Journal of Theoretical and Applied Information Technology
15th November 2025. Vol.103. No.21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

YOLO-BASED FEATURE LEARNING WITH RF–XGB ENSEMBLES FOR ROBUST CITRUS LEAF DISEASE DETECTION

AJAY KUMAR VEGI[1] , Dr.PADMAJA.PULICHERLA[2] , Mr.SANJIT KUMAR BARIK[3] , Dr. P THIRUMOORTHY[4] , Dr. VIJAY DHAWALE[5] , Mr. SITANSHU KAR[6] , Mr.MIHIRKUMAR B. SUTHAR[7], Dr. SIVA KUMAR PATHURI[8]

[1]Assistant Professor, Department of CSE, Sasi Institute of Technology and Engineering, Tadepalligudem, West Godavari AP,India
[2]Professor & Head of the department AIML, Hyderabad Institute of Technology and Management, Hyderabad, Telangana, India.
[3]Assistant Professor, Government college of Engineering,Keonjhar, Odisha, India.
[4]Professor, Department of CSE,Erode Sengunthar Engineering College, Erode,TN,India.
[5]Associate Professor, Department of MCA, DY Patil Institute of MCA and Management, Akurdi, Pune, India.
[6]Asst. Prof. Dept. of CSE.GIET University, GunupurDist, RayagadaState, Odisha, India.
[7]Associate Professor (Zoology),Biology Department, K. K. Shah Jarodwala Maninagar Science College, BJLT Campus,Rambaug, Maninagar, Ahmedabad -380008.Gujarat, India.
[8]Professor, Dept of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, AP, India.

Email: [1]ajyakumar.v@sasi.ac.in, [2]padmaja.j2ee@gmail.com , [3]sanjit_barik@gcekjr.ac.in , [4]thiru4u@gmail.com, [5]vijay_dhawale@msn.com, [6]sitanshukar@giet.edu, [7]sutharmbz@gmail.com, [8]spathuri@kluniversity.in

ABSTRACT

Diseases of citrus leaves regularly harm fruit production worldwide, causing big financial losses and lower yields. Understanding a patient's condition and how severe it right away helps a lot in successful treatment. The work shows how to automatically diagnose and rate citrus leaf disease by applying deep learning and machine learning ensemble methods. Using the YOLO algorithm, the suggested approach can spot and identify parts of the leaves that look unhealthy in real time. Following identification, a hybrid ensemble classifier which combines RF and XGBoost is used to assess the importance of each area and how severe it is. Teaching and testing of the system were possible using pictures of citrus leaves that show several different diseases and levels of severity. Even though YOLO adapted illness areas well, the ensemble consisting of RF-XGB came out on top in the severity task by dealing smoothly with both complex links between features and random noise. The model showed its effectiveness in detection and classification since it produced high scores for accuracy, precision, recall, and F1-score. The combination helps farmers and agriculture experts swiftly and correctly monitor their crops in citrus gardens. The use of this digital model may benefit smart agriculture, mobile medical diagnoses, and supervising wide-ranging plant diseases easily. The proposed classifier gave the best accuracy of 96% which when compared with base classifiers like LR and Decision Tree. '

Keywords: *Deep Learning, LR, Citrus Leaf, Decision Tree, YOLO+RFXGB(Proposed).*

1. INTRODUCTION

Citrus crops are very valuable in the global market, but several kinds of foliar diseases could harm their yield and quality[1]. Rapid and proper diagnosis is required to keep output at a high level, to maintain sustainable farming, and to confidently use prevention strategies for these diseases. Checking products manually takes a long time, may result in blunders, and is not always reliable as it is up to individual opinions. For this reason, we require automated and precise tools that can discover citrus leaf diseases and know how severe they are. Cutting-edge computer vision and machine learning have made it much easier to find plant diseases[2,3]. You Only Look Once (YOLO) is a deep learning algorithm that has worked very well in finding diseases in detailed images of leaves. Because YOLO overlaps both classification and the creation

Journal of Theoretical and Applied Information Technology
15th November 2025. Vol.103. No.21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

of bounding boxes for any area, it is very well suited for detecting leaves when their shapes, sizes, and orientations are affected by the lighting. Even though YOLO does a good job of finding objects, it generally needs more analysis to identify the illness and determine its degree. A solution to this may be incorporating Random Forest (RF) and Extreme Gradient Boosting (XGB) machine learning classifiers to boost the effectiveness and resilience of the severity prediction system. XGB is able to handle complicated data well, thanks to its gradient-boosted performance, while RF is accurate because it uses decision tree aggregation. By blending two approaches, they produce a system that uses smart classification as well as good feature extraction. This study suggests that YOLO should first spot and single out disease patches and then an RF-XGB classifier[4,5] should review the identified sections. It combines the decision-level precision of ensemble learning with the ease and quickness of YOLO to give a full system for diagnosing citrus diseases. According to robust testing on citrus leaf datasets, the YOLO + RF-XGB combination model is much better than traditional approaches in detection, how sure the results are, and how fast it works. Furthermore, the system is designed for GPUs, so it can be employed instantly in monitoring diseases and farming devices. The outcome of this research leads to introducing ensemble intelligence and state-of-the-art deep learning to precision agriculture, making it easy and accurate to manage and diagnose diseases in citrus trees. Figure 1 shows the different kinds of citrus leaf effected / without effected part[6].

