm the plant's growth, which can have a negative impact on overall crop production. Plant diseases are caused by a variety of conditions at various phases of plant development. As summarized in crop disease-causing variables are categorized into two: biotic factors and abiotic factors. Biotic factors such as viruses, fungi, bacteria, mites, and slugs emerge as a result of microbial infection in plants, whereas abiotic variables such as water, temperature, irradiation, and nutritional deprivation

Damage plant growth.

Accuracy The accuracy score of a model, often known as accuracy, is a classification statistic in DL and ML techniques that represents the proportion of correct predictions made by the model.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

IV. PERFORMANCE EVALUATION OF PLANT DISEASE DETECTION AND CLASSIFICATION

A. Sensitivity

$$\text{Sensitivity} = \frac{TP}{(TP + FN)}$$

B. Specificity

$$\text{Specificity} = \frac{TN}{(TN + FP)}$$

Where TP, TN, FP, and FN are for True Positive, True Negative, False Positive, and False Negative respectively.

In addition to the above performance evaluation metrics, determining the performances of the plant leaf and /or crop disease detection and classification performance should be evaluated in terms of the Logarithmic Loss (including the training loss, validation loss, and testing loss) and Area under Curve (AUC) metrics accordingly. As indicated in designing an effective logarithmic loss function is mandatory for the robust performance of plant leaf and /or crop disease detection and classification models.

V. PLANT DISEASE DETECTION AND CLASSIFICATION TECHNIQUES

Machine learning (ML) techniques or algorithms

A. The NB Technique

It is a probabilistic classifier variation built on the NB classifier idea. It is assumed that the patterns' prior probabilities are known to exist and that the class labels are assigned their posterior probabilities. In light of this premise, the maximum likelihood values of the data that belong to a specific class label are computed using the posterior probability. It is calculated by applying Baye's theorem to the product of each feature's conditional probability. This theory works fairly well in many classification problems, even though it usually does not hold in a real-life setting.

Fig. 5.1 some sample plant leaf images with different diseases from the Plant Village dataset

B. The KNN Technique

It is a nonparametric, supervised ML technique commonly applied to pattern recognition. It is predicated on the nearest neighbor rule, which is applied in ML applications to classify data. This method involves training the test pattern using the classifier, and then classifying the test pattern according to how similar it is to each training pattern. The KNN classifier produces a class membership value that it is a member of. The object is allocated to the most widely used class labels among its k-nearest neighbors based on the plurality vote of its neighbors. It functions similarly to an instance-based learning model, with locally approximated operations and distinct computations throughout the classification process.

C. The DT Technique

In supervised learning, it is a supervised classification and regression algorithm that creates classifiers by splitting the data into multiple smaller groups (tree structure) according to which division creates the greater disproportion One of the often-utilized attribute selection metrics that are frequently

Employed as disparity measurements is the Gini index, also known as entropy. One benefit of this method is that it may make it simple for humans to interpret the results. If the tree could have trained without being limited by its depth, a DT may generate very little training error. Several DT variations, including ID3, C4.5, and CART, are widely employed in various data mining and ML applications.

D. The SVM Technique

The separating hyperplane defines this supervised ML Classifier. In high-dimensional space, this technique determines the ideal hyperplane that maximizes the margin between the data points of the two classes. The kernel tricks that are helpful for nonlinear classification are an attribute of SVM. Obtaining more distinct features in the high-dimensional feature space is highly anticipated.

E. The RF Technique

It is a collection of learning techniques for randomized DT classifiers. During training, it is run by building several DTs. Based on each classification tree's vote, the class labels of the testing dataset are calculated. The class labels with the highest votes by the classification trees determine the classifier's final result. This approach attempts to produce an uncorrelated forest of trees that will predict performance more accurately than that of the individual tree by using bagging and randomness of features during the building of each tree.

F. Literature Survey

Nabobi et al. [5] (2022) had examined on thedetection of plant leaf diseases, highlighting the limitations of manual visual inspection. They addressed these challenges by utilizing image processing and artificial intelligence algorithms. Their research was organized into three parts: the first part explored various image processing and AI techniques, the second part evaluated different frameworks and their accuracy, and the third part provided a detailed explanation of the disease detection and classification performance. Through their work, they contributed valuable insights into automated approaches for disease recognition, promising to advance agricultural practices and crop management.

