iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** XII **Month of publication:** December 2023

DOI: https://doi.org/10.22214/ijraset.2023.57452

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue XII Dec 2023- Available at www.ijraset.com

Detection and Classification of Cassava Diseases using Machine Learning

Jagadish. K[1], Jagadhiswar[2], Sai Pavan. K[3], Vishnu Vardhan. B[4], Jahnavi. P[5], Janaki. M[6], Prof. Dr. K. Rajeshwar Rao[7]
[1, 2, 3, 4, 5, 6]*Department of AI-ML, Malla Reddy University*
[7]*Department of AI-ML, Malla Reddy University, Maisammaguda, Hyd*

Abstract: ***Cassava (Manihot esculenta) is a crucial food crop sustaining millions of people worldwide. However, the presence of diseases in cassava plants poses a significant threat to agricultural productivity. Traditional methods of disease detection are often labor-intensive, subjective, and prone to human error. To address these challenges, this research focuses on developing an automated system for cassava plant disease detection using machine learning. After the algorithms have been trained on the dataset, the accuracy of the algorithms is compared, the photos are categorised, and preventions for unhealthy plants are proposed. Apart from detection, this aims to support and help the greenhouse farmers in an efficient way. Plant disease identification by visual way is more laborious task and at the same time, less accurate and can be done only in limited areas. Whereas if automatic detection technique is used it will take less efforts, less time and become more accurate.***

I. INTRODUCTION

Cassava (Manihot esculenta) is a vital crop for food security, particularly in regions where it serves as a staple food. However, the sustainable cultivation of cassava faces significant challenges due to various diseases that affect the crop, leading to reduced yields and economic losses. Timely and accurate detection of these diseases is critical for implementing effective control measures and ensuring food security.Machine learning models can be trained on large datasets containing images of both healthy and diseased cassava leaves. By learning patterns and features from these images, the models can generalize their knowledge to accurately identify diseases in new, unseen images. This approach holds the potential to provide farmers with a rapid and reliable tool for early detection, allowing for timely intervention and the implementation of targeted management practices.This research seeks to harness the power of machine learning for cassava plant disease detection, addressing the limitations of traditional methods and offering a scalable solution that can be deployed in agricultural settings.. In this context, the following sections will delve into the methodology, challenges, and potential impacts of utilizing machine learning for cassava plant disease detection.

II. LITERATURE SURVEY

Research on the detection and classification of cassava diseases using machine learning has gained momentum in recent years. Numerous studies have explored the application of various algorithms and techniques to enhance the accuracy and efficiency of disease identification in cassava crops.Commonly utilized machine learning algorithms include convolutional neural networks (CNNs), support vector machines (SVMs), decision trees, and ensemble methods. CNNs, in particular, have demonstrated success in image-based disease recognition tasks by automatically learning hierarchical features from cassava plant images.Datasets play a crucial role in model training and evaluation. Researchers often employ datasets containing labeled images of healthy and diseased cassava plants. The availability of diverse and well-annotated datasets contributes to the robustness of machine learning models.Feature extraction and selection techniques also play a significant role in improving model performance. Extracting relevant features from cassava images helps capture essential patterns related to disease symptoms, contributing to accurate classification.Transfer learning is another approach that leverages pre-trained models on large datasets and fine-tunes them for cassava disease classification. This method is beneficial when limited labeled data for cassava diseases is available.

III. PROBLEM STATEMENT

Cassava (Manihot esculenta) is a vital staple crop for millions of people, particularly in Sub-Saharan Africa, providing a significant source of carbohydrates. However, cassava plants are susceptible to various diseases, and one of the critical challenges faced by farmers is the timely and accurate detection of diseases affecting cassava leaves.The traditional methods of disease detection in cassava plants often rely on visual inspection by farmers, which may not be reliable or timely. Additionally, the lack of expertise in identifying specific diseases can lead to misdiagnosis and inadequate treatment.

