LEAF DISEASE DETECTION USING MACHINE LEARNING

[1]Mamilla Maheswari, [2]P. Sai Likitha, [3]D.Amrutha, [4]Gvs.Bhavya Sri, [5]P.Shivani
[1]Assistant Professor, [2,3,4,5]UG Students, [1,2,3,4,5]Dept. of CSE (AI & ML), Malla Reddy Engineering College for Women (Autonomous), Hyderabad, India. E-Mail: munnellimaheswari@gmail.com

ABSTRACT

Although leaf diseases pose a serious threat to food security, it is still difficult in many parts of the world to quickly identify them because of the absence of the necessary foundation. Impressive results have been achieved with the development of precise techniques in the classification of leaf-based images. The information obtained from the data sets are used in this work to differentiate between healthy and sick leaves using Random Forest. Our proposed article includes several implementation phases, such as dataset preparation, extraction of features, classifier learning, and classifications. To distinguish between photos of damaged and healthy leaves, a collective Random Forest model is trained using the created datasets of infected and healthy leaves. To identify the features of an image, we employ the histogram of an Oriented Gradient (HOG). Overall, we can clearly identify illnesses that are present in plants on a massive scale by utilising machine learning techniques to train the vast data sets that are publically accessible.

INTRODUCTION

Because illness signs first develop on the leaves, plants are particularly susceptible to disease. Farmers should think about keeping an eye on their crops so they can minimise losses because plant diseases have negative effects on both the economy and the environment.

The key objective of the suggested system is not only to detect diseases of plants using image processing technology, but additionally to instruct user 12 (farmer) to utilise an application for mobile devices in which he can upload the image and receive information about the type of virus infection along with a recommendation of necessary pesticides. The latest technologies, namely image processing, have entered the agricultural area as a result of its digitalization. Because of this, the image processing technique is used to create our system, which is intended to be automated. Due of its advantages, which will be covered in the next sections, monoculture is preferred by farmers. Monoculture farming is therefore becoming more significant in modern times. This project would be helpful for the framers and will represent a sort of guidance.

LITERATURE SURVEY

Using back propagation to classify leaf diseases Colour and texture are employed as characteristics in neural networks as they work primarily on the segmentation of defective areas. In this particular case, a neural network-based classifier was used to perform the classification. The classification is 97.30% effective and transforms to L*a*b to obtain the image's chromaticity layers, which is the main advantage. The primary drawback is that it is only employed for a few crops. The BPNN classifier addresses the various class issues while the active contour model is employed to restrict the amount of energy inside the infection area. The categorization rate on average is 85.52%. The defective area is split using K-means clustering, textural information is gathered using GLCM, and the degree of severity of the issue is determined using fuzzy analysis. They used a classifier called an artificial neural network (ANN), which mainly helps to determine how severely the ill leaf is impacted. The following algorithms were used: Naive Bayes, Random forests, Decision trees, Especially a Random Tree, Most nearby Neighbours, and SV Classifier. Randomised trees give the programme flexibility, provide real-time data, and produce very high scores in seven classifiers.

Existing System

Experts can identify and detect plant illnesses by using simple naked-eye inspection to observe leaves for disease. When interacting with large farms, this requires a significant group of specialists and continuous plant observing, both of which can be fairly expensive.For the purpose of identifying diseases in wheat plants offline, they used a Support Vector Machine (SVM) classifier with MCS.

Proposed System

By seeing the results we can say that our proposal model SVM can predict leaf disease from plant. By adapting the method automatic location of disease can be more reliable by reducing the manual work. So that farmers can save their plants before loosing.

Certain procedures must be performed in order to figure out if the leaf is healthy or diseased. that is, Preprocessing, Feature Extraction, Classification, and Classifier Training. Preprocessing involves minimising all of the photos' sizes to a single, uniform value.

