Chelonian Conservation And Biology

Vol. 18 No. 2 (2023) | https://www.acgpublishing.com/ | ISSN - 1071-8443

DOI: doi.org/10.18011/2023.12(2).1892.1902

PLANT DISEASE CLASSIFICATION USINE MACHINE LEARNING

Kandula Vasudha

Department of CSE, Koneru Lakshmaiah education foundation, Vaddeswaram, India

Sandhya Naganaboyina

Department of CSE, Koneru Lakshmaiah education foundation, Vaddeswaram, India

Chintada Vasudharini

Department of CSE, Koneru Lakshmaiah education foundation, Vaddeswaram, India

Gollapudi Ashritha Sri Manjeera

Department of CSE, Koneru Lakshmaiah education foundation, Vaddeswaram, India

Dr. M. Madhusudhana Subramanyan

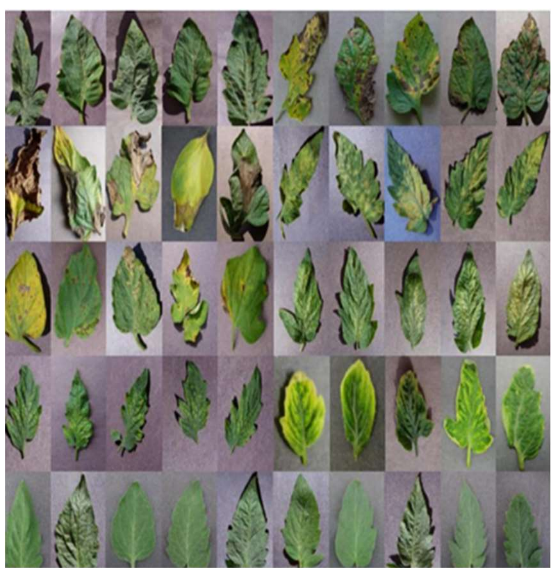
Associate professor, Department of CSE, Koneru Lakshmaiah education foundation,

Vaddeswaram, India

ABSTRACT: In India, a developing nation where agriculture supports nearly 58% of the rural population, tomatoes are a major crop. To prevent significant losses in tomato quantity and yield, it is essential to accurately identify and classify tomato plant diseases. Advanced technologies like image processing can be used to address this challenge using various techniques and algorithms. ASE is often leaf damage. This project uses four sequential stages to identify the specific type of disease: preprocessing, leaf segmentation, feature extraction, and classification. Preprocessing is used to remove noise from the images, while image segmentation isolates the affected areas of the leaflings, a guided, supervised, and advanced machine learning algorithm called k-nearest neighbours (KNN) is used. In the final stage, the user receives treatment recommendations. Diseases can have a devastating impact on living plants, highlighting the importance of early detection. This paper presents a novel approach to leaf disease detection using image processing, enabling farmers to identify tomato plant problems from images based on colour, boundaries, and texture. The goal is to provide farmers with timely and reliable results.

INTRODUCTION

In regions like India, the advancement of agricultural technologies is paramount for enhancing crop productivity. Ongoing research endeavours and exploratory studies in the crucial realm of precision farming aim to boost food crop yield and quality while keeping costs low and ensuring higher financial returns. The complex dynamics of agricultural ecosystems involve the interaction of soil, seeds, and chemicals to promote plant growth, with vegetables and fruits emerging as significant outputs. Achieving high-quality and cost-effective agricultural products necessitates a thorough assessment and enhancement of product value. Diseases profoundly disrupt the normal physiological functions of plants, impacting critical processes like transpiration, photosynthesis, fertilization, pollination, and germination. These debilitating diseases, caused by pathogens such as fungi, bacteria, viruses, and adverse environmental conditions, underscore the need for early-stage plant disease diagnosis. The initial stages of disease diagnosis often involve periodic monitoring by professionals, a process that proves both costly and time-consuming for farmers. Hence, there is a pressing need to develop rapid, cost-effective, and accurate methods for intelligently detecting diseases based on visible indicators present on plant leaves. Our study introduces a novel system designed to identify specific diseases that might affect tomato leaves, a vital crop in the agricultural landscape. Recognizing the type of disease that can afflict an important crop like tomatoes is crucial. Leveraging cutting-edge image recognition technologies, we make this identification process possible, offering visual insights into the application's functionality—a significant driver for the widespread adoption of digital technologies. Various individuals and technological groups are actively engaged in agricultural advancements to increase yield and throughput. While previous techniques have been employed to address the spread of diseases in tomato plants, our project stands out due to several merits.



