IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⩺ Peer-reviewed & Refereed journal ⩺ Vol. 13, Issue 9, September 2024

DOI: 10.17148/IJARCCE.2024.13924

A Study of Machine Learning Approaches and Models for Predicting Cotton Leaf Diseases: Identifying Research Gaps for Future Exploration

Mr. Parteek Singh[1], Smt. Sunita Rani[2], Er. Narinder Kumar[3]

Assistant Professor, University School for Graduate Studies, Chaudhary Devi Lal University, Sirsa-125055, India[1]

Research Scholar, Department of M.M. Institute of Computer Technology & BM (MCA), Maharishi Markandeshwar (Deemed to be University), Mullana-133203, India[2]

SBSBM University College, Sardulgarh, 151507, Constituent College of Punjabi University, Patiala, , Punjab, India[3]

Abstract: Plant diseases have been a source of concern for farmers and academics around the world from long time. It is critical to identify and manage these illnesses as soon as possible in order to avoid their spread and reduce their impact on crop output. The ability to analyse patterns and features from plant photos or data using Machine Learning algorithms allows us to diagnose diseases more quickly, accurately, and at scale. The paper presents a comprehensive review of machine learning techniques applied to plant disease prediction, emphasizing their effectiveness and limitations. An in-depth analysis on key factors such as the authors' approaches, datasets employed, specific problem statements addressed, and performance metrics used to evaluate model effectiveness. Through the analysis, various critical research gaps in the existing literature has been identified. The findings includes the need for standardized datasets, the integration of real-time data collection methods and the integration of ml and deep learning techniques for predicting the plant disease. The study provides a structured framework for future research, guiding the development of more robust and scalable machine learning solutions in plant disease management.

Keywords: Plant Disease Prediction, Machine Learning, Predictive Models, ML Techniques, Disease Detection, Agricultural AI

1. INTRODUCTION

Plant diseases have been a source of concern for farmers and academics around the world from long time. It is critical to identify and manage these illnesses as soon as possible in order to avoid their spread and reduce their impact on crop output. Traditional illness detection methods frequently rely on professional visual inspection, which takes time, is subjective, and is open to human mistake. However, the advent of Machine Learning techniques has created new opportunities for accurate and automated plant disease diagnosis. It is now possible to analyse enormous amounts of plant data and photos using the power of Machine Learning algorithms, allowing for faster and more reliable disease diagnosis. Machine Learning algorithms are particularly suited for detecting plant diseases because they are excellent at spotting intricate patterns and relationships in data. Machine Learning models may be trained to distinguish between distinct disease types and correctly categorize plants based on their health state using labelled datasets that contain photos of both healthy and diseased plants. Often invisible to the human eye, the models are able to extract useful features from the photos, such as colour changes, texture details, or spatial qualities. This makes it possible for Machine Learning algorithms to recognize subtle disease symptoms at an early stage, giving farmers and agronomists a useful tool to act quickly and efficiently [1].

The application of Machine Learning to plant disease diagnosis has many benefits for the agricultural industry. Firstly, it provides a scalable technology that can swiftly analyse huge volumes of plant data, enabling effective real-time crop monitoring. This makes it possible for farmers to identify disease outbreaks quickly, carry out focused interventions, and reduce crop losses. Second, by offering reliable and impartial assessments, Machine Learning-based illness detection systems can get around the constraints of human knowledge. The model's capacity to generalize and precisely detect diseases across many contexts is improved by the ability to train them on broad datasets that include a wide range of diseases and plant types. Additionally, Machine Learning models may continuously learn from fresh data and enhance their performance, keeping them abreast of new diseases. Image analysis is just one use of Machine Learning in the diagnosis of plant diseases. To improve the precision of disease detection, Machine Learning algorithms can also be used to additional data sources including sensor data, weather information, or genomic data. A complete view of plant health

DOI: 10.17148/IJARCCE.2024.13924

and disease dynamics can be obtained by integrating numerous data streams, enabling more thorough disease management tactics. Additionally, thanks to technological advancements like the accessibility of low-cost imaging devices, mobile apps, and cloud computing, farmers in remote or resource-limited areas now have easier access to Machine Learning-based disease detection, facilitating their access to timely and accurate information for disease control [2].

In conclusion, the use of Machine Learning approaches to plant disease detection has the potential to completely transform the agricultural industry. Our ability to analyze patterns and features from plant photos or data using Machine Learning algorithms allows us to diagnose diseases more quickly, accurately, and at scale. With the help of this technology, farmers, agronomists, and researchers may make wise decisions, carry out prompt interventions, and manage plant diseases successfully. The future of plant disease detection appears hopeful with further developments and ongoing advances in Machine Learning algorithms, as well as increasing data accessibility and availability, opening the door for improved food security and sustainable farming practices.

