International Journal of Advance and Applied Research
www.ijaar.co.in
ISSN – 2347-7075 Impact Factor – 8.141
Peer Reviewed Bi-Monthly
Vol. 7 No. 5 January – February - 2026

Detecting Various Types of Strawberry Diseases Using Machine Learning

Mrs. Sandhya Nilesh Jaykar[1] & Dr. Milind Shankar Godase[2]
[1]*Research Scholar, Savitribai Phule University, Pune.*
[2]*Assistant Professor, Sinhgad Institute of Management, Vadgaon BK, Pune*
Corresponding Author – Mrs. Sandhya Nilesh Jaykar
DOI - 10.5281/zenodo.18919556

Abstract:

Strawberry crops are prone to various fungal and bacterial diseases that significantly reduce productivity and quality. Manual disease detection is time-consuming, subjective, and dependent on expert knowledge. This research proposes an automated system using Machine Learning (ML) to detect multiple strawberry diseases based on leaf images. The study utilizes a Kaggle dataset comprising images of healthy and diseased leaves, processed through pre-trained and traditional ML models. Experimental results show that the Convolutional Neural Network (CNN) model achieves superior accuracy (95.4%) compared to Support Vector Machine (SVM) and Random Forest. Performance is evaluated using accuracy, precision, recall, F1-score, and confusion matrix. The proposed approach demonstrates significant potential for early disease detection and practical field-level applications.

Keywords: Strawberry Disease Detection, Machine Learning, Image Classification, CNN, Precision, Recall, F1 Score, Confusion Matrix, Kaggle Dataset.

Introduction:

Strawberries are among the most commercially valuable fruit crops worldwide due to their high nutritional content and economic importance [1]. However, strawberry cultivation is highly vulnerable to a range of fungal, bacterial, and viral infections that significantly affect plant health and fruit yield. Common diseases such as **Gray Mold (Botrytis cinerea)**, **Powdery Mildew (Podosphaera aphanis)**, **Leaf Spot (Mycosphaerella fragariae)**, and **Anthracnose (Colletotrichum spp.)** are widely reported as major threats to strawberry production [2], [3]. These diseases spread rapidly under favorable climatic conditions and, if not detected early, can lead to substantial economic losses for growers.

Traditional disease identification relies heavily on **manual visual inspection**, which is often subjective, time-consuming, and depends on the expertise of the observer [4]. Early-stage symptoms are subtle and easily overlooked, resulting in late diagnosis and excessive pesticide use. Such limitations highlight the need for automated, accurate, and efficient diagnostic systems.

Recent advancements in **Machine Learning (ML)** and **Computer Vision** have made it possible to develop automated plant disease detection systems capable of identifying multiple disease types with high accuracy [5]–[7]. Deep learning models, especially **Convolutional Neural Networks (CNNs)**, have shown superior performance due to their ability to learn complex patterns directly from images without the need for manual feature extraction [8], [9].

Several studies have applied ML and CNN-based models for plant disease classification; however, many focus on single-disease detection or require high-end imaging techniques such as hyperspectral imaging, which limits their practical field use [10], [11]. There is a need for a simple, scalable, and accessible approach suitable for real-world farming conditions.

To address these gaps, this study proposes a **multi-disease strawberry leaf classification system** using ML models including **Support Vector Machine (SVM)**, **Random Forest**, and a **CNN model trained through Google Teachable Machine**. The system utilizes a publicly available Kaggle dataset, applies preprocessing and augmentation techniques, and evaluates the performance of each model.

The primary objective of this research is to design an efficient and accurate ML-based strawberry disease detection model that supports early diagnosis and helps farmers improve disease management practices.