Utilization of Pre-existing Images: The project avoids the need for collecting inputs for laboratory studies by using readily available images of plant diseases.

Simulating Multiple Infections: It can simulate scenarios where a particular plant is simultaneously infected with more than one pest or disease in the same input.

Compatibility with Diverse Input Images: The system can seamlessly handle various input images captured by different cameras, such as mobile phones and other available devices, with different resolutions.

Versatility in Different Conditions: The project is systematically designed to manage diverse conditions related to illumination, object size in an image, background contrast, etc., across the neighbouring part of a specific plant.

Feasibility and Functionality: It presents a practical and functional approach suitable for field use without the requirement for expensive, complex, and sophisticated technologies.

LITERATURE SURVEY

Shruthi et al. proposed a new way to detect plant diseases using machine learning techniques. Their study showed that a Convolutional Neural Network (CNN) could detect many different diseases with high accuracy. P. Srinivasan et al. created a software program to classify groundnut leaf diseases. Their program used image acquisition, preprocessing, segmentation, feature extraction, and the K Nearest Neighbour (KNN) algorithm to categorize four distinct diseases that affect groundnut crops. L. Sherly reviewed different machine learning classification techniques for plant diseases. Their paper summarized different algorithms for classifying and detecting bacterial, fungal, and viral plant leaf diseases, and discussed the advantages and disadvantages of each approach. Gurleen Kaur et al. reviewed different methods for plant leaf disease detection, including BPNN, SVM, K-means clustering, Otsu's algorithm, CCM, and SGDM for image segmentation, feature extraction, and classification. Md. Selim et al. used eleven statistical features and the Support Vector Machine (SVM) classifier to accurately identify plant diseases. Their approach improved the efficiency of the detection, identification, and classification processes, and achieved a 93% accuracy in disease classification.

Monzurul Islam et al. combined image processing and machine learning to diagnose diseases from potato leaf images. Their approach achieved a 95% accuracy in disease classification by using Color Thresholder, GLCM, and multiclass SVM. Jobin Francis et al. calculated the damaged ratio of leaves to identify diseases in pepper plants. They used masking and threshold-based segmentation to separate leaves from the background, and then used backpropagation algorithms to identify two types of diseases.Vijai Singh et al. used a genetic algorithm for leaf image segmentation. Their method could identify diseases early with minimal computational effort, and produced good results.

Mrunmayee et al. outlined methods for disease detection and classification using image processing and neural networks. Their approach involved preprocessing color images, using k-means clustering for segmentation, and extracting texture features using the gray level co-occurrence matrix (GLCM) method. Their method achieved an overall accuracy of 90%. Sachin D. Khirade et al. discussed segmentation and feature extraction algorithms for plant disease detection. They proposed neural network methods such as self-organizing feature maps, backpropagation algorithms, and SVMs for disease classification. Usama Mokhtar et al. used color space transformation and gray-level co-occurrence matrix (GLCM) for preprocessing and feature extraction, respectively. They used the Support Vector Machine (SVM) algorithm with various kernel functions for the classification phase and achieved a classification accuracy of 99.83%.

Melike Sardogan et al. introduced a Convolutional Neural Network (CNN) model for detecting and classifying four types of tomato leaf diseases. Their model used three different input matrices for the R, G, and B channels, and the reLU activation function and max pooling for the output matrices. They used the Learning Vector Quantization (LVQ) algorithm for classification.

