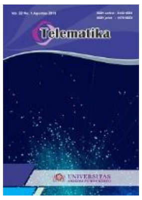
Telematika – Vol. 17, No. 2, August (2024) pp. 128-141 e-ISSN 2442-4528 | p-ISSN 1979-925X



Available online at :
http://ejournal.amikompurwokerto.ac.id/index.php/telematika/

Telematika

Accreditated SINTA "2" Kemenristek/BRIN, No. 85/M/KPT/2020

UNIVERSITAS AMIKOM PURWOKERTO

Detection and Classification of Banana Leaf Diseases: Systematic Literature Review

Ade Prasetyo[1], Ema Utami[2]

[1,2] Department Master of Informatics
Universitas AMIKOM Yogyakarta, Yogyakarta, Indonesia
E-mail:ade-prasetyo@students.amikom.ac.id[1], ema.u@amikom.ac.id[2]

ARTICLE INFO

History of the article:
Received November 08, 2024
Revised January 21, 2019
Accepted January 12, 2019

Keywords:
Banana Leaf Diseases
Classification
Image Processing
Machine Learning
Deep Learning

Correspondece:
E-mail: ade-prasetyo@students.amikom.ac.id

ABSTRACT

Bananas, a staple fruit globally, are essential for sustenance, employment, and income. However, diseases like Sigatoka, Bacterial Wilt, Bunchy Top, and Fusarium Wilt pose a threat to their cultivation, affecting both small-scale and large-scale production. This survey investigates methods for the early identification and classification of these banana leaf diseases using deep learning and machine learning techniques. A systematic review of 15 studies revealed that the majority of research concentrates on binary classification, which distinguishes healthy from diseased leaves. Common preprocessing steps include image resizing, color space conversion, and background removal to improve model accuracy. We utilize techniques such as ensemble approaches, support vector machines (SVM), random forests, K-means clustering, and convolutional neural networks (CNNs), with CNNs demonstrating superior performance, achieving accuracy rates ranging from 85% to 98.97%. CNNs excel in hierarchical feature extraction but require significant computational power. Traditional machine learning methods offer simplicity and resistance to overfitting but need careful parameter tuning. Advanced deep learning architectures, such as DenseNet and Inception V3, achieve high accuracy but with greater computational demands. Lightweight models like SqueezeNet balance performance and size, but ensemble methods, while improving generalization, add complexity. The choice of method depends on dataset characteristics, available computational resources, and desired trade-offs between performance and complexity. This study provides an overview of current research in banana leaf disease classification, discussing the strengths and limitations of various approaches and suggesting directions for future research to improve detection accuracy and robustness.

INTRODUCTION

Bananas are the most frequently eaten fruit in the world. They rank as the fourth largest food crop in the world, following rice, wheat, and maize(Voora et al., 2020). Bananas, which originated in Southeast Asia, are one of the most widely grown and consumed fruit crops in the world. More than 130 countries worldwide cultivate them in tropical and subtropical regions(Sau et al., 2023). Bananas play an important role in providing both employment and income, as well as serving as a vital source of nutrition. Nevertheless, banana plant growth faces significant challenges due to pests and diseases, serving as a major constraint on the output of both small-scale farmers and large corporations that aim to supply the global market. A range of diseases, primarily caused by fungi, bacteria, and viruses, can impact any part of the plant(Seetharaman & Mahendran, 2022). A primary hurdle in banana production is the presence of diseases that could significantly disrupt the cultivation of banana plants. These diseases include bacterial wilt, bunchy top virus, leaf spot disease, banana streak virus, and banana mosaic virus, posing a considerable threat to the cultivation process(Raja & Selvi Rajendran, 2022). Diseases affecting banana leaves can have a significant impact on banana productivity.

The proper detection and recognition of diseases affecting banana leaves is crucial in applying the required fertilizer. In the field, timely identification of pests and diseases is a critical first step in preventing and controlling them. Researchers have developed several methods to identify and categorize diseases in banana plants, such as using computerized image systems, thermal cameras, and machine learning methods(Gomez Selvaraj et al., 2020). These methods use different techniques, such as texture pattern analysis, multilevel thresholding, and deep transfer learning, to identify and classify diseases in banana plants. The use of these methods can help farmers and researchers detect and classify diseases in banana plants accurately, which can assist in the application of the required fertilizer and other control measures. Therefore, the proper detection and recognition of diseases affecting banana leaves are essential in ensuring the sustainability of banana production.

The classification of banana leaf diseases is critical in agricultural research because it facilitates early identification and control of plant diseases. Recent advancements in deep learning and machine learning techniques have significantly improved disease classification methods. This study explores the various methods utilized in classifying banana leaf diseases, including Artificial Neural Network (ANN), Support Vector Machine (SVM), K-Means Clustering, Random Forest (RF), K-Nearest Neighbors (KNN), Ensemble Approach, and Convolutional Neural Network (CNN). These techniques offer diverse approaches to analyzing image data and identifying patterns indicative of disease presence. Understanding the efficacy and applicability of these methods is critical for developing robust disease management strategies and ensuring banana farming techniques' sustainability.

This review paper aims to provide a comprehensive overview of the various methods and techniques used for detecting and categorizing banana leaf diseases, as well as the strengths and weaknesses of each algorithm employed. The paper discusses various diseases commonly impacting banana plants, including Fusarium wilt, Bunchy Top, Cordana, Sigatoka, Pestalotiopsis, cucumber mosaic virus, and bacterial wilt. Additionally, it summarizes different disease detection and classification techniques, such as pattern recognition, digital image processing, and machine learning algorithms, highlighting the importance of timely detection and precise recognition of banana leaf diseases to prevent significant economic losses. By evaluating the pros and cons of each algorithm, this paper intends to offer a deeper perspective and enhance the understanding of the effectiveness of various approaches in banana plant disease management.