SYSTEM ARCHITECTURE

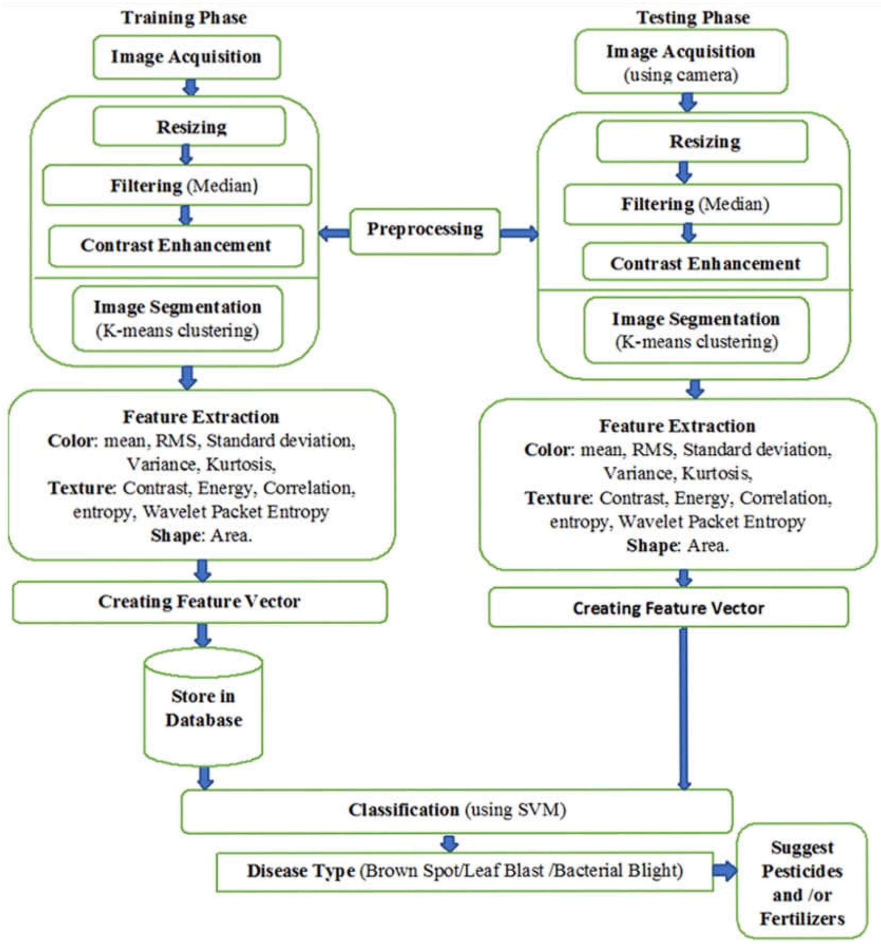

Fig.1.System Architecture

Advantages

a. Manual work is reduced.

b. Image processing helps in to get better results.
c. Computationally efficient works well for images having good contrast.

DEFINE PROBLEM

Disease identification is one of the key components of a successful farming system. Typically, a farmer will use eye observations to spot disease indications in plants that require ongoing attention. Different diseases harm a plant's leaves. Farmers encounter extra challenges while trying to diagnose these diseases. With the aid of photographs of plant leaves, image processing techniques are appropriate and effective for disease detection. Although ongoing plant health and disease monitoring improves crop quality and quantity, it is expensive. Algorithms for machine learning are being tested because they are more accurate. The goal of this project is to develop and test various models for the detection of plant diseases, address the issue of the lack of real-world representative data, experiment with various generative networks, produce greater quantities of leaf image data, and implement a segmentation pipeline to prevent incorrect classification due to unintended input.

Modules

While using the leaf detection, we use the following modules and components:

Preprocessing

Resize the images to a consistent input size suitable for the model.

Normalize the pixel values to a common range to ensure consistent input scaling

Training

Train the CNN model using the training set, optimizing the model's parameters (weights and biases) to minimize an appropriate loss function (e.g., cross-entropy loss).

Use the validation set to monitor the model's performance and make necessary adjustments (e.g., learning rate scheduling, early stopping) to prevent over fitting.

IMPLEMENTATION

Libraries

The libraries used for implementation are:

OpenCV: OpenCV is an essential open-source library for computer vision, machine learning, and image processing. OpenCV accepts a wide range of programming languages, like Python, C++, and Java. It may look for individuals, things, and even handwritten calligraphy in images and movies. When it is used with other libraries, such as Numpy, a highly productive library for numerical operations, the number of weapons in your armoury grows. OpenCV can be used in combination with Numpy to carry out any task.

The OpenCV training will teach you how to process images using a wide range of OpenCV projects and programmes, from basic to advanced operations on images and videos.

Django: Django, a Python-based web framework, provides the simple for developers to develop successful online projects. Django is often known as a "batteries included" framework because it provides built-in features for everything, such as the Django Admin Interface and the built-in database, SQLlite3. Building a website always requires the same set of elements: a method for managing authentication for users (signing up, signing in, and signing out), an administration system, and a content management system.

