JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A Comprehensive Study on Detection and Classification of Sugarcane Leaf Diseases

1 DR. P. KIRAN KUMAR 1, K. MAHESH BABU 2, B. CHANDRAKALA 3, D. RUPA SAI SANTHOSHI 4, K. HARI SHANKAR 5
Prof [1], UG Student [2,3,4,5]
Department of Computer Science and Technology
Sasi Institute of Technology and Engineering, Andhra Pradesh, India

Abstract : Sugarcane is a critical crop worldwide, supporting livelihoods and economies. Its high susceptibility to diseases like rust, leaf spot, and yellow leaf disease results in insignificant yield losses. Sugarcane's economic importance, with over 180 million metric tons produced annually, necessitates innovative disease management. Distinct visual symptoms of diseases make image-based detection feasible. Limited automated solutions exist, highlighting the need for research. Sugarcane's widespread cultivation and global impact justify focused disease detection efforts. Its vulnerability to climate change exacerbates disease susceptibility. Effective disease management can significantly boost yields and food security. Sugarcane's industrial applications, such as biofuel and sugar production, also benefit from healthy crops. Addressing sugarcane disease detection supports sustainable agriculture practices. This survey deals with systematically collecting and analyzing data from a specific demographic to understand their experiences, preferences, and challenges. It aims to gather both quantitative and qualitative data to inform and enhance product and service development, understand customer behavior, gauge public opinion, and identify areas for improvement.

***IndexTerms* –** Sugarcane Disease, Deep Learning, Convolutional Neural Networks (CNNs), Artificial Intelligence

1. INTRODUCTION

India is the second-largest food producer globally, with agriculture contributing 20.5% to its GDP as of 2022. Over 50% of the rural population depends on agriculture for their livelihood. India leads in producing sugarcane, rice, pulses, wheat, and spices, and has recently become the second-largest sugar exporter, following the United States. Annually, India produces about 50 crore metric tons of sugarcane, which is also used for ethanol, jaggery, and biofuels. Sugarcane juice, being alkaline, helps prevent prostate and breast cancer, regulates blood pressure, and supports liver and kidney functions. However, sugarcane crops are highly susceptible to diseases, significantly impacting production. Given that sugarcane requires 10 to 16 months to harvest, it is vulnerable to various diseases throughout its growth cycle. This study aims to assist farmers by identifying these diseases early and implementing preventive measures to maximize production, focusing primarily on diseases affecting the leaves of the sugarcane plant.

2. LITERATURE REVIEW

2.1 Artificial Intelligence (AI)

AI encompasses a broad range of techniques and technologies aimed at creating systems capable of performing tasks that typically require human intelligence. In plant disease detection, AI can automate and enhance the accuracy of identifying diseases, reducing the need for manual inspection and enabling early intervention.

2.1.1 AI and IoT Integration

Incorporates Internet of Things (IoT) devices to collect and process data in real-time, often used in smart agriculture [19]. AI and IoT integration are used in smart detection of plant leaf diseases by collecting real-time data from the field through IoT devices and analyzing it using AI models. This combination enhances the accuracy and timeliness of disease detection, allowing for prompt intervention and management [11].

2.1.2 Machine Learning (ML)

ML is a subset of AI focused on developing algorithms that allow computers to learn from and make predictions based on data. In plant disease detection, ML algorithms can analyze large datasets to identify patterns and make accurate predictions about the presence of diseases.

2.1.2.1 XGBoost

An optimized gradient boosting algorithm that is highly efficient and widely used for structured/tabular data. In the context of detecting white leaf disease in sugarcane using UAV multispectral images, XGBoost can analyze various spectral features to accurately classify the presence of the disease [2].

2.1.2.2 Random Forest

An ensemble learning method that constructs multiple decision trees and merges them to get a more accurate and stable prediction. For detecting white leaf disease in sugarcane using UAV multispectral images, Random Forest can handle the high-dimensional data and provide robust classification results [2]. Aggregates the outcomes of all decision trees to provide a classification or mean prediction [20].