RESEARCH METHODS

The systematic literature review (SLR) method gathers insights from various sources to tackle specific research inquiries or questions(Mangaroo-Pillay & Coetzee, 2022). An SLR is an independent academic method that identifies and evaluates all relevant literature on a topic in order to reach a conclusion about the particular question under investigation(Paul et al., 2021). It presents a methodical approach for collecting, rigorously assessing, merging, and showcasing discoveries derived from numerous research studies investigating a specific research question or topic. The SLR provides a tool to assess both the breadth and quality of the existing evidence related to a specific question or subject(Paul et al., 2021)

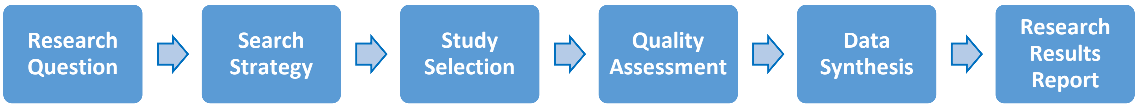


Figure 1. Categories of Banana Leaf Disease

This study used the SLR methodology to review research papers. One of the search results provides a guide on how to conduct systematic literature reviews in management research, which includes six steps, as shown in Figure 1. The literature review begins by formulating clear and focused research questions to guide the literature search process. Next, we develop the search strategy by designing a comprehensive approach to identify relevant literature across various scientific databases. Following this, we apply inclusion and exclusion criteria to select studies that meet the predetermined requirements for the review. We then evaluate the quality of the selected studies to confirm the reliability and validity of the synthesized information. We then integrate the data from the selected studies to address the research questions and objectives of the review. Finally, a research results report clearly and coherently presents the review's findings, including an overview of the main findings and implications for future research(Rodda et al., 2024).

1. Research Question

The main goal of this SLR is to fully understand the methodology used in previous studies that tried to classify diseases on banana leaves. Specifically, this study will focus on the dataset class, pre-processing techniques, methodologies, and classification results that were found. To ensure a thorough understanding of the utilized dataset categories, applied data pre-processing methods, various methodological approaches, and the achieved outcomes in disease classification, we formulated the questions outlined in Table I to provide guidance and maintain the review's concentration on the research objectives.

Table 1. Research Question

ID	*Research Question*	*Motivation*
RQ1	What types of data sets are detectable?	Determine the categories of datasets used to classify diseases affecting banana leaves.
RQ2	What methods are employed during data preprocessing?	Recognize the data preprocessing methods used to classify banana leaf diseases.
RQ3	What methods are used for detecting or classifying banana leaf diseases?	Identify the methods used in banana leaf disease classification.
RQ4	What are the detection or classification results of the reviewed studies?	Identify the evaluation and results from every method used in banana leaf disease classification.

2. Search Strategy

In order to comprehensively identify pertinent literature, our research encompassed a thorough exploration across a multitude of scientific databases, including ScienceDirect, IEEE, Sciendo, Springer, ACM Digital Library, arXiv, and Google Scholar. We conducted the searches exclusively in English, using a carefully selected set of keywords. Specifically, we strategically combined terms such as “detection” or “classification” with “banana leaf disease,” ensuring the incorporation of essential components within the search string, such as “classification of banana leaf disease” or “detection of banana leaf disease.” By adopting this rigorous approach, we aimed to meticulously gather scholarly works directly addressing the central themes of disease detection and classification in banana plants. We devised this comprehensive method to capture a diverse array of research endeavors, spanning various methodologies and offering multifaceted insights into the nuanced complexities of detecting and classifying diseases affecting banana leaves.

3. Study Selection

The paper screening process employs two types of standards: inclusion criteria and exclusion criteria. Inclusion criteria determine the studies suitable for inclusion in the review, while exclusion criteria identify which studies should be omitted.(Amin et al., 2022).

The criteria for inclusion in this study are outlined below:

a. The research collected includes data published from 2020 to October 2023.
b. Papers are composed in either English.
c. The research topic must pertain to the detection or classification of banana leaf disease.

The criteria for excluding studies in this research encompass the following:

a. Studies that do not align with the inclusion criteria.
b. Research lacking clear articulation of the methodology employed.
c. Studies unable to address the research objectives.

4. Quality Assessment

The meticulous evaluation of the reviewed papers is imperative to uphold the integrity and reliability of the research findings. To achieve this, a robust approach entails exclusive consideration of high-quality papers throughout the review process. During this research, we applied a stringent criterion that limited the selection to papers published in internationally recognized journals, sourced from reputable and esteemed outlets in the academic community. The research adheres to stringent criteria to enhance the credibility and trustworthiness of the synthesized information, ensuring the integration of only contributions from well-regarded sources into the overarching analysis.

5. Data Synthesis

During this pivotal phase, the paper selection process involved a comprehensive evaluation of each paper's ability to effectively address the overarching research question and align with the predetermined inclusion and exclusion criteria. Commencing with the collection of an initial thirty journals, the subsequent steps encompassed the application of meticulously crafted research questions, specific inclusion criteria, and strategic filters. Through this rigorous vetting process, a refined selection of fifteen papers emerged, meeting the stringent requirements set forth. Following this, the chosen papers underwent a thorough final review and in-depth analysis to ensure their relevance, methodological rigor, and contribution to the overarching research objectives.

6. Research Results Report

We meticulously selected papers during this critical phase, based on their contextual relevance in effectively addressing the core research question and satisfying the specific inclusion and exclusion criteria. We assembled an extensive collection of thirty journals to initiate this process and set the stage for a comprehensive inquiry. After using the carefully thought-out research questions, inclusion criteria, and exclusion filters in a planned way, a smaller group of fifteen papers was found that perfectly met the requirements. We then subjected these selected papers to a thorough and rigorous review, meticulously analyzing each to assess their methodological soundness, contribution to the research objectives, and alignment with the study's overarching goals.