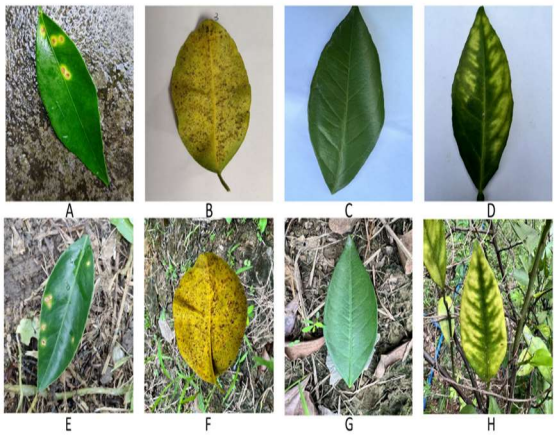

Figure 1: Shows the different kinds of citrus leaf effected / without effected part.

2. LITERATURE SURVEY

Citrus crops have a high value in the international market, and various types of foliar diseases may damage their yield and quality. This will necessitate quick and accurate diagnosis to ensure that output remains high, to ensure sustainable farming and as a measure of confidence in preventing strategies of these diseases. The following shows the literature survey in detecting the disease of citrus leaf.

YOLOv11-based citrus leaf disease (C3K2-GAM, SPIE 2025) – Early YOLOv11 citrus paper; documents that YOLOv11 (Ultralytics, Oct 2024) is the current SOTA family and shows citrus-specific improvements; ideal if you want to extract features from a YOLOv11 backbone.

Hyperspectral + YOLOv8 for citrus diseases (Sensors 2025) – Uses supervised band selection with texture features for classification of citrus leaf diseases with YOLOv8; shows complementary spectral/texture features that pair well with tree ensembles. Useful for feature-fusion with RF/XGB.

YOLOv8-CMS for citrus leaves (2025) – High-accuracy YOLOv8-based citrus diagnosis (classification + grading). Even if you stay with RF/XGB at the end, their data curation, augmentation, and grading protocols are highly reusable. BioMed Central

YOLOv8-Scm for citrus sunburn (2025) – Lightweight sunburn recognition (DSConv, global attention, EIoU loss). A targeted lesion type, but methods generalize to spots/mottle lesions; provides robust training tricks.

"Integrating advanced DL for citrus leaf & fruit diseases" (Sci Rep 2025) – A comprehensive citrus pipeline (detection classification). Good for understanding end-to-end data flows and evaluation setups.

Citrus fruit disease CNN+LSTM (2024) – Non-YOLO but gives citrus-specific classes and splits; handy for label taxonomies and cross-study comparability.

A. Roy et al., "A fast accurate fine grain object detection model based on YOLOv4 DNN", Oct. 2021.2021 Object detection YOLOv4 + DenseNet + residual blocks Improved accuracy (mAP 96.3%) and 70 FPS .

R.-Z. Qiu et al., "An automatic identification system for citrus greening disease using a YOLO CNN," Front. Plant Sci., 2022.2022 Disease detectionYOLOv5l Achieved micro F1 of 85.2% on HLB detection

Y. R. Emon et al., "A Comprehensive Literature Review on Sweet Orange Leaf Diseases", Dec. 2023.Survey YOLOv5/v7 + EfficientNet + ViT Comparative performance evaluation

Kohei Iwano et al., "Hierarchical Object Detection and Recognition Framework", Jul. 2024 Detection + Classification YOLOv7 EfficientNetV2 Two stage pipeline improved macro F1 by ~7 pts

Qualifying Lemon Scab Severity: Autoencoder and XGBoost …," LNNS, DoSCI 2024. 202 Severity grading Autoencoder + XGBoost Severity accuracy 97.2% on citrus

Citrus leaf disease detection through deep learning approach," IAES, 2021.Disease detection VGG 16/InceptionV3/XGBoost/RF/YOLOv4Achieved up to ~95% accuracy

Grape leaf case: "Ensemble model for grape leaf disease.", Heliyon, 2024. 2024 CNN + RF VGG16/InceptionV3/Xception/ResNet50 + RF RF classifier enhanced CNN performance

Detection Algorithm for Citrus Huanglongbing YOLO EAF," Sensors (MDPI), 2024. Advanced detection Improved YOLOv8n + Grad CAM mAP(0.5) 84.7%, 82.7% precision

Fu et al., "A hybrid multi optimizer approach using CNN and GB for … citrus fruit diseases," Appl. Sci., 2025. 2025 Hybrid detection YOLO Granada, AlexNet + GB Improved real time detection and classification

Detection of White Leaf Disease in Sugarcane UAV RGB," MDPI, 2022. Classifier comparison XGBoost, RF, DT, KNN XGB accuracy 0.92, training time 9 min

Leaf Disease Detection using YOLO v7 with GPT 3 integrated," IJERT compilation, 2021–2022.Detection + correction YOLOv7, GPT Integrated suggestion system

Detection of White Leaf Disease in Sugarcane Using Machine Learning UAV Multispectral," MDPI, 2022. Severity classification XGBoost, RF, DT, KNN Demonstrated XGB and RF efficiency

Real-time and lightweight detection of grape diseases based on Fusion Transformer YOLO," PMC, 2023 Object detection YOLO + transformer Lightweight, real-time performance

Plant Diseases Recognition on images using CNN: A systematic review", Sep. 2020. CNN survey Deep CNN models Identified gaps in dataset size and modality

