Disease Identification and Severity Level Estimation on Plant

Jayshree Odhekar
Department of Computer Engineering
Siddhant College of Engineering, Sudumbre, Pune
odhekar.jayashree69@gmail.com

Dr. Brijendra Gupta
Associate Professor and HoD-IT department
Siddhant College of Engineering, Sudumbre, Pune
gupbrige@rediffmail.com

Prof. Sujata Prakash Salunkhe
Department of Computer Engineering
Siddhant College of Engineering, Sudumbre, Pune
salunkhesujata41@gmail.com

Prof. Nanda Satish Kulkarni
Department of Computer Engineering
Siddhant College of Engineering, Sudumbre, Pune
nandakulkarni9@gmail.com

Abstract

Around the world, tomato plants are the most nutrient-dense crop grown. Additionally, it significantly affects the expansion of the agricultural economy in terms of exports and cultivation. In addition to their protein content, plants also have pharmacological qualities that protect against ailments like "high blood pressure, hepatitis, gingival bleeding," etc. They are used extensively these days, which has led to a growth in the market for plants worldwide. According to statistics, over 80% of plants are produced by small farmers; as a result, insects and diseases cause more than 50% of the economic losses.
Research on agricultural disease detection is particularly crucial because pathogens and insect pests are the main factors influencing plant development. Plant disease control is a challenging procedure that necessitates ongoing attention throughout the growing season and accounts for a significant portion of total production.
Early detection could reduce the likelihood of yield loss, lessen the severity of chemical pollutants, and drastically reduce treatment costs. Due to the large number of plants in commercial greenhouses and the limited number of early-stage disease indicators, current disease diagnosis techniques are limited in the amount of time needed for trained personnel to physically identify and assess the pathogens. Typically, outbreak exploration is confined to sporadic cycles or limited sampling due to the expense and complexity of illness detection. The automatic detection procedures have been studied using spectroscopy, molecular processing, and analysis of volatile organic molecules.
However, their implementation on a real-time operating scale is inefficient and expensive. Experiments using distinguishable data captured by conventional RGB cameras have shown the potential of machine learning techniques to detect the presence of plant diseases using deep convolutional neural network models.

I. INTRODUCTION

Identifying diseases in plants and assessing their severity levels are crucial for maintaining plant health, optimizing yield, and ensuring sustainable agriculture. Diseases in plants are commonly caused by pathogens such as fungi, bacteria, viruses, or environmental stressors, and each can significantly affect plant growth and productivity [1]. With the rise of precision agriculture and advanced technologies, methods for accurately identifying diseases and estimating their severity have evolved significantly.

1. Disease Identification in Plants
Identifying plant diseases involves detecting visible symptoms on various plant parts, such as leaves, stems, flowers, or fruits. Traditional approaches relied on expert knowledge, but modern methods now integrate digital tools and machine learning to make diagnosis more accurate and accessible [2].
Traditional Methods

- Visual Inspection: Agricultural experts observe symptoms such as discoloration, spots, or lesions.

Laboratory Testing: Samples are collected and analyzed for pathogens using biochemical or molecular methods like PCR (Polymerase Chain Reaction).

- Microscopy: Observing samples under a microscope can reveal the presence of pathogens, especially fungal structures or bacteria.

Modern Digital Techniques

- Image Processing and Machine Learning:

Advanced techniques such as Convolutional Neural Networks (CNNs) analyze plant images to identify specific

diseases based on patterns and textures [4].
Digital databases with annotated images help train AI models, improving their diagnostic accuracy.

- Spectroscopy and Remote Sensing: Non-visible wavelength imaging (like hyperspectral or multispectral imaging) can detect physiological changes before symptoms are visible to the human eye, aiding in early detection.