RESULTS AND DISCUSSION

The initial screening process involved evaluating abstracts and content for 35 studies. Subsequently, a meticulous selection process reduced the number of studies to 15. Then, these studies were carefully looked over, with a focus on four main areas that make up the core of this research: dataset classifications, data preprocessing techniques, topic classification or detection methodologies, and the results that came from the detection or classification processes within each topic. This rigorous approach aimed not only to ensure the inclusion of pertinent studies, but also to deliver a thorough analysis of the literature landscape about the identification and categorization of banana leaf diseases.

1. Banana Leaf Disease

In general, banana leaf detection or classification research consists of 2 classes in the dataset, namely detecting or classifying healthy leaves and leaves affected by disease. All studies used healthy leaf classes, while for leaf disease classes, there were 1 to 6 classes in each study. In this study, there were 7 types of banana leaf diseases listed in Figure 2, such as Sigatoka, Bacterial Wilt, Bunchy Top, Cordana, Fusarium Wilt, Cucumber Mosaic Virus, and Pestalotiopsis.

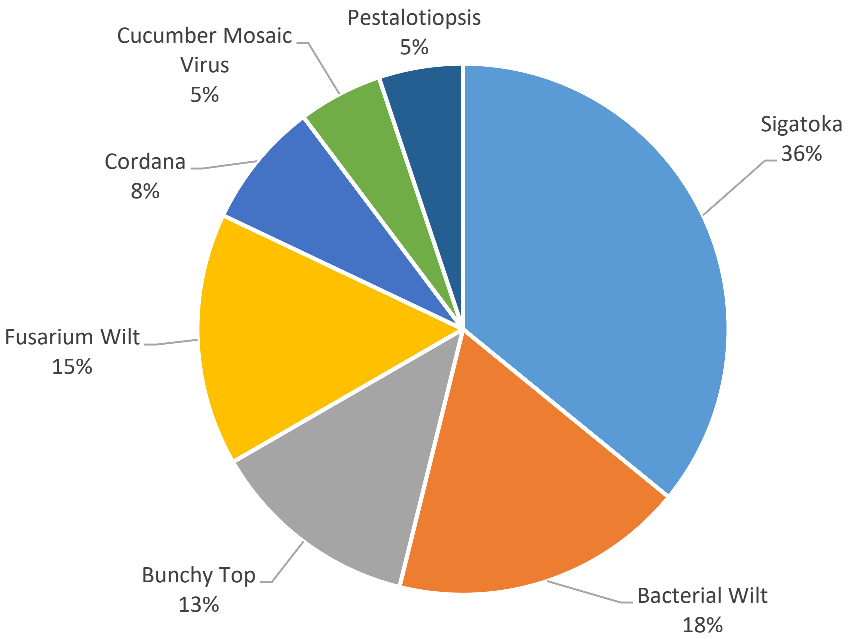


Figure 2. Categories of Banana Leaf Disease

a. Sigatoka

Sigatoka, a disease that impacts banana plants, results in small, yellow marks on their leaves. As this disease advances, these marks expand in size and change color to brown or black. The leaves also become distorted or develop holes(Upadhyay et al., 2021). The disease is caused by a fungus called Mycosphaerella Eumusae(George et al., 2022). The leaf disease that is often used in research is Sigatoka detection with a percentage of 36% which is divided into 3, namely black Sigatoka(Ani Brown Mary et al., 2020, 2021; Chaudhari & Patil, 2023; Criollo et al., 2020; Jadhav et al., 2023; Sanga et al., 2020; Sheena Basil & Brown Mary, 2022), yellow Sigatoka(Chaudhari & Patil, 2023; Jadhav et al., 2023) and mixed Sigatoka(Andreanov Ridhovan et al., 2022; Bhuiyan et al., 2023; Chaudhari & Patil, 2020; Mathew et al., 2023; Saranya et al., 2020).

http://dx.doi.org/10.35671/telematika.v17i2.2809

b. Bacterial Wilt

Bacterial wilt, stemming from the bacterium Ralstonia Solanacearum, is a plant disease that induces wilting, yellowing, and plant mortality. Additionally, it leads to brown staining in the vascular tissue within the plant's stems and roots(Dey & Sen, 2023). In research(Ani Brown Mary et al., 2020, 2021; Chaudhari & Patil, 2020, 2023; Criollo et al., 2020; Gomez Selvaraj et al., 2020; Sheena Basil & Brown Mary, 2022) the bacterial wilt leaf disease class was used.

c. Bunchy Top

Bunchy top disease causes reduced growth, leaf yellowing, and the development of a "bunchy" appearance in the leaves. The disease can also cause the death of the plant(Liu et al., 2022). The bunchy top leaf disease class was used in the study(Ani Brown Mary et al., 2020, 2021; Chaudhari & Patil, 2023; Gomez Selvaraj et al., 2020; Sheena Basil & Brown Mary, 2022).

d. Cordana

Cordana leaf spot causes leaf spots on banana plants. The indications of this illness vary from the symptoms seen in Sigatoka leaf spot disease(de Souza-Pollo & de Goes, 2020). The Cordana leaf spot class was used in the study(Andreanov Ridhovan et al., 2022; Bhuiyan et al., 2023; Mathew et al., 2023).

e. Fusarium Wilt

Fusarium wilt, resulting from the fungus Fusarium Oxysporum, is a plant ailment that induces wilting, yellowing, and plant demise. It also leads to the browning of the vascular tissue in the plant's stems and roots(Hay et al., 2022). The fusarium wilt leaf disease class was used in the study(Chaudhari & Patil, 2020; Jadhav et al., 2023; Sanga et al., 2020; Saranya et al., 2020; Yan et al., 2023; Zhang et al., 2022).

f. Cucumber Mosaic Virus (CMV)

CMV can lead to diverse symptoms in affected plants, such as mosaic-like patterns on leaves, inhibited growth, and decreased crop productivity. The virus can also cause necrosis and death of plant tissue(Abdelkhalek et al., 2022). The cucumber mosaic virus leaf disease class was used in the study(Chaudhari & Patil, 2020, 2023).

g. Pestalotiopsis

Diseases such as grey blight, leaf blight, canker, dieback, and several postharvest diseases are attributed to Pestalotiopsis species. These diseases significantly diminish the commercial production of ornamental plants and trees(Jiang et al., 2022). The Pestalotiopsis leaf disease class was used in the study(Andreanov Ridhovan et al., 2022; Bhuiyan et al., 2023).