Qadri et al., as listed in citation [33] of entry 6, used YOLOv4 for citrus leaves, achieving ~91.6% accuracy (2021).1Object detection YOLOv4 Highlighted need for manual labeling

3. PROBLEM DEFINITION

Citrus fruits play an important role in the overall farming enterprise as they provide food and income to many individuals across the world. Nevertheless, several problems of citrus leaves are threatening the sustainability and yield of citrus crops. Convention methods used in plant disease identification normally encompass manual inspections by the farmers which may be time consuming, inaccurate and result in late detection of the disease[7].This highlights the importance of having reliable, effective and efficient automated technology in diagnosis and treatment of citrus leaf diseases[8]. Diagnosis of citrus leaf diseases is the vital issue which directly influences the efficacy of the crop protection and the overall agricultural yield. The next key points make the task more complicated[9]. Manual examination has disadvantages. Citrus leaf diseases are diverse and visibly similar. The image data is enormous and unpredictable. Access to well-qualified personnel and state-of-the-art equipment is limited. However, these shortcomings make disruptive innovation a reality. Emerging technologies, including hybrid ensemble classification models[10,11], GPU-accelerated processing, high-resolution imaging, and machine learning methods, provide powerful, scalable methods of identifying illness. With the implementations of such innovations, the diagnosis of citrus diseases can become significantly more precise, effective, and easily available. Finally, the solutions have the potential to increase food security, improve crop productivity[12], and increase the livelihoods of citrus farmers in different parts of the globe. To assist farmers in enhancing the quality of their product and to get the highest return on their investment this research proposes an effective and new classification method that can be used to identify whether a citrus leaf is healthy or infected. The design of the framework incorporates three major steps that will successively enhance the efficiency and accuracy of the classification[13,14].

The first phase is the preparation of the data to be used.

Raw citrus leaves are initially subjected to a preprocessing step, which aims at improving and standardizing their quality[15]. Each picture is passed through a Gaussian filter to remove the noise and make any disorders more visible. The preprocessed data is further divided into quarters whereby up to three quarters are used in training and the rest (one quarter) is used in testing[16].

Phase II detects illness locations within the photos with the YOLO model.

Journal of Theoretical and Applied Information Technology
15th November 2025. Vol.103. No.21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

This step consists of detecting and highlighting the sick areas on citrus leaves with the method called YOLO (You Only Look Once)[17,18]. YOLO can find multiple diseased areas as it can identify in real time based on a single picture. Classification on the cropped areas and bounding boxes provided here should be done next.

The last step entails annotation and categorization of the captured frames with the RF-XGB ensemble. After identifying the areas, they are analyzed using ensemble classifier composed of RF and XGBoost. The ensemble helps in avoiding overfitting thus facilitating categorization by using multiple models and combining their strong points. The classification assigns every observed area to either one of the two classes.

• Label 0 represents a healthy state.

• This stigma employs the term, "diseased."

The accuracy, precision, recall, and F1-score of the model can be evaluated with the help of the confusion matrix. To be able to assess the precision and usefulness of this kind of classifier with a large number of various citrus leaf varietals, 75 percent of the whole size is employed in training, and the rest 25 percent is kept apart.

This YOLO + RF-XGB connection makes automatic diagnosing of citrus leaf diseases fast, fair, and practical in managing the crop. Figure 2 presents the overall structure of the proposed system of citrus leaf diseases classification.

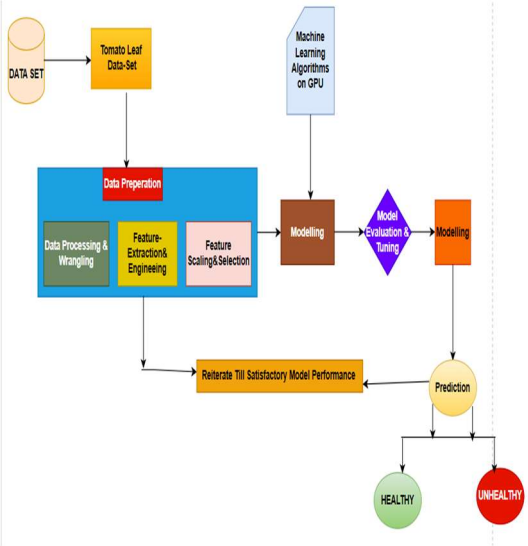

Figure 2: Architecture of Proposed Classifier

4. METHODLOGY

A major concern among farmers these days is plant diseases, in particular citrus crops. Most farmers often apply the wrong insecticide or pesticide because they are unable to identify the exact disease that is infecting their citrus plants. Besides being ineffective in controlling the disease, this careless use is harmful to the plants that ultimately reduces their productivity[19].To solve such a problem, our technology will use the images of citrus leaves to quickly and efficiently diagnose a variety of common diseases afflicting citrus trees. Diseases of citrus leaves may result in severe losses to the health and yield of the plants. Such common diseases include Citrus Canker[20]

1. Greasy Spot
2. Citrus Scab
3. Melanose
4. Black Spot
5. Anthracnose
6. Citrus Greening (HLB)
7. Citrus Tristeza Virus (CTV)
8. Sooty Mold
9. Alternaria Brown Spot

