STM JOURNALS

Trends in
Machine Design

ISSN: 2455-3352
Volume 11, Issue 2, 2024
May–August
DOI (Journal): 10.37591/TMD

http://engineeringjournals.stmjournals.in/index.php/TMD/index

Review TMD

Crop Disease Prediction by Machine Learning

Sanika Deshmukh[1], Yash Katkar[2], Atharva Ubale[3], Mangesh Kondhalkar[4], Santosh S Shinde[5]

Abstract

The classification of Crop can be classified into several methods. The data set of crop leaf illnesses, notably Bacterial Leaf Blight disease (BLB), a crop leaf disease with significant outbreaks throughout Thailand, and Brown Spot Crop disease (BSR), is classified employing image classification in this study. Additionally, image processing technology is used for identifying different types of crop leaf disease. These algorithms include the Random Forest, Decision Tree, Gradient Boost, and Naive-Bayes algorithms, and their accuracy, precision, and recall are tested. A number of image processing techniques and classification algorithms are investigated, such as Naive Bayes, Random Forest, Decision Tree, and Gradient Boosting. Metrics like accuracy, precision, and recall are used to gauge how well each algorithm performs in terms of differentiating between various crop leaf diseases. The best result of performance in the image classification of Crop leaf diseases is CNN algorithm equal to69.44 percent. Crop diseases pose significant threats to global food security by affecting crop yield and quality. Early and accurate detection of these diseases is crucial for timely intervention and effective management. Machine learning (ML) techniques have emerged as powerful tools in agricultural research, offering predictive capabilities that can assist farmers in identifying and mitigating crop diseases promptly. This research article explores various ML approaches used for crop disease prediction, highlighting their benefits, challenges, and future directions.

Keywords: Crop leaf disease: algorithm, image processing, classification, cnn algorithm.

INTRODUCTION

Image classification is a statistical data collection approach. Its goal is to classify each image based on its distinguishing characteristics. This is a resemblance of picture points to the same class determined by using mathematical ideas and statistics to develop a model to determine differentials in features of the image points. The data classification technique is divided into two sections: supervised learning classification and unsupervised learning classification. This study strives to extract behaviors or traits from visual data. The information is used by a supervised learning algorithm to classify photos. Image learning based on information and labelling, which includes data mining. We can learn from image classification properties. To forecast that, we can use a classification model with new images. The two primary frameworks in which picture classification operates are supervised learning and learning without supervision.

In supervised learning, labeled data—images that are each assigned to a predetermined class or category—are used to train the classification model. The model learns about patterns and characteristics in the images that correspond with specific labels. Although this approach constantly yields greater classification task accuracy, training with a significant amount of labeled data is expected [1].

***Author for Correspondence**

Email:sanikadeshmukh242@gmail.com

[1]Student, Department of Computer Engineering, Savitribai Phule Pune University, Pune, India
[2]Student, Department of Computer Engineering, Savitribai Phule Pune University, Pune, India
[3]Student, Department of Engineering, Savitribai Phule Pune University, Pune, India
[4]Student, Department of Engineering, Savitribai Phule Pune University, Pune, India
[5]Student, Department of Engineering, Savitribai Phule Pune University, Pune, India

Received Date: May 27, 2024
Accepted Date: August 06, 2024
Published Date: September 02, 2024

Citation: Sanika Deshmukh, Yash Katkar, Atharva Ubale, Mangesh Kondhalkar, Santosh S Shinde, Crop Disease Prediction by Machine Learning. Trends in Machine Design. 2024; 11(2): 21–25p.

Conversely, unsupervised learning entails the grouping of images according to shared characteristics without the need for predetermined labels. Although it could need more interpretation and confirmation, this method is helpful for finding patterns and structures in data.

The Significance of Classifying Images Applications for image classification can be found in many different fields: Using medical scans and photographs to diagnose illnesses is known as medical imaging.

Agriculture: Using leaf photos to identify agricultural diseases and improve treatment plans. Identifying people and things in surveillance footage is a security and surveillance concern. Autonomous vehicles: seeing and responding to their surroundings using visual cues [2-4].