2. Pre-processing Data

Pre-processing data involves readying unprocessed data for analysis or machine learning by converting it into a more comprehensible and analyzable format for algorithms. Techniques in pre-processing encompass actions like data cleaning, normalization, transformation, feature selection, and engineering(Mallikharjuna Rao et al., 2023). Resizing is one of the most frequently used methods when pre-processing data, including all reviewed studies.

Color conversion is also widely used in the research reviewed, for example, research(Ani Brown Mary et al., 2020, 2021) has converted RGB to YCbCr, while conversion of RGB to L*a*b has been used in research(Chaudhari & Patil, 2020, 2023), and other research has converted RGB to

http://dx.doi.org/10.35671/telematika.v17i2.2809

grayscale(Saranya et al., 2020). Apart from that, other studies also removed background(Criollo et al., 2020), removed noise, and enhanced contrast(Chaudhari & Patil, 2023).

3. Classification Method

Machine learning and deep learning are the two primary methods typically employed to classify diseases affecting banana leaves. Machine learning employs methods such as Ensemble Approach, K-Means Clustering, Support Vector Machine (SVM), Random Forest, and K-Nearest Neighbors (KNN). Meanwhile, deep learning encompasses techniques like ANN and CNN. Table 1 offers a summary of the techniques used to classify diseases found on banana leaves.

a. Convolutional Neural Network (CNN)

CNN is a widely utilized type of deep neural network for processing images and videos. It includes various layers, such as convolutional layers, pooling layers, and fully connected layers(Saleem et al., 2022). They are typically trained using backpropagation with stochastic gradient descent and data augmentation techniques(Zheng et al., 2020). CNN is used in research(Jadhav et al., 2023; Sheena Basil & Brown Mary, 2022; Yan et al., 2023; Zhang et al., 2022). Other studies also use CNN but with different architectures, such as HEAP autoencoder(Ani Brown Mary et al., 2020), Inception V3(Andreanov Ridhovan et al., 2022; Sanga et al., 2020), AlexNet (Criollo et al., 2020), DenseNet 121(Andreanov Ridhovan et al., 2022; Mathew et al., 2023), and SqueezeNet(Bhuiyan et al., 2023). CNN methods can be combined with other methods such as DenseNet 121 and Inception V3 in learning(Andreanov Ridhovan et al., 2022).

b. Artificial Neural Network (ANN)

An artificial neural network (ANN) emulates the architecture and functioning of the human brain within the context of machine learning. Comprising interconnected nodes, or neurons, these networks manage the processing and transmission of information. Researchers apply the ANN technique(Saranya et al., 2020).

c. K-Means Clustering

The K-means algorithm operates by initially choosing K centroids randomly, with K indicating the specified number of clusters. The algorithm then allocates every data point to the nearest centroid, updating the centroids based on the average of these assigned data points. We repeat this sequence until the centroids stay unchanged or we reach a set maximum number of iterations(Sharma, 2020). K-Means Clustering method is used in research(Chaudhari & Patil, 2020)

d. Random Forest (RF)

For tasks like classification and regression, we employ the Random Forest algorithm. It operates as an ensemble learning method by combining multiple decision trees to improve the model's accuracy and robustness. In research, the RF method is used(Gomez Selvaraj et al., 2020).

e. Support Vector Machine (SVM)

SVM functions by finding a hyperplane that separates the data into different classes. We choose this hyperplane to maximize the margin between the classes, which measures the distance from the hyperplane to the closest data points of each class. Additionally, SVM can use kernel functions to transform the data into a higher-dimensional space, facilitating a clearer separation between the classes(Cervantes et al., 2020). The SVM method is used in research(Gomez Selvaraj et al., 2020).

f. K-Nearest Neighbors (KNN)

K-Nearest Neighbors works by locating the K closest data points to a new data point and assigning it to the class that is most prevalent among those K data points. You can adjust the value of K as a hyperparameter to improve the algorithm's performance. KNN can also use metrics other than Euclidean distance, such as Manhattan distance or Mahalanobis distance.(Zhou et al., 2021). The KNN method is used in research(Ani Brown Mary et al., 2021).

g. Ensemble Approach

The ensemble approach is a method in machine learning that merges various models to improve the accuracy and robustness of the predictions. Researchers employ the ensemble approach method(Chaudhari & Patil, 2023).

Table 2. Method Classification

Research	*Method Classification*
(Andreanov Ridhovan et al., 2022; Ani Brown Mary et al., 2020; Bhuiyan et al., 2023; Criollo et al., 2020; Jadhav et al., 2023; Mathew et al., 2023; Sanga et al., 2020; Sheena Basil & Brown Mary, 2022; Yan et al., 2023; Zhang et al., 2022)	CNN
(Saranya et al., 2020)	ANN
(Chaudhari & Patil, 2023)	K-Means Clustering
(Gomez Selvaraj et al., 2020)	Random Forest (RF)
(Gomez Selvaraj et al., 2020)	Support Vector Machine (SVM)
(Ani Brown Mary et al., 2021)	KNN
(Chaudhari & Patil, 2023)	Ensemble Approach

4. Detection and Classification Result

We conducted a systematic literature review (SLR) on disease classification in banana plants, which involved an analysis of 15 research studies. Each study employed various methodologies and techniques to classify diseases affecting banana plants, focusing on achieving high accuracy rates. Table 3 contains the data examined in the research.