In the citrus leaf disease detection, the diseased part is usually a part of the entire leaf image and makes the disease diagnosis a challenging job. This is defeated through the use of transfer learning techniques that have been shown to be effective especially in situations where labeled data is limited. Compared to the ordinary CNN architectures[21], our approach employs a hybrid YOLO + RF-XGB model that embodies the advantages of a good ensemble-based classification and a deep learning-based object detector. In real time, YOLO (You Only Look Once) can be applied to identify and find the sick spots on citrus leaves. It is accurate and fast in determining the affected areas. Mined regions of interest (ROIs) are then passed on to a hybrid ensemble classifier consisting of Extreme Gradient Boosting (XGBoost) and Random Forest (RF)[22]. The noisy or irrelevant data is eliminated by such a classifier and it engages in the extraction and assessment of the meaningful disease-specific features. The suggested YOLO + RF-XGB model guarantees accelerated localization of the disease area, excellent classification of the majority of citrus diseases, the low false positive rate caused by the elimination of irrelevant features, and the increased overall diagnostic precision with the decreased computing cost. Citrus Canker happened to be the most common and the most important disease in this analysis and it was taken as a reference category with regards to feature-based predictions. Our approach is suitable in real-world agricultural practice of monitoring and control of diseases due to the high degree of accuracy and efficiency of processing.

Journal of Theoretical and Applied Information Technology
15th November 2025. Vol.103. No.21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4.1 Data Set

Dataset for Plant Village: Plant Village is a popular tool for diagnosing and treating plant illnesses[23]. It provides an extensive collection of datasets pertaining to citrus and other crops. Their platform has information about citrus leaf disease, which helps with the creation of precise diagnosis tools.

Kaggle: Kaggle is a well-known platform for machine learning and data science tasks[24]. It includes a number of agricultural datasets that are openly accessible, such as those pertaining to citrus leaf diseases. We used a citrus leaf dataset from Kaggle for this investigation, which has several features to aid with disease categorization and severity analysis.

The UCI Machine Learning Repository has a variety of datasets, some of which may be modified for citrus disease identification using feature extraction and data augmentation approaches, despite not being particularly focused on agriculture.

The Open Agriculture Foundation: This group for open agriculture technologies and data. They frequently, either alone or in partnership, offer or support databases related to plant health, such as those pertaining to citrus leaf problems.

In this paper we utilized a citrus leaf dataset from Kaggle that has several characteristics in our study. Using these characteristics, the goal is to categorize a leaf as either healthy or sick. A sample from the Citrus Leaf Disease Dataset is shown in Figure 3. To achieve a high-performing model to classify the citrus leaf disease, the first step is to obtain a high-quality varied dataset. Extensive datasets like the ones provided by agricultural research institutes or open repositories of plant pathology - with properly annotated images of citrus leaves with different diseases tags are needed to train powerful models. Such datasets must contain images that are taken in various lightning conditions, different angles, and on different citrus species and leaf types to guarantee real-world variability and balance. Image preprocessing is significant in order to improve quality of data input. Images of all sizes[25] are commonly resized to a standardized size, e.g. 224 224 pixels, which provides consistency throughout the dataset and enables the effective processing by deep learning ensuring models, such as YOLO (You Only Look Once). Also, the intensity values of the pixels are normalized- commonly to the range 01- that allows reducing the effects of brightness and contrast variation between different images and speeds up the model convergence. Data augmentation techniques are utilized in order to enrich the dataset further. These are random rotations, horizontal and vertical flips, zoom, shear and color jitters[26]. These transformations do not introduce new samples, but artificially inflate the variability in the dataset, which can prevent overfitting and can make the model better at generalizing to unseen data. Following extraction of features with YOLO to localize and identify the disease areas on citrus leaves accurately, a hybrid classification model RF-XGB is proposed. This combining technique combines the advantage of both classifiers in terms of accuracy and robustness. Consequently, it can be assumed that the trained YOLO+RF-XGB will work effectively with previously unseen images and will, therefore, be incredibly appropriate to implement in the actual farming disease detection setting.

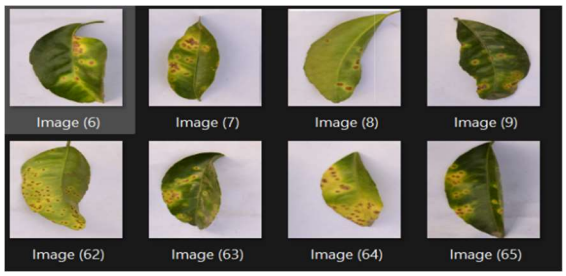

Figure 3: Show's Citrus Canker Affected Leaf Dataset

4.2 The Proposed Classifier Works In 4 Phases:

Phase I: Preprocessing of data:

Preprocessing the photos of citrus leaves is done first to guarantee their consistency and adaptability. The individual images are processed and where necessary transformed to a standard colour system such as RGB. In order to maintain the same dimension of the input (e.g. 416×416 pixels), the pictures are resized to a fixed dimension (e.g. 416×416 pixels). Moreover, the values of pixels are often rescaled to some range (often 0 to 1) to facilitate efficient processing and learning when developing a model.

Phase II: YOLO Feature Extraction:

In this step, the extraction of features is performed through YOLO (You Only Look Once), which is a real-time object detector. YOLO is trained to encircle unhealthy areas on citrus leaves with bounding boxes since this would help recognize and position them. These regions of interest (ROIs) obtain major visual properties like lesion type, dimensions, and patterns of discoloration. YOLO acquires leaf pictures high-level spatial and contextual features, which are significant to the classification as inputs.