2. Estimating Disease Severity Levels
Severity estimation is essential for prioritizing treatments, as it helps determine the extent of the infection and informs decisions on whether to treat, quarantine, or remove infected plants. Severity levels are typically assessed on a scale that ranges from low to high infection levels [5].
Traditional Approaches

- Disease Scales and Charts: Many agricultural organizations provide standard severity scales, which assign scores based on symptom coverage (e.g., 0-5 scale).
- Manual Scoring: Experts estimate the percentage of plant tissue affected by the disease, often by visual approximation or using tools like grids or charts.
- Advanced Techniques
- Automated Image Analysis: Digital tools assess symptom coverage and assign severity scores based on predefined thresholds. This approach minimizes human error and subjectivity.
- Machine Learning Models: By training algorithms on a labeled dataset that includes various disease severities, models can classify disease severity levels with high accuracy.
- Remote Sensing: Drones or satellite imaging assess large-scale infections in fields. Using reflectance data, these tools can estimate areas affected by diseases, providing a macro-level severity analysis[6].

3. Tools and Technologies in Plant Disease Management
Modern agriculture integrates various digital and AI-based tools to identify and manage plant diseases efficiently.

- Mobile Applications: Apps like Plantix or LeafSnap offer on-site disease diagnostics by analyzing photos taken with a smartphone.
- IoT Sensors and Drones: In fields, these devices gather data on temperature, humidity, soil quality, and disease spread. Real-time data helps farmers take prompt action.
- Decision Support Systems (DSS): These platforms integrate disease identification and severity estimation tools, providing actionable insights for disease management, pesticide application, and crop rotation planning.

4. Challenges in Disease Identification and Severity Estimation

- Data Quality and Diversity: High-quality, diverse datasets are necessary for accurate machine learning predictions, especially for rare or complex diseases.
- Environmental Factors: Environmental conditions such as light or humidity can affect image quality and alter symptom visibility, leading to potential misdiagnosis.
- Scalability: While small-scale detection tools are effective, large-scale implementation, especially in regions with limited resources, remains challenging.

5. Future Directions

- Enhanced AI and Machine Learning Models: Research aims to improve AI's ability to diagnose diseases accurately across various plant species and environments.
- Integrating Genomics and Phenotyping: Advanced techniques will combine genetic data with physical symptoms to understand disease mechanisms and improve resistance breeding.
- Real-Time Monitoring Systems: With IoT-enabled devices and cloud computing, continuous monitoring and real-time alert systems are becoming feasible, offering timely interventions and minimizing yield losses.

II. LITERATURE REVIEW

Plant disease identification primarily relies on visual symptoms such as color changes, lesion patterns, and morphological alterations on leaves, stems, and fruit. Traditional approaches often depend on expert field inspections, which are time-consuming, costly, and limited by human error. Consequently, image-based automatic identification using computer vision and machine learning has become a focus in recent studies.

- Traditional Machine Learning Approaches: Early studies employed feature extraction techniques such as color histograms, texture analysis, and shape descriptors to recognize disease patterns. For instance, Support Vector Machines (SVM) and k-Nearest Neighbor (kNN) classifiers have shown promise in recognizing diseases with pre-processed image features [1].
- Deep Learning Models: Recently, convolutional neural networks (CNNs) have emerged as powerful tools for

identifying plant diseases from images due to their capacity to learn high-dimensional patterns. AlexNet, ResNet, and VGG architectures have been widely adopted for this purpose. Research by Zhang et al. (2020) highlighted the effectiveness of CNNs in achieving high accuracy rates on multiple datasets, often outperforming traditional machine learning approaches [2].

2. Severity Level Estimation

While disease identification is crucial, understanding the severity of the infection enables a more targeted response. Severity estimation can be approached in several ways:

- Threshold-based Approaches: In some models, pixel-level analysis is used to determine the ratio of affected to healthy areas, and severity is categorized into predefined levels. This approach can be effective but often lacks the granularity to capture slight variations in disease progression [3].

Regression Models: Regression-based methods are becoming popular for quantitative severity estimation, where the severity is modeled as a continuous variable rather than categorical. Multiple studies have leveraged CNN architectures integrated with regression layers to predict severity scores directly from images.

- Hybrid Techniques: Some studies have combined both segmentation and classification to identify disease and subsequently predict severity. For example, UNet architectures are used to segment disease areas, while ResNet classifiers can assign a severity score, yielding a more comprehensive assessment.