ARCHITECTURE

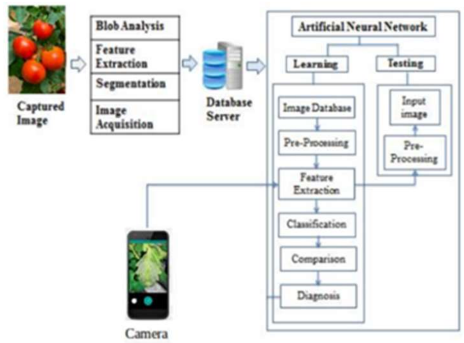
The Architectural Structure illustrates how the system interacts and how control flows through various stages. In this discussion, we will explore the hardware control flow, starting from image capture to disease detection and display. Diseased plant leaf images are captured using a mobile or digital camera. Subsequently, the captured image is transmitted to the system, initiating the image processing stage. The resulting image becomes input for this software system, forming our testing data, which determines the accuracy of our system. Before subjecting the image to testing with our Machine Learning model, it undergoes a series of preprocessing steps. These steps involve resizing, recoloring, and various techniques to prepare the image for testing, making it ready for evaluation.

Dataset Collection and Creation

The dataset is a substantial compilation of unprocessed data that requires training to extract meaningful information. Our system's dataset comprises images of various plant leaves, categorized as either healthy or falling into one of the diseased categories, including:

- Early Blight
- Bacterial Spot
- TYLCV

Achieving optimal accuracy and precision in our disease identification model is contingent upon the availability of a substantial dataset. Our dataset is comprised of 200 leaf images, with 50 images assigned to each of the following categories: healthy, early blight, bacterial spot, and TYLCV. A crucial step in maintaining uniformity involves resizing and refining all images to consistent quality and dimensions. This dataset, functioning as our training data, serves as the foundation for initiating the digital image processing phase.



Digital Image Processing

The digital image processing phase is a multifaceted mechanism, integrating various stages and algorithms. Its primary objectives encompass Training on Pre-Composed Images: The system undergoes training on pre-composed and collected images stored in our dataset. Testing Captured Plant Leaf Image, The system tests images captured by the camera to ascertain whether the plant leaf is diseased or healthy.

Web Interface

The collected dataset, constituting the training data, undergoes training against our image processing model. Subsequently, this model is preserved and employed to test images captured through the camera. The incorporation of a web interface is imperative when a user-friendly platform is needed for uploading images to the front-end. The model is pre-trained on the dataset in the back-end, and the results are generated on the user interface. This design eliminates the necessity for users to navigate between the training and testing phases of the system, ensuring a seamless flow of control. Users can interact effortlessly with the Plant Disease Identification System without requiring an in-depth understanding of the entire underlying mechanism. The intricate process of training the model and evaluating its performance occurs seamlessly in the backend. To facilitate user interaction without friction, we employ a frontend application. This application plays a pivotal role in analyzing images sourced from various outlets, be it external cameras or downloads from the internet. The test images uploaded undergo a comprehensive examination through backend procedures, and the outcomes are promptly displayed through the user interface. This interface serves as the conduit for efficiently presenting results as we analyze these images.

The Web Interface assumes a critical role by establishing a seamless connection between our backend model and the user's camera or phone in the frontend. This integration guarantees a user-friendly and efficient experience, enabling users to effortlessly engage with the Plant Disease Identification System. The frontend acts as the portal for users to submit images for analysis, while the backend, where intricate processing takes place, remains transparent to the user. The Web Interface functions as a mediator that encapsulates the elaborate processes within the backend, shielding users from the intricacies involved in model training, image processing, and disease identification. Users encounter a user-friendly environment that streamlines the entire process, from submitting images to presenting results. The interface serves as an intuitive platform, sparing users from delving into the complexities of the underlying technology. Through this pioneering integration of frontend and backend components, our system ensures accessibility and usability. Users can harness the capabilities of our disease identification model without grappling with technical intricacies.