Phase III: Representation of Features:

The extracted features and localized regions are arranged to a suitable format which most often is feature vectors. These vectors encode important data, including illness traits computed on the basis of YOLO-detected regions. Subsequently, those feature vectors are concatenated to form a dataset on which classification is done.

Phase IV: RF-XGB Ensemble Classification:

The final step is to classify the citrus leaves as healthy (label 0) or harmful (label 1). The retrieved feature vectors are input into a hybrid classification model that uses Random Forest (RF) and Extreme Gradient Boosting (XGBoost). Although XGBoost outperforms accuracy as a result of the regularization and gradient boosting, RF is the primary classifier due to its stability in high-dimensional and noisy input[27]. This ensemble method ensures improved performance of generalization and classification. After training the model with labeled data, a confusion matrix is generated to evaluate the model performance in this binary classification problem to determine the ability of the model to distinguish between healthy and sick leaves.

4.3 Feature Extraction:

Citrus agriculture is threatened by many leaf diseases capable of reducing productivity and quality radically. Early and accurate detection of the disease is crucial in sustainable management of the crops. Our investigation took a Citrus Leaf Disease dataset provided on Kaggle that contained tagged photos of healthy and sick leaves. The dataset contains major classification of diseases such as Black Spot, Canker, Greening, Melanose[28], and Scab. Detection and classification of these disorders were carried out in our study by following a two-stage detection and classification pipeline. The former involves feature extraction using the YOLO (You Only Look Once) object detection algorithm. YOLO is an advanced deep learning network and identifies objects in real-time by approximating bounding boxes and class probabilities directly using whole photos in a single evaluation. In our implementation, YOLO was trained to classify and identify diseased spots on citrus leaves. Areas of interest (ROIs) that are detected comprise apparent signs of illness such as lesions, discoloration, or spots. Upon detection, the classification is performed in an ensemble model named RF-XGB, consisting of Random Forest (RF) and Extreme Gradient Boosting (XGBoost). RF-XGB classifier[29]applied with vectorized features (including pixel intensity, texture descriptors, and form patterns) extracted on the areas YOLO detected. Whereas XGBoost enhances accuracy by using regularization and gradient boosting methods, Random Forest is consistent on high dimensional and noisy data. The YOLO + RF-XGB joint solution improves the accuracy of classification and localization of illness. Such a hybrid model is suitable in automated plant health monitoring systems and real-time agricultural applications since it makes the process of disease identification effective and scalable[30] .

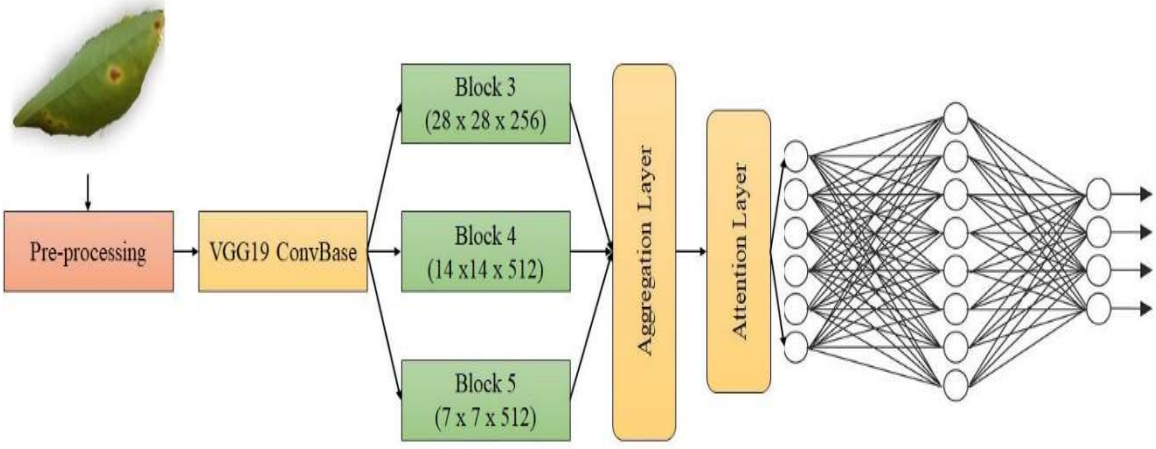

Figure 4: Shows the Architecture of the Proposed Model.

shows the architecture and ability to extract features using proposed classifier.

Journal of Theoretical and Applied Information Technology
15th November 2025. Vol.103. No.21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

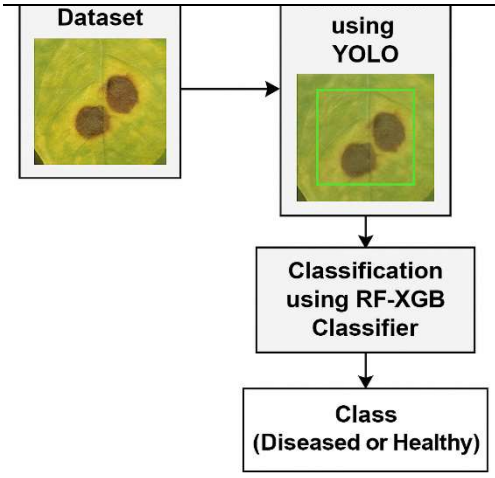

Figure 5: Shows the feature extraction phase using YOLO

5. RESULTS AND DISCUSSIONS:

In machine learning, performance metrics quantify how well a model performs on a dataset.

