International Journal of Scientific Research & Engineering Trends
Volume 10, Issue 5, Sept-Oct-2024, ISSN (Online): 2395-566X

Agri Shield: Identify Plant Disease Using Machine Learning

Aditya Bathre, Aajinkya Ingalkar, Awanish Srivastava, Anurag Patel
Computer Science and Business System
Rajiv Gandhi Prodyogiki Vishwavidhalay

Abstract- This paper introduces Agri Shield, an innovative approach using machine learning, particularly convolutional neural networks, for predicting plant diseases and recommending sustainable individualized remedies. Agri Shield embodies early-stage disease detection with ecologically friendly solutions, making it easier for farmers and plant enthusiasts to care for plants, as such information would be sourced from a multiplicity of sources. The proposed system is able to detect 20 different diseases of 5 common plants with 93% accuracy.

Index Terms- Digital image processing, Plant health, machine learning, CNN, disease prediction, personalized treatment, sustainable agriculture.

I. INTRODUCTION

Plant diseases have incessantly remained one of the primary threats towards the yield and quality of crops in this highly dynamic world of agriculture today. Most of the current methods still rely on either the visual inspection or professional analysis that takes considerable time and expense to determine such diseases in crops. It is not very practical for a mass scale implementation, particularly in small and medium scale farming operations. While the world continues to be increasingly populated and stressed through food production, the whole world needs urgent answers that will lead to early and accurate disease detection while sustaining agriculture.

Agri Shield presents a challenge with the latest advanced cutting-edge solution using algorithms from machine learning, including Convolutional Neural Networks, predicting diseases in plants more precisely. More than just indicating the beginning of disease, Agri Shield takes the opportunity to hint at something that is so personalized, ecological, and amiable to individual cases of health issues in plants. It implements environmental sustainability principles merged with advanced technologies to give insight into ways of managing disease for farmers and plant enthusiasts.

The system can recognize 20 diseases on five common species with a very impressive record of 93% and makes the identification process smooth and helps in recommending treatment to the diseased plants. Using machine learning models for prediction data obtained from various sources and hence does disease prediction and prevention on crops hence protecting against losses while also minimizing the impacts of agricultural practices on the environment. This project bridges the gap between state-of-the-art AI technology and actionable, scalable solutions for modern agriculture with a focus on productivity and sustainability.

This paper documents the development and deployment of AgriShield including its machine learning architecture, performance evaluation, and the promise it holds to revolutionise plant disease management by providing farmers with tools to carry out eco-friendly farming practices in an efficient manner.
Thesis Statement

This research will further the development and application of Agri Shield: a machine learning-driven system using Convolutional Neural Networks (CNNs) in providing high accuracy predictions of plant diseases with sustainable and individualized recommendations. The study intends to demonstrate how Agri Shield can change the face of plant disease management through early detection and solutions that are not only ecological-friendly but also lead to crop yield increase while promoting sustainable agriculture practices.

II. REVIEW OF LITERATURE

Singh and others. (2021) [1] built up a machine learning methodology-based model for disease detection and classification of potato plant leaves. In the proposed methodology, the following activities are utilized: image collection, image preprocessing, segmentation, feature extraction, and the use of SVM with a linear kernel for classification. The results show 95.99% accuracy and an F1-score of 96.16%, thus illustrating the efficacy of machine learning techniques in plant disease classification.

© 2024 IJSRET

Devi and others. in 2019 [2] proposed an IoT-enabled system for the effective detection and classification of plant diseases.

They experimented the system with 80 images of hill banana plants targeting two diseases: bunchy top and Sigatoka leaf spot. In their system, they used K-means clustering and GLCM features, which got an accuracy of almost 99% by using an RFC, suggesting the sight of how IoT can be beneficial to the management of agricultural diseases.

Sharma and others. (2020) [3] established the performance analysis of CNN models for detecting plant diseases on the basis of image segmentation and reported nearly fivefold increase in performance metrics, from 42.3% to 98.6%, while trained by the segmented images based S-CNN model rather than a lower performance full images training-based CNN model and emphasized that such improvements are highly dependent on the quality of the pre-processed image.