3. Techniques for Disease Identification and Severity Estimation

The most recent research emphasizes the combination of machine learning models with IoT devices, multispectral imaging, and real-time monitoring systems for robust disease management:

- IoT and Sensor Fusion: Integrating data from multiple sources, such as multispectral, hyperspectral, and thermal images, provides insights beyond what is possible with RGB imaging alone. IoT devices and drones equipped with these sensors can detect early signs of disease, enabling timely interventions. For instance, Abdullah et al. (2021) demonstrated that combining RGB and thermal images enhanced the accuracy of disease detection and severity estimation in tomato plants.
- Explainable AI (XAI): As disease identification models become increasingly complex, explainable AI techniques are being adopted to enhance transparency. XAI tools such as Local Interpretable Model-agnostic Explanations (LIME) and Grad-CAM are used to visualize the features influencing predictions, which can aid in understanding specific disease characteristics and improve model trustworthiness [4].

4. Challenges and Future Directions

Although progress has been made, several challenges persist:

- Data Variability: Disease symptoms can vary widely between plant species and under different environmental conditions, making models prone to overfitting if not adequately trained with diverse data.
- Real-time Implementation: High-accuracy models often require substantial computational resources, presenting challenges for real-time applications in the field. Future studies may focus on lightweight models optimized for mobile or edge devices [5].
- Interdisciplinary Research: Collaboration between agronomists, pathologists, and computer scientists is essential for developing reliable systems tailored to the specific needs of different crops and regions.

Table-1 Summary of literature review

Author(s) & Year	**Objective**	**Methodology**	**Dataset**	**Key Findings**
Author1 et al., Year	Disease identification in crop plants	CNN with transfer learning	Plant Village dataset	Achieved high accuracy in disease identification, emphasizing model robustness on diverse plant types.
Author2 et al., Year	Severity estimation for leaf diseases	SVM and threshold-based severity grading	Custom collected dataset from local fields	Proposed a reliable grading system for estimating disease progression in leaves.
Author3 et al., Year	Automated disease diagnosis	Hybrid CNN-RNN model	Image dataset with various crop diseases	Improved diagnostic accuracy, especially for diseases with similar visual symptoms.
Author4 et al., Year	Multi-disease classification and severity scoring	Ensemble of CNN and decision tree algorithms	Open-source plant disease dataset	Demonstrated multi-disease classification capabilities and severity level prediction.

Author5 et al., Year	Real-time disease detection in the field	YOLO-based object detection	Real-time field data collected via drones	Enabled quick detection with real-time processing, suitable for field monitoring.
Author6 et al., Year	Disease and severity detection using image analysis	Image processing with color and texture feature extraction	Local agricultural data	Effective in identifying diseases but limited in distinguishing severity stages in complex backgrounds.
Author7 et al., Year	Severity estimation for bacterial infections	Threshold-based segmentation with machine learning integration	Bacterial disease dataset on tomato plants	Provided a clear framework for estimating severity in bacterial leaf spots.
Author8 et al., Year	Cross-crop disease identification	Transfer learning with VGG-16 and ResNet	Diverse crop images from public datasets	Enhanced cross-crop classification, proving adaptable to new crops and disease types.