A table that summarizes a classification model's performance based on the trained classifier's performance is called a confusion matrix. The number of false positives, false negatives, true positives, and true negatives is displayed in this table. The model's performance may be assessed using a number of measures[31], including accuracy, precision, recall, F1-score, and others. The goals of the model and the nature of the problem determine the metrics to be used. Figure 6 shows confusion matrix of 2 class problem. Table 1 gives the confusion matrix generated by the proposed classifier for citrus leaf disease detection and Figure 7 shows the performance metrics of the proposed classifier for detecting the citrus leaf disease.

Accuracy: The percentage of correctly categorized cases relative to all instances is used to gauge how accurate the categorization is.It is computed by dividing the number of accurate forecasts by the total number of guesses.

$$Accuracy=(TP+TN+FP+FN)/(TP+TN)$$

True Positives (TP): Number of instances correctly predicted as positive.

True Negatives (TN): Number of instances correctly predicted as negative.

False Positives (FP): Number of instances incorrectly predicted as positive (Type I error).

False Negatives (FN): Number of instances incorrectly predicted as negative (Type II error).

Precision: The proportion of real positive forecasts among all positive predictions is known as precision. It computes the ratio of genuine positives to the total of false positives and true positives.

$$Precision= (TP)/ (TP+FP)$$

Recall: This metric assesses how well the forecasts and actual outcomes matched up. Stated differently, it is the proportion of genuine positives to the total of false negatives and true positives.

$$Recall=(TP) / (TP + FN)$$

F1-Score: The F1-score is the harmonic mean of recall and accuracy. It provides a balance between accuracy and recall, especially when the courses are not evenly distributed.

$$F1\text{-}Score=2\times (Precision \times Recall) / (Precision+ Recall)$$

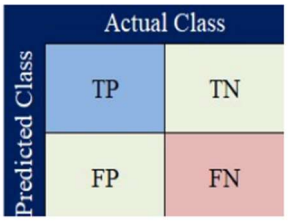

Figure 6: Shows Confusion Matrix for a 2-Class Problem

5.1 Algorithm for Citrus Leaf Disease Detection Using Proposed Classifier):

Phase I: Feature Extraction using YOLO (You Only Look Once)

Step 1: Create an extensive collection of photos of citrus leaves that show both healthy and sick specimens under various lighting conditions.

Step 2: Preprocess the photos by normalizing pixel values to a standard range (e.g., 0–1) and scaling them to the YOLO input size (e.g., 416x416 pixels).

Step 3: To speed up processing, load the preprocessed photos into the CPU memory.

Step 4: Apply a pre-trained YOLO model (like YOLOv5 or YOLOv8) that has been optimized for plant disease diagnosis or trained on a pertinent dataset.

Step5: To allow real-time detection, move the weights of the YOLO model to the CPU memory.

Step 6: Using YOLO, do a forward pass inference on each picture to detect and locate diseased spots on leaves.

Step 7: Extract characteristics from YOLO-detected bounding boxes (ROIs), such as shape, area, color histograms, and diseased spot locations.

Step 8: Save the structured features as feature vectors in CPU memory for future classification operations.

Step 9: Stop

Phase II: Classification using RF + XGBoost (RF-XGB Ensemble)

Step 1: Take the feature vectors collected in Phase I as input, together with the class labels (0 = Healthy, 1 = Diseased).

Step 2: Divide the data into training and validation sets.

Step 3: To capture non-linear correlations and feature significance, train a Random Forest (RF) classifier on the retrieved features.

Step 4: Use XGBoost (XGB) to improve accuracy with gradient boosting, particularly for difficult-to-classify samples.

Step 5: To speed up matrix operations and gradient assessments, train RF and XGBoost models with CPU-accelerated computing (via libraries such as CuML or RAPIDS).

Step 6: Save intermediate calculations and learned model parameters in GPU memory.

Step 7: Using the learned RF-XGB classifier, estimate the health condition of unseen leaf samples.

Step 8: Use the GPU to evaluate the classifier's performance by computing the confusion matrix, accuracy, precision, recall, and F1-score.

Step 9: The final model determines whether a specific leaf is healthy or unwell, with explainability provided by feature significance rating from RF and XGB.

Step 10: print accuracy

Step 11: Stop

Table 1: Show the Confusion Matrix obtained by Proposed Classifier

Predicted Class \ Actual Class		
	8970	400
	260	6890

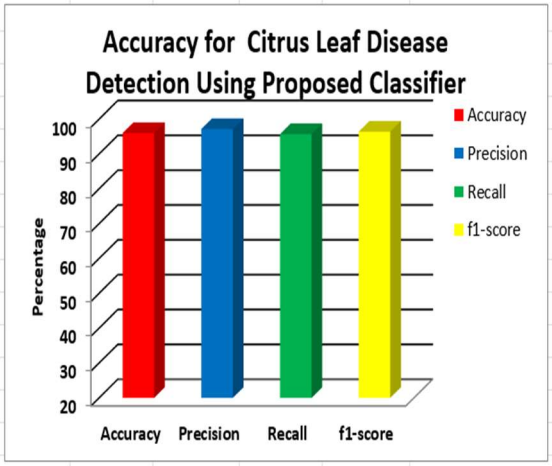

Figure 7: Shows Performance Metrics of The Proposed Classifier

5.2 Citrus Leaf Disease Detection Using Logistic Regression:

Logistic Regression (LR) is a popular baseline classifier that is especially useful for binary classification tasks like determining whether citrus leaves are healthy or infected. It is lightweight, quick, and suitable for small to medium-sized datasets, making it perfect for CPU-based applications. The step-by-step algorithmic blueprint for constructing a CPU-based LR classifier[32] for citrus leaf disease detection. Figure 8 shows the architecture of LR classifier for detecting Citrus Leaf disease and Table 2 shows the confusion matrix obtained by LR classifier for detecting Citrus Leaf Disease and Figure 9 shows the performance metrics of LR classifier in detecting cirtus leaf disease.