Durmus and others. (2017) [4] investigated tomato leaf disease detection with deep learning. A deep CNN model was trained using the error-correcting output code with the support of the Plant Village Dataset, with an accuracy of 98.8%. It has brought into consideration that deep learning techniques really do a great job in classification - classifying different types of diseases that occur on the tomato plant.

Garima Shrestha and others. used CNNs to identify plant diseases [5]. The proposed model was applied on classifying 12 plant diseases with the accuracy of 88.80% using the dataset of 3000 high-resolution RGB images.

The proposed CNN consisted of three convolution and pooling blocks, which made it computationally intensive. The F1 score of the model was 0.12, meaning there was a very high false-negative prediction rate.

III. METHODOLOGY

1. Dataset Collection

The dataset used in this study is the "New Plant Diseases Dataset" in Kaggle, containing about 87,000 RGB images of healthy and diseased crop leaves, categorized into 38 different classes. The classes used involve several species of plants and various diseases they may encounter.

The train/validate split at an 80/20 ratio was followed to ensure that the directory structure is maintained for automatic label inference. Another file, a directory containing 32 test images, is created specifically for prediction purposes.

Thus, real-world testing occurs after the phases of training and validation to test the performance of the model. So, proper evaluation and robust model development is ensured.

Table 1: Plant Diseases Dataset

Plant	Disease Name	No. of Images
Apple	Healthy Diseased Scab Diseased: Black rot Diseased: Cedar apple rust	2008 2016 1987 1760
Corn	Healthy Diseased: Cercosporin leaf spot Diseased: Common rust Diseased: Northern Leaf Blight	1859 1642 1907 1908
Grapes	Healthy Diseased: Black rot Diseased: Esca (Black Measles) Diseased: Leaf blight (Isariopsis)	1692 1888 1920 1722
Potato	Healthy Diseased: Early blight Diseased: Late blight	1824 1939 1939
Tomato	Healthy Diseased: Bacterial spot Diseased: Early blight Diseased: Late blight Diseased: Leaf Mold Diseased: Septoria leaf spot Diseased: Two-spotted Diseased: Target Spot Diseased: Yellow Leaf Curl Virus Diseased: Tomato mosaic virus	1926 1702 1920 1851 1882 1745 1741 1827 1961 1790

2. Data Preprocessing

Tensor flow and Keras is used to preprocess the dataset in such a way that both the training set elements and the validation set elements are prepared in an appropriate order for a CNN model. The labels are assigned as "inferred", so the

© 2024 IJSRET

labels depend upon the directory structure of the dataset, where in every sub-folder a specific kind of plant disease is included. This label inference technique saves a lot of efforts for preparing the appropriate data.

In this situation, the label_mode is categorized as "categorical" as there exist more than two classes, thereby classification nature being multi-class classification. This mode of categorical encoding transforms the labels to one-hot encoded vectors, where each label appears as a binary vector; the active bit in such cases turns to be the correct class.

The color_mode is set to "rgb", thus leaving the images in their three-channel color information, which is important for the identification of the subtle differences in plant leaf colors as in the signs of diseases.

Let's set the batch_size to 32, indicating that the model reads through 32 images before updating its weights. Optimizing this memory usage ensures that training goes efficiently even though the model still learns appropriately.

The image_size is set standard to 128x128 pixels. This makes sure all images are uniformly in the same dimensions, so the neural network processes them accordingly.

Finally, shuffle is set to True, and thus images are presented in random order at each epoch for the model, preventing any possible bias or effects due to an ordering of images and enhancing the model's generalization and accuracy.

Figure 1: Training and Validation

3. Model Architecture

A Convolutional Neural Network (CNN) architecture was selected for plant disease prediction due to its efficiency in image recognition tasks.

Convolutional Neural Networks are widely applied on tasks such as images and text. They are used on various applications, including classification, object detection and natural language processing, among others. CNNs follow a structured approach with no human intervention to learn features from raw input. They are divided into three foundational layers: Convolution, Pooling, and Fully Connected (FC) layers.