The first study (Ani Brown Mary et al., 2020) utilized a HEAP autoencoder (HAE) after preprocessing techniques involving resizing and conversion to YCbCr color space. With an impressive accuracy of 98.97%, this approach demonstrated the potential of autoencoder-based methods in disease classification. The benefit of employing autoencoders is their capacity to learn intricate patterns and reduce data dimensionality, thereby boosting the model's performance. However, the computational resources needed for training autoencoders and interpreting the encoded features might pose challenges.

In contrast, (Chaudhari & Patil, 2020) employed K-Means Clustering with resizing and transformation to L*a*b* color space. Despite achieving a moderate accuracy of 85%, this method displayed simplicity in implementation but was sensitive to noise. One advantage of K-Means Clustering is its simplicity in implementation and computational efficiency, which makes it well-suited for handling large datasets. However, its performance is significantly influenced by the initial choice of centroids and is sensitive to outliers, which can impact classification accuracy.

(Gomez Selvaraj et al., 2020) opted for SVM and RF methods following resizing preprocessing. Their approach yielded a commendable accuracy of 93.43% and showcased resilience against overfitting. Random Forest and SVM are robust machine learning algorithms that can manage high-dimensional data and nonlinear relationships. These algorithms are less susceptible to overfitting than other methods, making them well-suited for classification tasks involving complex datasets. However, achieving optimal performance with these methods necessitates careful parameter tuning.

(Sanga et al., 2020) utilized Inception V3 after resizing, achieving an accuracy of 95.5%. This method, known for capturing abstract features, exhibited robust performance in disease classification. Inception V3, being a deep learning model, has the advantage of automatically learning hierarchical features from images, which can improve classification accuracy. However, training deep neural networks like Inception V3 demands substantial amounts of labeled data and computational resources.

(Criollo et al., 2020) applied AlexNet with background removal and resizing preprocessing, achieving an accuracy of 87.5%. While effective in recognizing disease patterns, the interpretability of features remained a challenge. AlexNet is a groundbreaking deep-learning architecture for image classification tasks. Its strength is in its capacity to detect detailed patterns in images and attain high accuracy. However, the deep layers of AlexNet make it difficult to interpret the learned features, limiting its transparency.

(Saranya et al., 2020) employed an Artificial Neural Network (ANN) with various preprocessing steps. Although the accuracy was not provided, the study highlighted the network's ability to handle complex and nonlinear data. Artificial Neural Networks, especially deep neural networks, can learn intricate relationships within data and adapt to various types of input features. However, training ANNs requires extensive computational resources and suffers from overfitting if not properly regularized.

(Ani Brown Mary et al., 2021) adopted K-Nearest Neighbors (KNN) after resizing and transforming to YCbCr color space, achieving an accuracy of 98%. This method offered simplicity in implementation but was sensitive to the choice of neighbors. K-Nearest Neighbors is an easy-to-understand and straightforward algorithm for classification tasks, making it easy to implement and understand. It does not require training, making it well-suited for small datasets. However, the selection of the number of neighbors (K) significantly affects the algorithm's performance, and it does not perform well in high-dimensional spaces or with imbalanced data.

(Andreanov Ridhovan et al., 2022) utilized DenseNet 121 and Inception V3 models after resizing, achieving an accuracy of 84.73%. While capable of processing images of varying sizes, significant computational resources were required. DenseNet and Inception V3 are deep learning architectures known for their ability to capture intricate features from images. DenseNet's densely connected layers enhance feature reuse and gradient flow, whereas Inception V3's inception modules support efficient learning of features at multiple scales. However, training these models can be resource-intensive, necessitating powerful hardware and extended training periods.

(Sheena Basil & Brown Mary, 2022) employed a Convolutional Neural Network (CNN) after resizing, achieving an impressive accuracy of 98.8%. CNNs excel in capturing hierarchical features, which contributes to their exceptional performance. Convolutional Neural Networks are cutting-edge models for image classification tasks because they can automatically extract pertinent features from images using convolutional and pooling layers. They exhibit translation invariance and can handle spatial hierarchies present in images. However, training CNNs typically requires large datasets and computational resources, and model interpretability is limited.

(Zhang et al., 2022) also utilized CNN after resizing, achieving an accuracy of 95%. The study highlighted CNN's effectiveness in disease classification tasks. CNNs provide flexibility in managing different types of image data and can adjust to various image sizes and resolutions. Their hierarchical architecture allows them to learn complex patterns at multiple levels of abstraction. However, CNNs suffer from overfitting if not properly regularized, and training deep models can be computationally expensive.

(Bhuiyan et al., 2023) employed SqueezeNet after resizing, achieving an accuracy of 96.25%. Despite its lightweight architecture, a compromise between model size and performance was observed. SqueezeNet is designed to be computationally efficient while maintaining competitive performance in image classification tasks. Its compact design makes it well-suited for deployment on devices with limited resources. However, reducing model size results in a loss of representational capacity, potentially affecting classification accuracy.

(Chaudhari & Patil, 2023) implemented an ensemble approach after several preprocessing steps, achieving an accuracy of 96%. This method effectively reduced variance and improved generalization. Ensemble methods aggregate predictions from multiple models to generate a final prediction, frequently resulting in improved performance compared to individual models. They are resistant to overfitting and can identify a range of patterns within the data. However, constructing an effective ensemble requires careful selection of diverse base models and aggregation techniques.

(Mathew et al., 2023) utilized DenseNet 121 after resizing, achieving an accuracy of 91.7%. The study emphasized the model's capability to handle complex datasets with diverse scales. DenseNet's densely connected layers enhance feature reuse and improve gradient flow, enabling efficient learning of intricate patterns. However, DenseNet's performance degrades on datasets with high levels of noise or class imbalance, and training deep models requires significant computational resources.

