(J I D P T S)
ISSN(O):2581-6934

DESIGN AND ANALYSIS OF A MACHINE LEARNING MODEL TO DIAGNOSE CROP DISEASES

Pradeep Byrapaka[1], Balusupati Anil kumar[2], P. Deekshith Chary[3]

[1 & 2]Assistant professor, Dept. of CSE (Artificial Intelligence & Machine Learning)
[3]Assistant professor, Dept. of CSE (Data Science)
[1, 2&3]CMR Institute of Technology, Medchal Rd, Hyderabad, Telangana

Abstract : An approach referred to as "smart farming" involves high-end technological advances in modern farming by collecting data from various sources such as robotics, sensors, broadcasts on social media, etc.Plant diseases are usually passed through viruses and insects that are pests that may substantially decrease production when they aren't immediately handled. Additionally to keeping track of soil quality, this study proposes a system to detect and remove diseases on cotton plants. Though there have been many advances in smart farming down to this moment using image processing, data mining, IOT, etc., machine learning has emerged as the area that is developing most rapidly. In real-time, machine learning applications utilizing supervised or unsupervised approaches are currently in more significant popularity. The conventional farming method includes gathering data, managing unpredictable weather, covering diseases in pesticides, and other behaviors that threaten agricultural workers, particularly in drought-stricken areas. Considering the present state of affairs with conventional farming, there has been an essential demand for predictive data in agriculture which could help farmers to recognize their existing issues and implement necessary measures.

Keywords: Machine Learning, Convolution neural networks, Support Vector Machine

Introduction

Many farmers and agri-businesses may benefit from the technological possibilities provided by smart farming. Various applications include agricultural precision farming, weather prediction, quality assurance, data collection, analysis, etc. Crop disease detection is a part of smart farming in regards to it. The development of the crops depends significantly on the prompt identification of crop diseases. Farmers nowadays use their own eyes as their primary method of identification. It requires money and time as it requires continuous review and monitoring [1].Many farmers use their expertise to recognize diseases; others seek expert advice. Still, medical professionals commonly employ their own eyes to assess disease indications. Consequently, such a disease may have been diagnosed with similar symptoms. Any errors made during disease identification could occasionally end in inadequate treatment and unnecessary application of pesticides [2]. Developing creative methods for disease control and automatic assessment will therefore be essential. Additionally, various pesticide varieties are available to control disease and increase production. However, selecting the most suitable and effective pesticide for handling the diseased condition can be complex and costly. Symptoms on leaves are the primary indicator of disease on cotton farms. Therefore, an artificial based on the vision that can detect diseases from images of cotton leaves and propose an appropriate pesticide as an option is required. According to that point, farmers can expect immediate advantages from automatically predicting numerous farming diseases, saving consumers time, money, and crop life. A machine-learning model that can predict the frequency of cotton crop disease based on the temperature in the natural environment and from the soil moisture temperature is proposed in this paper to handle a growing annoyance provided to farmers [3].

Literature Survey

This module includes several crop disease detection systems developed and utilized by other researchers. To understand more about their method and approach. The identification and classification of three cotton leaf diseases, Myrothecium, Alternaria, and Bacterial Blight, was proposed by Rothe, P. R., et al. [4]. Using the active counter model, image segmentation is carried out, and Hu's moments are obtained for the training of the adaptive neuro-fuzzy inference system. For classification, back propagation neural network models (BPNN) are utilized. Images of Cotton Leaf Spot were utilized in Viraj A. Gulhane et al. [5] to demonstrate technological solutions and categories the diseases. The machine learning algorithm is trained to use smart farming, such as early disease detection in the groves, customized fungicide application, etc. The proposed research is based on edge the identification of images utilizing segmentation techniques, requiring first manipulating the acquired images for development. The system takes a single plant leaf as input and divides the leaf after reducing the picture frame. The sick part of the leaf can be identified by applying the high-pass filter to the segmented leaf image. The segmented leaf image can be obtained using a unique texture feature based on fragments [6]. They provide a superior fabric model because fractal-based elements are locally invariant. The author has presented an SVM-based Multiple Classifier System for the pattern recognition of wheat leaf diseases. The performance of disease recognition in the wheat plant has also been improved by using a layered generation structure and mid-level generation of features.

