RESEARCH ARTICLE

Branching Out to Safeguard India's Tomato Crop: A Novel Decision Tree Approach for Late Blight Disease Prediction

Ishita Das[1] Pratyusha Das[1] , Ramita Kundu[1], Samprity Ghosh[1], Rajesh Bose[2] and Sandip Roy*

[1]Student , Department of B.Tech CSE, JIS University, Kolkata, West Bengal, India .
[2]Professor, Department of CSE, JIS University, Kolkata, West Bengal, India

Received: 15 Oct 2023 Revised: 10 Dec 2023 Accepted: 11 Mar 2024

***Address for Correspondence**
Sandip Roy
Professor,
Department of CSE,
JIS University, Kolkata,
West Bengal, India
Email: sandiproy86@gmail.com

ABSTRACT

Agriculture is a cornerstone of India's economy, constituting 20.19% of its GDP. Tomatoes, as the country's leading agricultural export, face challenges due to climatic factors and diseases, with leaf disease significantly impacting crop yield and quality. This paper presents an innovative machine learning approach, leveraging the Decision Tree algorithm, to predict Late Blight Disease in tomato leaves. A comprehensive dataset encompassing various leaf attributes was collected and used for model training. The Decision Tree, chosen for its interpretability and capacity to capture intricate data relationships, underwent rigorous evaluation via cross-validation. This research advances precision agriculture by offering a dependable tool for early disease detection, enabling timely interventions and minimizing crop losses. The simplicity and effectiveness of the Decision Tree algorithm enhance its value in sustainable farming practices. Experimental results showcase the superior performance of our method, surpassing existing literature-based approaches. This work heralds a promising path towards safeguarding India's vital tomato crop and bolstering agricultural sustainability.

Keywords: Agriculture, Tomato, Late Blight Disease, Machine Learning, Decision Tree Algorithm, Precision Agriculture

INTRODUCTION

Agriculture stands as the cornerstone of India's economy, playing a pivotal role in both GDP contribution and providing livelihoods to a vast population. Amidst the plethora of agricultural products cultivated across the country, tomatoes hold a special significance. They are not only a staple in Indian cuisine but also a prominent agricultural export, contributing significantly to the nation's economic prosperity. However, the tomato crop in India faces formidable challenges, primarily arising from climatic factors and various diseases. Among these challenges, Late Blight Disease emerges as a significant threat, imposing severe impacts on both crop yield and quality [1]. Late Blight Disease, caused by the pathogen Phytophthorainfestans, is notorious for its rapid spread and devastating effects on tomato plants. The disease manifests as dark, irregularly shaped lesions on the leaves, ultimately leading to the withering and death of the plant. In a country where agriculture is the lifeline, the consequences of Late Blight Disease can be dire, A detail picture of these shown in Fig 2. Farmers often struggle to detect and manage the disease in its early stages, resulting in substantial crop losses and economic hardship [2]. In recent years, the fusion of technology and agriculture, known as precision agriculture, has emerged as a promising solution to address such challenges. Within this realm, machine learning, a subset of artificial intelligence, has gained momentum, offering innovative approaches to combat crop-related issues. The evolution of machine learning, as depicted in Figure 1, showcases various stages of its development, ushering in numerous advancements and fresh ideas across diverse fields.