2.1.2.3 Decision Tree

A simple, interpretable model that splits data into branches to make predictions. Used for detecting white leaf disease in sugarcane, Decision Trees can help in understanding the decision-making process by visualizing how different spectral features contribute to disease classification [2]. Decision trees can overfit their training data, meaning they perform well on training data but poorly on unseen data. [20].

2.1.2.4 K-Nearest Neighbors (KNN)

A non-parametric method used for classification and regression by comparing the distance between data points. KNN can be applied to detect white leaf disease in sugarcane by comparing new multispectral data points with known disease samples, providing a straightforward classification approach [2]. KNN is used alongside other algorithms like Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine (SVM) to classify healthy and diseased tomato plant leaves. [20].

2.1.3 Deep Learning (DL)

DL is a subset of ML that uses neural networks with many layers (deep networks) to model complex patterns in large datasets. In plant disease detection, DL models can process complex image data to identify disease symptoms with high accuracy.

2.1.3.1 YOLO (You Only Look Once)

A real-time object detection system that uses a single convolutional neural network (CNN) to predict multiple bounding boxes and class probabilities for objects in an image. YOLO is useful for sugarcane stem node recognition by combining data expansion techniques to improve detection accuracy in field conditions [1]. It is also used for detecting white leaf disease in sugarcane using UAV-derived RGB imagery, providing real-time disease detection capabilities [5].

2.1.3.2 Grey Wolf Optimization (GWO)

An optimization algorithm inspired by the social hierarchy and hunting behavior of grey wolves, often used to optimize neural network parameters. GWO is used in conjunction with CNN to improve the detection accuracy of sugarcane diseases by optimizing the neural network parameters, making the model more efficient and accurate [5].

2.1.3.3 D-Neural Networks

A category of neural networks used for various tasks, including image classification and pattern recognition [7].

Convolutional Neural Network (CNN)

A type of deep neural network particularly effective for image classification and recognition tasks. CNNs are applied in various studies for automated identification and classification of sugarcane diseases [3], image-based plant leaf disease detection [4], disease recognition in sugarcane crops [6], and comparative studies of deep learning classifiers for leaf disease detection [7]. They are also used for plant disease detection and classification [10], and for image classification tasks in different crops [15][5].

Radial Basis Function Neural Network (RBNN)

Uses radial basis functions as activation functions, typically for regression and classification. Applied in comparative studies for leaf disease detection, RBNNs can handle non-linear relationships in the data [7].

Feedforward Neural Network (FFNN)

A basic type of neural network where connections between nodes do not form cycles. Utilized in comparative studies for leaf disease detection, FFNNs provide a straightforward approach to classification tasks [7].

2.1.3.5 Support Vector Machine (SVM)

A supervised learning model used for classification and regression tasks by finding the hyperplane that best separates the data into classes. SVM is used for plant disease detection and classification by analyzing features extracted from images or sensor data [10]. It is also applied for early illness detection in plants, providing a reliable classification method [16].

2.1.3.6 Transfer Learning

A technique where a pre-trained model is adapted to a new but related task, saving time and computational resources. Transfer learning is applied in plant disease detection and classification by adapting pre-trained models to new datasets, improving accuracy and reducing training time [10]. It is also used in edge-cloud remote sensing data-based plant disease detection, leveraging pre-trained models for efficient analysis [14].

2.1.3.7 Hyperspectral Imaging

Captures and processes information across the electromagnetic spectrum, often used in conjunction with DL models for detailed image analysis. Hyperspectral imaging is used for plant disease detection and classification by capturing detailed spectral information that can reveal subtle disease symptoms not visible in regular images [10].