1.1 Machine Learning Outline

Artificial intelligence (AI) has an area called Machine Learning that focuses on creating algorithms and models that let computers learn and make predictions or judgments without having to be explicitly programmed. Its foundation is the notion that systems have the capacity to learn from and adapt to data, enhancing their performance over time. Traditionally, computers have been programmed with explicit instructions written by humans. In Machine Learning, we provide the computer data and allow it to learn patterns and relationships from that data rather than manually programming it. The computer then employs this newly acquired knowledge to formulate forecasts, categorize fresh data, or resolve challenging issues. Machine Learning approaches have been used in a variety of fields, such as healthcare, financial modelling, recommender systems, audio and picture identification, and natural language processing. With improvements in algorithms, data accessibility, and computer power, the discipline is still developing and allowing for more complex models and applications [3].

Machine Learning algorithms can be broadly categorized into three types: supervised learning, unsupervised learning, and reinforcement learning.

1.1.1 Supervised Learning

In this type of Machine Learning, the algorithm learns from labelled examples or training data. The data consists of input features and corresponding output labels or target values. The algorithm learns to map the input features to the correct output based on the provided examples. Supervised learning is commonly used for tasks like classification (predicting categories) and regression (predicting numerical values) [4].

1.1.2 Unsupervised Learning

Unsupervised learning involves learning from unlabelled data, where the algorithm does not have specific target labels. The algorithm explores the patterns and structures in the data to find meaningful insights, group similar data points together (clustering), or reduce the dimensionality of the data. Unsupervised learning can be useful for tasks like anomaly detection, customer segmentation, and recommendation systems [4].

1.1.3 Reinforcement Learning

Reinforcement learning involves an agent interacting with an environment and learning to make decisions based on trial and error. The agent receives feedback in the form of rewards or penalties for its actions, and it aims to maximize cumulative rewards over time. Reinforcement learning is commonly used in scenarios like game playing, robotics, and autonomous driving [4].

1.2 Machine Learning Classification Techniques

These are the five classification techniques that cover a wide range of scenarios and provide a good starting point for various classification tasks. However, it's important to evaluate and choose the appropriate technique based on the specific requirements and characteristics of the dataset

1.2.1 Logistic Regression

Logistic regression is a linear model used for binary classification problems. It models the relationship between the features and the probability of an instance belonging to a certain class using the logistic function. The model estimates the coefficients of the features to maximize the likelihood of the observed data. Logistic regression is interpretable, computationally efficient, and works well with linearly separable data but it has a linear decision surface, hence it cannot address non-linear issues. Real-world situations rarely involve linearly separable data [5].

1.2.2 Random Forest

Random forest is an ensemble learning method that combines multiple decision trees. It creates a set of decision trees by random Machine Learning selecting features and data samples. Each decision tree is trained on a bootstrapped subset of the data and votes for the class label. Random forest reduces overfitting and provides robust predictions by averaging the results of the individual trees. It handles high-dimensional data, captures feature interactions, and can handle both categorical and numerical features but it is non-interpretable and as it integrates numerous decision trees to decide the class, training takes a lot of time [3].

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 13, Issue 9, September 2024
DOI: 10.17148/IJARCCE.2024.13924

1.2.3 Support Vector Machines (SVM)
SVM is a powerful algorithm for binary classification that finds an optimal hyperplane to separate classes. It maps the data into a high-dimensional space and finds the hyperplane that maximizes the margin between the classes. SVM can use different kernel functions to handle linearly inseparable data by transforming it into a linearly separable space. It is effective in cases where the number of features is much greater than the number of instances. SVM works well with both linear and non-linear data and overfitting problem is not as much as other methods, but it can be computationally expensive for large datasets [6].

1.2.4 Naive Bayes
Naive Bayes is a probabilistic classifier based on Bayes' theorem with the assumption of feature independence. It calculates the probability of an instance belonging to each class and assigns it to the class with the highest probability. Naive Bayes is simple, computationally efficient, and requires a small amount of training data. It performs well in text classification and spam filtering tasks. However, the independence assumption may not hold in some cases, which can affect the accuracy of the model [7].

1.2.5 K-Nearest Neighbours
K-Nearest Neighbours (KNN) is a simple and intuitive Machine Learning algorithm used for both classification and regression tasks. It is a type of instance-based or lazy learning algorithm, meaning it memorizes the training data and makes predictions based on the similarity between new data points and the training instances. But KNN is sensitive to the scale of features. If the features have different scales, some features may dominate the distance calculation, leading to biased predictions.

1.2.6 Neural Network
In numerous areas, such as speech and picture recognition, natural language processing, gaming, and many more, neural networks have shown to function at the cutting edge. They can be computationally expensive though, and they need a lot of data to train on. Choosing an appropriate design, avoiding overfitting, and properly tweaking hyperparameters are crucial for getting the most performance out of neural networks.