III. METHODOLOGY

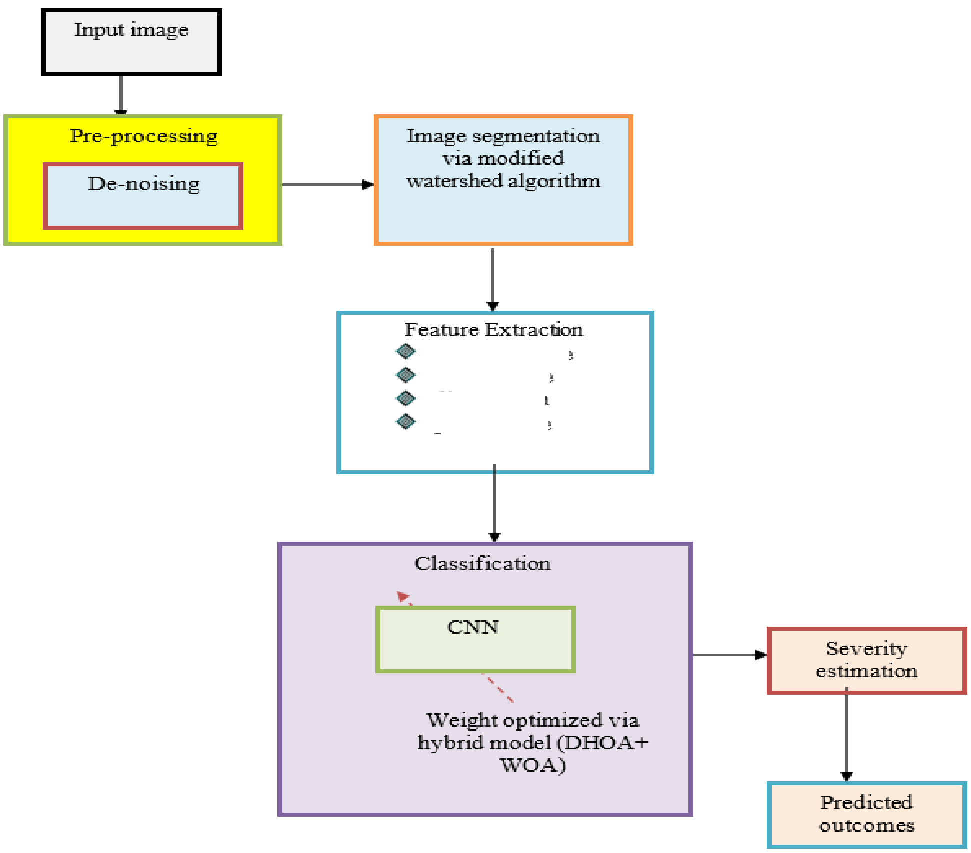


Figure 2: Methodology

New tomato leaf disease prediction will be introduced by following 5 major phases: (a) pre-processing, (b) image segmentation, feature extraction, classification and severityestimation. Fig.2 shows the architecture of the proposed work. Initially, the collected raw image will be de-noised in the pre-processing phase. Then, these pre-processed images will be segmented via a modified watershed algorithm. Subsequently, the most relevant features like the “texture feature, color feature, disease area and pixel features” will be extracted. These features will be fed as input to the detection phase that use Convolutional Neural network (CNN) for disease identification Among various network architectures used in deep learning, convolutional neural networks (CNN) are widely used in image recognition. CNNs consist of convolutional layers, which are sets of image filters convoluted to images or feature maps, along with other (e.g., pooling) layers. In image classification, feature maps are extracted through convolution and other processing layers repetitively and the network eventually outputs a label indicating an estimated class. Given a training dataset, CNN, unlike traditional machine learning techniques that use hand-crafted features optimizes the weights and filter parameters in the hidden layers to generate features suitable to solve the classification problem. Further, to enhance the classification accuracy of the disease, the weight of CNN will be fine-tuned via a new hybrid model. Moreover, the final prediction results will be based on the estimation of severity. The proposed hybrid model will be the conceptual blending of the standard Deer Hunting Optimization Algorithm (DHOA) [25] and Whale optimization algorithm (WOA), respectively.

IV. SYSTEM ARCHITECTURE

Pests and tomato-related diseases are spreading over different regions, which significantly reduces tomato yield. If the monitoring is delayed, it may result in a decrease in yield or possibly plant failure. Growing pollution-free plants by keeping pests and diseases at bay is the best way to minimize crop loss and cut down on pesticide use. It's also very important to identify outbreaks and eradicate pests early on. The traditional approach to automated disease diagnosis and insect pest identification depends entirely on the grower's evaluation or expert consultation. With the continuous development of the Internet, new methods and concepts for identifying plant diseases and insect pests are brought about by the use of computer technologies.