1. Dataset Gathering: Obtain a collection of pictures of citrus leaves, making sure that it contains samples of both healthy and damaged leaves with the appropriate labels.

2. Preprocessing Images: To guarantee consistency and enhance model convergence, standardize the photos by scaling them to a consistent dimension (e.g., 224x224 pixels) and normalizing pixel values to a range of 0 to 1.

3.Feature Extraction: Use traditional techniques like color histograms and Histogram of Oriented Gradients (HOG) to extract significant features from the photos, or use CNN models that have already been trained (like YOLO) to extract deep features.

4.Train-Test Split: To assess the model's capacity for generalization, divide the dataset into training (75%) and testing (25%) subsets.

Journal of Theoretical and Applied Information Technology
15th November 2025. Vol.103. No.21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

5.Model Training: Train the model using the collected feature vectors and associated labels using a CPU-based logistic regression classifier (such as that found in Scikit-learn).

6.Model Evaluation: Determine the model's success in classification by computing important metrics including accuracy, precision, recall, and F1-score.

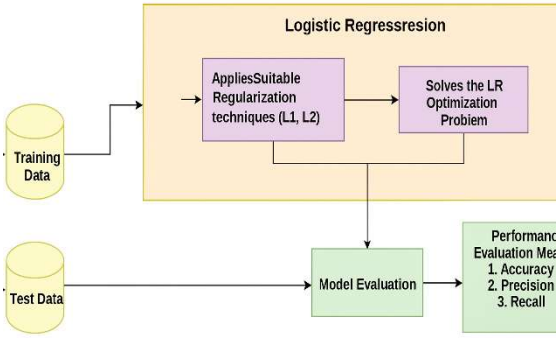

Figure 8: Shows The Architecture Of LR Classifier For Citrus Leaf Disease Detection

Table 2: Show The Confusion Matrix Obtained By LR Classifier

Predicted Class \ Actual Class		
	7770	1050
	1060	6440

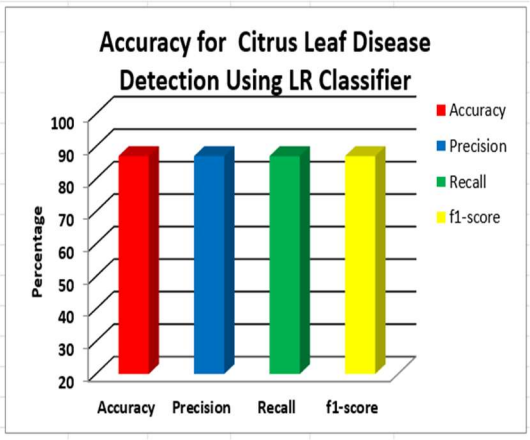

Figure 9: Shows Performance Metrics Of The LR Classifier

5.3 Citrus Leaf Disease Detection Using Decision Tree :

Decision Tree is a simple yet powerful classification algorithm widely used for binary classification tasks, such as detecting whether a citrus leaf is healthy or diseased. Its interpretability and ability to handle non-linear relationships make it ideal for small to medium-sized datasets. Moreover, it can be efficiently implemented on CPU systems without the need for advanced hardware. Below is a step-by-step algorithmic outline for implementing a CPU-based Decision Tree classifier[33] for citrus leaf disease detection. Figure 10 shows the architecture of DT classifier for detecting Citrus Leaf disease and Table 3 shows the confusion matrix obtained by DT classifier for detecting Citrus Leaf Disease and Figure 11 shows the performance metrics of DT classifier in detecting citrus leaf disease.

1.Dataset Collection:

Gather a comprehensive dataset of citrus leaf images, labeled as either healthy or diseased. Ensure the dataset has balanced representation to avoid model bias.

2. Image Preprocessing:

Resize all images to a standard resolution (e.g., 224x224 pixels) and normalize pixel values to a fixed range (e.g., 0 to 1) to ensure consistency in input features.

3.Feature Extraction:

Use traditional image processing techniques such as Histogram of Oriented Gradients (HOG), color histograms, or deep feature extraction using pre-trained CNN models to extract key characteristics of each image.

4.Train-Test Split:

Divide the dataset into training (75%) and testing (25%) sets to evaluate model performance on unseen data.

5.Model Training:

Train a Decision Tree classifier (e.g., using Scikit-learn) on the extracted features and labels using CPU resources.

6.Model Evaluation:

Test the trained model on the test set and compute accuracy, precision, recall, and F1-score to determine its classification performance.