The Web Interface not only elevates user experience but also contributes to the widespread adoption and applicability of the Plant Disease Identification System. It stands as a testament to our dedication to user-centric design, making cutting-edge technology accessible and advantageous for a broader audience.

METHODOLOGY

The digital image processing phase in our plant disease identification system is a complex and nuanced process that involves a sequence of steps and algorithms orchestrated in a controlled flow. The following is a more detailed explanation of each stage.

Image Resizing

The first step is to resize the test image to match the dimensions of the training images. This ensures that the image is processed in a consistent manner and that the features extracted from the image are comparable to those extracted from the training images. We use the imresize() function in MATLAB to resize the image. This function ensures that the pixel values in the image are not altered during resizing.

Smoothing

Once the image has been resized, it is smoothed to reduce noise and improve the visual clarity of the image. We use the RGB2GRAY() function to convert the image to grayscale before smoothing. This is because grayscale images are less susceptible to noise than colored images. We use the median filter to smooth the image. This filter works by replacing each pixel value with the median of the pixel values in its neighborhood. This helps to reduce noise without blurring the image too much.

Noise Filtering

Even after smoothing, there may still be some noise present in the image. To remove this noise, we use a noise filtering technique called the median filter. The median filter works by replacing each pixel value with the median of the pixel values in its neighbourhood. This helps to reduce noise without blurring the image too much.

Feature extraction is the process of identifying and extracting the most important features from an image. These features can then be used to classify the image into different categories. We use two feature extraction techniques in our system. Histogram of Oriented Gradients (HOG): HOG is a feature extraction technique that is commonly used for object detection. It works by dividing the image into small segments and computing a histogram of the gradient orientations in each segment. This histogram provides a unique representation of the image that can be used to identify the object in the image. Gray Scale Co-occurrence Matrix (GLCM): GLCM is a feature extraction technique that is commonly used for texture analysis. It works by computing the frequency of occurrence of different pairs of pixel values in the image. This matrix provides a representation of the texture of the image that can be used to classify the image into different categories. Once the features have been extracted, they are used to train a machine learning classifier. We use the K-nearest neighbors (KNN) classifier. KNN is a simple but effective classifier that works by finding the K most similar images in the training dataset to the test image and then assigning the test image to the class that is most common among the K nearest neighbours. The overall process of the digital image processing phase is as follows, Resize the test image to match the dimensions of the training images. Smooth the image to reduce noise and improve visual clarity. Remove noise from the image using a noise filter. Extract features from the image using HOG and GLCM. Train a KNN classifier on the extracted features from the training images. Classify the test image using the trained KNN classifier. This process is repeated for each test image. The results of the classification are used to determine whether the plant leaf is diseased or not.

Advantages of the Proposed Approach

The proposed approach has several advantages over other approaches to plant disease identification, It is robust to noise in the images, It is able to identify a wide range of plant diseases, It is relatively fast and efficient, It is easy to implement. The digital image processing phase is a critical component of our plant disease identification system. It allows us to extract features from the images that can be used to classify the images into different disease categories. The proposed approach is robust, accurate, and efficient, making it a promising solution for plant disease identification.

RESULTS

This study utilizes a dataset comprising ten types of diseases commonly found on tomato plants, including bacterial spot, early blight, late blight, leaf Mold, and others. These diseases typically manifest on the leaf's surface. The dataset contains a total of 10,000 images, with 70% (7,000 images) forming the training dataset and the remaining 30% (3,000 images) serving as the testing dataset. Figure 6 specifically illustrates images portraying bacterial spot, early blight, late blight, and leaf Mold. To eliminate the image background and preserve essential leaf characteristics, a bitwise operation is performed using a mask derived from the saturation thresholding of the image. This process ensures minimal background colour, allowing for the extraction of crucial data features. A comprehensive set of 298 features is extracted from each individual image, including

HSV Histogram, Hara lick textures, and colour moments within the RGB colour space. These features provide valuable information for subsequent disease classification.