Cotton Leaf Diseases

Bacterial Blight

The bacteria "Xanthomonas campestris pv. Malvacearum" is the principal cause of the bacterial disease known as "bacterial blight." Initial indications of bacterial blight are dark green and watery leaves. Regarding 1 to 5 mm-long, angular spot with a crimson to brown border on a leaf. These pointed leaf spots appear as wet areas that eventually become dark brown to black [7]. As the petioles and stems get diseased and the leaves prematurely fall off, the spots on the lesion area of the leaves may extend over the primary veins of the leaf affected by Bacterial Blight, as shown in Fig. 1(a).

Cerespora

The Cercospora Gossypina is the root cause of Cercospora [8]. The impacted leaf has red spot blemishes on the leaves that measure up to 2 cm across. The dots are circular or erratic, with white centers and yellowish, purple, dark brown, or blackish edges. The rakish leaf spot manifests as a result of the leaf's tiny veins limiting the region of injury. This disease affects mature plants' more experienced leaves. A Cerespora-infected leaf is shown in Fig. 1(b).

(a)

(b)

Methodology

Convolution Neural Network

Convolution neural networks consist of neurons capable of training weights and biases, identical to regular neural networks. Each neuron takes in some inputs and then applies some to train the network's neurons and generate an output that exceeds a linear path. From the processed image pixels at one end to the class scores at the other, every network component represents a single differentiable scoring function. In addition, they still contain a loss function (such as SVM/Softmax) on the last (completely connected) layer, and all the techniques we established for learning conventional neural networks remain. With this instance, the CNN object classification model analyses images as input of digits to analyze, evaluate and classify them. CNN models can be trained and tested attributable to supervised training. A combination of filters (Kernals), pooling, fully connected layers (FC), convolution layers, and Softmax will be applied

 ISSN:-2581-6934 |Available at :www.jidps.com | pp (22-27)

to the input image before identifying an object utilizing probabilistic values between 0 and 1[9].The number of parameters is then further decreased by including pooling layers. The prediction is made when several convolution and pooling layers were recently added. Convolution layers assist in feature extraction. Compared to an external network, where the characteristics recovered are more general, particular features are extracted as deep in the network.

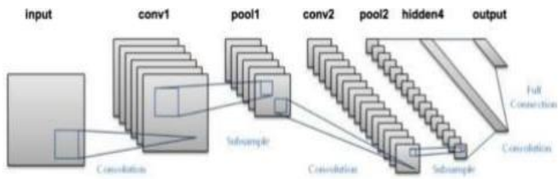

Figure 2: Convolutional Neural Network

Image Acquisition

The primary aspect of the process is collecting images of the infected leaves to create a database. The cotton farming in the Buldhana area was used to create the RGB color photos of cotton leaves, which were then captured in JPEG format with the necessary resolution for disease identification [10]. A database of 900 images has been compiled. Preprocessing removes noise that is produced during this process of creation.

Figure 3: Infected cotton leaf RGB input image preprocessed image

Feature Extraction

Feature extraction aims to gather the appropriate information that differentiates different categories [11]. Feature extraction seeks to obtain features that maximize the recognition rate while using the fewest elements possible for every character. The region of interest (ROI) in feature extraction, which is utilized to extract different characteristics that are used to diagnose the disease, is the sick area that has been split. Using partial least square regression (PLSR), the current approach extracts eight color and texture features.