1.2.7 Gradient Boosting
Gradient boosting is an ensemble learning method that combines multiple weak classifiers in a sequential manner. It starts with a weak classifier and then fits subsequent classifiers to the residual errors of the previous ones. The final prediction is made by aggregating the predictions of all the weak classifiers. Gradient boosting is known for its high accuracy and ability to handle complex interactions between features. It is robust to outliers and missing data but can be computationally expensive and prone to overfitting if not properly tuned [8].

2. LITERATURE REVIEW

Singh *et al.* (2023) focused on applying deep learning paradigms to detect cotton leaf diseases in India, which have a significant impact on agricultural production. The paper underscored the importance of agriculture in India's economy and the challenges faced by cotton plants, including pests, climate variations, and nutrient deficiencies. To overcome the limitations of previous studies that concentrated on a limited number of cotton leaf diseases, the researchers proposed a model that utilized convolutional neural networks (CNN) for disease detection. The model demonstrated outstanding performance, achieving an impressive accuracy of 99.39% on the test set. This surpassed the performance of existing approaches, showcasing the potential of the proposed model for real-time disease detection systems. The research concluded by recognizing the need for larger datasets and employing data augmentation techniques to enhance the performance of deep learning models. The authors emphasized the significance of addressing issues such as weak learning, regularization, and bias-variance balance. Despite the challenges, the proposed approach exhibited promise for future applications and could be adapted to tackle new problems with the availability of additional data instances [9].
Rai and Pahuja (2023) introduced an innovative approach to identifying and predicting cotton plant diseases using deep learning techniques. The study explored the utilization of a Deep Convolution Neural Network (DCNN) based model for the classification of diseases in cotton leaves and plants. The authors conducted experiments to examine the impact of different data split ratios, pooling layer choices, and epoch sizes on the performance of the model. The proposed model achieved an impressive accuracy of 97.98% in classifying cotton leaf and plant diseases, surpassing the performance of previous approaches. This technique holds potential benefits in terms of reducing the time and human error associated with identifying and assessing the severity of cotton leaf diseases in major production regions. The research highlighted the broader applications of intelligent farming, Machine Learning, and image processing in the agricultural sector. It emphasized the significance of incorporating these technologies to enhance disease detection and management practices in farming. It is observed that the research made a significant contribution to the field of automated plant disease detection and classification by presenting an innovative approach that achieved high accuracy using deep learning techniques. The findings demonstrated the potential for further advancements in the field, providing valuable insights for researchers and practitioners working on automated disease detection in agriculture [10].

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 13, Issue 9, September 2024

DOI: 10.17148/IJARCCE.2024.13924

Qian *et al.* (2022) presented a preliminary study on the utilization of deep learning and the Ant Colony Optimization (ACO) algorithm for the detection and management of Cotton Root Rot (CRR) in cotton fields. The authors emphasized the significance of effectively managing CRR-infected fields, which incurred substantial economic losses, with Texas alone experiencing an average annual loss of approximately 29 million USD. Traditional methods for mapping CRR regions were limited in their real-time detection capabilities. The study introduced a deep-learning-based approach using YOLOv5 for real-time detection of CRR-infected regions and showcased its deployment on an edge-computing platform. The results demonstrated a moderate level of detection accuracy, with a promising average inference speed of 11 frames per second (FPS). Furthermore, the study illustrated how the detected CRR region locations could be utilized to generate an optimal path for efficient management practices employing the ACO algorithm. The total distance covered based on the optimal path for four detected regions of CRR was 160 m. These findings showcased the potential of combining the deep learning-based approach with the ACO algorithm to expedite CRR management using multispectral aerial imagery. Future work will focus on enhancing detection accuracies beyond 70% by implementing image pre-processing methods and exploring the implementation of Mask R-CNN. Field tests accompanied by cost-benefit analyses will be conducted, taking into account the optimal path generated by the ACO algorithm. Overall, the study provided valuable insights into a promising approach to address the challenges posed by CRR in cotton fields. By leveraging deep learning and the ACO algorithm, this research contributes to the development of effective strategies for CRR management and offers a foundation for further advancements in the field [11].

Patil and Burkpalli (2022) focused on the classification of leaf diseases in cotton crops that played a crucial role in preventing yield loss in India, where cotton is a prominent crop. The researchers employed image-processing techniques to identify diseases such as Alternaria, bacterial blight, and grey mildew. They utilized a modified factorization-based active contour method to extract cotton leaf images with complex backgrounds. Various features, including segmented images, colour, gray-level co-occurrence matrix, and local binary patterns, were extracted to train machine-learning classifiers. The experimental results demonstrated that the ensemble classifier proposed in the study outperformed other state-of-the-art classifiers, such as support vector machines, random forests, and multilayer perceptron's. The accuracy of the proposed model was reported to be 92.29%, indicating its effectiveness in disease classification. The results confirmed the superior performance of the model compared to other single and hybrid classifiers, with precision and Matthews correlation coefficient (MCC) scores of 0.905 and 0.872, respectively. Overall, the study demonstrated that the proposed model yielded promising and comparatively good results in the classification of cotton leaf diseases. The findings affirmed the effectiveness of the approach and its superiority over other classification methods, highlighting its potential for practical application in disease management in cotton crops [12].