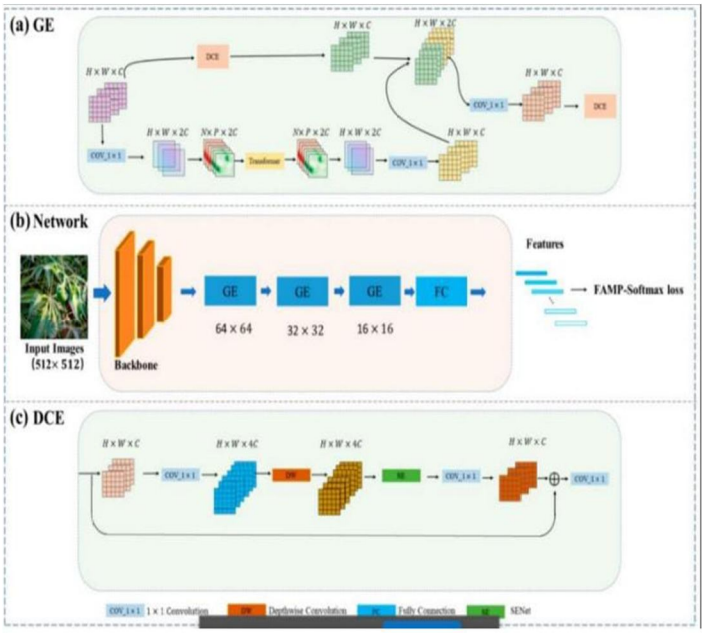
©IJRASET: All Rights are Reserved

Therefore, there is a need for an automated and accurate system for cassava leaf disease detection using machine learning.By addressing these challenges, the proposed machine learning solution aims to enhance the efficiency and accuracy of cassava leaf disease detection, contributing to improved crop management practices, increased agricultural productivity, and ultimately, food security for communities relying on cassava as a primary food source.

IV. SYSTEM DESIGN

Research on the detection and classification of cassava diseases using machine learning has gained momentum in recent years. Numerous studies have explored the application of various algorithms and techniques to enhance the accuracy and efficiency of disease identification in cassava crops. Commonly utilized machine learning algorithms include convolutional neural networks (CNNs), support vector machines (SVMs), decision trees, and ensemble methods. CNNs, in particular, have demonstrated success in image-based disease recognition tasks by automatically learning hierarchical features from cassava plant images. Datasets play a crucial role in model training and evaluation. Researchers often employ datasets containing labeled images of healthy and diseased cassava plants. The availability of diverse and well-annotated datasets contributes to the robustness of machine learning models. Feature extraction and selection techniques also play a significant role in improving model performance. Extracting relevant features from cassava images helps capture essential patterns related to disease symptoms, contributing to accurate classification. Transfer learning is another approach that leverages pre-trained models on large datasets and fine-tunes them for cassava disease classification. This method is beneficial when limited labeled data for cassava diseases is available. Evaluation metrics such as accuracy, precision, recall, and F1 score are commonly used to assess the performance of machine learning models. Studies often compare the results of different algorithms to identify the most effective approach for cassava disease detection and classification.

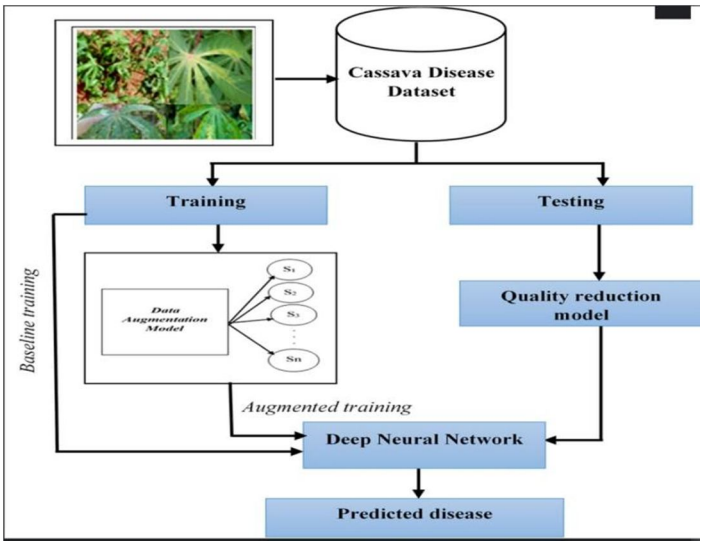
A. Architecture Diagram



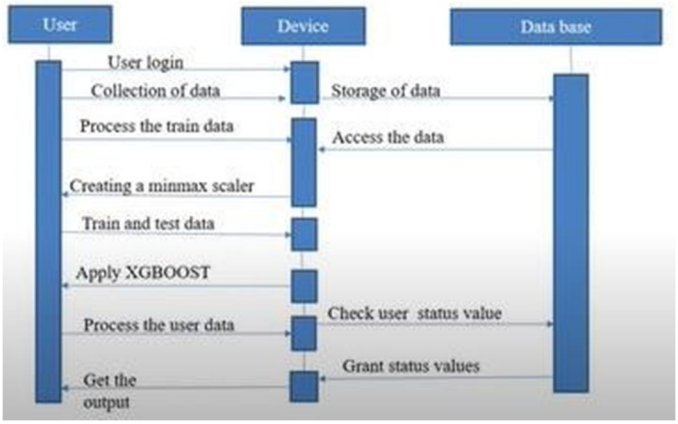
B. Training Data

Splitting the dataset into Training set and testing set:

In machine learning data preprocessing, we have to break our dataset into both training set and test set. This is often one among the crucial steps of knowledge preprocessing as by doing this, we will enhance the performance of our machine learning model. Suppose, if we've given training to our machine learning model by dataset and that we test it by a totally different dataset. Then, it'll create difficulties for our model to know the correlations between the models. If we train our model alright and its training accuracy is additionally very high, but we offer a replacement dataset there to, then it'll decrease the performance. So we always attempt to make a machine learning model which performs well with the training set and also with the test dataset.



C. Sequence Diagram





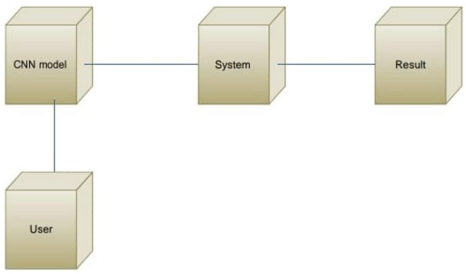
D. Data Pre-Processing

In analyzing plant diseases using machine learning techniques, effective data preprocessing is crucial. Common techniques include:

1) *Data Cleaning:* Remove missing values and outliers to ensure data quality. Impute missing values using methods like mean, median, or machine learning algorithms. Normalization/Standardization: Scale numerical features to a standard range to prevent dominance of certain features.
2) *Feature Selection:* Identify and keep relevant features to reduce dimensionality and computational complexity. Techniques include correlation analysis, recursive feature elimination, or using domain knowledge.

3) *Data Augmentation:* Generate additional training samples by applying transformations like rotation, flipping, or zooming to address limited data issues.
4) *Image Preprocessing:* For image-based data, techniques such as resizing, cropping, and color normalization can enhance model performance. Label Encoding/One-Hot Encoding: Convert categorical variables into a format suitable for machine learning models.
5) *Handling Imbalanced Data:* Address class imbalance issues by oversampling minority classes, undersampling majority classes, or using synthetic data generation techniques.
6) *Temporal Considerations:* If dealing with time-series data, account for temporal dependencies and trends in the preprocessing steps.
7) *Cross-Validation:* Implement techniques like k-fold cross-validation to assess model generalization performance.
8) *Data Splitting:* Divide the dataset into training and testing sets to evaluate model performance on unseen data.
9) *Handling Skewed Distributions:* If the target variable has a skewed distribution, consider transformations (e.g., log transformation) to normalize it.
10) *Noise Reduction:* Apply filters or smoothing techniques to reduce noise in the data, particularly relevant in image processing.
11) *Data Compression:* Use techniques like Principal Component Analysis (PCA) to reduce dimensionality while retaining key information.

E. Data Flow Diagram



V. EXPERIMENT RESULT

To demonstrate the results of our project, we take the remaining test data and it is tested using three algorithms. After that our trained model to ready to predict the disease is present or not. The test accuracy is done in the Google colab which is our python notebook.First, KNN algorithm is
trained with the training dataset and later it was tested with the remaining test data Second, the Naïve Bayes algorithm is trained with the training dataset and later it was tested with the remaining test data Third, the Logistic Regression algorithm is trained with the training dataset and later it was tested with the remaining test data.
From this three techniques, we got Naïve Bayes with more accuracy and this model is used in front end. The model is loaded into the pickle file and that file is opened in the frontend and compares the user input values with this corresponding model. Finally it results with a text message displaying that either the patient having Parkinson's disease or not.

VI. CONCLUSION

By this we can conclude that plant leaf disease detection is done successfully with the help of the CNN and the open CV through the raspberry pi. We can achieve with in in very less time. Work will be reduced when we use drone across the field to identify the leaf diseases. for any other handling we can use the server. The core goal of the given project is to detect the plant leaf diseases and display it on the device and accurately identify the diseases and yield more output and prevent plants from the diseases.