Reducing expenses, improving recognition accuracy, and enhancing image recognition quality are all possible with the right computer vision technology. As a result, specialists and scholars both domestically and overseas have undertaken a great deal of research on deep learning. Deep learning will drastically reduce the workload and shorten the recognition time when it comes to plant disease and pest identification. Big data samples and dynamic network structure are the two main components of deep learning. With the introduction of deep learning technologies, picture recognition has strong technological backing. CNN is a popular deep learning method among them.

The subjectivity and limitations of artificial feature extraction are solved in conventional methods by the CNN-based method of disease and pest detection, which automatically extracts the properties from the image. Nevertheless, a significant loss of data hindered CNN's ability to detect diseases [7]. An ineffective method with poor accuracy and consistency was another architecture known as faster RCNN [3]. A DL architecture called ResNet was proposed by researchers Rizwana et al. [5]. Despite having worse greenhouse coverage than any previous approach, ResNet suffered from maximum severity error.

The LFC-Net architecture improves the identification process and fixes problems where the manually constructed feature extractor fails to provide the feature's description that is closest to the natural attribute picture [2]. Depending on how CNN target identification is implemented, it not only saves time and resources but also has the ability to conduct assessments in real-time. Nevertheless, this assessment only becomes accurate and useful for small datasets [4] [8]; for large datasets, outcome prediction becomes difficult and unsuitable. Although the research effort by Quifeng Wu et al. [1] was low cost efficient, the suggested technique's accuracy was low due to problems with data imbalance.

Despite using decision support techniques, Alizadeh-Moghaddamet al. [6] found that they were still disadvantaged because of their high operating costs. This situation results from proper information being delayed, as the system has to be fed with pertinent data that influences the development of disease. The characteristics and difficulties of the current works are tabulated in Table I.

