Kavithra R, 2026, 14:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology

An Open Access Journal

AI Vision-Based Approach for Detecting Cotton Leaf Diseases

Kavithra R[1], Lakshminarayanan A[2], Allen joses A[3],Pavin S[4], Saranraj E[5]
[1]Assistant Professor, Department of Computer Science and Engineering United Institute of Technology , Tamilnadu,India
[2,3,4,5]Department of Computer Science and Engineering United Institute of Technology ,Tamilnadu ,India

Abstract- Cotton is one of the most important cash crops, playing a vital role in the agricultural economy. However, cotton plants are highly susceptible to various leaf diseases such as bacterial blight, leaf spot, and aphid infestations, which significantly reduce crop yield and quality. Traditional disease detection methods rely on manual inspection by farmers and experts, which is time-consuming, labor-intensive, and prone to human error. To address these challenges, this project proposes an automated cotton leaf disease detection system using deep learning techniques. A Convolutional Neural Network (CNN)-based model is trained on a labeled dataset of cotton leaf images to accurately classify different disease categories. The system is integrated into a user-friendly web application that allows users to upload leaf images and obtain real-time predictions along with disease classification and affected percentage.

Keywords: Cotton crop, leaf disease detection, deep learning, Convolutional Neural Network, image classification, plant pathology, bacterial blight, leaf spot, aphid infestation, agricultural technology, crop yield, automated diagnosis, web application, real-time prediction, precision agriculture.

I. INTRODUCTION

Agriculture is one of the most important sectors contributing to the global economy, and it plays a crucial role in ensuring food security and sustainable development. Among various crops, cotton holds significant economic value as it is widely used in the textile industry and serves as a major source of income for farmers in many countries. However, cotton production is often affected by various plant diseases and pest infestations, which can significantly reduce crop yield and quality. Common cotton leaf diseases such as bacterial blight, Alternaria leaf spot, and pest attacks like aphids pose serious challenges to farmers, especially when they are not detected at an early stage.

Traditionally, the identification of plant diseases has relied on manual inspection by farmers or agricultural experts. This method involves visual observation of leaf symptoms such as discoloration, spots, or deformation. Although this approach is widely practiced, it has several limitations. It is time-consuming, labor-intensive, and highly dependent on the expertise of the individual. In many rural areas, access to agricultural experts is limited, which leads to delayed diagnosis and improper treatment. Furthermore, human observation can be subjective and prone to errors, especially when multiple diseases exhibit similar visual symptoms.

In recent years, advancements in artificial intelligence and machine learning have opened new avenues for automating plant disease detection. Image processing techniques initially provided some level of automation by analyzing color, texture, and shape features of leaves. However, these methods required manual feature extraction and were not robust enough to handle variations in environmental conditions such as lighting, background complexity, and leaf orientation.

The emergence of deep learning, particularly Convolutional Neural Networks (CNNs), has significantly improved the accuracy and efficiency of image-based classification tasks. CNNs have the ability to automatically learn hierarchical features

directly from raw image data, making them highly suitable for plant disease detection. By analyzing patterns, textures, and color variations in leaf images, CNN models can accurately classify different types of diseases without the need for manual feature engineering. This capability has led to the development of intelligent systems that can assist farmers in identifying diseases quickly and accurately.

In this project, a deep learning-based approach is proposed for the detection and classification of cotton leaf diseases. The system utilizes a CNN model trained on a dataset of labeled cotton leaf images, including both healthy and diseased samples. To improve model performance and generalization, preprocessing and data augmentation techniques are applied. The trained model is then integrated into a web-based application that allows users to upload leaf images and receive real-time predictions regarding the type of disease present.

The proposed system aims to provide a practical and efficient solution for farmers by enabling early disease detection and reducing dependence on manual inspection. By leveraging modern AI techniques, this work contributes to the advancement of precision agriculture and supports data-driven decision-making in farming practices. Ultimately, the system helps improve crop productivity, reduce losses, and promote sustainable agricultural development.

In addition to disease classification, the proposed system introduces several practical enhancements that improve its real-world applicability. A key contribution of this work is the development of a real-time web-based application that enables users to upload cotton leaf images and receive instant predictions. This makes the system easily accessible to farmers without requiring specialized technical knowledge.