Convolution Layer (CONV)

The convolution layer forms an input image as a 2D matrix of pixel values; the filter scans across the input image to identify edges, textures, or color patterns important for the features of an image, usually 3x3 or 5x5. In this context, stride refers to how many pixels the filter would shift after each step. This shift of the filter across the image is how the filter moves and thus ends up picking important features that could be then represented as a feature map output. This process can extract more than just one feature using multiple filters.

Pooling Layer (POOL)

Pooling is a downsampling operation, reducing the dimensionality of feature maps, which makes the model more efficient and less likely to overfit. The two main types are: Max Pooling: It chooses the maximum value within the pool of the sub-region from the feature map. Average Pooling: This calculates the average value within the pool. Pooling layers decrease the size of the feature map but retain most crucial features.

Fully Connected Layer (FC)

The output of the previous layers, usually after pooling, is flattened to a 1D vector in the fully connected layer and passed to a regular neural net. In this layer, each neuron gets connected to all of those from the previous layer. This is where the model learns to make the final predictions based on the features extracted.

Dropout layers are also used extensively in fully connected layers to avoid overfitting. In this case, at the time of training, dropout randomly drops out some neurons and helps the model generalize more.

This combination of convolution, pooling, and fully connected layers helps CNNs learn complex patterns from images, with reduced computational complexity and increased accuracy.

4. Model Training

The CNN architecture contains five convolutional and max pooling layers. After applying the layers, flattening is carried out to reduce the dimensionality of 2D feature maps into a vector that is acceptable to a fully connected neural network. Before the network, dropout was applied with a drop rate of 0.25. This prevented the problem of overfitting by randomly deactivating 25% of the neurons while training.

The model has several hidden layers with 1,500 units. The hidden layers help the model in learning complex data patterns.

© 2024 IJSRET

International Journal of Scientific Research & Engineering Trends
Volume 10, Issue 5, Sept-Oct-2024, ISSN (Online): 2395-566X

To avoid overfitting dropout is used with a rate of 0.4.

For the output layer softmax activation function is used. It will first compute the probability distribution over all of the 38 classes, then it picks the class with maximum value.

The model trains on both training_set and validation_set for 10 epochs The results are as follows: -Training accuracy is 0.99, showing very good performance on the training data. Validation accuracy is 0.96, which represents good generalization to unseen data.

5. Model Evaluation

To evaluate the mode several evaluation metrics were used Accuracy: overall accuracy was calculated as the ratio of correctly classified images to the total number of images. The visualization graph is as follow

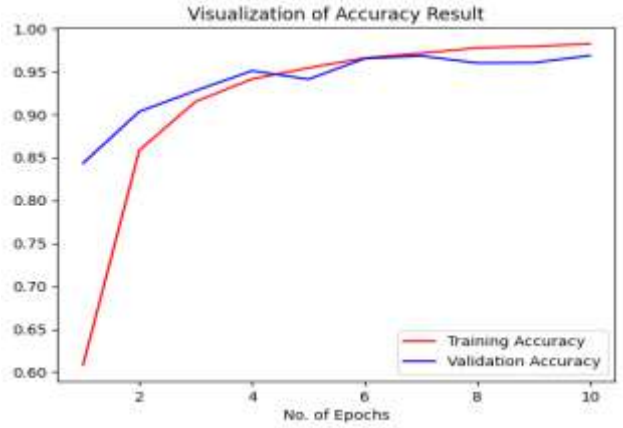

Figure 2: Accuracy Result

Precision, Recall, and F1-score: These metrics were used to evaluate the model's ability to classify diseased vs. healthy plants, considering both false positives and false negatives.

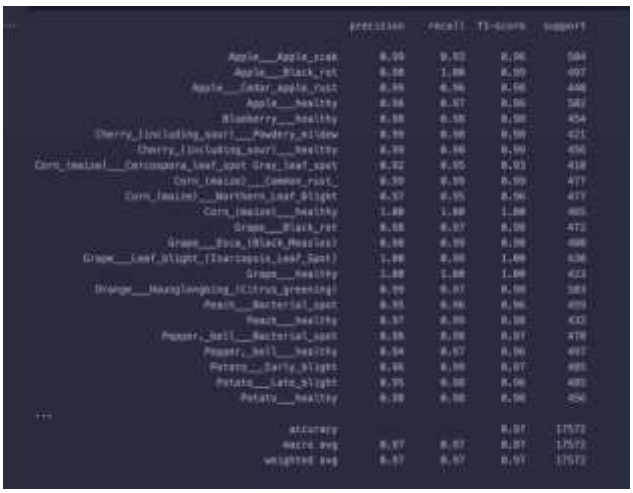

Figure 3: Precision, Recall, and F1-score

Confusion Matrix: A confusion matrix was used to visualize the model's predictions and better understand misclassifications.