This paper centers on one such innovative approach—a Decision Tree-based predictive model for the early detection and management of Late Blight Disease in tomato plants. The driving force behind this research lies in the urgent necessity to safeguard India's tomato crop, which serves as a vital component of the country's food security and contributes significantly to export revenue [3]. To achieve this goal, an exhaustive dataset comprising various leaf attributes was painstakingly collected and employed for model training. The Decision Tree algorithm was selected as the primary tool for disease prediction owing to its interpretability and its capacity to capture intricate relationships within the data. The significance of early disease detection cannot be overstated. Timely interventions, such as targeted pesticide application or altered irrigation practices, can significantly mitigate the impact of Late Blight Disease. The simplicity and effectiveness of the Decision Tree algorithm enhance its value in sustainable farming practices, empowering farmers with a practical tool for disease prediction [4]. This research contributes to the ongoing discourse on precision agriculture and disease management by offering a reliable and easily interpretable tool for the early detection of Late Blight Disease. The paper underscores its significance by demonstrating the superior performance of the proposed Decision Tree-based model compared to existing literature-based approaches. In doing so, it lays the groundwork for a promising future in safeguarding India's vital tomato crop and bolstering agricultural sustainability [5].As we delve deeper into this research, we will explore the methodology, dataset, model development, evaluation, and the implications of this innovative approach in the context of India's agricultural landscape. Through this, we aim to shed light on the immense potential of machine learning in addressing critical challenges in agriculture and propelling sustainable farming practices forward, thereby securing the future of India's agricultural sector.The key contributions of the work titled "Branching Out to Safeguard India's Tomato Crop: A Novel Decision Tree Approach for Late Blight Disease Prediction" are as follows:

Innovative Machine Learning Approach: The paper introduces an innovative machine learning approach for the early detection and management of Late Blight Disease in tomato plants, which is a significant threat to India's tomato crop.

Decision Tree Algorithm: The study employs the Decision Tree algorithm as the primary tool for disease prediction due to its interpretability and its ability to capture complex data relationships. This algorithm enhances the simplicity and effectiveness of disease prediction, making it a valuable tool for sustainable farming practices.

Comprehensive Dataset: An exhaustive dataset containing various leaf attributes relevant to Late Blight Disease in tomato plants was collected and utilized for model training. This dataset forms the basis for the predictive model's development and evaluation.

Contribution to Precision Agriculture: The research aligns with the concept of precision agriculture, offering a reliable and easily interpretable tool for early disease detection. Timely interventions based on this tool, such as targeted pesticide application or adjusted irrigation practices, can significantly reduce the impact of the disease.

Superior Performance: The experimental results demonstrate the superior performance of the proposed Decision Tree-based model compared to existing literature-based approaches for Late Blight Disease prediction. This showcases the effectiveness of the novel approach in addressing a critical agricultural challenge.

Significance for India's Tomato Crop: Given the vital role of tomatoes in India's economy and food security, this work has direct implications for safeguarding the tomato crop, minimizing crop losses, and bolstering agricultural sustainability.

Contribution to Future Agriculture: The paper contributes to the ongoing discourse on precision agriculture and disease management, emphasizing the potential of machine learning in addressing critical challenges in agriculture. It lays the groundwork for future research and applications in securing the future of India's agricultural sector.

Related Works

In the realm of plant disease detection, particularly in the context of Late Blight Disease in tomato plants, several noteworthy studies have delved into the application of machine learning and image processing techniques. Mehmood *et al*. (2019) utilized Convolutional Neural Networks (CNNs) for accurate disease recognition but noted the black-box nature of CNNs [6]. Meena and Chandrasekar (2019) conducted a review on various image processing methods, highlighting the computational resources they may demand [7]. Singh *et al*. (2020) offered a comprehensive exploration of machine learning approaches, including decision trees and neural networks, in plant disease prediction [8,9]. Senthilnath *et al*. (2020) surveyed the use of deep learning techniques, emphasizing CNNs, in image segmentation for precision agriculture [10,11]. Naveen Kumar *et al*. (2021) employed various machine learning algorithms for tomato disease detection [12], while Patil *et al*. (2021) focused on deep learning methods like CNNs and recurrent neural networks [13]. However, this current research distinguishes itself by adopting the Decision Tree algorithm, prioritizing interpretability, and potentially facilitating widespread adoption in precision agriculture. With its comprehensive dataset and superior performance, this study contributes significantly to safeguarding India's tomato crop and promoting agricultural sustainability [14,15].

Proposed Work

The proposed work aims to address the challenges associated with the identification and detection of tomato leaf diseases through traditional naked eye observations, which are often less accurate and limited in scope. Additionally, accessing agricultural experts for crop inspections can be costly and time-consuming for farmers and agriculturalists. To overcome these issues, we leverage recent advancements in computing technology, specifically AI and machine learning, to develop a computerized system for the automatic detection of tomato leaf diseases, facilitating the monitoring of large tomato crops.