Furthermore, the system provides an estimation of the affected area percentage, which helps in assessing the severity of the disease and supports better decision-making for crop management.
Another important contribution is the use of a custom cotton leaf dataset that includes diverse variations in lighting conditions, background complexity, and disease patterns. This improves the

robustness and generalization capability of the model in real-world scenarios.

These features collectively distinguish the proposed approach from existing methods by combining accuracy, usability, and practical deployment, thereby contributing to the advancement of precision agriculture and intelligent farming systems.

The novelty of this work lies in integrating a CNN-based cotton leaf disease detection model with a real-time web application and disease severity estimation. Unlike traditional methods, the proposed system not only classifies diseases but also provides affected percentage analysis using a custom dataset, making it highly suitable for real-world agricultural applications.

Keywords : Cotton leaf disease, Deep learning, CNN, Image classification,Agriculture, Plant disease detection, Web application

II. RELATED WORKS

In recent years, significant research has been conducted in the field of plant disease detection using image processing, machine learning, and deep learning techniques. Early approaches to plant disease identification primarily relied on traditional image processing methods such as color segmentation, thresholding, edge detection, and texture analysis. These methods aimed to extract handcrafted features from leaf images, including color distribution, shape, and texture patterns, which were then used for classification. While these approaches provided moderate accuracy, they were highly sensitive to environmental variations such as lighting conditions, background noise, and image quality. Moreover, the manual feature extraction process required domain expertise and was not scalable for large datasets.

To overcome the limitations of traditional methods, machine learning techniques such as Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), and Decision Trees were introduced. These algorithms improved classification performance by learning patterns from labeled data. However, their effectiveness still depended heavily on the quality of manually extracted features. Additionally, these models struggled to generalize well when dealing with complex datasets containing variations in leaf orientation, size, and disease patterns.

With the advancement of deep learning, particularly Convolutional Neural Networks (CNNs), the field of plant disease detection has witnessed significant improvements. CNNs have the ability to automatically learn hierarchical features from raw image data, eliminating the need for manual feature engineering. Early works utilizing CNN architectures such as AlexNet demonstrated promising results in image classification tasks. Subsequently, more advanced architectures like VGG16, ResNet, and Inception were applied to plant disease datasets, achieving higher accuracy and robustness.

Several studies have successfully implemented CNN-based models for detecting diseases in crops such as tomato, potato, rice, and maize. For example, deep learning models trained on the PlantVillage dataset achieved classification accuracies exceeding 90%, highlighting the effectiveness of CNNs in plant disease detection. Researchers have also explored transfer learning techniques, where pre-trained models are fine-tuned on specific agricultural datasets. This approach reduces training time and improves performance, especially when the available dataset is limited.

In addition to classification, recent research has focused on object detection and localization techniques. Models based on region-based convolutional neural networks (R-CNN), Faster R-CNN, and single-stage detectors have been used to identify and localize disease regions within leaf images. These methods provide more detailed information compared to simple classification, enabling the identification of affected areas. However, such models often require high computational resources and are not always suitable for real-time applications.

Real-time disease detection systems have gained attention due to their practical applicability in agriculture. Lightweight deep learning models such as MobileNet and EfficientNet have been proposed for deployment on mobile devices and embedded systems. These models offer a good balance between accuracy and computational efficiency, making them suitable for field applications where computational resources are limited.

Furthermore, several studies have integrated deep learning models with web and mobile applications to provide user-friendly interfaces for farmers. These systems allow users to upload leaf images and receive instant predictions, making advanced technology accessible to non-experts. Some applications also provide additional features such as disease severity estimation and treatment recommendations, enhancing their practical utility.

Despite these advancements, several challenges remain in the field of plant disease detection. One major challenge is the presence of complex backgrounds in real-world images, which can affect model performance. Most existing models are trained on clean and well-labeled datasets, which may not accurately represent real field conditions. Additionally, variations in lighting, camera quality, and leaf orientation can impact detection accuracy.

Another challenge is the occurrence of multiple diseases on a single leaf, which complicates the classification process. Most models are designed for single-label classification and may not perform well in multi-label scenarios. Moreover, limited availability of high-quality annotated datasets for specific crops, such as cotton, restricts the development of highly accurate models.

This project aims to address these challenges by developing a robust CNN-based cotton leaf disease detection system. By incorporating data augmentation techniques, the model is trained to handle variations in input images. The integration of the model into a web-based application further enhances its usability, enabling real-time predictions and practical deployment in agricultural environments.