Pranesh et al. [6] (2021) had presented an intelligent and effective method for crop disease detection, employing computer vision and machine learning techniques. Their system demonstrates an impressive 93% accuracy in detecting 20 diseases across 5 common plants. To accomplish this, the researchers utilized the Plant Village dataset, a publicly available collection of 87,000 RGB images of healthy and diseased plant leaves. The algorithm was rigorously tested across 25 different classes to validate its performance.

Rahul et al. [7] (2022) had described the significance of India's agriculture industry. Their research study demonstrated methods for identifying plant illnesses using image processing in leaves in an effort to provide a solution to the query of whether the grains and crops are chemical-free and healthy.

Kelothu et al. [8] (2023) had utilized a dataset comprising of 9127 images to perform a study on the detection of plant diseases. The study employed VGG16, VGG19 and CNN models. The outcome of the study favors VGG16 model with an accuracy of 0.96. Their work stated both the advantages and limitations of deep learning like automatic feature extraction and scalability, overfitting, need for sophisticated computational resources, and high-quality annotated data.

Bharath et al. [9] (2020) had carried out smart farming systems to improve agricultural production quality and quantity by addressing plant leaf disease, which is an important threat to food safety. They employed deep learning and computer vision for disease diagnosis in plant leaves. CNNs have been successful in classifying various plant leaf diseases using neuron-wise and layer-wise visualization methods.

VI. EXISTING SYSTEM

Previous studies in this area used the Chan – Vese (CV) algorithm for detecting plant illnesses. Plants can showcase illness by considerable versions of their signs, such as color changes, spots, lesions, and deformations. These variations make it tough to develop strong system learning models that could correctly locate and classify illnesses throughout distinct plant species and environmental situations. The algorithm utilizes the standards of lively contours and level set techniques to iteratively optimize the segmentation based on a predefined electricity useful. The extracted features are used to classify the detected sicknesses. Machine gaining knowledge of strategies inclusive of SVM, decision trees, or neural networks can be hired for this cause. The class version is educated by the use of classified statistics that accomplice

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 11 Issue 7 July 2024

particular features with known diseases. The extracted features are used to classify the detected sicknesses. Machine getting to know strategies along with SVM, selection timber, or neural networks can be employed for this reason. The type of model is skilled the usage of categorized facts that accomplice particular capabilities with recognized diseases.

A. Drawbacks of existing system

In most of the existing systems for plant sickness detection the use of gadget learning regularly war with generalizing their performance to unseen facts and different environmental conditions. The models skilled on particular datasets and may have difficulty detecting illnesses in flowers that exhibit versions in symptoms or come from exclusive regions or climates.

Some current structures may also require specialized hardware or technical understanding to set up and function successfully. This restricted accessibility can pose obstacles for farmers, agricultural workers, or people in aid-confined settings who may also advantage from such disease detection systems.

Most existing structures depend on outstanding snap shots for accurate disease detection. However, in actual-international scenarios, taking pictures wonderful pix might not always be feasible because of elements which includes lighting situations, digicam barriers, or photograph noise. This dependency on terrific photographs can restrict the sensible applicability of those systems.

VII. PROPOSED METHOD

To overcome the issues of the prevailing methodologies, machine regression and classification algorithms may be used. The CNN considering the system learning model. This System Architecture is shown in Figure 1. The CNN model shows the accuracy level properly and the set of rules additionally enables to distinguish between the affected leaves and unaffected leaves. The proposed system's intention is to overcome the shortcomings of the existing one. The gadget's requirements were advanced primarily based on input from prior metrics device users in addition to flaws that had been previously documented. The following dreams for the cautioned gadget are listed:

- Farmers can easily recognise the name of the plant disease, accuracy level and lifetime.
- There is no need to look for a solution for the affected leaves separately since the web application shows the ailment and prevention.
- It saves the time for the farmers by providing the required information from their whereabouts without going to the agriculture office to test.