(Yan et al., 2023) utilized CNN after resizing, achieving an accuracy of 98%. The study highlighted CNN's effectiveness in capturing hierarchical features from images. CNNs provide versatility in processing diverse types of image data and can accommodate different image sizes and resolutions. Their hierarchical architecture allows them to learn complex patterns at multiple levels of abstraction. However, CNNs suffer from overfitting if not properly regularized, and training deep models can be computationally expensive.

Finally, (Jadhav et al., 2023) employed CNN after resizing, achieving an accuracy of 88.32%. While CNN demonstrated superior performance, it necessitated large training datasets and intricate parameter tuning.

Table 3 presents a concise summary of influential research in the field of plant disease detection, including dataset categories, preprocessing techniques, and outcomes. This summary elucidates the methodologies and precision of different approaches, providing valuable perspectives on present trends and efficacy in the field.

Table 3. Overview of Reviewed Studies

No.	*Research (Author, Year)*	*Dataset Class*	*Preprocessing*	*Methods*	*Result*
1	(Ani Brown Mary et al., 2020)	Bunchy Top, Black Sigatoka, Bacterial Wilt, Healthy	Resizing, Convert RGB to YCbCr	HEAP autoencoder (HAE)	Accuracy (98.97%)
2	(Chaudhari & Patil, 2020)	Sigatoka, Cucumber Mosaic Virus, Bacterial Wilt, Fusarium Wilt	Resizing, Convert RGB to L*a*b*	K-Means Clustering	Accuracy (85%)
3	(Gomez Selvaraj et al., 2020)	banana cluster, Bacterial wilt, Bunchy Top, and Healthy	Resizing	Random Forest (RF) and support vector machine (SVM)	Accuracy (93.43%)
4	(Sanga et al., 2020)	Black Sigatoka, Fusarium wilt, Healthy leaves	Resizing	Inception V3	Accuracy (95.5%)
5	(Criollo et al., 2020)	Black Sigatoka, Bacterial Wilt, Healthy	Resizing, Background removal, and resizing	AlexNet	Accuracy (87.5%)
6	(Saranya et al., 2020)	Sigatoka, Fusarium Wilt, Healthy	Cropping, Resizing, color conversion	ANN	
7	(Ani Brown Mary et al., 2021)	Bunchy Top, Bacterial Wilt, Black Sigatoka, Healthy	Resizing, Transformed from RGB to YCbCr	KNN	Accuracy (98%)
8	(Andreanov Ridhovan et al., 2022)	Healthy, Sigatoka, Cordana, Pestalotiopsis	Resizing	DenseNet 121 & Inception V3	Accuracy (84.73%)
9	(Sheena Basil & Brown Mary, 2022)	Bunchy Top, Bacterial Wilt, Black Sigatoka	Resizing	CNN	Accuracy (98,8%)
10	(Zhang et al., 2022)	Fusarium Wilt	Resizing	CNN	Accuracy (95%)
11	(Bhuiyan et al., 2023)	Healthy, Pestalotiopsis, Sigatoka, Cordana	Resizing	SqueezeNet	Accuracy (96.25%)
12	(Chaudhari & Patil, 2023)	Healthy, Bunchy Top, Cucumber Mosaic Virus, Yellow Sigatoka, Black Sigatoka, Bacterial Wilt	Resizing, Eliminate Noise, Convert RGB to L*a*b* color, contrast enhancement	Ensemble Approach	Accuracy (96%)
13	(Mathew et al., 2023)	Sigatoka, Cordana	Resizing	DenseNet 121	Accuracy (91.7%)
14	(Yan et al., 2023)	Healthy, Fusarium Wilt	Resizing	CNN	Accuracy (98%)
15	(Jadhav et al., 2023)	Healthy, Yellow Sigatoka, Black Sigatoka, Fusarium Wilt	Resizing	CNN	Accuracy (88.32%)

Comparing all the methods utilized in the reviewed studies, it's evident that convolutional neural networks (CNNs) consistently yielded high accuracy rates across various disease classification tasks. CNNs are highly effective at capturing hierarchical features from images, making them well-suited for complex and varied datasets. However, their training requires significant computational resources and extensive parameter optimization to prevent overfitting. In contrast, traditional machine learning algorithms like SVM, RF, and KNN offer simplicity in implementation and resilience against overfitting. They perform well on smaller datasets, but careful parameter tuning is required for optimal performance.

Deep learning architectures such as DenseNet and Inception V3 achieve competitive accuracy rates by capturing intricate features from images, but training these models can be computationally intensive. Lightweight models like SqueezeNet strike a good balance between size and performance, making them ideal for deployment in resource-constrained environments. Ensemble methods, on the other hand, combine the advantages of several models to improve overall generalization and decrease variance, but they require more computational power. The appropriate method should consider the dataset's specifics, the available computational resources, and the balance between model performance and complexity.

CONCLUSIONS AND RECOMMENDATIONS

These reviewed studies provided comprehensive insights into four key aspects: dataset classes, data pre-processing, classification or detection techniques, and the results achieved in banana leaf disease classification research. The research primarily consisted of a binary classification of healthy and diseased leaves. Studies identified various diseases such as Sigatoka, Bacterial Wilt, Bunchy Top, Cordana, Fusarium Wilt, Cucumber Mosaic Virus, and Pestalotiopsis. Pre-processing methods were heavily inclined towards resizing images, enhancing their suitability for analysis. Additionally, we commonly employed color space conversions such as RGB to YCbCr, RGB to Lab*, and background removal to prepare the data.