VII. FUTURE WORK

Future enhancements in research papers related to automatic plant leaf disease detection can explore several areas to improve the accuracy, efficiency, and practicality of the detection systems. Here are some potential avenues for future research

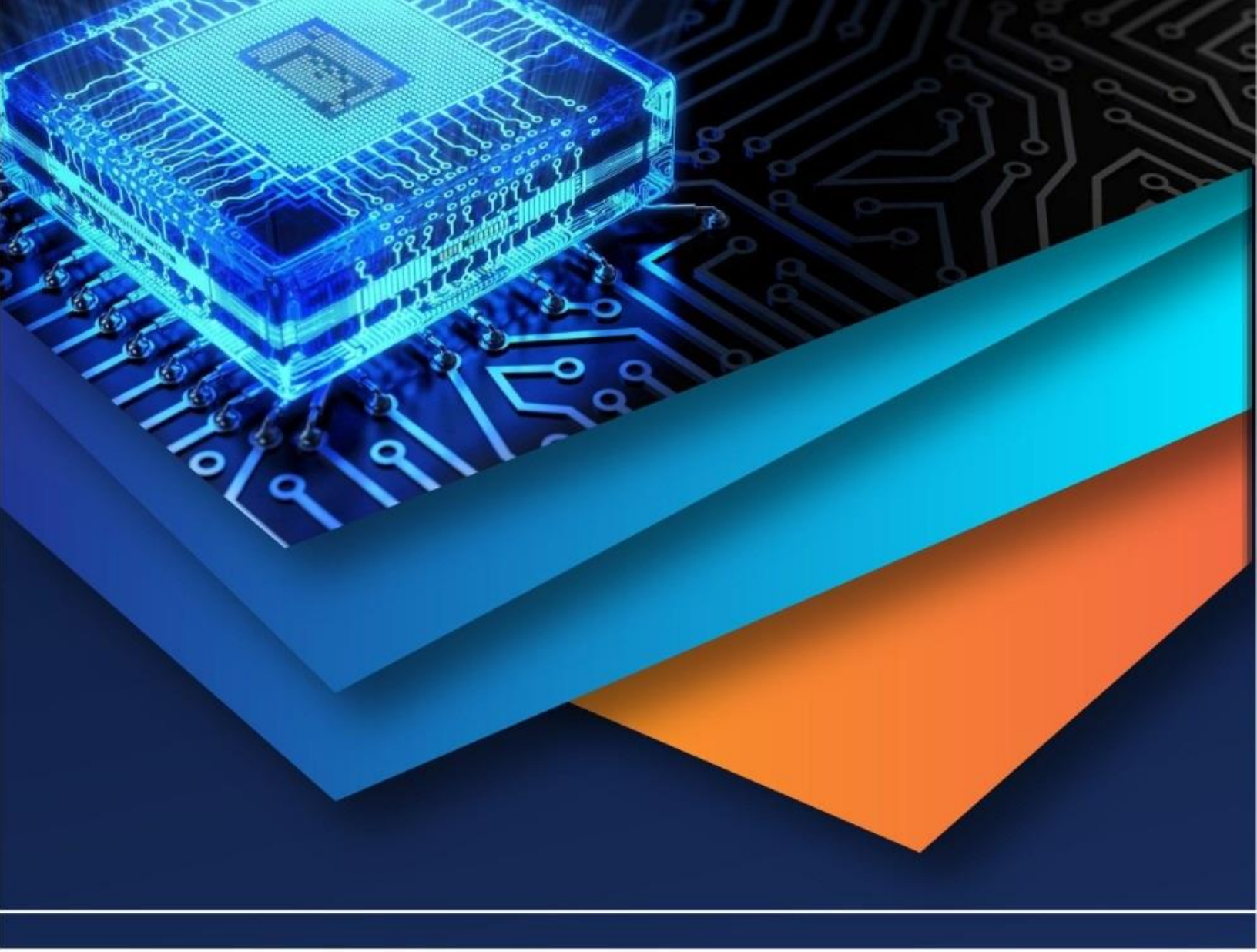
Incorporation of Multimodal Data

1) Transfer Learning and Domain Adaptation
2) Real-Time and Edge Computing
3) Dataset Expansion and Annotation
4) Explainability and Interpretability
5) Incorporation of Advanced AI Techniques

These future research directions can contribute to the advancement of automatic plant leaf disease detection systems and enable more accurate, efficient, and scalable solutions for disease management in agriculture.

REFERENCES

[1] Dr.S Murugan, et al. “Social Ramification of Fire on Forest using IoT”, CSI communication, vol. 42, issue.3, pp.3234,2018.
[2] G rama mohan reddy “Internet of Things: Power controlling through in smart mobiles” international journal of pure and applied mathematics (IJPAM)) vol 118 issue 17 pg.No 781-800,2017
[3] Smith J S., Camargo A., "An algorithm based on image processing to automatically identify visual symptoms of plant disease."



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)