A. System Architecture

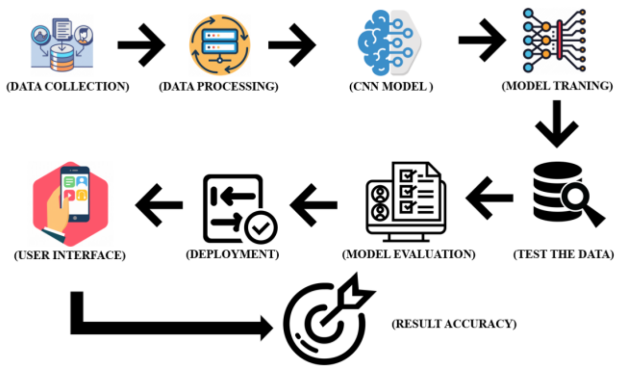

Figure 7. 1. System Architecture

B. Data collection

The dataset would not encompass any duplicate rows or lacking cells; hence no education is vital to dispose of duplicates and impute missing data.

C. Data Pre-Processing

Pre-processing involves cleaning raw information by amassing statistics within the real world and then changing them into error-free datasets. This part of the conversation is known as statistics pre-processing," while a specific operation is performed to transform the data into a free information set that is barely decreased and appropriate for evaluation.

The following are some examples of crucial pre-processing strategies:

- Missing data handling
- Dimensionality Reduction
- Dimensionality Reduction
- Data Partitioning.

If the set of rules' facts is simply too large to ever be processed it will very quickly become an extra complex feature set. To reduce the complexity, the feature selection, which specifies a subset of the authentic features. Relevant records from the statistics must be added to the selected functions to complete a better task. The use of this truncated representation rather than the entire starting records curtails the difficulty in handling the complexity of the dataset. The reduced description of complicated record structures is an advantage of the function.

Withdrawals: This avoids clustering estimates with the intention to over-converge for guidance and poor convergence for other activities. Highlight extraction is a common time period for techniques of combining factors to resolve those troubles, but to build a successful version, the use of more advantageous element extraction display records extra accurately. When testing with a couple of factors, it calls for loads of memory and processing electricity.

D. Model Selection

The process of selecting the final AI model from a list to train for plant disease from images is called model selection.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 11 Issue 7 July 2024

Model selection is a cycle that can be used to compare different models and unique models set with different model hyperparameters, so, it is used a CNN to train the data set.

E. CNN

A CNN is a deep learning algorithm worn for image recognition and computer vision tasks. It integrates multiple layers, including convolutional layers, pooling layers, and fully connected layers. Convolutional layers detect local patterns and features in images, while pooling layers reduce spatial dimensions and improve robustness to fetching variations. Activation functions introduce non-linearities, such as ReLU and sigmoid functions, and fully connected layers make predictions based on learned

```
history = model.fit_generator(
    aug.flow(x_train, y_train, batch_size=BS),
    validation_data=(x_test, y_test),
    steps_per_epoch=len(x_train) // BS,
    epochs=EPOCHS, verbose=1
    )
```

Figure 7.5. Training and testing data.

F. Evaluation of the Proposed Model

A classification model's performance is measured using a variety of metrics, including precision, F1-score, recall, AUC, confusion matrix, cross-validation, ROC curve, and accuracy. These metrics can be used to identify the strengths and weaknesses of a model, and to select the right metric for the specific application. By understanding how these metrics work, it is possible to improve the performance of a model.

G. Performance metrics

The performance metrics provide insights into the model's accuracy, erudition to identify diseased samples, and its robustness in handling imbalanced datasets. It is important to consider multiple metrics and assess them collectively to get a comprehensive understanding of the CNN algorithm's performance in plant disease detection.

- Convolution operation: The output feature map size can be calculated using the formula: ((input_size - kernel_size + 2 * padding) / stride) + 1.
- Max Pooling operation: The output size is calculated similarly to the convolution operation using the formula above.
- Softmax activation: The softmax function is defined as softmax(x) = e^x / sum(e^x) for each class, where e is the base of the natural logarithm.