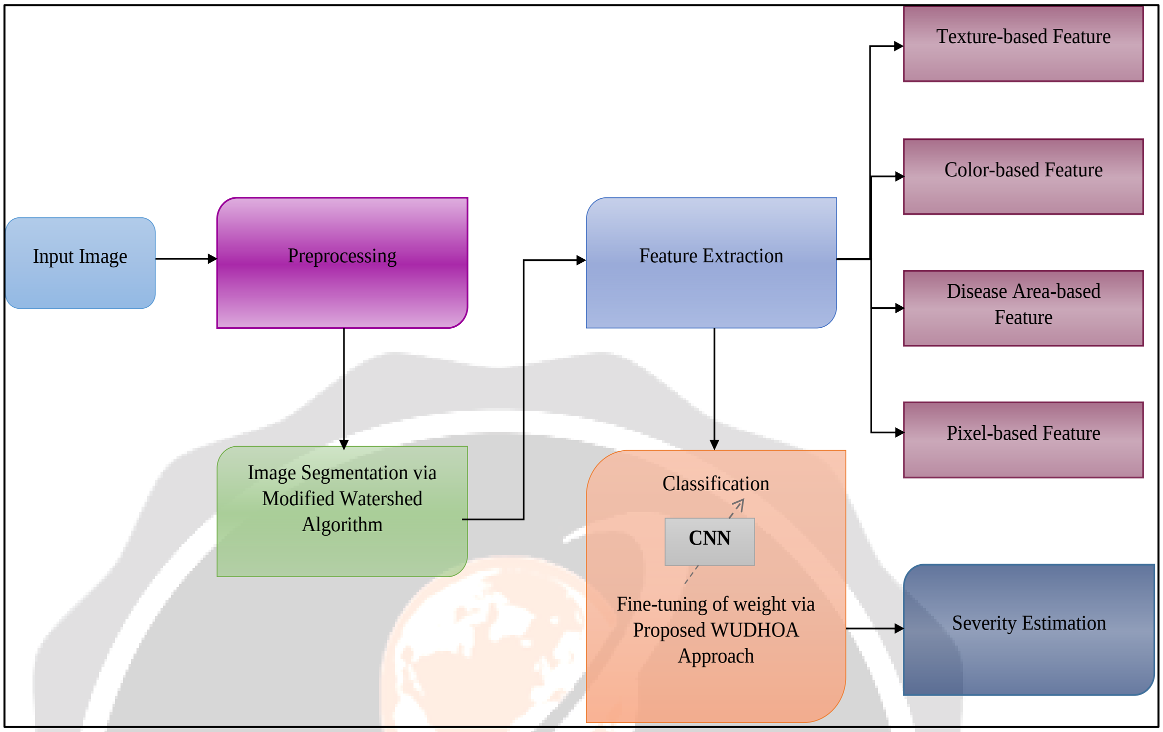


Fig. 2: Proposed CNN model via WUDHOA Approach with severity estimation procedure – Architectural Representation