K-nearest neighbor

The supervised machine learning approach, the "k-nearest neighbor" (KNN), is primarily used for classifying. It has been widely applied to disease prediction. Analyzing the features and labels of the initial data, KNN, a supervised algorithm, predicts the unlabeled categorization of data. The k nearest training data points (neighbors), those closest to the test issue, are usually utilized by the KNN method to classify datasets using a training model similar to the testing query. The initiative then employs a majority voting rule for deciding which category to utilize. The KNN algorithm is one of the many primary forms of machine learning algorithms. It is often used in classification problems due to its incredibly adaptable and simple understanding of architecture. The algorithm's use in regressing and classification situations for data with various label numbers, frequency ranges, and contexts is widely recognized. As disease prediction is a practical challenge, this research is developing a study around an approach based on classifying datasets. Identifying solutions to this problem's adaptation needs is appealing. The computations and workings of the algorithm are simple. It allows itself to be modified in various methods to reduce drawbacks and obstacles and raise its accuracy and applicability for use in a broader range of scenarios [12].

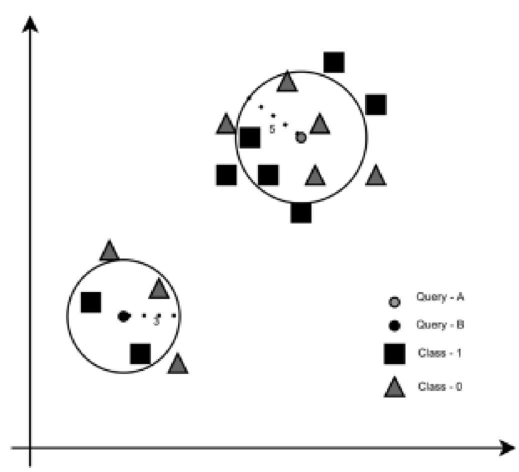

Figure 4: Visual illustration of the KNN algorithm.

Support Vector Machine (SVM)

The method most commonly used in the supervised machine learning field for classification tasks is the Support Vector Machine (SVM). While SVM algorithms can be used for statistical regression tasks, they are most usually used for

 ISSN:-2581-6934 |Available at :www.jidps.com | pp (22-27)

applications requiring classification, such as separating binary data into two categories. In addition, more machine learning algorithms are used for classification and regression analysis tasks, including decision trees, random forests, and K-NN. However, the fundamental method used by each algorithm type—including SVM—to solve classification issues ranges. The Support Vector Machine (SVM) technique searches out the closest data point (support vector) across the class boundary lines [13]. SVM also aims to maximize the margin between the support data points (vectors) by optimizing a hyper planet. Support vectors, or vectors for individual data points in the plotted graph, have been referred to as such in SVM. The numerical value of the particular coordinate is the value of each feature in a data point. SVM transforms labeled binary data (linearly separable data) into the required formats for output from its input.

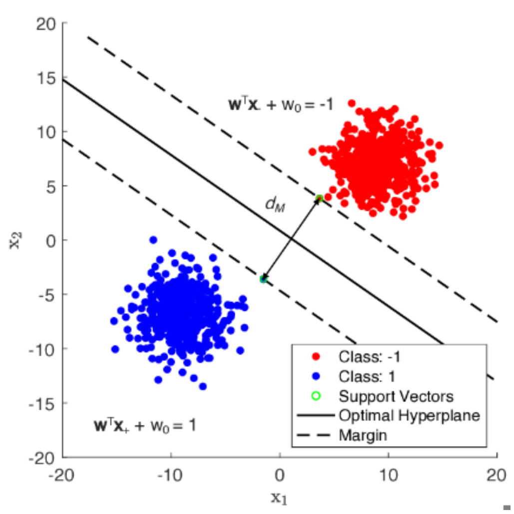

Figure 5: Visual illustration of Support Vector Machine Classification

It is used to identify the type of leaf disease. Combining the provided data vectors with one of the trained data from several classes is what classification is all about. Different classifier types can be used in machine learning for variety. SVM is trained and evaluated with several kernels to obtain high accuracy, and it is discovered that the Gaussian kernel provides increased accuracy. In the current system, a non-linear Gaussian kernel and SVM-based regression technique are utilized to classify the diseases that affect cotton leaves. By locating the optimum hyper plane, the SVM-based regression determines the nonlinear relationship between input vectors and response variables [14].