VIII. RESULTS AND DISCUSSION

The output of the performance metrics of the model:

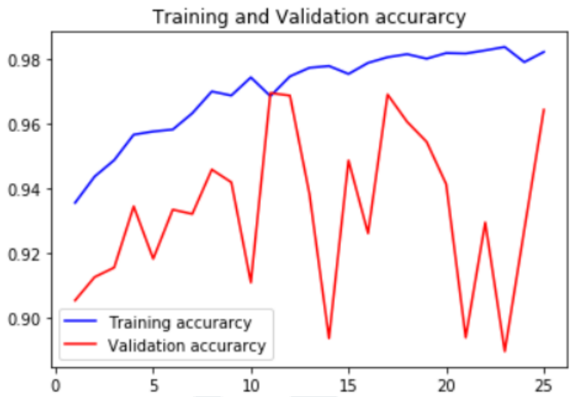

Figure 8.1. Accuracy of CNN model

Training accuracy is a measure of how well a machine learning model performs on the training dataset. It is calculated by comparing the model's predicted labels with the true labels in the training dataset. Great precision while training indicates that the model is able to fit the training data well and learn the underlying patterns. The Figure 3 represents Training and validation accuracy.

Utilized a validation dataset to determine the validation accuracy.

A. Model accuracy

```
scores = model.evaluate(x_test, y_test)

print(f"Test Accuracy: {scores[1]*100}")

Test Accuracy: 96.67230919129554
```

Figure 8.2. Model accuracy.

Model accuracy is a metric used to measure how well a machine learning model performs on a given task. It represents the proportion of correct predictions made by the model out of all predictions. Model accuracy refers the ratio of the accurate prediction to the total number of predictions made an accuracy of 96.67 using the model which was shown in Figure 7.4.

B. Deployment of the model

The web interface is developed using Flask, a popular web application framework written in Python. The integration of the model and the web page is achieved through the use of the pickle library.

For example, a specimen's chemical composition need not be reported if the main purpose of a paper is to introduce a new measurement technique. Authors should expect to be challenged by reviewers if the results are not supported by adequate data and critical details.

IX. CONCLUSION

The paper utilized a CNN algorithm to accurately predict plant disease. A web application is created utilizing loads of technologies, including python, HTML, CSS, JavaScript, sklearn, matplot, numpy, pandas, Flask and other libraries, to make the consumer interface (UI) extra accessible and easy. With the deployment of the proposed version, farmers can access via web application easily which is user friendly.

The application lists out the name of the plant disease, its lifetime, preventive and treatment measures to the farmers after they had uploaded the picture of the affected leaves. Finally, by detecting illnesses at an early stage, farmers can take appropriate remedial steps to curb sicknesses and crop losses.

REFERENCES

[1] Patel, Sneha, Jaliya, Pranay: A Survey on Plant Leaf Disease Detection. International Journal for Modern Trends in Science and Technology, 2020, Vol. 6, pp. 129-134.

[2] Monigari, Vaishnavi. Plant Leaf Disease Prediction. International Journal for Research in Applied Science and Engineering Technology.2021. Vol. 9. pp. 1295-1305.

[3] Reddy, B, Neeraja, S.: Plant leaf disease classification and damage detection system using deep learning models. Multimed Tools Appl 2022. Vol. 81, pp. 24021–24040.

[4] Li, Lili, Zhang, Shujuan, Wang, Bin. Plant Disease Detection and Classification by Deep Learning—A Review. IEEE Access.2021. PP. 1-1.

[5] Hasan, Nabobi, Md, Mustavi.: Plant Leaf Disease Detection Using Image Processing: A Comprehensive Review. Malaysian Journal of Science and Advanced Technology. 2022. pp. 174-182.

[6] Pranesh, Atharva Karwande, Tejas Kolhe.: Plant Disease Detection Using Image Processing and Machine Learning. 2021.

[7] Kundu, R, Chauhan, U, Chauhan, S, P, S.: Plant Leaf Disease Detection using Image Processing. Proceedings of 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM), 2022. pp. 393-396.

[8] Moupojou, E.: FieldPlant: A Dataset of Field Plant Images for Plant Disease Detection and Classification with Deep Learning. Proceedings of IEEE Access.2023. vol. 11. pp. 35398-35410.

[9] Shivaprasad, K, Wadhawan, A.: Deep Learning-based Plant Leaf Disease Detection. Proceedings of 7th International Conference on Intelligent Computing and Control Systems (ICICCS).2023. pp. 360-365.

[10] Bharath, S, Vishal, K, Pavithran, P, Malathi, T.: Detection of Plant Leaf Diseases using CNN. International Research Journal of Engineering and Technology (IRJET).2020. pp. 1-1.