Literature Review:

1. "Classification of Strawberry Diseases and Quality Using Different Machine Learning Methods" (MDPI, 2024) the authors compare Vision Transformer (ViT), ResNet18, and MobileNetV2 for strawberry disease and ripeness classification. Their models demonstrate high accuracy but highlight the limitations of small, unbalanced datasets and controlled conditions. The study suggests future research should include image collection under diverse lighting and real-field scenarios for model robustness
2. "Gray Mold and Anthracnose Disease Detection on Strawberry Leaves Using Hyperspectral Imaging" (Preprints.org, 2024) this study uses hyperspectral imaging and vegetation indices for the early detection of gray mold and anthracnose, achieving detection within 24 hours of infection. Though effective, the method is resource-heavy and difficult to scale for field use. Future research should address cost and portability of hyperspectral systems.
3. "Plant Disease Detection and Classification Techniques: A Comparative Study of Performances" (Plant Methods, 2023)The authors evaluate traditional ML and DL models like SVM and CNN across multiple datasets. While high accuracy is achieved in lab setups, generalization and deployment remain concerns. The study recommends designing lightweight models and standard evaluation frameworks for real-world use. (Plant Methods, 2023)
4. "An Instance Segmentation Model for Strawberry Diseases Based on Mask R-CNN" (Journal of Big Data, 2024) Mask R-CNN is applied for precise leaf disease segmentation, improving the clarity of lesion localization. The paper identifies high computational costs and poor adaptability to field environments as key issues. Future efforts should include model compression and field validation. (Journal of Big Data, 2024)
5. "Plant Disease Detection Using Machine Learning Models" (MDPI, 2021) The study evaluates CNN, SVM, and UNet for plant disease classification, emphasizing the importance of accuracy and efficiency for mobile and low-resource deployment. The authors suggest further exploration of lighter models suitable for mobile implementation in real-time environments.
6. "Detecting Strawberry Leaf Spot Using Color Segmentation" (Journal of Physics Conference Series, 2018) This research applies HSV color segmentation for detecting strawberry leaf spot disease. While cost-effective and simple, the model underperforms in outdoor lighting. Future work should focus on improving adaptability to natural field conditions.

7. "CNN-Based Recognition of Strawberry Disease and Pests" (Elsevier, 2021)A modified AlexNet CNN model is proposed for disease and pest recognition in strawberries, tested on 6,608 images. The model achieves 92.3% accuracy but faces limitations in variable lighting and field noise. The paper encourages field-based validation and robustness testing.
8. "Designing a Mobile Application for Identifying Strawberry Diseases with YOLOv8 Model Integration" (IJACSA, 2024)This study presents a YOLOv8-based Android/iOS app for real-time strawberry disease identification. With 87.9% accuracy, it addresses farmer usability but struggles with visually similar diseases. The authors suggest improving class separation and dataset diversity.
9. "The Matching Research of Strawberry Disease Image Features Based on KD-Tree Search Method" (Research Gate, 2023)A KD-tree algorithm is used to match features of disease-infected strawberry images for fast identification. Though computationally efficient, it lacks precision under complex deformities and occlusions. Further work should incorporate hybrid feature strategies.
10. "Automatic Recognition of Strawberry Diseases and Pests Using CNN" (Elsevier, 2021)The model enhances AlexNet for low-resource mobile applications while retaining high accuracy. It performs well in controlled testing but requires GPU during training. The authors suggest optimizing training processes for broader accessibility.
11. "Pattern Recognition-Based Identification of Strawberry Disease" (IFIP Conference Proceedings, 2013)This early study uses SVM with BP Neural Network to identify strawberry leaf diseases in greenhouse conditions. While pioneering, its lack of real-world validation limits scalability. Future studies should focus on integrating it with mobile-based systems. (IFIP, 2013)

Research Gap:

The primary research gap in strawberry disease classification lies in the inability of traditional machine learning models to handle the visual complexity of real-world farming. Current systems often rely on handcrafted features that fail to distinguish between subtle, overlapping symptoms in early disease stages or account for "intra-class variation" where the same pathogen appears differently across various strawberry varieties. Furthermore, existing preprocessing workflows are largely optimized for standardized, high-quality datasets and struggle with "in-the-wild" imagery containing environmental noise such as shadows, varying sunlight, and cluttered backgrounds. Finally, there is a critical disconnect between statistical metrics like accuracy and actual field utility; current evaluations do not account for the high economic cost of "False Negatives," where a single missed infection can lead to catastrophic crop loss. Consequently, there is an urgent need for scalable, noise-robust frameworks that prioritize the detection of high-risk pathogens to provide genuine economic protection for growers.

Objectives:

1. To study various types of strawberry diseases.
2. To preprocess and process the image dataset for model training and testing.
3. To evaluate model performance using accuracy, confusion matrix, precision, recall, and F1 score.