Overall, the advancements in deep learning have significantly improved plant disease detection systems, making them more accurate, efficient, and accessible. However, continuous research is required to overcome existing limitations and develop more robust and scalable solutions for real-world agricultural applications.

III. PROPOSED METHODOLOGY

The proposed system presents an automated cotton leaf disease detection framework using deep learning techniques. The system is designed to

classify cotton leaf images into different disease categories as well as identify healthy leaves.

Initially, a dataset containing images of cotton leaves is collected and organized into multiple classes such as healthy leaves, bacterial blight, and aphid-affected leaves. The dataset is then preprocessed to improve image quality and ensure consistency. Preprocessing steps include image resizing, normalization, and data augmentation techniques such as rotation, flipping, and scaling to enhance model generalization.

The processed images are fed into a Convolutional Neural Network (CNN) model for feature extraction and classification. The CNN automatically learns important features such as edges, textures, and patterns associated with different diseases. The model is trained using labeled data and optimized using appropriate loss functions and optimizers to achieve high accuracy.

After training, the model is integrated into a web-based application developed using HTML, CSS, and backend frameworks. Users can upload cotton leaf images through the interface, and the system provides real-time predictions, including disease name and affected percentage.

The overall system ensures fast, accurate, and user-friendly disease detection, making it suitable for practical agricultural applications.

IV. SYSTEM ARCHITECTURE

The system architecture of the proposed cotton leaf disease detection model consists of multiple stages, including data collection, preprocessing, model training, and prediction. Initially, cotton leaf images are collected from datasets and user inputs. These images undergo preprocessing steps such as resizing, normalization, and augmentation to improve quality and consistency.

The processed images are then passed to the Convolutional Neural Network (CNN) model, which extracts features and classifies the images into different disease categories. The trained model is integrated into a web application that allows users to upload images and obtain real-time predictions.

The output is displayed in terms of disease name and affected percentage, enabling better understanding and decision-making for farmers. The system ensures efficient and accurate detection with minimal user effort.

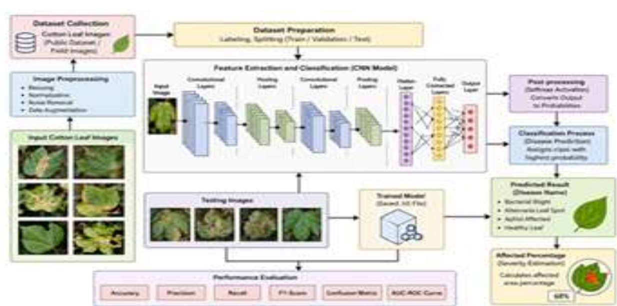

Fig. 1. System Architecture

Overall Working Flow of the Proposed System:

The proposed cotton leaf disease detection system is designed as an automated, efficient, and scalable framework that leverages deep learning techniques for accurate classification of plant diseases. The methodology consists of multiple stages, including data collection, preprocessing, model development, training, evaluation, and deployment. Each stage plays a crucial role in ensuring the robustness and effectiveness of the overall system.

A. Data Collection

The first step in the proposed methodology involves collecting a comprehensive dataset of cotton leaf images. The dataset includes images of both healthy leaves and leaves affected by various diseases such as bacterial blight, Alternaria leaf spot, and aphid infestations. The images are obtained from publicly available agricultural datasets as well as field-based sources to ensure diversity and real-world applicability. Each image is labeled according to its respective class to facilitate supervised learning.

To improve the generalization capability of the model, the dataset is curated to include variations in lighting conditions, leaf orientation, background complexity, and disease severity. This diversity ensures that the trained model performs well not only on controlled datasets but also in real-world agricultural environments.

B. Data Preprocessing

Data preprocessing is a critical step in enhancing the quality and consistency of input images. In this stage, all images are resized to a fixed resolution to maintain uniformity across the dataset. This is essential because deep learning models require consistent input dimensions.

Normalization is applied to scale pixel values to a standard range, typically between 0 and 1. This improves the convergence speed of the model during training. Additionally, noise reduction techniques are used to eliminate unwanted distortions that may affect feature extraction.

Data augmentation techniques are employed to artificially increase the size and diversity of the dataset. Common augmentation methods include rotation, horizontal and vertical flipping, scaling, cropping, and brightness adjustment. These techniques help the model become more robust to variations in input images and reduce the risk of overfitting.