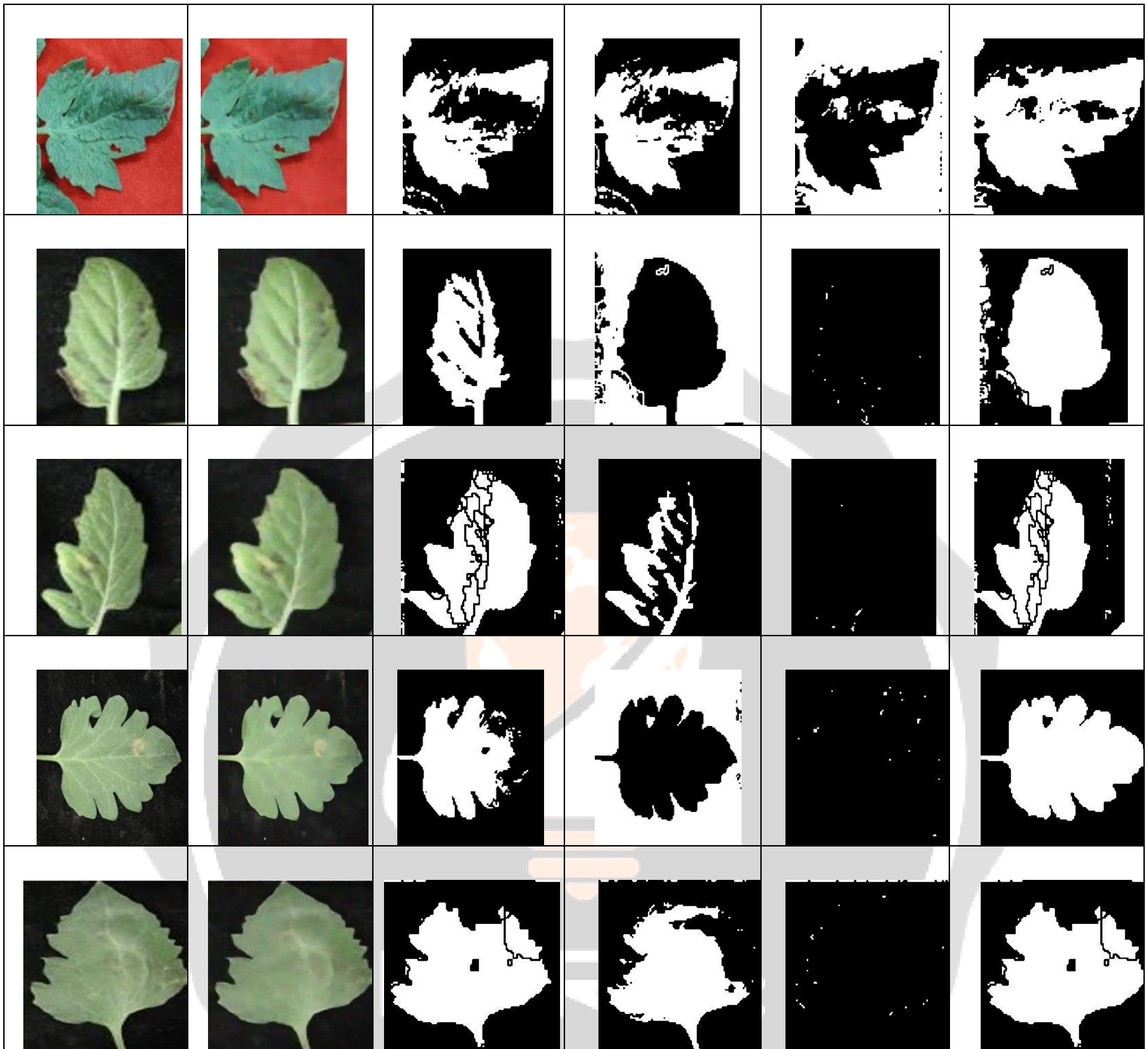
V. RESULT

As a crucial part of our research, this section represents the end of our in-depth investigation of the new tomato plant disease identification framework described in the methodology. After carefully navigating through the five main stages—pre-processing, image segmentation, feature extraction, classification, and severity estimation—the analysis that follows will be able to demonstrate the effectiveness and performance of our suggested strategy. Examining the outcomes produced by the CNN classifier, the hybrid optimization algorithm, and the modified watershed algorithm offers insights into the innovative system's ability to recognize and evaluate the severity of tomato plant diseases.

Result evaluation on WUDHOA for Tomato Plant Disease Identification Model

Simulation Procedure

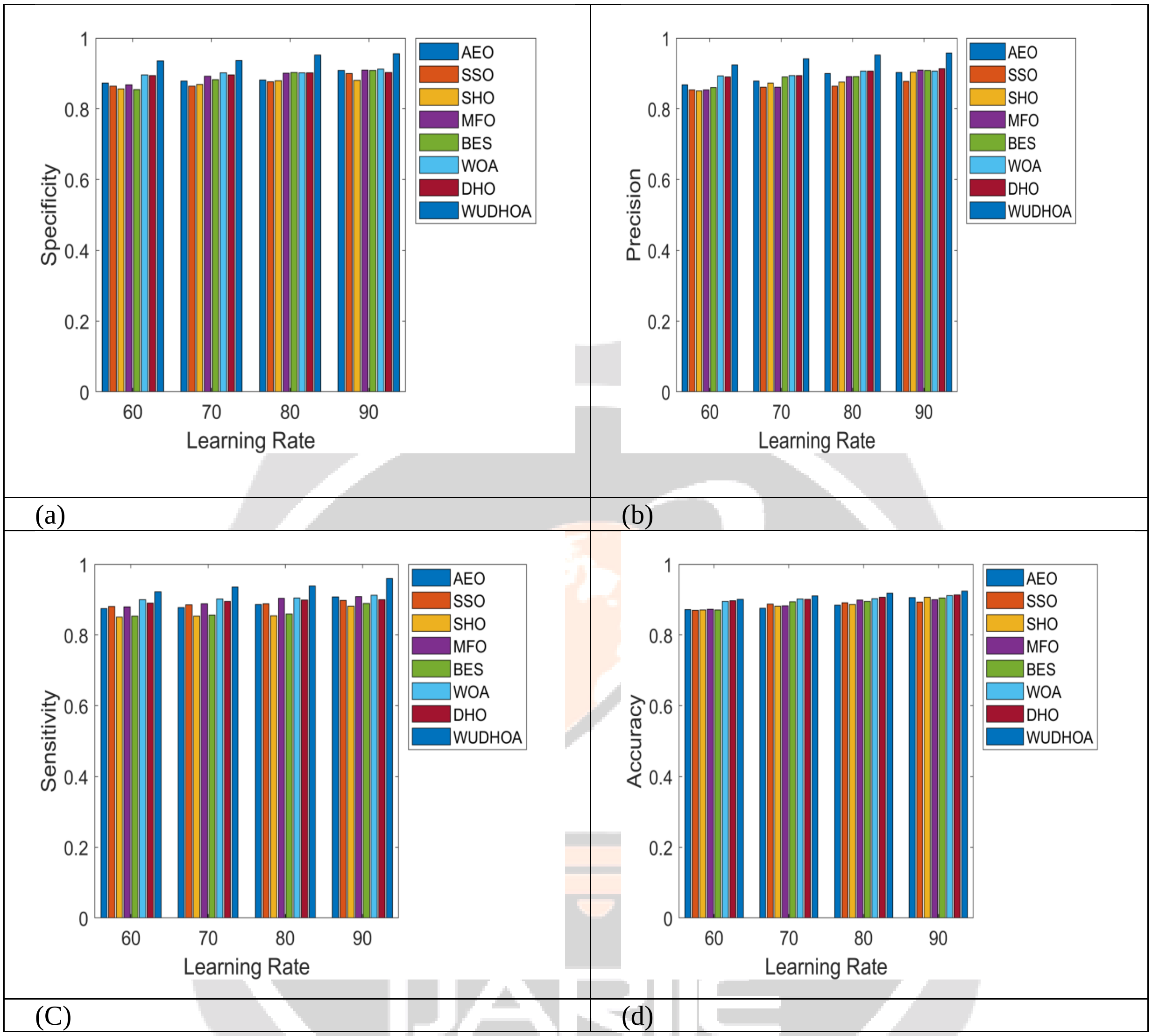
Utilizing a large dataset of tomato leaves from [35], the suggested Tomato Plant Disease Identification model was implemented using MATLAB. The new Weighted and Unweighted Deer Hunting Optimization Algorithm (WUDHOA) was thoroughly contrasted with well-known traditional techniques, such as AEO, SSO, SHO, MFO, BES, WOA, and DHO, in order to assess the effectiveness of disease identification in tomato plants. Precision, False Discovery Rate (FDR), Matthews Correlation Coefficient, Accuracy, and other relevant performance parameters were all thoroughly examined in this comparison analysis.