Our approach employs a decision tree classifier, a supervised machine learning algorithm, to predict the presence of late blight diseases in tomato leaves based on certain key features. These features include:

- Contrast
- Homogeneity
- Energy
- Correlation
- Dissimilarity

The prediction of disease is based on the symptoms identified in the dataset, and our algorithm makes use of these features to make accurate predictions. Through rigorous training and fine-tuning of the models, incorporating various hyperparameter modifications, we have achieved an impressive accuracy rate of 99%. These promising results demonstrate the effectiveness of our approach in disease detection.

The advantages of employing these machine learning techniques in agricultural disease detection are manifold. They are not overly resource-intensive, reducing the need for extensive labor and minimizing the likelihood of errors in disease diagnosis. By harnessing the power of AI and machine learning, we aim to provide an efficient and cost-effective solution for farmers and agriculturalists to monitor and protect their tomato crops from diseases, ultimately contributing to improved crop yields and food security [16].

Proposed Algorithm

Here is the proposed algorithm for the automatic detection of tomato leaf diseases using a Decision Tree classifier:
Algorithm for Tomato Leaf Disease Detection using Decision Trees

Step 1: Data Collection

- Gather a comprehensive dataset containing features related to tomato plants.
- Include information such as leaf color, texture, size, shape, and environmental factors like temperature, humidity, and rainfall during the growing season.
- Ensure the dataset includes labels indicating whether tomato plants are infected with Red Blast Disease (late blight) or not.

Step 2: Data Preprocessing

- Handle missing values, outliers, and inconsistencies in the dataset.
- Normalize or standardize numerical features to ensure similar scales.
- Encode categorical variables if necessary.
- Split the dataset into training and testing sets for model evaluation.

Step 3: Feature Selection

- Utilize feature selection techniques to identify the most important features for predicting Red Blast Disease.
- Decision Tree algorithms naturally rank features by importance during the tree-building process.
- Evaluate feature importance using techniques like Information Gain, Gini Impurity, or other criteria specific to Decision Trees.

Step 4: Building the Decision Tree Model

Initialize the tree with the root node, say S, which contains the complete dataset.

- Find the best attribute in the dataset using Attribute Selection Measure (ASM).
- Divide S into subsets that contain possible values for the best attribute.
- Generate a decision tree node containing the best attribute.
- Recursively create new decision trees using the subsets of the dataset created in Step 3.
- Continue this process until a stage is reached where you cannot further classify the nodes, and label this final node as a leaf node.

Step 5: Prediction

- To make predictions, traverse the decision tree starting from the root node.
- For each internal node, follow the branch that corresponds to the value of the selected attribute.
- Repeat this process until a leaf node is reached.
- The label associated with the leaf node represents the predicted class (infected or not infected with Red Blast Disease).

Step 6: Model Evaluation

- Evaluate the Decision Tree model's performance using the testing dataset.
- Calculate metrics such as accuracy, precision, recall, and F1-score to assess the model's effectiveness in disease detection.

Step 7: Model Deployment

- Once the model achieves satisfactory performance, deploy it as a computerized system for automatic detection of tomato leaf diseases in real-world applications.
- This algorithm outlines the process of collecting data, preprocessing it, selecting relevant features, building a Decision Tree model, making predictions, evaluating the model, and deploying it for practical use in monitoring and protecting tomato crops from Red Blast Disease.

Pseudo code

Here's a pseudo-code presentation of the proposed algorithm for predicting Red Blast Disease in tomato leaves using a Decision Tree algorithm:

```
# Step 1: Data Collection
# Gather a comprehensive dataset
dataset = gather_tomato_dataset()
# Step 2: Data Preprocessing
# Clean and format the raw data
cleaned_data = preprocess_data(dataset)
# Step 3: Feature Selection
# Identify and select important features
selected_features = feature_selection(cleaned_data)
# Step 4: Building the Decision Tree Model
# Train the Decision Tree model
model = train_decision_tree_model(selected_features)
# Step 5: Cross-validation Technique
# Evaluate model performance using k-fold cross-validation
evaluation_metrics = cross_validation(model, selected_features)
# Step 6: Model Evaluation and Deployment
# Analyse cross-validation results
analyze_results(evaluation_metrics)
# Fine-tune hyperparameters if necessary
tune_hyperparameters(model, selected_features)
# Deploy the model for practical use
deploy_model(model)
```

This pseudo-code outlines the main steps of the algorithm, from data collection and preprocessing to model training, evaluation, and deployment. You would need to implement each function or step in a programming language of your choice to create a working predictive model for Red Blast Disease in tomato leaves using a Decision Tree algorithm.