RELATED WORKS

Name 1: Image Classification of Crop Leaf Diseases

- *Author:* Panuwat Mekha, Nutnicha Teeyasuksaet
- *Abstract:* This research presents the algorithm that can be used to classify Crop leaf diseases by images. It was found that the random forest algorithm, which corresponds to 69.44 percent of the image, has the greatest performance [11].

Classification. Name 2: An Improved Random Forest Classifier for Image Classification

- *Author:* Baoxun Xu and Yunming Ye
- *Abstract:* We present an improved random forest algorithm by simultaneously taking into account of a new feature weighting method [12].

Name 3: Image Classification using Random Forests and Ferns

- *Author*: Andrew Zisserman
- *Abstract*: We have demonstrated that using random forests/ferns with an appropriate node test reduces training and testing costs significantly over a multi- way SVM, and has comparable performance [13].

Name 4: Crop Leaf Disease Detection and Crop Yield Prediction Using Random Forest

- *Author*: Shilpa Ajeesh M,Anagha
- *Abstract:* This proposed system provides the disease name, its symptoms, cause, and remedies following the inputted image and also provides the prediction of crop yield for the coming year [14].

Name 5: Enriched Random Forest for High Dimensional Genomic Data

- *Author:* Debopriya Ghosh, and Javier Cabrera
- *Abstract:* We applied weighted random sampling instead of simple random sampling, that chances of selecting less informative features are reduced [15].

SOFTWARE REQUIREMENT SPECIFICATION

Assumption and Dependencies

A number of factors that affect the requirements of the system are:

- The system the application is executing on will have the required resources available as necessary.
- Another assumption is that the software and hardware components work in the same way as used while developing this project [5].

The Requirements and Operation of The System Are Profoundly Affected by Several Factors:

- *Resources available:* The system makes the assumption that there are enough resources (processing power, memory, and storage) on the hardware and infrastructure where the program

is installed, available as needed. This guarantees scalability and optimal performance in a range of operational scenarios.

- *Hardware and software component consistency:* It is assumed that the hardware and software components function in accordance with how they did during the project's development phase. Compatibility with operating systems, libraries, frameworks, and other dependencies that are essential to the operation of the system are included in this.
- *Consistent network access:* For the system to provide smooth communication between dispersed components and outside services, network connectivity must be steady. Ensuring ongoing operation and data integrity requires making assumptions about bandwidth availability and network dependability.
- *Security and integrity of data:* It is essential to make assumptions about security procedures and data integrity. The system anticipates that protective measures like encryption, access control, and backup plans will be successfully put into place to safeguard confidential data and stop illegal access or data loss.
- *Respect for regulatory standards:* The system is predicated on adhering to pertinent industry best practices and regulatory criteria. This entails following software licensing terms, data protection legislation, and industry-specific operational standards.
- *User authorization and access***:** There are stated presumptions regarding user permissions and access rights. The system restricts illegal access to protect system integrity and confidentiality, while assuming that users have the necessary degrees of access to carry out their assigned activities [6].

Function Requirement Specification

Registration and authentication

No registration and Authentication is required.

Taking user's inputs

Users will have to choose one Crop leaf image for detection . After that users have to upload the Crop leaf image for disease detection.

External Interface Requirements

User Interface

The software's graphical user experience, or UI, ought to work with any common operating system, like Window. The user can access the system by using this UI. There are several tools and software packages available to construct the user interface, or UI.

Hardware Interfaces

A hardware interface is needed to run the software. Python compatible hardware is required which is minimal requirement [7].

Software Interfaces

It uses Python as the front-end programming tool. Windy

SYSTEM DESIGN

*System architectur*e: By employing the right insecticides are used and medicines, we may lessen the infestation of pests. By using the right size reduction strategies, we may shrink the photos without dramatically sacrificing quality. We can broaden the scope of the authors' prior research such that the system also displays the disease's cure. The main aim is to use image processing to identify plant diseases. Additionally, it suggests the name of the pesticide to be used when the illness has been discovered.

It also demonstrates the pests and insects that are causing the pandemic. Aside from these dual purposes, this drone saves an incredible amount of time. Although the model's budget is quite expensive

for small-scale farming, it will be cost-effective for large-scale farming [8]. It accomplishes all processes step by step, resulting in the achievement of every the result.

Thus, the main objectives are:

1. To create a system that is accurate in identifying parasites and agricultural problems.
2. Establish an online registry of pesticides for each type of insect and illness.
3. To offer therapy to the illness that has been discovered.