Studies utilized a mix of machine learning and deep learning algorithms. CNNs were notably prevalent, implementing different architectures (e.g., Inception V3, DenseNet 121, SqueezeNet). Additionally, researchers employed machine learning algorithms such as K-Means Clustering, RF, SVM, KNN, and ensemble approaches for classification. Studies achieved various accuracy levels, with some notably surpassing the 90% accuracy mark. Techniques like HEAP autoencoder, Inception V3, CNNs, and ensemble approaches produced highly accurate results in disease classification.

Further exploration and combination of different classification techniques enhance accuracy and robustness in disease detection. The inclusion of more diverse instances of diseases and healthy samples could increase the model's ability to generalize. Given CNNs' success, exploring novel deep-learning architectures and hybrid models could potentially improve disease classification. Standardized evaluation metrics and validation methodologies would aid in comparative analysis and enhance the robustness of models.

REFERENCES

Abdelkhalek, A., Király, L., Al-Mansori, A.-N. A., Younes, H. A., Zeid, A., Elsharkawy, M. M., & Behiry, S. I. (2022). Defense Responses and Metabolic Changes Involving Phenylpropanoid Pathway and PR Genes in Squash (Cucurbita pepo L.) following Cucumber mosaic virus Infection. *Plants*, *11*(15), 1908. https://doi.org/10.3390/plants11151908

Amin, Z. M., Anwar, N., Mohd Shoid, M. S., & Samuri, S. (2022). Method for Conducting Systematic Literature Review (SLR) for Cyber Risk Assessment. *Environment-Behaviour Proceedings Journal*, *7*(SI10), 255–260. https://doi.org/10.21834/ebpj.v7iSI10.4130

Andreanov Ridhovan, Aries Suharso, & Chaerur Rozikin. (2022). Disease Detection in Banana Leaf Plants using DenseNet and Inception Method. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *6*(5). https://doi.org/10.29207/resti.v6i5.4202

Ani Brown Mary, N., Robert Singh, A., & Athisayamani, S. (2020). Banana leaf diseased image classification using novel HEAP auto encoder (HAE) deep learning. *Multimedia Tools and Applications*, *79*(41–42). https://doi.org/10.1007/s11042-020-09521-1

Ani Brown Mary, N., Robert Singh, A., & Athisayamani, S. (2021). *Classification of Banana Leaf Diseases Using Enhanced Gabor Feature Descriptor* (pp. 229–242). https://doi.org/10.1007/978-981-15-7345-3_19

Bhuiyan, Md. A. B., Abdullah, H. M., Arman, S. E., Saminur Rahman, S., & Al Mahmud, K. (2023). BananaSqueezeNet: A very fast, lightweight convolutional neural network for the diagnosis of three prominent banana leaf diseases. *Smart Agricultural Technology*, *4*, 100214. https://doi.org/10.1016/j.atech.2023.100214

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, *408*, 189–215. https://doi.org/10.1016/j.neucom.2019.10.118

Chaudhari, V., & Patil, M. (2020). Banana leaf disease detection using K-means clustering and Feature extraction techniques. *2020 International Conference on Advances in Computing, Communication & Materials (ICACCM)*, 126–130. https://doi.org/10.1109/ICACCM50413.2020.9212816

Chaudhari, V., & Patil, M. P. (2023). Detection and Classification of Banana Leaf Disease Using Novel Segmentation and Ensemble Machine Learning Approach. *Applied Computer Systems*, *28*(1), 92–99. https://doi.org/10.2478/acss-2023-0009

Criollo, A., Mendoza, M., Saavedra, E., & Vargas, G. (2020). Design and Evaluation of a Convolutional Neural Network for Banana Leaf Diseases Classification. *2020 IEEE Engineering International Research Conference (EIRCON)*, 1–4. https://doi.org/10.1109/EIRCON51178.2020.9254072

de Souza-Pollo, A., & de Goes, A. (2020). Banana Pathology and Diseases. In *Handbook of Banana Production, Postharvest Science, Processing Technology, and Nutrition*. https://doi.org/10.1002/9781119528265.ch3

Dey, P., & Sen, S. K. (2023). A REVIEW ON SOLANACEOUS PLANT DISEASES CAUSED BY RALSTONIA SOLANACEARUM HAVING SERIOUS ECONOMIC IMPACT. *PLANT ARCHIVES*, *23*(2). https://doi.org/10.51470/plantarchives.2023.v23.no2.072

George, M., Anita Cherian, K., & Mathew, D. (2022). Symptomatology of Sigatoka leaf spot disease in banana landraces and identification of its pathogen as Mycosphaerella eumusae. *Journal of the Saudi Society of Agricultural Sciences*, *21*(4), 278–287. https://doi.org/10.1016/j.jssas.2021.09.004

Gomez Selvaraj, M., Vergara, A., Montenegro, F., Alonso Ruiz, H., Safari, N., Raymaekers, D., Ocimati, W., Ntamwira, J., Tits, L., Omondi, A. B., & Blomme, G. (2020). Detection of banana plants and their major diseases through aerial images and machine learning methods: A case study in DR Congo and Republic of Benin. *ISPRS Journal of Photogrammetry and Remote Sensing*, *169*. https://doi.org/10.1016/j.isprsjprs.2020.08.025

Hay, W. T., Anderson, J. A., McCormick, S. P., Hojilla-Evangelista, M. P., Selling, G. W., Utt, K. D., Bowman, M. J., Doll, K. M., Ascherl, K. L., Berhow, M. A., & Vaughan, M. M. (2022). Fusarium head blight resistance exacerbates nutritional loss of wheat grain at elevated CO2. *Scientific Reports*, *12*(1), 15. https://doi.org/10.1038/s41598-021-03890-9

Jadhav, S., Gandhi, S., Joshi, P., Choudhary, V., & Walunjkar, S. (2023). Banana Crop Disease Detection Using Deep Learning Approach. *International Journal for Research in Applied Science and Engineering Technology*, *11*(5), 2061–2066. https://doi.org/10.22214/ijraset.2023.51827