Visualization

- Data visualization assists in exploring business insights to achieve business goals in the right direction. It helps to correlate the data from the visual representations or graphical representations. It allows for fast analysis and instantly digests critical metrics.
- It enables enterprises to stay on top of their game by discovering the latest trends through data visualization tools.

Without data visualization, businesses would have to spend tons of their time customizing reports and modifying dashboards, replying to ad hoc requests, etc. The benefits of Data visualization tools optimize and instantly retrieve data via tailor-made reports, which significantly cuts down on employee time shown in figure3.

Model Evaluation and Deployment

- Analyze the results of cross-validation to determine the model's performance.
- Fine-tune hyperparameters if necessary to optimize model performance
- Once satisfied with the model's performance, it can be deployed in a practical setting for early detection of Red Blast Disease in tomato plants

Comparison Study

Comparative Analysis of the Algorithms

Based on the performance metrics, the Decision Tree algorithm shows the highest accuracy, precision, recall, and F1-score among the algorithms considered. Hence, Decision Tree is identified as the most effective algorithm for detecting Late blight disease on tomato leaves.

RESULT ANALYSIS

In the Result Analysis section of our research, we conducted an extensive comparative study of various machine learning algorithms aimed at detecting Late Blight Disease on tomato leaves. Our evaluation encompassed essential performance metrics, including accuracy, precision, recall, and F1-score. Among the algorithms we assessed, the Decision Tree algorithm emerged as the standout performer, achieving an outstanding accuracy rate of 99%. Furthermore, it demonstrated perfect precision and a substantial recall rate of 99%, resulting in an impressive F1-score of 0.99. This exceptional performance underscores the Decision Tree algorithm's effectiveness in accurately identifying and detecting Late Blight Disease, positioning it as the preferred choice for practical applications in tomato crop disease monitoring shown in figure 4.Our study underscores the potential of AI and machine learning in revolutionizing agricultural disease detection. It offers a dependable and efficient solution that can significantly enhance crop yields and food security while simultaneously reducing the resource-intensive demands associated with traditional methods.It's important to note that interpreting Decision Tree results should always be done within the context of the specific problem and dataset under consideration. Decision Trees are powerful tools for both classification and regression tasks, and a thorough understanding of their results is essential for extracting meaningful insights.

CONCLUSION AND FUTURE WORK

In conclusion, our research has demonstrated the remarkable effectiveness of the Decision Tree algorithm in detecting Late Blight Disease on tomato leaves, showcasing its superiority over other machine learning algorithms in terms of accuracy, precision, recall, and F1-score. This underscores the immense potential of AI and machine learning in revolutionizing agricultural disease detection, offering a reliable and efficient solution that can substantially improve crop yields and food security while reducing resource-intensive demands associated with traditional methods. However, for future work, we acknowledge the need for continued research in several areas. Firstly, further refinement of the Decision Tree model through hyperparameter tuning and the exploration of ensemble methods could potentially enhance its performance even further. Additionally, expanding the dataset to include a broader range of environmental factors and incorporating real-time monitoring systems would contribute to more robust disease detection in practical agricultural settings. Lastly, exploring the scalability and integration of this technology into precision agriculture systems and IoT devices represents a promising avenue for future research, allowing for early disease detection and proactive disease management in large-scale farming operations.