Experimental Results

As mentioned, the main objective of the current study is to identify and manage diseases that affect cotton leaves. Maintaining the soil's quality is the secondary objective. We examined 900 pictures of cotton leaves to find conditions. 629 of these are used for training, and 271 are utilized for testing [15]. The accuracy for identifying specific cotton diseases of the leaves can be observed in Table 1.

Name of Disease	Correctly Classified (CC)	Incorrectly Classified (IC)	Accuracy of CC
BacterialBlight	65	12	87.89%
Alternaria	57	13	85.63%
Cerespora	42	06	83.95%
Grey Mildew	34	07	84.75%
Fusarium Wilt	22	08	84.32%
Healthy leaf	05	01	82%

Table 1: Accuracy of Cotton Leaf Disease Detection

The model's accuracy has improved as the number of epochs rises. According to the graph below, it will be somewhere around 85%. The loss decreases as the number of epochs rises, in an inverse relationship with the number of epochs. For the test, it is roughly 0.2, and for the train, it is about 0.3.

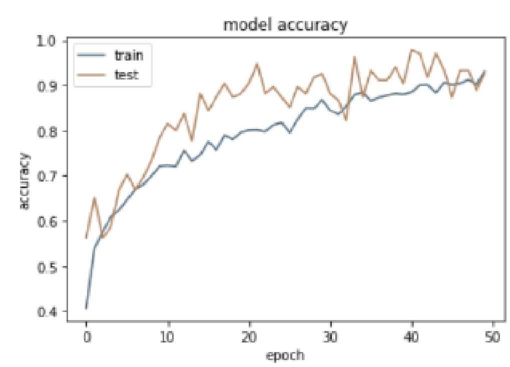

Figure 6: Model Accuracy Test Result

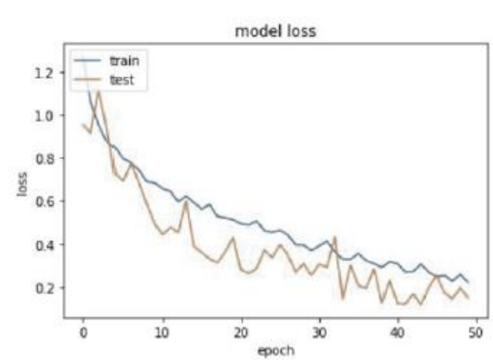

Figure 7: Model Loss Test Result

 ISSN:-2581-6934 |Available at :www.jidps.com | pp (22-27)

CONCLUSION

Produce a method to identify the illness affecting a cotton plant leaf. Since utilizing to detect a disease present in cotton crops can result in some errors, doing the same task with the assistance of a machine will reduce the possibility of error. The system's fundamental module will help the user through disease evaluation, forecasting diseases, and disease treatment ideas. The system's other modules will provide users with a list of agriculture studies that concentrate just on the cotton crop and additionally provide suggestions for items they could enjoy and might be able to purchase at the moment. The system will be developed to assist the user as they analyze the state of the cotton crop at the moment. The CNN model, which forms the system's foundation, will aid with the identification of the diseases harming the cotton crop. The model will be accurate 85% of the time, losing about 0.2 percent of every time. The accuracy can be improved by increasing the number of epochs and the amount of the data.

References:

[1]. V.A.Gulhane and A.A.Gurjar,"Detection of diseases on cotton leaves and its possible diagnosis."International Journal of Image Processing(IJIP),vol.5 , no. 5, pp. 590-598

[2]. P.Revathi and M.Hemalatha, "Classification of cotton leaf spot diseases using image processing" in Emerging Trends in Science Engineering and Technology(INCOSET) International Conference on , IEEE, pp. 169-173.