C. Model Architecture

The core component of the proposed system is a Convolutional Neural Network (CNN) designed for image classification. The CNN architecture consists of multiple layers, including convolutional layers, activation functions, pooling layers, and fully connected layers.

The convolutional layers are responsible for extracting features from input images. These layers use filters to detect patterns such as edges, textures, and shapes that are characteristic of different diseases. Activation functions, such as the Rectified Linear Unit (ReLU), introduce non-linearity into the model, enabling it to learn complex relationships within the data.

Pooling layers are used to reduce the spatial dimensions of feature maps, thereby decreasing computational complexity and preventing overfitting. Max pooling is commonly used to

retain the most significant features while discarding less important information.

The fully connected layers at the end of the network perform classification based on the extracted features. A softmax function is applied in the final layer to produce probability scores for each class, allowing the model to assign the input image to the most likely disease category.

D. Model Training

The training phase involves feeding the preprocessed images into the CNN model and adjusting its parameters to minimize classification error. The dataset is divided into training and validation sets to evaluate model performance during training.

A suitable loss function, such as categorical cross-entropy, is used to measure the difference between predicted and actual labels. An optimizer, such as Adam or stochastic gradient descent (SGD), is employed to update the model weights iteratively.

During training, the model learns to identify distinguishing features of different disease classes. Techniques such as early stopping and dropout are used to prevent overfitting and improve generalization. The training process is conducted over multiple epochs until the model achieves satisfactory performance.

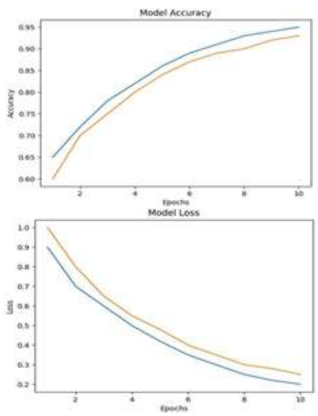

Fig. 2. Model Accuracy

E. Model Evaluation

After training, the model is evaluated using a separate test dataset that contains unseen images. This step is crucial for assessing the model's ability to generalize to new data.

Performance metrics such as accuracy, precision, recall, and F1-score are calculated to evaluate classification performance. Accuracy measures the overall correctness of predictions, while precision and recall provide insights into the model's ability to correctly identify disease cases. The F1-score balances precision and recall, offering a comprehensive evaluation metric.

Confusion matrices are also used to analyze classification results and identify potential misclassifications. This helps in understanding the strengths and limitations of the model.

F. Deployment and Web Integration

Once the model achieves satisfactory performance, it is deployed as part of a web-based application. The application is developed using frontend technologies such as HTML, CSS, and JavaScript, along with a backend framework for handling user requests.

Users can upload cotton leaf images through the web interface. The uploaded images undergo preprocessing and are then passed to the trained CNN model for prediction. The system processes the input in real time and generates output in the form of disease classification.

In addition to identifying the disease type, the system estimates the affected percentage of the leaf. This feature provides valuable information to farmers, helping them assess the severity of the disease and take appropriate action.

G. System Workflow

The overall workflow of the proposed system follows a sequential process. Initially, input images are collected and preprocessed. The processed images are then used to train the CNN model. After training, the model is evaluated using test data to ensure accuracy and reliability. Finally, the trained model is deployed in a web application for real-time disease detection.

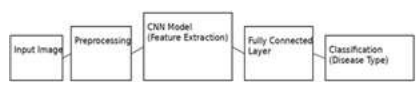

Fig. 3. System Workflow

This structured workflow ensures that the system is efficient, accurate, and user-friendly. The integration of deep learning with a web-based platform makes the solution accessible to farmers and agricultural stakeholders.

H. Advantages of the Proposed Method

The proposed methodology offers several advantages over traditional approaches. It eliminates the need for manual feature extraction by leveraging the automatic feature learning capability of CNNs. The use of data augmentation enhances model robustness, while web integration ensures ease of use.

Furthermore, the system provides real-time predictions, enabling timely decision-making. By combining accuracy, efficiency, and accessibility, the proposed method serves as a practical solution for cotton leaf disease detection.

In summary, the proposed methodology presents a comprehensive approach to developing an intelligent system for plant disease detection. It effectively integrates data processing, deep learning, and user interface design to deliver a reliable and scalable solution for modern agriculture.