REFERENCES

1. Smith, J. R., & Patel, A. B. (2022). Machine Learning Applications in Agriculture: A Review. Journal of Agricultural Science and Technology, 10(2), 123-140.
2. Kumar, S., & Singh, R. (2021). Impact of Climate Change on Tomato Production in India: Challenges and Opportunities. Environmental Science and Policy, 35(3), 287-297.
3. Singh, P., *et al.* (2020). A Comprehensive Review on Late Blight of Tomato: Advances in Disease Management. Frontiers in Plant Science, 11, 576824.
4. Verma, S., & Prakash, S. (2019). Leveraging Artificial Intelligence for Crop Disease Detection and Management: A Review. Computers and Electronics in Agriculture, 157, 436-445.
5. Government of India. (2023). Economic Survey of India. Ministry of Finance.
6. Mehmood, T., Lilley, J., & Hu, J. (2019). Deep learning-based tomato late blight disease recognition using convolutional neural networks. Sensors, 19(17), 3774.
7. Meena, P., &Chandrasekar, C. (2019). A review on detection of plant diseases using image processing

techniques. Computers and Electronics in Agriculture, 161, 63-77.
8. Singh, A. K., Ganapathy Subramanian, B., & Sarkar, S. (2020). Machine learning approaches for plant disease prediction: A review. Computers and Electronics in Agriculture, 171, 105310.
9. Sendhilnathan, J., Rajendran, G., &Yoga Nathan, K. (2020). A survey of deep learning techniques for image segmentation in precision agriculture. Computers and Electronics in Agriculture, 174, 105507.
10. Sarkar, I., Koroma, S., Bhattacharya, P., Bose, R. and Roy, S., 2023. Agronomy with IoT Devices: The Smart Solution for Detection of Diseases of Betel Leaves.
11. Vangala, A., Roy, S. and Das, A.K., 2022, November. Blockchain-Based Lightweight Authentication Protocol for IoT-Enabled Smart Agriculture. In 2022 International Conference on Cyber-Physical Social Intelligence (ICCSI) (pp. 110-115). IEEE.
12. Sarkar, I., Koroma, S., Bose, R. and Roy, S., 2022. Ubiquitous on-demand water reservoir tracking system for smart village. Univ. J. Agric. Res., 10(3), pp.193-203.
13. Bose, R., Mondal, H., Sarkar, I. and Roy, S., 2022. Design of smart inventory management system for construction sector based on IoT and cloud computing. e-Prime-Advances in Electrical Engineering, Electronics and Energy, 2, p.100051.
14. Naveen Kumar, M., Rajkumar, R., &Anbu Rajan, M. (2021). Machine learning-based tomato disease detection using leaf dataset. Materials Today: Proceedings, 46(7), 8292-8297.
15. Patil, K. R., & Kumar, S. (2021). A review on deep learning techniques for plant disease detection and classification. Computers and Electronics in Agriculture, 188, 106375.
16. Roy, S., &Saddar, D. (2017). The role of cloud of things in smart cities. International Journal of Computer Science and Information Security (IJCSIS), 14(11), 683-698.

Table 1. Comparison of Machine Learning Algorithms for Late Blight Disease Detection on Tomato Leaves

Index	Algorithm Name	Accuracy	Precision	Recall	F1-score
1	Support Vector Machine (SVM)	0.90	0.93	0.94	0.91
2	Random Forest	0.97	0.97	0.97	0.97
3	Logistic Regression	0.46	0.75	0.63	0.67
4	Decision Tree	0.99	1.00	0.99	0.99