Kumar*et al.* (2022) cantered on the development of a prediction model using the CNN algorithm to identify cotton diseases in India. The researchers utilized TensorFlow's Keras API to create the prediction model, which was then integrated into a mobile app. This app aimed to assist farmers in identifying cotton diseases and provide recommendations for suitable pesticides. The Machine Learning model was prepared using the TensorFlow open-source platform and subsequently converted into the Core Machine Learning model for seamless integration into an iOS app. The reported accuracy of the model was approximately 90%. At present, the app was capable of detecting boll rot and fungal leaf spot disease, with the potential for expansion to include other cotton diseases. Overall, this paper presented an innovative approach that empowered farmers and aimed to mitigate production loss caused by cotton diseases in India. By leveraging the power of Machine Learning and mobile technology, the app provided a valuable tool for farmers to identify diseases and make informed decisions about disease management, ultimately contributing to improved crop health and productivity [13].

Udawant and Srinath (2022) concentrated on the utilization of the Mask RCNN object detection algorithm with transfer learning for the detection of cotton leaf diseases in practical situations. The study aimed to address the limitations of previous research, which primarily relied on controlled environment images and faced challenges in providing accurate results in real-world scenarios. A dataset comprising 2000 photos of cotton leaves affected by various diseases and pests, with diverse backgrounds and multiple leaves in the frame, was collected by the researchers. By applying transfer learning, the model attained a training accuracy of 94% and effectively identified the specific disease or pest present on the diseased region. The study concluded that Mask RCNN exhibited promise as a suitable algorithm for the detection of diseases and pests on cotton leaves. Future work involves the expansion of the dataset by incorporating additional images encompassing various diseases, as well as the inclusion of additional characteristics such as soil type and temperature. Furthermore, the researchers proposed leveraging Generative Adversarial Networks (GANs) for data augmentation to enhance the training data. In summary, the research presented a comprehensive exploration of the utilization of the Mask RCNN algorithm for cotton leaf disease detection, demonstrating its potential in real-world scenarios. The findings pave the way for further advancements in the field by suggesting future directions for research and improvement in dataset diversity and data augmentation techniques [14].

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⸎ Peer-reviewed & Refereed journal ⸎ Vol. 13, Issue 9, September 2024
DOI: 10.17148/IJARCCE.2024.13924

Patil and Patil (2021) designed a novel system for the identification and detection of diseases in cotton plants, utilizing deep learning techniques and an IoT-based platform. The study collected a dataset of infected and healthy cotton leaf images through an IoT-based system that incorporated cameras and sensors deployed in crop fields and infected areas. The collected images were pre-processed, labelled, and subjected to image augmentation techniques to mitigate overfitting during the model training phase. A deep convolutional neural network (CNN) was trained using popular deep learning frameworks such as Theano and Torch7. The accuracy of the model was evaluated using the F1 score, with the authors reporting an accuracy of approximately 98.0% after training the model for 10k iterations. The performance of the developed model was tested through various test cases to ensure its effectiveness and economic feasibility. The proposed system demonstrated promising results in accurately identifying and classifying cotton plant diseases, leading to improved crop production. The integration of an IoT platform and sensor data collection enabled the detection of climatic changes, further enhancing the system's efficiency. It is observed that the study presented an innovative approach with high accuracy and efficiency in detecting diseases in cotton plants, leveraging deep learning and an IoT-based platform. The findings of this research provided valuable insights for researchers and contributed to advancements in the field of agricultural technology, particularly in the realm of disease detection and monitoring in cotton farming [15].

Caldeira *et al*. (2021) developed a digital image processing system to detect and identify lesions on cotton plant leaves. The system utilized an algorithm for texture extraction from images and employed a neuro-fuzzy classifier trained to discriminate between soil, healthy leaves, and lesioned leaves. The algorithm demonstrated the ability to recognize all three classes, with better performance in identifying background information compared to healthy leaf areas. The overall accuracy of the system was 71.2%, indicating that the unbalanced data of different classes affected the results. Despite challenges such as variations in illumination, colour, texture, and image complexity, the model's behaviour was considered satisfactory. The hybrid nature of the model was expected to contribute to the advancement of intelligent systems in agriculture. The algorithm was made available as a tool in integrated management of cotton crops to assist in decision-making for pest and disease control. The system achieved a hit rate of 71.1% in correctly identifying lesions, which was deemed satisfactory for practical implementation [16].