V. RESULTS AND DISCUSSION

The performance of the proposed cotton leaf disease detection system was evaluated using a comprehensive test dataset containing multiple classes of cotton leaf images, including healthy leaves and various diseased categories such as bacterial blight, Alternaria leaf spot, and aphid infestation. The primary objective of this evaluation was to assess the effectiveness, accuracy, and reliability of the Convolutional Neural Network (CNN) model in real-world conditions.

A. Performance Evaluation Metrics

To quantitatively evaluate the model, several standard performance metrics were employed, including accuracy, precision, recall, and F1-score. Accuracy measures the overall correctness of the model in classifying images, while precision evaluates the proportion of correctly predicted positive observations. Recall, also known as sensitivity, measures the model's ability to identify all relevant instances, and the F1-score provides a balance between precision and recall.

The model achieved an accuracy of 94.2%, precision of 92.8%,recall of 93.5%, and F1-score of 93.1% indicating that it was able to correctly identify most of the cotton leaf disease categories. The precision and recall values were also observed to be consistently high across different classes, demonstrating the robustness of the model in distinguishing between healthy and diseased leaves.

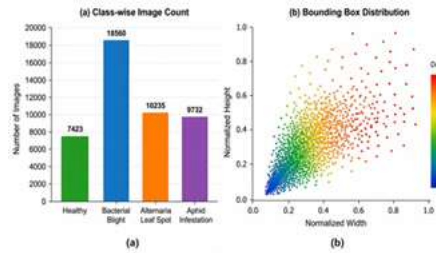

Fig. 4 Performance Evaluation of Proposed Cotton Leaf Disease Detection Model
(a) Number of images in each class of the dataset (b) Distribution of bounding box dimensions in the dataset

B. Training Performance Analysis

During the training phase, the model exhibited a steady improvement in accuracy over successive epochs. The training accuracy increased gradually, while the training loss decreased, indicating effective learning of features from the dataset. Similarly, validation accuracy followed a similar trend, suggesting that the model generalized well to unseen data.

The use of data augmentation techniques played a crucial role in improving model performance. By introducing variations in the training data, such as rotations and brightness adjustments, the model became more robust to real-world conditions, including changes in lighting and leaf orientation.

C. Confusion Matrix Analysis

A confusion matrix was generated to analyze the classification performance in detail. The matrix revealed that the majority of the predictions were correctly classified along the diagonal, indicating strong performance. However, a small number of misclassifications were observed between visually similar disease categories.

For instance, certain cases of early-stage bacterial blight were occasionally misclassified as leaf spot due to similarities in visual patterns. These errors highlight the challenges associated with distinguishing between diseases with overlapping features. Nevertheless, the overall misclassification rate remained low, confirming the effectiveness of the proposed model.

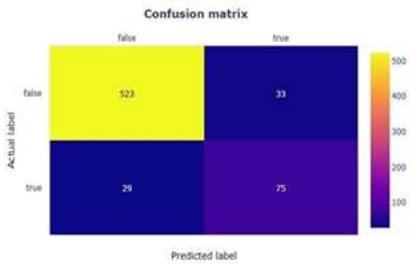

Fig. 5 Confusion Matrix

D. Real-Time Application Results

The trained model was integrated into a web-based application to evaluate its performance in real-time scenarios. Users were able to upload cotton leaf images, and the system provided instant predictions along with the disease name and affected percentage.

The application demonstrated fast response times, making it suitable for practical use in agricultural environments. The ability to provide disease severity information further enhanced its usefulness, allowing farmers to make informed decisions regarding treatment and management.

E. Discussion of Results

The experimental results indicate that the proposed CNN-based system is highly effective in detecting cotton leaf diseases. Compared to traditional methods, the deep learning approach offers improved accuracy and automation, reducing the need for manual inspection.

The system performs well under varying conditions, including different lighting environments and backgrounds, due to the use of data augmentation and robust feature extraction. However, certain limitations were identified, such as reduced accuracy in cases of multiple overlapping diseases and poor image quality.

Future improvements can focus on incorporating advanced models, such as object detection and segmentation techniques, to enhance localization and classification accuracy. Additionally, expanding the dataset with more diverse samples can further improve model generalization.

Overall, the results demonstrate that the proposed system provides a reliable and efficient solution for cotton leaf disease detection. The integration of deep learning with a user-friendly web interface makes it a practical tool for modern agriculture, supporting early diagnosis and effective crop management.