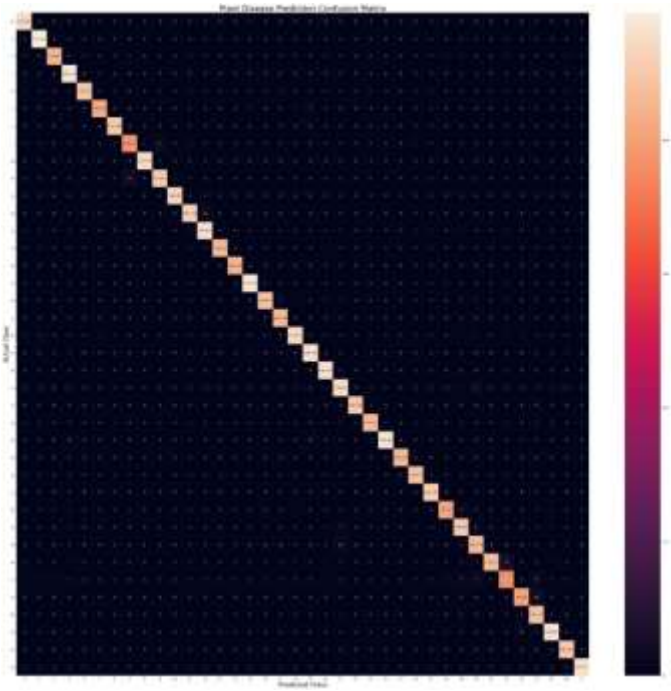

Figure 4: Confusion Matrix

IV. CONCLUSION AND FUTURE SCOPE

Agriculture remains a crucial sector in India, with a substantial portion of the population dependent on it for their livelihood. Consequently, identifying crop diseases is vital for economic growth. This study focuses on detecting and classifying various diseases affecting important crops.

The research employs a Convolutional Neural Network (CNN) model to categorize leaf diseases in these crops using data from the Plant Village dataset. A basic CNN architecture with minimal layers was implemented to classify the leaf diseases of tomatoes, potatoes, and bell peppers into distinct categories.

Future Scope

Future work may also involve experimenting (24x7) monitoring System that can help farmers to manage various crops simultaneously without any difficulty.

It can also involve experiments with more modern architectures to improve the model's functionality on the train set. As a result, farmers can utilize the aforementioned model as a decision-making tool to assist and support them in recognizing diseases that can affect their crops.

© 2024 IJSRET

REFERENCES

1. Singh, A. and Kaur, H., 2021. Potato plant leaves disease detection and classification using machine learning methodologies. In IOP Conference Series: Materials Science and Engineering (Vol. 1022, No. 1, p. 012121). IOP Publishing.
2. Devi, R.D., Nandhini, S.A., Hemalatha, R. and Radha, S., 2019, March. IoT enabled efficient detection and classification of plant diseases for agricultural applications. In 2019 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET) (pp. 447- 451). IEEE.
3. Sharma, P., Berwal, Y.P.S. and Ghai, W., 2020. Performance analysis of Deep Learning CNN models for disease detection in plants using image segmentation. Information Processing in Agriculture, 7(4), pp.566-574.
4. We Durmuş, H., Güneş, E.O. and Kırcı, M., 2017, August. Disease detection on the leaves of the tomato plants by using Deep Learning. In 2017 6th international conference on agro-geoinformatics (pp. 1-5). IEEE.
5. G. Shrestha, Deepsikha, M. Das and N. Dey, "Plant Disease Detection Using CNN," 2020 IEEE Applied Signal Processing Conference (ASPCON), 2020, pp. 109-113, doi: 10.1109/ASPCON49795.2020.9276722.

© 2024 IJSRET