- The insect genus at the larval stage are recognized as leaf miners. They eat on the leaf's sides.
- Disease caused by mining for leaves Plants that have a lot of insects on them suffer serious harm. There might be up to six maggots on one leaf. As a result, it could seriously harm a plant leaf. It can limit growth in plants, which results in less.

Design of Machine Learning Model

A model of machine learning is constructed through two stages: the testing phase, when the model is trained, and the training phase, where the model is tested using an input known as the test data. The model includes many image processing stages, namely feature extraction, segmentation, acquisition, pre-processing, and SVM classifier for illness assessment [9,10]. Image acquisition: The camera is used to take a picture of the sick leaf. It is taken at a consistent distance and with enough light to allow for classification and learning. The algorithm is trained using the example photos of the sick leaves that are retrieved. Images involving healthy and damaged leaves are captured in order to train and test the system. A standard arrangement will be used to store the photographs (Figure 1).

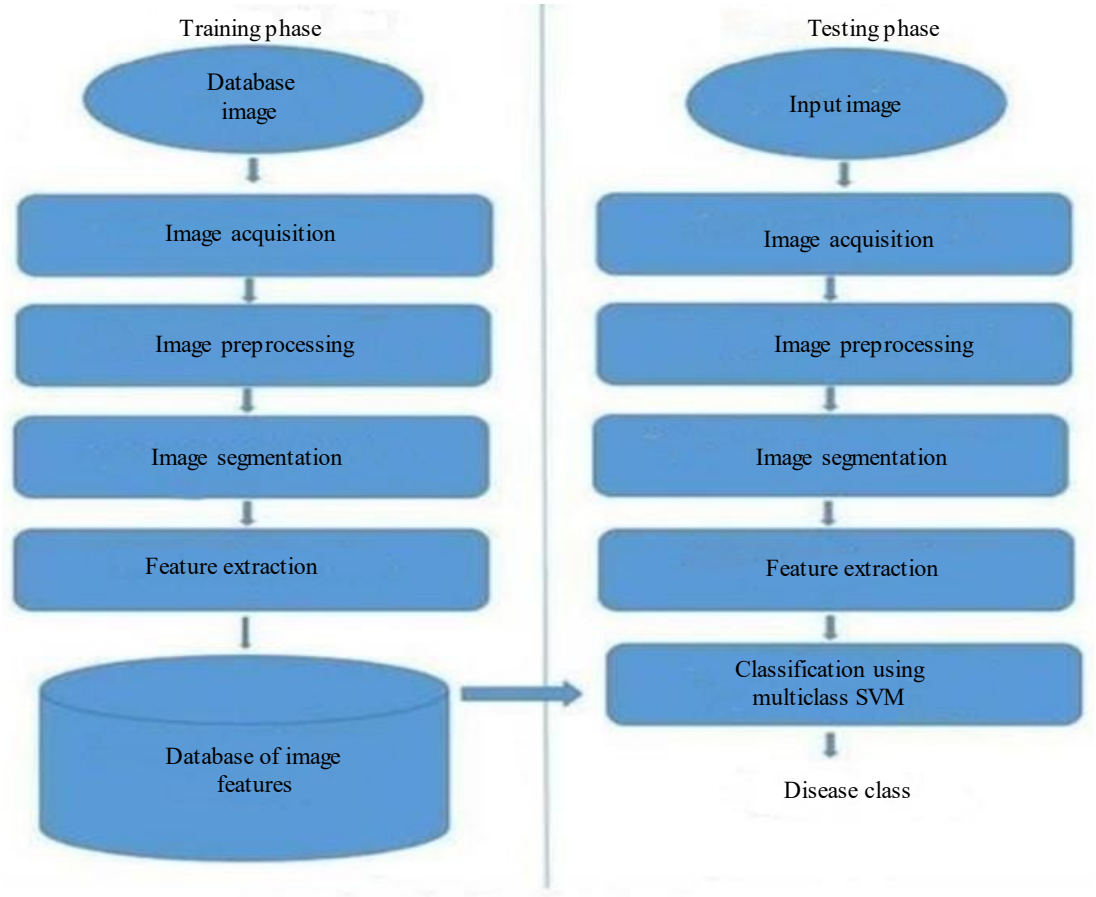

FIG: ML model with two phases

Figure 1. Machine learning model.