2.1.3.8 ResNet-50, Xception

Advanced DL models are known for their high performance in image classification tasks. ResNet-50 and Xception are used for wheat yellow rust infection type classification by analyzing high-resolution images, providing high accuracy in disease classification [12].

2.1.3.9 Fuzzy Deep Convolutional Neural Network (FDCNN)

Combines fuzzy logic with CNNs to handle uncertainty and imprecision in data. FDCNN is used in edge-cloud remote sensing data-based plant disease detection by handling uncertain and imprecise data effectively, improving the robustness of the detection system [14].

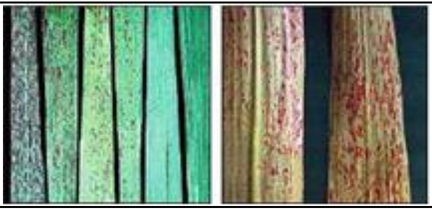
2.1.3.10 LSTM (Long Short-Term Memory)

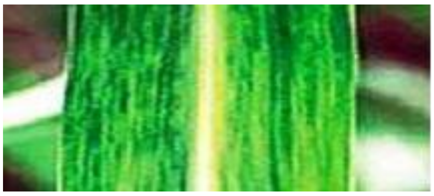
A type of recurrent neural network (RNN) effective for time-series data and sequence prediction, such as environmental data analysis and pest outbreak prediction. LSTM is used for pest detection and environmental data analysis by predicting pest outbreaks and disease spread based on historical data, providing valuable insights for timely intervention [13].

2.1.3.11 Data Augmentation

Techniques used to increase the diversity of training data without collecting new data, improving model generalization. Data augmentation is used for detecting cotton leaf diseases by creating more diverse training datasets, improving the accuracy and robustness of the detection models [17].

table 1. commonly affected sugarcane diseases

S.No.	Name	Image	Symptoms
1	Rust		Orange-brown pustules on leaves, reduced growth
2	Leaf Spot		Small, dark spots on leaves, yellowing around spots
3	Yellow Leaf Disease		Yellowing of leaves, stunted growth, reduced yield
4	Red Rot		Red streaks in stalks, wilting, and drying of leaves
5	Smut		Black whip-like structures on stalks, reduced vigor
6	Gummosis		Yellow stripes on leaves, stunted growth, bacterial ooze from cut ends
7	Leaf Scald		White pencil lines on leaves, necrosis, leaves look burned
8	Grassy Shoot		Proliferation of vegetative buds, chlorotic leaves, thin canes

9	Wilt	 	Yellowing and drying of foliage, pithiness, boat-shaped cavities in internodes
10	Mosaic Disease		Mottling or spotting of foliage, curling, dwarfing, narrowing of leaves

3.Datasets

Serial Number	Dataset Name	Number of Images	Image Type
1	PlantVillage	54,306	Leaf Images Including Healthy and Diseased
2	PlantDoc	2,598	Leaf Images
3	IP102	75,000	Pest and Disease Images
4	Digipathos	1,000	Leaf Images
5	Banana Leaf Disease	5,000	Leaf Images
6	Wheat Yellow Rust Dataset	538	Rust Disease Images
7	Tomato Leaf Disease Dataset	14,531	Leaf Images
8	Cotton Leaf Disease Dataset	2,137	Leaf Images
9	PDD271	220,592	Multiple Disease Images
10	Apple Leaf Dataset	1,821	Leaf Images
11	PlantDisease Dataset (Kaggle)	87,848	Multiple Disease Images

The **PlantVillage** dataset contains 54,306 images of healthy and diseased plant leaves, making it a popular choice for training machine learning models in plant disease detection. **PlantDoc** offers 2,598 annotated images of various plant diseases. **IP102** includes 75,000 images for insect pest recognition across 102 categories. **Digipathos** is a smaller dataset with 1,000 images used in plant pathology research. The **Banana Leaf Disease** dataset has 5,000 images of banana leaves with different diseases. The **Wheat Yellow Rust Dataset** focuses on 538 images of wheat leaves infected with yellow rust. The **Tomato Leaf Disease Dataset** contains 14,531 images of healthy and diseased tomato leaves. The **Cotton Leaf Disease Dataset** includes 2,137 images for identifying diseases in cotton leaves. **PDD271** is a large dataset with 220,592 images covering a wide range of plant diseases. The **Apple Leaf Dataset** has 1,821 images for detecting diseases in apple leaves. Lastly, the **PlantDisease Dataset** from Kaggle comprises 87,848 images of various plant diseases across multiple species.