Wang *et al.* (2020) discussed the utilization of remote sensing (RS) techniques for delineating cotton root rot (CRR) infested areas in cotton fields. The researchers conducted their study in three Texas fields known to be infested with CRR, employing an unmanned aerial vehicle (UAV) to collect high-resolution RS images. They evaluated various classification methods, including supervised, unsupervised, and combined unsupervised methods, along with two new automated methods specifically designed for UAV RS images. The results demonstrated that the new automated methods outperformed conventional approaches, with the best method achieving an overall accuracy of 88.5%. This method combined k-means segmentation and morphological opening and closing, resulting in the lowest errors of omission (11.44%) and commission (16.13%) compared to all other methods considered. Overall, the findings highlighted the potential of utilizing UAV-based RS techniques for accurately identifying CRR-infested areas in cotton fields. This precise identification enables targeted application of fungicides, thereby minimizing treatment costs and optimizing disease management strategies. The study contributes to the advancement of CRR detection and demonstrates the effectiveness of UAV-based RS methods in agricultural practices [17].

Prashar *et al*. (2020) addressed the economic losses faced by the cotton agriculture industry in India due to pest infections and bacterial or viral contagions. It proposed a solution by using visual features to recognize cotton leaf diseases. An expert system was implemented, utilizing multi-layered perceptrons (MLP) with overlapping pooling and a flexible layering mechanism to classify plant leaves as infected or healthy. The graph-based MLP network model was employed to dynamically transform feature orientation and discover similar components in the database. The k-nearest neighbour (kNN) and support vector machine (SVM) were utilized for double-layered classification to minimize errors. The model combined various techniques to accurately localize disease regions with over 96% accuracy. The paper emphasized the importance of this system in helping farmers detect and analyse emerging disease symptoms, leading to early precautions and disease prevention for higher yield. The authors suggested improving SVM and KNN for real-time systems and exploring deep learning classification for efficient handling of growing data volumes. Overall, the work held significant importance for the agriculture industry in reducing losses caused by crop diseases on both national and international scales [18].

Shah and Jain (2019) proposed a Machine Learning-based approach that aimed to assess the health and detect diseases in cotton plants by utilizing artificial neural networks (ANN) and image pre-processing techniques to analyse leaf images and identify disease types based on colour changes. The neural network tool used in this research was implemented in MATLAB, which classified the quality of cotton leaf diseases by analysing the RGB and HSV components of the images. The results of the study demonstrated that the ANN-based approach outperformed traditional visual inspection methods, effectively reducing errors caused by varying visual perception and lighting conditions. The study utilized a dataset of

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 9, September 2024
DOI: 10.17148/IJARCCE.2024.13924

18 leaf images, representing six different types of diseases, and the results were stored in a binary format, with '1' indicating disease presence and '0' indicating disease absence. Statistical tests, including t-tests, were conducted to validate the accuracy and reliability of the ANN data. The obtained results aligned with the initial expectations, further indicating the reliability of the data derived from the artificial neural network. It is observed that the research highlighted the potential of using Machine Learning techniques, specifically artificial neural networks, to enhance disease detection in cotton farming. By offering a more efficient and accurate approach to monitor the health of cotton plants, this research provides valuable insights for improving crop management practices and optimizing disease control strategies in cotton production [19].

Udawant and Srinath (2019) developed a deep Convolutional Neural Network (CNN) model that accurately classified diseased and healthy cotton leaves. The authors demonstrated that the CNN algorithm outperformed other classification algorithms in terms of speed. The model was trained for 1000 epochs, and the results showed accurate classification of cotton plant leaf images. The training accuracy graph depicted the progressive improvement of the model with each epoch, while the validation accuracy confirmed its ability to predict outputs for new inputs. The proposed model successfully identified both diseased and healthy cotton leaves. The paper also acknowledged the potential for further improvements by collecting more data and exploring advanced models such as faster RCNN and SSD. Overall, this research paper presented an enhanced method for accurately classifying cotton leaves and provided valuable insights for future enhancements [20].

Dubey *et al.* (2018) presented a technique for detecting and classifying cotton leaf diseases using the concepts of roughness measure and simple linear iterative clustering. The proposed algorithm achieved an average classification accuracy of 94% for cotton leaf diseases, with impressive results for individual disease categories. Alternaria diseases and White flies were classified with 100% accuracy, while Bacterial diseases and Healthy plants achieved 88% accuracy. The algorithm utilized superpixel-based roughness measure for accurate image segmentation and feature extraction using gray level co-occurrence matrix. The classification was performed using Support Vector Machine, a supervised Machine Learning algorithm. The research demonstrated the effectiveness of the proposed approach, establishing it as a valuable contribution to the field of cotton leaf disease analysis [21].