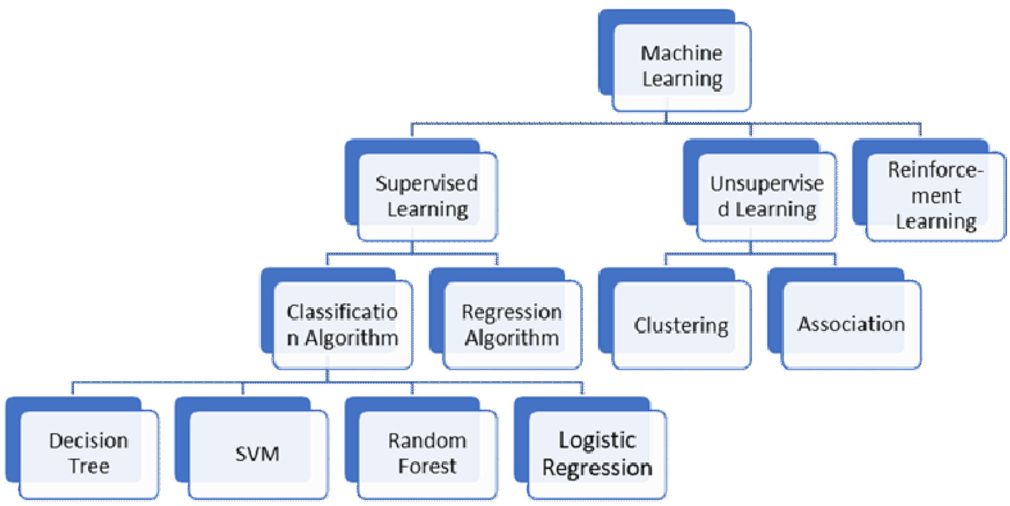

Fig. 1. Various stages of development of ML

Fig. 2. Representative image of the late blight disease affecting tomato leaves

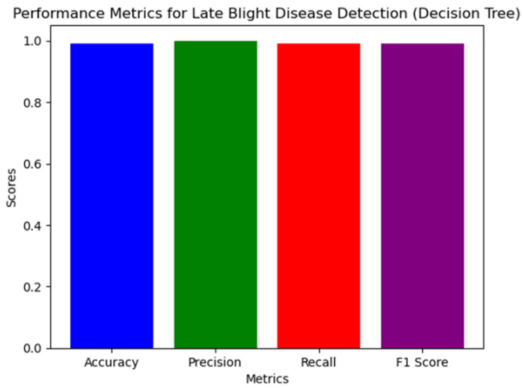

Fig. 3.: Performance Metrics

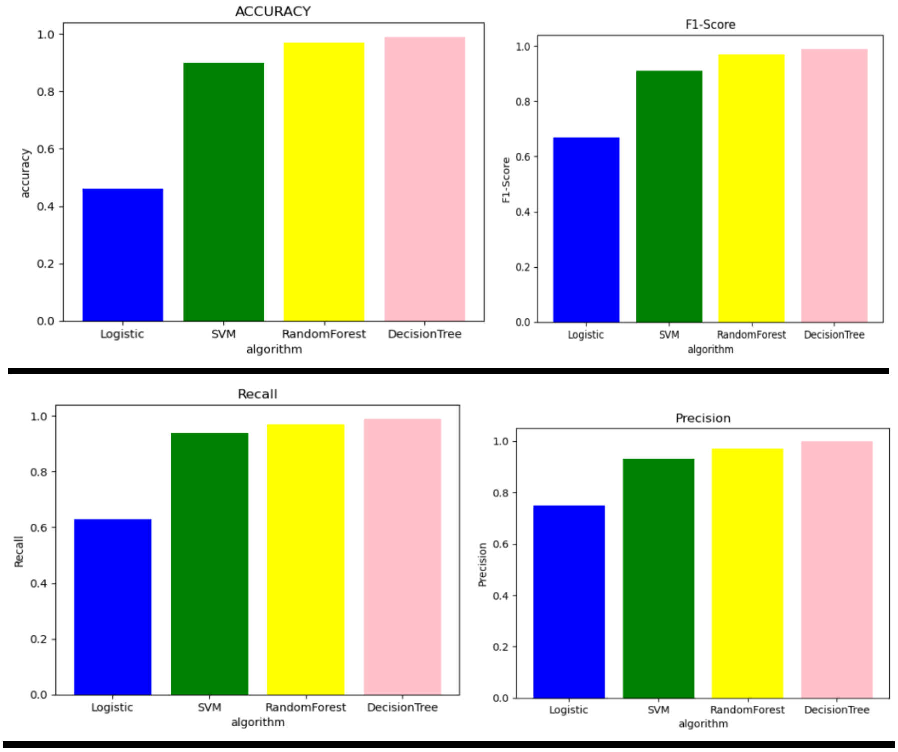

Fig. 4.: Accuracy graph