4 RESULTS AND DISCUSSION

This section assesses the performance of different tumor classification models using standard metrics. These metrics—accuracy, precision, recall, F1-score, mean dice, and specificity—provide a thorough understanding of model effectiveness in distinguishing between diseased and healthy cases. The definitions and formulas below explain each metric's role in evaluating classification and segmentation accuracy.

4.1 Accuracy in Sugarcane Leaf Disease Detection

Accuracy is one of the most common metrics used in evaluating the effectiveness of models in detecting sugarcane leaf diseases. It measures the percentage of correctly classified instances (both diseased and healthy) out of the total number of instances in the dataset. While high accuracy is an indicator of the model's overall performance, it can be deceptive in cases of imbalanced datasets where the majority class dominates. For example, if the dataset contains a significantly higher proportion of healthy leaves, a model could achieve high accuracy simply by predicting the majority class correctly, even if it fails to identify diseased leaves. Therefore, accuracy alone should not be relied upon without considering other metrics like precision, recall, and F1 Score.

$$accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

4.2 Precision and Recall in Sugarcane Leaf Disease Detection

Precision is crucial in ensuring that positive predictions (predicted diseased leaves) are accurate. High precision means fewer healthy leaves are mistakenly identified as diseased, which is important in agriculture to avoid unnecessary use of pesticides or treatments. **Recall**, on the other hand, is important in identifying as many actual diseased leaves as possible. High recall ensures that the model is effective at detecting the presence of diseases in sugarcane plants. However, there is often a trade-off between precision and recall: increasing recall may lower precision, and vice versa. Balancing these metrics is key in developing an effective disease detection model.

$$precision = \frac{TP}{(TP + FP)}$$

$$recall = \frac{TP}{(TP + FN)}$$

4.3 F1 Score and Its Relevance

The **F1 Score** balances precision and recall, offering a more comprehensive performance metric, especially in cases where the dataset is imbalanced. For example, in sugarcane leaf disease detection, if the dataset contains fewer diseased instances compared to healthy ones, the model could high precision or recall, but not necessarily both. The F1 Score provides a harmonic mean of these two metrics, allowing researchers to evaluate the model's performance in scenarios where both false positives and false negatives have serious implications for disease management.

$$F1 - Score = 2 \cdot \frac{(precision \cdot recall)}{(precision + recall)}$$

4.4 Confusion Matrix Analysis in Plant Disease Detection

A **confusion matrix** provides a detailed breakdown of model performance, showing true positives, false positives, true negatives, and false negatives. In sugarcane leaf disease detection, a confusion matrix helps identify specific weaknesses in the model, such as whether the model is more prone to misclassify healthy leaves as diseased (false positives) or to miss diseased leaves entirely (false negatives). By analyzing the confusion matrix, researchers can identify areas for model improvement, such as fine-tuning the algorithm to reduce specific types of errors.