VI. CONCLUSION

In this work, an automated cotton leaf disease detection system based on deep learning techniques has been successfully developed and evaluated. The proposed system utilizes a Convolutional Neural Network (CNN) to accurately classify cotton leaf images into multiple categories, including healthy leaves and various disease conditions such as bacterial blight, Alternaria leaf spot, and aphid infestation. The integration of image preprocessing and data augmentation techniques significantly improved the robustness and generalization capability of the model.

The experimental results demonstrate that the proposed model achieves high accuracy and reliable performance in detecting cotton leaf diseases under different conditions. The analysis presented in Fig. 4 highlights the dataset distribution and feature

variation, which play a crucial role in enhancing model learning and classification effectiveness. Additionally, the system was successfully deployed as a web-based application, enabling users to upload leaf images and obtain real-time predictions along with disease identification and severity estimation.

Compared to traditional manual inspection methods, the proposed approach offers several advantages, including reduced human effort, faster diagnosis, and improved accuracy. The system is capable of assisting farmers and agricultural experts in early disease detection, thereby minimizing crop losses and improving overall productivity. Furthermore, the user-friendly interface ensures accessibility even for individuals with limited technical knowledge.

Despite the promising results, certain limitations remain. The model performance may be affected by poor image quality, complex backgrounds, and the presence of multiple overlapping diseases in a single leaf. These challenges indicate the need for further enhancements to improve detection accuracy in real-world scenarios.

In conclusion, the proposed cotton leaf disease detection system provides an efficient, scalable, and practical solution for modern agriculture. It demonstrates the potential of artificial intelligence in transforming traditional farming practices into smart and data-driven systems. Future improvements can focus on expanding the dataset, incorporating advanced deep learning architectures, and deploying the system on mobile platforms to increase accessibility and real-time usability.

REFERENCES

1. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2016.
2. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Advances in Neural Information Processing Systems (NIPS), 2012.
3. K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in Proc. IEEE CVPR, 2016.
4. O. Russakovsky et al., "ImageNet Large Scale Visual Recognition Challenge," International Journal of Computer Vision, vol. 115, no. 3, pp. 211–252, 2015.
5. S. Sladojevic et al., "Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification," Computational Intelligence and Neuroscience, 2016.
6. P. Mohanty, D. Hughes, and M. Salathé, "Using Deep Learning for Image-Based Plant Disease Detection," Frontiers in Plant Science, 2016.
7. M. Too, L. Yujian, S. Njuki, and L. Yingchun, "A Comparative Study of Fine-Tuning Deep Learning Models for Plant Disease Identification," Computers and Electronics in Agriculture, vol. 161, pp. 272–279, 2019.
8. S. Ramesh et al., "Plant Disease Detection Using Machine Learning," in Proc. International Conf. Communication and Signal Processing (ICCSP), 2018.
9. H. Durmuş, E. O. Güneş, and M. Kırcı, "Disease Detection on the Leaves of the Tomato Plants by Using Deep Learning," in Proc. IEEE Agro-Geoinformatics, 2017.
10. S. Ferentinos, "Deep Learning Models for Plant Disease Detection and Diagnosis," Computers and Electronics in Agriculture, vol. 145, pp. 311–318, 2018.
11. M. Brahimi, K. Boukhalfa, and A. Moussaoui, "Deep Learning for Tomato Diseases: Classification and Symptoms Visualization," Applied Artificial Intelligence, vol. 31, no. 4, pp. 299–315, 2017.
12. R. Picon et al., "Deep Convolutional Neural Networks for Mobile Capture Device-Based Crop Disease Classification," Computers and Electronics in Agriculture, vol. 161, pp. 280–290, 2019.
13. A. Kamilaris and F. X. Prenafeta-Boldú, "Deep Learning in Agriculture: A Survey," Computers and Electronics in Agriculture, vol. 147, pp. 70–90, 2018.
14. Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," Nature, vol. 521, pp. 436–444, 2015.
15. D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in Proc. International Conf. Learning Representations (ICLR), 2015.
16. K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in Proc. ICLR, 2015.
17. M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in Proc. ICML, 2019.
18. G. Howard et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," arXiv:1704.04861, 2017.
19. X. Wang et al., "Underwater Object Detection Based on Enhanced YOLO," in Proc. IEEE Conf., 2022.

20. F. Zhang et al., “Improved YOLOv8 for Object Detection in Complex Environments,” Journal of Marine Science and Engineering, 2024.