Chopda *et al.* (2018) reviewed the existing crop disease detection systems and presented a system that aimed to predict cotton crop diseases using a decision tree classifier. The authors highlighted the increasing demand for real-time applications in Machine Learning for agriculture and emphasized the importance of predicted data in farming to enable timely addressing of issues by farmers. The research discussed the integration of image processing, data mining, IoT, and Machine Learning techniques in smart farming. The proposed system utilized temperature and soil moisture parameters as input to the decision tree classifier for predicting cotton crop diseases. Additionally, the paper mentioned the development of an Android application that provided real-time output to farmers. Overall, the research addressed an important issue in agriculture and proposed a solution using Machine Learning techniques [22].

Prajapatii *et al.* (2016) emphasized the significance of cotton as a cash crop in India and highlighted the detrimental impact of diseases on cotton productivity. The researchers presented a comprehensive approach for disease detection, which encompassed various image processing steps such as acquisition, pre-processing, segmentation, and feature extraction. They discussed different segmentation techniques and classification methods, including SVM, K-NN, and NN. To carry out the research, provided a thorough survey of related work, comparing different methodologies and highlighting the importance of texture features and image processing techniques. Overall, the research served as a valuable resource for researchers and practitioners working on cotton disease detection [23].

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 13, Issue 9, September 2024

DOI: 10.17148/IJARCCE.2024.13924

Table 1. Existing Machine Learning Techniques in Cotton Leaf Disease Detection

Sr. No	Author(s)	Technique Used	Dataset	Problem Statement	Performance Metrics	Shortcomings or Limitations
01	Singh et al. (2023) [9]	Convolutional Neural Networks (CNN), deep learning.	Manually collected images of cotton leaves from various regions of Punjab (22 classes).	Detection of cotton leaf diseases in India using deep learning to improve agricultural production.	Accuracy of 99.39%.	Limited to a specific region and crop, may not generalize to other plants.
02	Rai and Pahuja (2023) [10]	Deep Convolutional Neural Network (DCNN), deep learning, image processing.	2,293 in-field images of cotton leaves from Kaggle.	Identification and prediction of cotton plant diseases using deep learning, aiming to reduce time and human error.	Classification accuracy of 97.98%.	Dataset size is relatively small for deep learning.
03	Qian et al. (2022) [11]	Deep learning, ant colony optimization (ACO), YOLOv5, edge computing, image preprocessing.	Github repository of MicaSense.	Detection and management of cotton root rot (CRR) in cotton fields using optimization algorithms.	Detection accuracy above 70%, inference speed of 11 FPS.	Lower accuracy, limited to a single disease (CRR).
04	Patil and Burkpalli (2022) [12]	Image-processing techniques, modified factorization-based active contour method, machine learning classifiers.	Cotton leaf images with complex backgrounds containing diseases like Alternaria, bacterial blight, etc.	Classify cotton leaf diseases in India to prevent yield loss.	Accuracy of 92.29%, precision of 0.905, MCC score of 0.872.	Complex backgrounds may reduce accuracy.
05	Kumar et al. (2022) [13]	CNN, TensorFlow's Keras API, mobile app development.	Cotton disease images from India.	Develop a CNN-based model to identify cotton diseases and integrate into a mobile app for farmers.	Accuracy of approximately 90%.	Limited accuracy, possible issues with mobile app performance in real-world conditions.
06	Udawant & Srinath (2022) [14]	Mask RCNN, transfer learning, data augmentation (GANs).	2000 photos of cotton leaves with various diseases and pests.	Utilize Mask RCNN and transfer learning for detecting cotton leaf diseases.	Training accuracy of 94%.	Requires large datasets, performance on unseen data not clearly reported.
07	Patil & Patil (2021) [15]	Deep learning, IoT platform, CNN, image preprocessing, image augmentation.	Infected and healthy cotton leaf images collected via IoT.	Design a system for cotton disease detection using deep learning and IoT.	Accuracy of 98% after 10k iterations (F1 score).	Reliance on IoT infrastructure might limit scalability in rural areas.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 13, Issue 9, September 2024