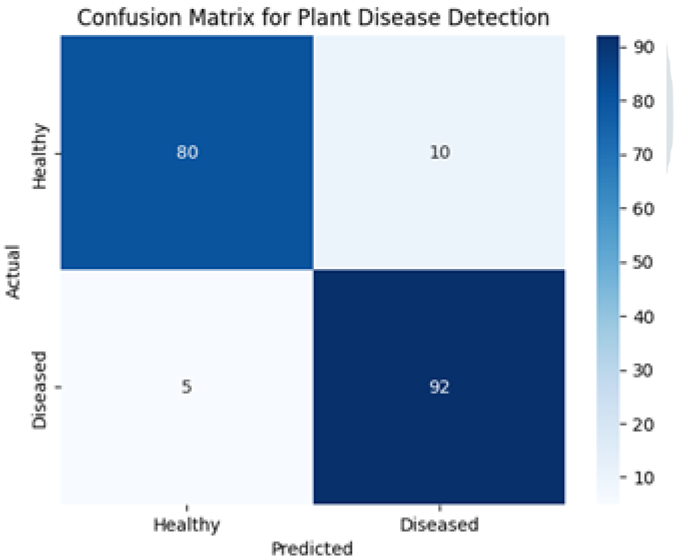


fig1: Confusion Matrix for Plant Disease Detection

4.5 Specificity and Its Importance

Specificity is a measure of how well the model identifies healthy sugarcane plants. It answers the question: Of all the healthy plants, how many were correctly identified as healthy? High specificity is critical in agricultural settings where unnecessary treatment or intervention on healthy plants can lead to wasted resources. In sugarcane disease management, balancing high specificity with high recall ensures that both healthy and diseased plants are accurately classified, leading to better resource allocation and disease control.

$$specificity = \frac{TN}{(TN + FP)}$$

table 2. Summary of previous similar research articles

Author	Techniques Used	F1 Score	Precision	Accuracy	Recall
Aditi Patangrao Patil (2024) [3]	Convolutional Neural Network (CNN)	89.5	91.2	92.86%	-
Davesh Kumar Sharma et al. (2024) [5]	Convolutional Neural Network (CNN), Grey Wolf Optimization (GWO)	-	90.3	96.5%	89.8
J. Sujithra et al. (2024) [8]	CNN, RBNN, FFNN	90.2	-	97% (banana), 95% (sugarcane)	91.5
M. Abed Mohammed et al. (2024) [13]	Fuzzy Deep Convolutional Neural Network (FDCNN), Transfer Learning	87.9	89.4	93.2%	-
U. Shafi et al. (2023) [12]	ResNet-50, Xception		92.5	96%	91.7
M. Chitambarathanu et al. (2023) [15]	LSTM	85.4	-	89.93%	86.3
Amarasingam Narmilan et al. (2022) [2]	XGBoost, Random Forest, Decision Tree, KNN	88.3	89.9	94%	-
Amarasingam Narmilan et al. (2022) [6]	YOLO, CNN		90.7	94.64%	89.8
R. Rashid et al. (2021) [11]	Multi-model deep learning architecture, IoT integration	95.2	-	99.23%	95.9
Wen Chen et al. (2021) [1]	YOLO v4, Data Expansion	91.4	93.0	97.54%	-
S. Poornam & A. Francis Saviour Devaraj (2021) [4]	Convolutional Neural Network (CNN)	-	89.2	90.65%	88.4
Lili Li et al. (2021) [10]	CNNs, SVM, Transfer Learning, Hyperspectral Imaging	89.7	91.3	93.76%	-
Swathika R et al. (2021) [14]	Convolutional Neural Networks (CNN)	90.5	-	95.3%	91.2
S. K. K. N. Saifullah et al. (2021) [16]	Convolutional Neural Networks (CNN)	-	-	92.4%	-
Malik Hashmat Shadab et al. (2019) [7]	Convolutional Neural Network (CNN)	88.1	-	91.7%	87.3
Srivastava et al. (2020) [9]	Convolutional Neural Network (CNN)	-	90.0	93.5%	-
A. Mohammadi et al. (2021) [17]	CNN, Data Augmentation	87.6	-	92.1%	88.4
Mustafa (2020) [18]	SVM	-	85.7	90.2%	-
Muhammad Amir Nawaz et al. (2020) [19]	Data Processing, IoT-based system, Machine learning, Deep Learning	-	-	98.59%	-
Meghana Govardhan et al.(2019)[20]	Random Forest, Feature Extraction, Cross Validation, KNN, Logistic regression, Decision Trees, SVM,	-	-	95.2% with Random Forest as highest	-