CONCLUSIONS

This experiment suggested utilizing a mix of texture and color feature extraction to classify leaf picture sequences in order to diagnose diseases in leaves. Farmers first send in an online picture of a sick leaf on a plant. Python then reads the image and uses SVM to analyze it automatically, delivering the findings. The project's findings comprise the identification of suitable traits for leaf disease of certain diseases that typically affect plants. First, a collection and pre-processing of both healthy and sick photos are made. Next, attributes like form, hue, and texture are taken out of these photographs. Following that, the support vector machine classifier is used to categorize these pictures. In order to identify which traits are most suitable for identifying leaf disease, a variety of features are combined. Shape features are the least reliable when utilized alone, whereas texture features are the most accurate. For maximum classification accuracy, texture and color extracted characteristics are combined. Good classification accuracy is achieved by combining the polynomial kernel with texture and color extraction.

A short text message was delivered to the project user based on the disease's categorized a sincere manner

REFERENCES

1. Mrinalini R. et al., An application of K-means clustering and artificial intelligence in pattern recognition for crop diseases ,2011.
2. Sarraj Kumar, S.Sowrirajan,‖ Automatic Leaf Disease Detection and Classification usingHybrid Features and Supervised Classifier‖, International Journal of Advanced Research inElectrical, Electronics and Instrumentation Engineering, vol. 5, Issue 6,2016..
3. Totem, D. J. Rogers, and S. I. Hay, ―Global transport networks and infectious disease spread, ‖ Advances in Parasitology, vol. 62, pp. 293–343, 2006.
4. J. R. Rohr, T. R. Raffle, J. M. Romanic, H. McCallum, and P. J. Hudson,Evaluating the links between climate, disease spread, and amphibian declines,Proceedings of the National Academy of Sciences of the United States of America, vol. 105, no. 45, pp. 17436–17441, 2008.
5. T. Van der Wet, ―Present worldwide distribution of fire blight, ‖ in Proceedings of the 9th International Workshop on Fire Blight, vol. 590, Napier, New Zealand, October 2001.
6. H. Cartwright, Ed., Artificial Neural Networks, Humana Press, 2015.
7. Steinhart and A. Chrisman, Support Vector Machines, Springer Science & Business Media, New York, NY, USA, 2008.
8. stein wart and A. Chrisman, Support Vector Machines, Springer Science & Business Media, New York, NY, USA, 2008. .
9. S. Sankaran, A. Mishra, R. Ehsan, and C. Davis, ―A review of advanced techniques for detecting plant diseases, ‖ Computers and Electronics in Agriculture, vol. 72, no. 1, pp.1–13, 2010.
10. P. R. Reddy, S. N. Diya, and R. Vijayalakshmi, ―Plant disease detection technique tool— a theoretical approach, ‖ International Journal of Innovative Technology and Research, pp. 91–93, 2015.
11. Panuwat Mekha, Nutnicha Teeyasuksaet. Image Classification of Rice Leaf Diseases Using Random Forest Algorithm. 2021 ; IEEE. Cha-am, Thailand. DOI: 10.1109/ECTIDAMTNCON51128.2021.9425696.
12. Xu B, Ye Y, Nie L. An improved random forest classifier for image classification. 2012 ; 2012 IEEE International Conference on Information and Automation. Shenyang, China. DOI: 10.1109/ICInfA.2012.6246927.
13. Bosch A, Zisserman A, Munoz X. Image Classification using Random Forests and Ferns. 2007 ; 2007 IEEE 11th International Conference on Computer Vision. Rio de Janeiro, Brazil. DOI: 10.1109/ICCV.2007.4409066
14. Shilpa Ajeesh M ,Anagha C , Fathima Sifa , Hilma T , Deepthi P , Jijina M T. Rice Leaf Disease Detection and Crop Yield Prediction Using Random Forest. International Journal of Engineering Research in Computer Science and Engineering (IJERCSE). Vol 9, Issue 8, August 2022
15. Ghosh D, Cabrera J. Enriched Random Forest for High Dimensional Genomic Data. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2022 ;19(5):2817–28.