DOI: 10.17148/IJARCCE.2024.13924

Sr. No	Author(s)	Technique Used	Dataset	Problem Statement	Performance Metrics	Shortcomings or Limitations
08	Caldeira et al. (2021) [16]	Digital image processing, texture extraction, neuro-fuzzy classifier.	60,659 images, including soil, healthy leaves, and damaged leaves.	Develop an automatic detection system for lesions on cotton plant leaves.	Overall accuracy of 71.2%, hit rate of 71.1%.	Limited accuracy, especially with complex lesions.
09	Wang et al. (2020) [17]	Remote sensing (RS), UAV, supervised and unsupervised classification, k-means segmentation.	RS images collected by UAV from Texas fields with CRR.	Utilize remote sensing to delineate CRR-infested areas in cotton fields.	Overall accuracy of 88.5%, errors of omission (11.44%) and commission (16.13%).	Dependent on UAV and remote sensing technology, which may not be cost-effective.
10	Prashar et al. (2020) [18]	Multi-layer perceptrons (MLP), kNN, SVM.	40 infected images.	Use visual features and machine learning to recognize and classify cotton leaf diseases.	Accuracy of over 96% in localizing disease regions.	Small dataset limits generalization.
11	Shah & Jain (2019) [19]	Artificial Neural Networks (ANN), image preprocessing techniques, RGB and HSV analysis.	18 leaf images representing six different disease types.	Assess health and detect diseases in cotton plants using image analysis and machine learning.	Validation using statistical tests (t-tests).	Very small dataset, leading to potential overfitting.
12	Udawant & Srinath (2019) [20]	CNN model.	Cotton plant images.	Classify diseased and healthy cotton leaves using CNN.	Training and validation accuracy reported.	Lacks detailed performance metrics for practical deployment.
13	Dubey et al. (2018) [21]	Roughness measure, SLIC, gray level co-occurrence matrix, SVM.	ikisan cotton disease management.	Detect and classify cotton leaf diseases using roughness measure and SLIC algorithm.	Classification accuracy of 94%.	Limited to specific image characteristics, may not generalize well.
14	Chopda et al. (2018) [22]	Decision tree classifier for disease prediction.	ikisan cotton disease management.	Develop a system to predict cotton crop diseases using decision tree classification.	Accuracy not clearly reported.	Missing performance metrics, potentially less accurate with larger datasets.

3. RESEARCH GAPS FOR FUTURE STUDY

A research gap refers to a specific area or aspect of a field of study that has not been sufficiently explored It represents a knowledge or understanding deficit within the current body of literature and signifies the need for further investigation or research to fill that gap. Identifying research gaps is crucial for future research. Researchers often conduct literature reviews to identify these gaps and formulate research questions that can contribute to filling those gaps.
After reviewing the relevant published literature, the following research gaps were discovered and are as under:-

DOI: 10.17148/IJARCCE.2024.13924

Table 2: Research Gap of Various Existing Methods of Cotton Leaf Disease Detection

Sr. No	Author's	Year	Research Gaps identified
1	Singh *et al.*, [9]	2023	Limited coverage of cotton leaf diseases in previous studies. Smaller datasets. Addressing issues related to weak learning, regularization, and bias-variance balance.
2	Rai and Pahuja,[10]	2023	Limited exploration of disease types. Lack of analysis on real-world scenarios. The absence of a detailed comparative analysis with other existing methods.
3	Qian *et al.*, [11]	2022	There is need to improve the detection accuracy beyond 70%, optimize image pre-processing methods. There is a need to explore the implementation of Mask R-CNN. There is a need to evaluate the efficiency of the generated optimal paths to conduct field tests with cost-benefit analyses.
4	Patil and Burkpalli, [12]	2022	Limited exploration of deep learning techniques for plant disease classification.
5	Kumar*et al.*, [13]	2022	There is a need to expand the app's capabilities to detect a broader range of cotton diseases. There is a need to validate and evaluate the accuracy of disease identification.
6	Udawant and Srinath, [14]	2022	There is a need for a more diverse dataset. Lack of evaluation on real-world data. There is a need of further exploration of data augmentation techniques.

4. CONCLUSION

The review emphasizes the significant advancements in utilizing machine learning for predicting plant diseases, while also identifying key areas where further research is needed. By systematically presenting various approaches, datasets, and performance metrics, the study provided a valuable reference for ongoing and future research. The success of machine learning in agriculture hinges not only on optimized algorithms but also on access to diverse and high-quality datasets. Future research should focus on refining algorithms, enhancing data quality, and incorporating diverse datasets to improve model accuracy and reliability. Lastly, harnessing the power of machine learning can lead to more sustainable agricultural practices and improved food security worldwide.