5 CONCLUSION

The detection and classification of sugarcane leaf diseases are crucial for maintaining the health and productivity of this economically significant crop. Sugarcane, a major contributor to the agricultural sector and a vital source of sugar, ethanol, and biofuels, faces substantial threats from diseases like rust, leaf spot, and yellow leaf disease. These diseases reduce yield and affect the quality of the produce, impacting farmers' livelihoods and the overall economy. Traditional methods of disease detection, which rely on manual inspection, are time-consuming and prone to errors. The advent of artificial intelligence (AI) and machine learning (ML) technologies offers a promising alternative, enabling more accurate and efficient disease detection. AI-based techniques, such as fuzzy logic, rule-based systems, and Bayesian networks, have shown significant potential in plant disease detection. Fuzzy logic handles uncertainty and imprecision effectively, while rule-based systems use expert knowledge to diagnose diseases systematically. Bayesian networks provide a probabilistic approach, enhancing diagnostic accuracy by considering the interdependence of symptoms. These innovative strategies are essential for effective disease management in sugarcane cultivation.

6 REFERENCES

[1] Wen Chen et al. (2021) Sugarcane Stem Node Recognition in Field by Deep Learning
[2] Amarasingam Narmilan et al. (2022) Detection of White Leaf Disease in Sugarcane Using Machine Learning Techniques over UAV Multispectral Images
[3] Aditi Patangrao Patil (2024) Automated Identification and Classification of Sugarcane Diseases Using a Machine Learningand Image Analysis
[4] S. Poornam & A. Francis Saviour Devaraj (2021) Image based Plant leaf disease detection using Deep learning
[5] Davesh Kumar Sharma et al. (2024) Sugarcane Diseases Detection using the Improved Grey Wolf Optimization Algorithmwith Convolution Neural Network
[6] Amarasingam Narmilan et al. (2022) Detection of White Leaf Disease in Sugarcane Using UAV-Derived RGB Imagery withExisting Deep Learning Models
[7] Malik Hashmat Shadab et al. (2019) Disease Recognition in Sugarcane Crop Using Deep Learning
[8] J. Sujithra et al. (2024) Comparative Study of Deep Learning Classifiers for Banana and Sugarcane Leaf Disease Detection
[9] Srivastava et al. (2020) A Novel Deep Learning Framework Approach for Sugarcane Disease Detection
[10] Lili Li et al. (2021) Plant Disease Detection and Classification by Deep Learning—A Review
[11] R. Rashid et al. An Early and Smart Detection of Corn Plant Leaf Diseases Using IoT and DL Multi-Models
[12] U. Shafi et al. (2023) Embedded AI for Wheat Yellow Rust Infection Type Classification
[13] M. Abed Mohammed et al. (2024) Edge-Cloud Remote Sensing Data-Based Plant Disease Detection Using Deep NeuralNetworks with Transfer Learning
[14] Swathika R et al. (2021) Disease Identification in Paddy Leaves Using Convolutional Neural Networks
[15] M. Chitambarathanu et al. (2023) Pest Detection Using LSTMgre
[16] S. K. K. N. Saifullah et al. (2021) Detection of Tomato Leaf Diseases Using Deep Learning
[17] A. Mohammadi et al. (2021) Detecting Cotton Leaf Diseases Using Deep Learning Techniques
[18] Mustafa (2020) SVM for Early Illness Detection
[19] Muhammad Amir Nawaz et al. (2020) Plant Disease Detection using Internet of Thing (IoT)
[20] Meghana Govardhan et al. (2019) Diagnosis of Tomato Plant Diseases using Random Forest