An analysis comparing ELM with models such as SVM and decision trees has been conducted. The dataset's features and labels remain unchanged before this comparison. Detailed parameters for these models are provided in Table 4. The accuracy distribution across each model is visually represented through a heatmap, presenting a two-dimensional view in the form of a confusion matrix. This heatmap uses colors to visually summarize information. Disease classes are represented numerically, starting from '0' for healthy leaves and incrementing for other diseases, for bacterial spot. Post-testing, the classification accuracy of ELM is observed to be 84.94%, calculated by combining the accuracy of each class depicted in the heatmap, where darker shades indicate higher accuracy.

CONCLUSION

A groundbreaking plant disease detection system has emerged, promising heightened accuracy, efficiency, and user-friendliness compared to its predecessors. This innovation has the potential to revolutionize the tomato farming industry, empowering farmers to achieve higher yields while simplifying disease detection processes. Leveraging advanced technologies such as image processing and machine learning, the system excels in identifying the three most prevalent tomato leaf diseases: early blight, bacterial spot, and curl. Its user-friendly interface allows farmers to capture an image of a tomato leaf, and the system promptly provides information on the leaf's health status and specifies the type of disease, if present. Structured around four integral modules, the system seamlessly integrates various stages: pre-processing, segmentation, feature extraction, and classification. The pre-processing module optimizes the image for analysis, the segmentation module precisely identifies the leaf region, the feature extraction module discerns relevant features from the leaf region, and the classification module employs a machine learning classifier to accurately categorize the leaf disease. In a comprehensive comparative study against existing systems, this innovative approach demonstrated superior accuracy, precision, and expeditious implementation. Its potential to simplify the lives of farmers and enhance tomato yields is evident, marking a significant advancement in agricultural technology. Beyond its immediate benefits for tomato farmers, this system holds promise for the broader agricultural sector, potentially contributing to increased tomato production and mitigating concerns surrounding food insecurity. In summary, the proposed tomato plant leaf disease detection system stands out as a transformative technology with the capacity to make a substantial impact on the tomato farming industry and, by extension, the entire agricultural sector.

REFERENCES

[1]. Sankaran, S.; Mishra, A.; Ehsani, R. “A review of advanced techniques for detecting plant diseases”. Comput.Electron. Agric. 2010, 72, 1–13. [CrossRef]

[2]. Chaerani, R.; Voorrips, R.E. “Tomato early blight (Alternariasolani): The pathogens, genetics, and breeding for resistance”, J. Gen. Plant Pathol. 2006, 72, 335–347. [CrossRef]

[3]. Alvarez, A.M.” Integrated approaches for detection of plant pathogenic bacteria and diagnosis of bacterial diseases”, Annu. Rev. Phytopathol. 2004, 42, 339–366. [CrossRef] [PubMed]

[4]. Gutierrez-Aguirre, I.; Mehle, N.; Delic, D.; Gruden, K.; Mumford, R.; Ravnikar, M. “Real-time quantitative PCR based sensitive detection and genotype discrimination of Pepino mosaic virus”,. J. Virol. Methods 2009, 162, 46–55. [CrossRef] [PubMed]

[5]. Martinelli, F.; Scalenghe, R.; Davino, S.; Panno, S.; Scuderi, G.; Ruisi, P.; Villa, P.; Stropiana, D.; Boschetti, M.; Goudart, L.; et al. “Advanced methods of plant disease detection. A review” ,Agron.Sust. Dev. 2015, 35, 1–25. [CrossRef]

[6]. Bock, C.H.; Poole, G.H.; Parker, P.E.; Gottwald, T.R. “Plant Disease Sensitivity Estimated Visually, by Digital Photography and Image Analysis, and by Hyperspectral Imaging”. Crit. Rev. Plant Sci. 2007, 26, 59–107.