REFERENCES

[1] Liu, J., & Wang, X. (2021). Retrieved from https://plantmethods.biomedcentral.com/articles/10.1186/s13007-021-00722-9
[2] (N.d.-a). Retrieved from https://ieeexplore.ieee.org/document/9388488
[3] Saikat Dutt, Subramanian Chandramouli and Amit Kumar Das, "Machine Learning", Pearson India Education Services Pvt. Ltd, India, (2019).
[4] Alpaydin, E. (2020). Introduction to machine learning. Cambridge, MA: The MIT Press
[5] Gholami, A., & Srivastava, A. K., Comparative analysis of ML techniques for data-driven anomaly detection, classification and localization in Distribution System. 2020 52nd North American Power Symposium (NAPS). https://doi.org/10.1109/naps50074.2021.9449712
[6] Keita, Z. (2022). Classification in machine learning: A guide for beginners. Retrieved from https://www.datacamp.com/blog/classification-machine-learning
[7] Brownlee, J. (2020). Retrieved from https://machinelearningmastery.com/types-of-classification-in-machine-learning/
[8] Keita, Z. (2022). Classification in machine learning: A guide for beginners. Retrieved from https://www.datacamp.com/blog/classification-machine-learning.
[9] P. Singh, P. Singh, U. Farooq, S. S. Khurana, J. K. Verma, and M. Kumar, 'CottonLeafNet: cotton plant leaf disease detection using deep neural networks', Multimed. Tools Appl., 2023, doi: 10.1007/s11042-023-14954-5.
[10] C. K. Rai and R. Pahuja, 'Classification of Diseased Cotton Leaves and Plants Using Improved Deep Convolutional Neural Network', Multimed. Tools Appl., 2023, doi: 10.1007/s11042-023-14933-w.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 13, Issue 9, September 2024
DOI: 10.17148/IJARCCE.2024.13924

[11] Q. Qian et al., 'Cotton crop disease detection on remotely collected aerial images with deep learning', no. June, p. 5, 2022, doi: 10.1117/12.2623039.
[12] B. M. Patil and V. Burkpalli, 'Cotton Leaf Disease Classification by Combining Color and Texture Feature-based Approach', Proc. - 2022 6th Int. Conf. Intell. Comput. Control Syst. ICICCS 2022, no. May, pp. 1783–1789, 2022, doi: 10.1109/ICICCS53718.2022.9788405.
[13] S. Kumar, R. Ratan, and J. V. Desai, 'Cotton Disease Detection Using TensorFlow Machine Learning Technique', Adv. Multimed., vol. 2022, 2022, doi: 10.1155/2022/1812025.
[14] P. Udawant and P. Srinath, 'Cotton Leaf Disease Detection Using Instance Segmentation', J. Cases Inf. Technol., vol. 24, no. 4, pp. 1–10, 2022, doi: 10.4018/JCIT.296721.
[15] B. V. Patil and P. S. Patil, 'Computational method for cotton plant disease detection of crop management using deep learning and internet of things platforms', Lect. Notes Data Eng. Commun. Technol., vol. 53, no. January, pp. 875–885, 2021, doi: 10.1007/978-981-15-5258-8_81.
[16] R. F. Caldeira, W. E. Santiago, and B. Teruel, 'Soft computational techniques to identify cotton leaf damage', Aust. J. Crop Sci., vol. 15, no. 8, pp. 1177–1185, 2021, doi: 10.21475/ajcs.21.15.08.p3256.
[17] T. Wang, J. A. Thomasson, C. Yang, T. Isakeit, and R. L. Nichols, 'Automatic classification of cotton root rot disease based on UAV remote sensing', Remote Sens., vol. 12, no. 8, 2020, doi: 10.3390/RS12081310.
[18] K. Prashar, R. Talwar, and C. Kant, 'CNN based on Overlapping Pooling Method and Multi-layered Learning with SVM KNN for American Cotton Leaf Disease Recognition', 2019 Int. Conf. Autom. Comput. Technol. Manag. ICACTM 2019, no. January, pp. 330–333, 2019, doi: 10.1109/ICACTM.2019.8776730.
[19] N. Shah and S. Jain, 'Detection of Disease in Cotton Leaf using Artificial Neural Network', Proc. - 2019 Amity Int. Conf. Artif. Intell. AICAI 2019, pp. 473–476, 2019, doi: 10.1109/AICAI.2019.8701311.
[20] P. Udawant and P. Srinath, 'Diseased portion cassification & recognition of cotton plants using convolution neural networks', Int. J. Eng. Adv. Technol., vol. 8, no. 6, pp. 3492–3496, 2019, doi: 10.35940/ijeat.F9526.088619.
[21] Y. K. Dubey, M. M. Mushrif, and S. Tiple, 'Superpixel based roughness measure for cotton leaf diseases detection and classification', Proc. 4th IEEE Int. Conf. Recent Adv. Inf. Technol. RAIT 2018, pp. 1–5, 2018, doi: 10.1109/RAIT.2018.8388993.
[22] J. Chopda, H. Raveshiya, S. Nakum, and V. Nakrani, 'Cotton Crop Disease Detection using Decision Tree Classifier', 2018 Int. Conf. Smart City Emerg. Technol. ICSCET 2018, no. c, pp. 1–5, 2018, doi: 10.1109/ICSCET.2018.8537336.
[23] B. S. Prajapati, V. K. Dabhi, and H. B. Prajapati, 'A survey on detection and classification of cotton leaf diseases', Int. Conf. Electr. Electron. Optim. Tech. ICEEOT 2016, pp. 2499–2506, 2016, doi: 10.1109/ICEEOT.2016.7755143.