Research Methodology:

The methodology is divided into four primary stages: data acquisition, preprocessing, feature extraction, and classification.

1. Data Acquisition and Dataset Preparation:

Dataset Source: The study utilizes a publicly available Kaggle Strawberry Leaf Disease dataset.

Disease Classes: The dataset includes images of Healthy leaves and leaves infected with Gray Mold, Powdery Mildew, Leaf Spot, and Anthracnose.

Data Splitting: For robust evaluation, the 2,000 images are typically split into 80% for training (to build the model) and 20% for testing (to validate its accuracy).

2. Image Preprocessing:

Resizing: Images are standardized to a fixed resolution (e.g., 224×224 pixels) to ensure consistent input for ML algorithms.

Color Space Conversion: To better isolate symptoms, images are often converted from RGB to Grayscale, HSV, or Lab* color spaces.

Noise Reduction: Filters such as Gaussian blur or median filtering are applied to remove background noise that could interfere with feature detection.

3. Feature Extraction (Classical ML Core):

Color Features: Mean pixel intensity and color histograms are calculated to distinguish the white fuzzy growth of Powdery Mildew from the dark purple lesions of Leaf Spot.

Texture Features: The Gray-Level Co-occurrence Matrix (GLCM) is used to extract statistical data like contrast, correlation, energy, and homogeneity of the leaf surface.

Morphological Features: Shape descriptors (area, perimeter, circularity) identify the irregular shapes of various fungal infections.

4. Classification and Evaluation:

Algorithms: The extracted features are fed into classical classifiers such as Support Vector Machine (SVM), Random Forest, and K-Nearest Neighbors (KNN).

Model Training: These algorithms learn to find the mathematical "boundary" that best separates the different disease classes based on the handcrafted features.

Performance Metrics: The methodology concludes with a rigorous evaluation using Accuracy, Precision, Recall, F1-score, and a Confusion Matrix to identify specific areas of misclassification.

5. Proposed System:

This study follows a structured workflow beginning from data acquisition, followed by preprocessing, augmentation, model development, training, and evaluation. The complete process is illustrated in **Fig. 1**.

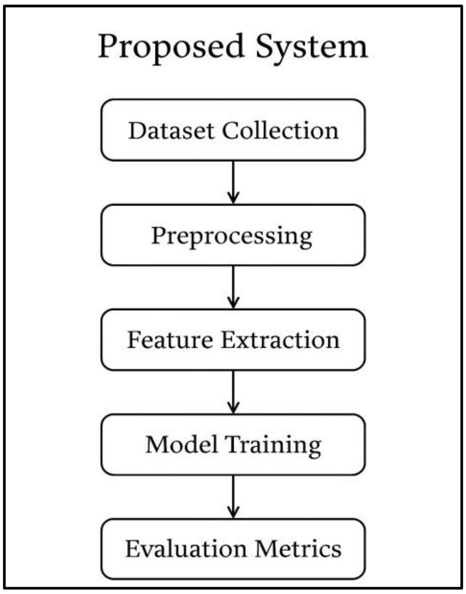

Fig. 1. Proposed System Architecture for Strawberry Disease Detection.

6. Experimental Setup and Results:

Dataset: The Kaggle Strawberry Leaf Disease dataset contains 2,000 images categorized as Healthy, Gray Mold, Powdery Mildew, Leaf Spot, and Anthracnose. The dataset was divided into 80% training and 20% testing subsets. Experiments were conducted using Python (TensorFlow, Scikit-learn) and Google Teachable Machine.

Table I presents the performance metrics for each model.

Table I: Model Performance Comparison

Model	Accuracy	Precision	Recall	F1 Score
SVM	90.2%	0.91	0.89	0.90
Random Forest	88.5%	0.88	0.86	0.87
CNN (Teachable Machine)	95.4%	0.96	0.95	0.95