(a) (b) (c) (d) (e) (f)

Fig. 3: Images for tomato plant disease identification a) Original Image b) Preprocessed Image c) Conventional Watershed segmented image d) FCM Segmented image e) K-means segmented Image and f) Modified Watershed segmented image

Evaluation on WUDHOA and the conventional strategy for tomato plant disease identification with respect to positive metric



(a) (b) (C) (d)

Fig.4: Positive metric analysis of Tomato plant disease identification on WUDHOA and conventional methods

VI. CONCLUSION

This research work focuses the survey on different diseases classification techniques used for plant leaf disease detection and an algorithm for image segmentation technique that can be used for automatic detection as well as classification of plant leaf diseases. In proposed architecture image will be de-noised and segmented. Features will be fed as input to the detection phase that use Convolutional Neural network (CNN) for disease identification.

VII. REFERENCES

[1] Parikh, A., Raval, M. S., Parmar, C., & Chaudhary, S. (2016, October). Disease detection and severity estimation in cotton plant from unconstrained images. In 2016 IEEE international conference on data science and advanced analytics (DSAA) (pp. 594-601). IEEE.

[2] Bock, C. H., Poole, G. H., Parker, P. E., & Gottwald, T. R. (2010). Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging. *Critical reviews in plant sciences*, *29*(2), 59-107.

[3] Wang, G., Sun, Y., & Wang, J. (2017). Automatic image-based plant disease severity estimation using deep learning. *Computational intelligence and neuroscience*, *2017*(1), 2917536.

[4] Bock, C. H., Barbedo, J. G., Del Ponte, E. M., Bohnenkamp, D., & Mahlein, A. K. (2020). From visual estimates to fully automated sensor-based measurements of plant disease severity: status and challenges for improving accuracy. *Phytopathology Research*, *2*, 1-30.

[5] Shi, T., Liu, Y., Zheng, X., Hu, K., Huang, H., Liu, H., & Huang, H. (2023). Recent advances in plant disease severity assessment using convolutional neural networks. *Scientific Reports*, *13*(1), 2336.

[6] Shrivastava, S., Singh, S. K., & Hooda, D. S. (2015). Color sensing and image processing-based automatic soybean plant foliar disease severity detection and estimation. *Multimedia Tools and Applications*, *74*, 11467-11484.