Jiang, N., Voglmayr, H., Xue, H., Piao, C.-G., & Li, Y. (2022). Morphology and Phylogeny of *Pestalotiopsis* (*Sporocadaceae* , *Amphisphaeriales*) from *Fagaceae* Leaves in China. *Microbiology Spectrum*, *10*(6). https://doi.org/10.1128/spectrum.03272-22

Liu, S., Niu, K., Chen, S., Sun, X., Liu, L., Jiang, B., Chu, L., Lv, X., & Li, M. (2022). TiO_2 bunchy hierarchical structure with effective enhancement in sodium storage behaviors. *Carbon Energy*, *4*(4), 645–653. https://doi.org/10.1002/cey2.172

Mallikharjuna Rao, K., Saikrishna, G., & Supriya, K. (2023). Data preprocessing techniques: emergence and selection towards machine learning models - a practical review using HPA dataset. *Multimedia Tools and Applications*, *82*(24), 37177–37196. https://doi.org/10.1007/s11042-023-15087-5

Mangaroo-Pillay, M., & Coetzee, R. (2022). Lean frameworks: A systematic literature review (SLR) investigating methods and design elements. *Journal of Industrial Engineering and Management*, *15*(2), 202. https://doi.org/10.3926/jiem.3677

Mathew, D., Kumar, C. S., & Anita Cherian, K. (2023). Classification of leaf spot diseases in banana using pre-trained convolutional neural networks. *2023 International Conference on Control, Communication and Computing (ICCC)*, 1–5. https://doi.org/10.1109/ICCC57789.2023.10165629

Paul, J., Lim, W. M., O'Cass, A., Hao, A. W., & Bresciani, S. (2021). Scientific procedures and rationales for systematic literature reviews (SPAR-4-SLR). *International Journal of Consumer Studies*. https://doi.org/10.1111/ijcs.12695

Raja, N. B., & Selvi Rajendran, P. (2022). Comparative Analysis of Banana Leaf Disease Detection and Classification Methods. *Proceedings - 6th International Conference on Computing Methodologies and Communication, ICCMC 2022*. https://doi.org/10.1109/ICCMC53470.2022.9753840

Rodda, S. N., Bijker, R., Merkouris, S. S., Landon, J., Hawker, C. O., & Dowling, N. A. (2024). How to Peer Review Quantitative Studies, Qualitative Studies, and Literature Reviews: Considerations from the 'Other' Side. *Current Addiction Reports*, *11*(5), 771–782. https://doi.org/10.1007/s40429-024-00594-8

Saleem, M. A., Senan, N., Wahid, F., Aamir, M., Samad, A., & Khan, M. (2022). Comparative Analysis of Recent Architecture of Convolutional Neural Network. *Mathematical Problems in Engineering*, *2022*. https://doi.org/10.1155/2022/7313612

Sanga, S. L., Machuve, D., & Jomanga, K. (2020). Mobile-based Deep Learning Models for Banana Disease Detection. *Engineering, Technology & Applied Science Research*, *10*(3). https://doi.org/10.48084/etasr.3452

Saranya, N., Pavithra, L., Kanthimathi, N., Ragavi, B., & Sandhiyadevi, P. (2020). Detection of Banana Leaf and Fruit Diseases Using Neural Networks. *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, 493–499. https://doi.org/10.1109/ICIRCA48905.2020.9183006

Sau, S., Bhattacharjee, P., Kundu, P., & Mandal, D. (2023). Banana. In *Tropical and Subtropical Fruit Crops* (pp. 1–62). Apple Academic Press.

Seetharaman, K., & Mahendran, T. (2022). Leaf Disease Detection in Banana Plant using Gabor Extraction and Region-Based Convolution Neural Network (RCNN). *Journal of The Institution of Engineers (India): Series A*, *103*(2). https://doi.org/10.1007/s40030-022-00628-2

Sharma, P. (2020). Advanced image segmentation technique using improved K means clustering algorithm with pixel potential. *PDGC 2020 - 2020 6th International Conference on Parallel, Distributed and Grid Computing*. https://doi.org/10.1109/PDGC50313.2020.9315743

Sheena Basil, D., & Brown Mary, D. (2022). Classification of Diseases in Banana Leaves using Diagonal Path Value Pattern. *Ijsdr.Org International Journal of Scientific Development and Research*, *7*, 76. www.ijsdr.org

Upadhyay, A., Oommen, N. M., & Mahadik, S. (2021). *Identification and Assessment of Black Sigatoka Disease in Banana Leaf* (pp. 237–244). https://doi.org/10.1007/978-981-15-5421-6_24

Voora, V., Larrea, C., & Bermudez, S. (2020). *Global market report: bananas*.

Yan, K., Shisher, M. K. C., & Sun, Y. (2023). *A Transfer Learning-Based Deep Convolutional Neural Network for Detection of Fusarium Wilt in Banana Crops*. https://doi.org/10.20944/PREPRINTS202309.1681.V1

Zhang, S., Li, X., Ba, Y., Lyu, X., Zhang, M., & Li, M. (2022). Banana Fusarium Wilt Disease Detection by Supervised and Unsupervised Methods from UAV-Based Multispectral Imagery. *Remote Sensing*, *14*(5), 1231. https://doi.org/10.3390/rs14051231

Zheng, Q., Yang, M., Tian, X., Jiang, N., & Wang, D. (2020). A full stage data augmentation method in deep convolutional neural network for natural image classification. *Discrete Dynamics in Nature and Society*, *2020*. https://doi.org/10.1155/2020/4706576

Zhou, N.-R., Liu, X.-X., Chen, Y.-L., & Du, N.-S. (2021). Quantum K-Nearest-Neighbor Image Classification Algorithm Based on K-L Transform. *International Journal of Theoretical Physics*, *60*(3), 1209–1224. https://doi.org/10.1007/s10773-021-04747-7