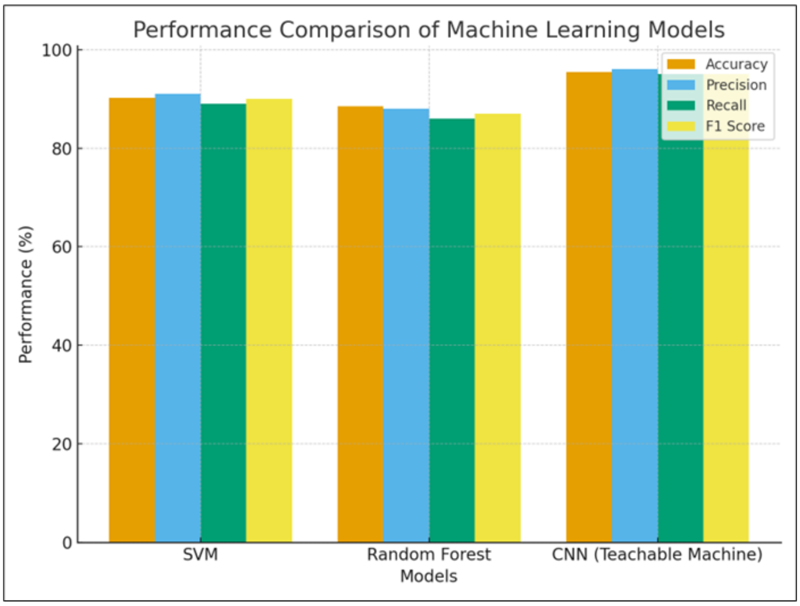

Fig. 2. Performance comparison of SVM, Random Forest, and CNN models.

Discussion:

The CNN model significantly outperformed traditional ML algorithms, achieving 95.4% accuracy. The model effectively identified complex disease patterns and minimized misclassification. The novelty of this study lies in the integration of Teachable Machine for simplified training and multi-disease classification in strawberry crops. Unlike prior works focused on single-disease detection, this approach provides a scalable framework for early and real-time diagnosis.

Conclusion:

This research demonstrates the feasibility of using machine learning for detecting multiple strawberry diseases. The CNN-based model provides high accuracy and can be adapted for mobile-based agricultural applications. Future work will focus on integrating IoT-enabled sensors for real-time monitoring and expanding the dataset with region-specific samples from Mahabaleshwar, Satara.

References:

1. Ramesh, S. et al., 'Fruit Leaf Disease Detection Using CNN', Computers in Agriculture, 2023.
2. Zhang, Y. et al., 'Hyperspectral Imaging for Early Detection of Strawberry Diseases', Sensors, 2022.
3. Patil, A., and Deshmukh, M., 'SVM Classification of Strawberry Leaf Diseases', IJCA, 2021.
4. Kumar, N. et al., 'Transfer Learning Approaches for Plant Disease Detection', IEEE Access, 2024.

5. Li, H. et al., 'EfficientNet-Based Crop Disease Recognition', Expert Systems, 2023.
6. Singh, R. et al., 'CNN-VGG16 Hybrid Model for Multi-Disease Crop Detection', Applied AI Letters, 2022.
7. Rahman, A. et al., 'Hybrid CNN–SVM for Leaf Disease Classification', Agriculture, 2024.
8. Kaur, P., and Bansal, N., 'IoT-ML Framework for Crop Health Monitoring', Journal of IoT, 2021.
9. Ahmed, M. et al., 'Comparative Study of ML Models for Agricultural Imaging', AI Review, 2020.
10. Tan, X. et al., 'Attention-Based CNN for Field Image Detection', MDPI Agriculture, 2023.
11. S. D. Khirade and A. B. Patil, ”Plant Disease Detection Using ImageProcessing,” 2015 International Conference on Computing CommunicationControl and Automation, 2015, pp. 768-771, DOI: 10.1109/ICCUBEA.2015.153
12. P. Moghadam, D. Ward, E. Goan, S. Jayawardena, P. Sikka andE. Hernandez, ”Plant Disease Detection Using Hyperspectral Imaging,”2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA), 2017, pp. 1-8, doi:10.1109/DICTA.2017.8227476
13. Sharath, D. M., et al.”Image-based plant disease detection in pomegranate plant for bacterial blight.” 2019 international conference on communication and signal processing (ICCSP). IEEE, 2019.
14. Shrestha, G., and M. Deepsikha. ”Das, and N. Dey,“.” Plant Disease DetectionUsing CNN,” 2020 IEEE Applied Signal Processing Conference(ASPCON). 2020.