A dashboard for your website, forms, a system to submit files, etc. Django offers you with ready-to-use components to employ, and this is too for quick development.

CMake: A Makefile generator for all platforms is CMake. Simply simply, CMake creates your project's Makefiles automatically. Although it is capable of much more (such as creating MS Visual Studio solutions), in this presentation I will concentrate on its auto-generation of Makefiles for C/C++ applications.

Asgiref: The ASGI standard, which is positioned as the asynchronous replacement for WSGI, enables communication between Python asynchronously web apps and servers.
Tkinter: The approach most frequently employed is Tkinter. The Tk GUI toolkit offered by Python has a traditional Python interface. Python and Tkinter offer the fastest and least complicated approach to creating GUI applications. The GUI design procedure has been simplified less complicated by Tkinter.
Numpy: When interacting with arrays, capitalise on the NumPy Python module. It additionally covers matrices, the Fourier transform, and linear algebra working algorithms.
Time: Working using time in Python is made possible by the time module. Among other functions, it makes effectiveness like getting the present time and disrupting the program's execution possible. So, while we start, we need to import this module.
Argparse: It is simple to create user-friendly command-line interfaces thanks to the argparse module. The programme specifies the arguments it needs, then argparse works out how to extract those arguments from sys.argv.Additionally, the argparse module automatically creates use and help messages. When users supply the programme with erroneous arguments, the module will likewise produce errors.
Logging: This module defines the classes and functions that make up an adaptable logging of events system for libraries and applications.The main advantage of having a conventional library module provide the logging API is the fact that all modules written in Python can participate in logging, allowing your application log to incorporate messages from both your own modules and modules from other libraries.
Date-Time: Modules for directing dates and times are available in the datetime module.Although date and time calculations are enabled, the implementation's main goal is to efficiently extract attribute data for outputs formatting and manipulation.
Algorithm

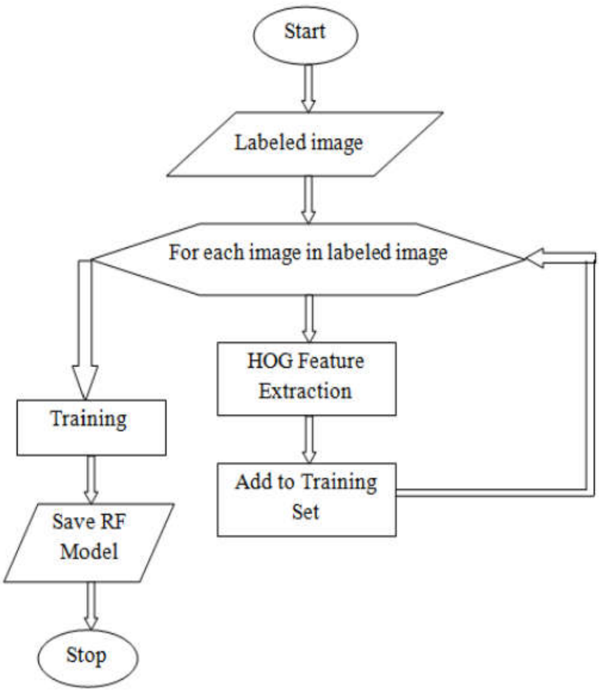

Fig.2. Flow chart for training.

In this particular case, the theory is put into practise using a random forests classifier. They may be utilised for both classification and regression techniques and are versatile. In comparison with SVM, Gaussian Naive Bayes, logistic regression, and linear discriminate analysis, random forests were more precise for smaller sets of data from images. The illustration below shows how our suggested approach is structured. The datasets with labels are separated from testing and training data. The feature vector for the training dataset is generated using HoG feature extraction..