[7]. Down To Earth's print edition dated April 1-15, 2019

[8]. Lefeuvre, Pierre, Darren P. Martin, Gordon Harkins, Philippe Lemey, Alistair JA Gray, Sandra Meredith, Francisco Lakay et al. "The spread of tomato yellow leaf curl virus from the Middle East to the world." PLoS pathogens 6, no. 10 (2010).

[9]. India Agriculture GDP report from Ministry of Statistics and Programme Implementation (MOSPI)

[10]. L. S. Puspha Annabel, T. Annapoorani and P. Deepalakshmi, “Machine Learning for Plant Leaf Disease Detection and Classification – A Review,” 2019 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 2019, pp. 0538-0542.

[11]. J. D. Pujari, R. Yakkundimath and A. S. Byadgi, “Identification and classification of fungal

[12]. disease affected on agriculture/horticulture crops using image processing techniques,” 2014 IEEE International Conference on Computational Intelligence and Computing Research, Coimbatore, 2014, pp. 1-4.

[13]. R. Hari Shankar, A. K. Veeraraghavan, Uvais, K. Sivaraman and S. S. Ramachandran, "Application of UAV for Pest, Weeds and Disease Detection using Open Computer Vision," 2018

[14]. International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India,2018, pp. 287-292.

[15]. Kumar, Meeta, et al. “Survey on techniques for plant leaf classification.” International Journal of Modern Engineering Research 1.2 (2011): 538-544.

[16]. Fuentes, Alvaro & Lee, Yujeong& Hong, Youngki& Yoon, Sook& Park, Dong. (2016). Characteristics of Tomato Plant Diseases - A study for tomato plant disease identification. Babatunde, Oluleye H., et al. "A genetic algorithm-based feature selection." (2014).

[17]. F. Y. A. Rahman, S. R. M. S. Baki, A. I. M. Yassin, N. M. Tahir and W. I. W. Ishak, "Monitoring of Watermelon Ripeness Based on Fuzzy Logic," 2009 WRI World Congress on Computer Science and Information Engineering, Los Angeles, CA, 2009, pp. 67-70.

[18]. P. Pawar, V. Turkar and P. Patil, "Cucumber disease detection using artificial neural network," 2016 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, 2016, pp. 1-5.

[19]. Zhang, Harry. "The optimality of naive Bayes." AA 1.2 (2004): 3.

[20]. Z. N. Reza, F. Nuzhat, N. A. Mahsa and M. H. Ali, "Detecting jute plant disease using image processing and machine learning," 2016 3rd International Conference on Electrical Engineering

[21]. S. Ramesh et al., "Plant Disease Detection Using Machine Learning," 2018 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C), Bangalore, 2018, pp. 41-45.

[22]. Meenakshi K, Safa.M, Krthick T, Sivaranjani N.” A novel study of machine learning algorithms for classifying health care data”, Research Journal of Pharmacy and Technolgy, Research Journal of Phamacy and Technology, 2017

[23]. Meenakshi K, Maragatham, G, Agarwal, N., Ghosh, I, “ A Data mining Technique for analysing and predicting the success of movie”, Journal of Physics: Conference Series, Vol 1000, Issue 1, 2018

[24]. Saranya, G., & Pravin, A. (2020). A comprehensive study on disease risk predictions in machine learning. International Journal of Electrical and Computer Engineering (IJECE), 10(4), 4217.

[25]. Saranya, G., Geetha, G., & Safa, M. (2017). E-antenatal assistance care using decision tree analytics and cluster analytics based supervised machine learning. 2017 International Conference on IoT and Application (ICIOT)

[26]. G Geetha, M Safa, C Fancy, D Saranya “A hybrid approach using collaborative filtering and content based filtering for recommender system”, Journal of Physics: Conference Series Vol 1000, Issue 1, 2018.

[27]. M Srivastava, S Pallavi, S Chandra, G Geetha,"Comparison of optimizers implemented in Generative Adversarial Network (GAN)" International Journal of Pure and Applied Mathematics Volume 119 Issue 12,2018