[3]. MeleseZekiwos and Abey Bruck," Deep Learning-Based Image Processing for Cotton Leaf Disease and Pest Diagnosis",Journal of Electrical and Computer Engineering , vol.2021.

[4]. Viraj A. Gulhane et al. "Identifying PV module mismatch faults by a thermography-based temperature distribution analysis," IEEE Transactions on Device and Materials Reliability, vol. 14, no. 4, pp. 951–960, 2014.

[5]. Poornachander Vadicherla, Dhanalakshmi Vadlakonda," Study on energy efficient routing protocols scheme in heterogeneous wireless sensor networks (network & mobility)", Materials Today: Proceedings, Volume 47, Part 15, 2021, Pages 4955-4958, ISSN 2214-7853,

[6]. Ahmad and K. Pothuganti, "Design & implementation of real-time autonomous car by using image processing & IoT," 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2020, pp. 107-113, doi: 10.1109/ICSSIT48917.2020.9214125.

[7]. M. Aghaei, U. E. Madukanya, A. K. V. de Oliveira, and R. Ruther, "Fault inspection by aerial infrared thermography in a PV plant after a meteorological tsunami," in VII Congresso Brasileiro de Energia Solar–Gramado, 17 a 20 de Abril de, Brazil, 2018.

[8]. M. Muttilo, I. Nardi, V. Stornelli, T. de Rubeis, G. Pasqualoni, and D. Ambrosini, "On field infrared thermography sensing for PV system efficiency assessment: results and comparison with electrical models," Sensors, vol. 20, no. 4, p. 1055, 2020.

[9]. L. Schirone, P. F. Califano, U. Moschella, and U. Rocca, "Fault finding in a 1 MW photovoltaic plant by reflectometry," in Proceedings of 1994 IEEE 1st World Conference on Photovoltaic Energy Conversion - WCPEC (A Joint Conference of PVSC, PVSEC and PSEC), Waikoloa, HI, USA, 1994.

[10]. S.Ramana, M.Pavan Kumar, N.Bhaskar, S. China Ramu, & G.R. Ramadevi. (2018). Security tool for IOT and IMAGE compression techniques. Online International Interdisciplinary Research Journal, {BiMonthly}, 08(02), 214–223. ISSN Number: 2249-9598

[11]. S. Ramana, N. Bhaskar, S. China Ramu, M. V. Ramana Murthy." A Two-Level Authentication Protocol for Secure M-Commerce Transactions using AMQP Protocol", Vol 2021: Issue 06, pp:844– 861.

[12]. Peddyreddy. Swathi. (2022). Industry Applications of Augmented Reality and Virtual Reality. Journal of Environmental Impact and Management Policy(JEIMP) ISSN:2799-113X, 2(02), 7–11. Retrievedfromhttp://journal.hmjournals.com/index.php/JEIMP/article/view/453

[13]. Keerthi Vuppula, "An advanced machine learning algorithm for fraud financial transaction detection"

 ISSN:-2581-6934 |Available at :www.jidps.com | pp (22-27)

Journal For Innovative Development in Pharmaceutical and Technical Science (JIDPTS) Volume:4, Issue:9, Sep:2021

[14]. S. G. Wu, F. S. Bao, E. Y. Xu, Y.-X. Wang, Y.-F. Chang, Q.-L. Xiang, "A leaf recognition algorithm for plant classification using probabilistic neural network," In Proceedings of 2007 IEEE International Symposium on Signal Processing and Information Technology, pp. 11-16, 2007.

[15]. N.-H. Pham, T.-L. Le, P. Grard, V.-N. Nguyen, "Computer aided plant identification system," In Proceedings of 2013 International Conference on Computing, Management and Telecommunications (ComManTel), pp. 134-139, 2013.

[16]. Ahmad and K. Pothuganti, "Design & implementation of real-time autonomous car by using image processing & IoT," 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2020, pp. 107-113, doi: 10.1109/ICSSIT48917.2020.9214125.