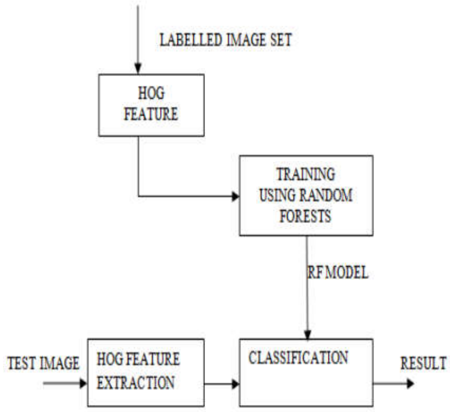

Fig.3. Architecture of the proposed model

RESULTS

To produce grayscale version of any image, RGB images have to be transformed. As it is only possible to calculate Hu moments structured adjectives and Haralick features over a single channel, this is done. Therefore, RGB must be converted to grey scale when computing Hu moments and Haralick characteristics..Finally, the major goal of our study is to use a random forest classification system to determine if a leaf is sick or healthy.
This algorithm's goal is to detect anomalies on plant in their natural or greenhouse environments. To avoid occlusion, the image is usually captured with a plain background.
The procedure has been compared with different machine learning models for accuracy. Using a classifier based on Random Forests and 160 images of papaya leaves, the machine learning system was built. The model's classification accuracy was about 70%. The accuracy can be increased by using a large training dataset and adding other regional data with the global features.

CONCLUSION

We concentrated on how an image from a particular dataset (a trained dataset) was used in the field and historical data sets to forecast the pattern of plant diseases. The following information regarding predicting plant leaf disease is provided by this. This method will cover the greatest variety of plant leaves, allowing farmers to learn about leaves that may have never been cultivated. By listing all potential plant leaves, it aids farmers in choosing which crop to grow. Additionally, this technology takes historical data production into account, giving the farmer information into market prices and demand for specific plants. Compared to AlexNet and GoogleNet LeNet got the best results in the experiment.
The aim of this algorithm is to find anomalies on plants in greenhouses or uncultivated environments. The image is often taken with an unadorned backdrop to prevent occlusion. The algorithm has been compared with different machine learning models for accuracy. A random forest classification algorithm and 160 images of papaya leaves were utilised to create the model. The model's categorising accuracy was about 70%. Accuracy can be improved by employing more regional information in along with global characteristics and practising on many different images.

FUTURE SCOPE

The suggested model employs preprocessing methods such as RGB to grayscale conversion, HE, K-means clustering, and contour tracing. The descriptive characteristics of the leaf samples are extracted using a variety of descriptors, including the Nonlinear Wavelet Transform, the Principal Component Analysis, and GLCM. SVM, K-NN, and CNN are three machine learning techniques that are used to differentiate between sick and healthy leaves. In comparison to other state-of-the-art methods, the analysis of the suggested framework is appropriate for CNN machine learning segmentation methodology with the needed accuracy. In the future, the algorithm can be enhanced by applying fusion approaches to extract important features and by testing it with different dataset leaf samples.

REFERENCES

[1] S. S. Sannakki and V. S. Rajpurohit,” Classification of Pomegranate Diseases Based on Back Propagation Neural Network,” International Research Journal of Engineering and Technology (IRJET), Vol2 Issue: 02 | May-2015

[2] P. R. Rothe and R. V. Kshirsagar,” Cotton Leaf Disease Identification using Pattern Recognition Techniques”, International Conference on Pervasive Computing (ICPC),2015.

[3] Aakanksha Rastogi, Ritika Arora and Shanu Sharma,” Leaf Disease Detection and Grading using Computer Vision Technology &Fuzzy Logic” 2nd International Conference on Signal Processing and Integrated Networks (SPIN)2015.

[4] Godliver Owomugisha, John A. Quinn, Ernest Mwebaze and James Lwasa,” Automated Vision-Based Diagnosis of Banana Bacterial Wilt Disease and Black Sigatoka Disease “, Preceding of the 1’st international conference on the use of mobile ICT in Africa ,2014.

[5] Uan Tian, Chunjiang Zhao, Shenglian Lu and Xinyu Guo,” SVM-based Multiple Classifier System for Recognition of Wheat Leaf Diseases,” Proceedings of 2010 Conference on Dependable Computing (CDC’2010), November 20-22, 2010.

[6] S. Yun, W. Xianfeng, Z. Shanwen, and Z. Chuanlei, “Pnn based crop disease recognition with leaf image features and meteorological data,” International Journal of Agricultural and Biological Engineering, vol. 8, no. 4, p. 60, 2015.

[7] J. G. A. Barbedo, “Digital image processing techniques for detecting, quantifying and classifying plant diseases,” Springer Plus, vol. 2, no.660, pp. 1–12, 2013.

[8] Caglayan, A., Guclu, O., & Can, A. B. (2013, September). “A plant recognition approach using shape and color features in leaf images.” In International Conference on Image Analysis and Processing (pp. 161-170).