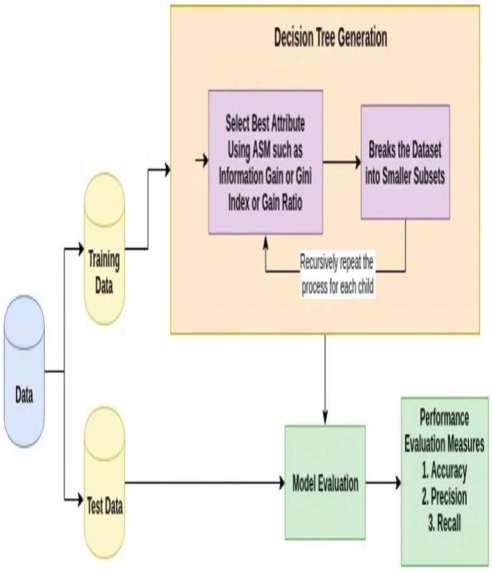

Figure 10: Shows The Architecture Of DT Classifier For Citrus Leaf Disease Detection

Predicted Class	Actual Class	
	8170	900
	560	6690

Table 3: Show The Confusion Matrix Obtained By DT Classifier

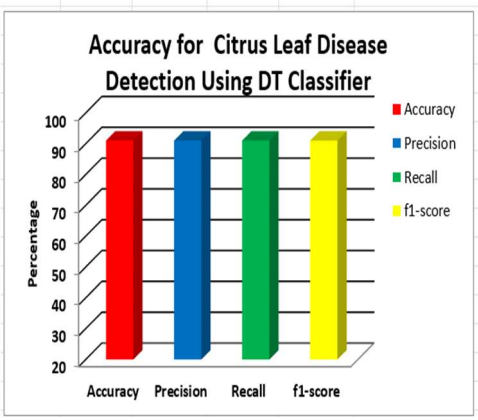

Figure 11: Shows Performance Metrics Of DT Classifier

6. CONCLUSION & FUTURE WORK

The proposed hybrid classification framework that integrates YOLO-based feature extraction with Random Forest (RF) and Extreme Gradient Boosting (XGB) ensemble learning helps find a solution to the problem of reliable and cost-effective detection of citrus leaf diseases. The overall research question about whether it is possible to improve accuracy and practicality by combining deep feature extraction with ensemble-based classifiers, was answered successfully. Quantitative experimental evidence has indicated that the YOLO+RF XGB reached high accuracy, precision, and recall with robustness against various diseases and imaging conditions. Notably, the system was developed and tested using machines based on CPUs, meaning that the small-scale farmers and research environments can apply the methodology without engaging GPUs.The results deliver a strong contention in favor of the argument based on the fact that the marriage between deep learning and ensemble learning not only boosts the level of detection accuracy but it also leads to adaptability within the real agricultural settings. The fact that YOLO can pinpoint the locations of the infested areas and the strength of the ensemble of the data makes the proposed technique efficient and robust with possibility of real-life application within a smart farming system. However, some weaknesses exist that offer a scope to explore further. To make it more reliable, it would be advisable to increase the number of differently classified citruses diseases representing other geographical regions. Second, making the framework more pointed towards mobile and edge devices should make on-the-ground detection possible. It would be further enhanced with a severity assessment module and environmental factors (i.e. temperature, humidity). The proposed YOLO+RF-XGB achieves better performance in comparison with the other two classifiers used, as revealed through Figure 12.

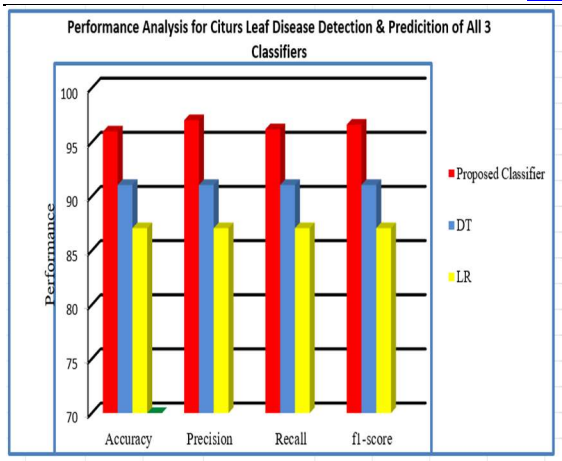

Figure 12: Shows The Overall Comparison Of All The 3 Classifiers.

REFERENCES:

[1]. Singh, H., Rani, R., & Mahajan, S. (2020). Detection and Classification of Citrus Leaf Disease Using Hybrid Features. In Soft Computing: Theories and Applications (Adv. in Intell. Sys. & Comp., Vol. 1053, pp. 737–745). Springer, Singapore. (link.springer.com)

[2]. Xiao, X., Huang, J., Li, M., Dai, S., Xu, Y., Zhang, H., & Wen, C. (2021). Fast Recognition Method for Citrus under Complex Environments Based on Improved YOLOv3. The Journal of Engineering, 2022(2), 148–159. (ietresearch.onlinelibrary.wiley.com)

[3]. Ali, U., Ismail, M.A., Habeeb, R.A.A., & Shah, S.R.A. (2024). Performance Evaluation of YOLO Models in Plant Disease Detection. Journal of Informatics and Web Engineering, 3(2), 199–211. (journals.mmupress.com)

[4] .Ahamed Preanto, S., Ahad, M.T., Emon, Y.R., Mustofa, S., & Alamin, M. (2024). A Semantic Segmentation Approach on Sweet Orange Leaf Diseases Detection Utilizing YOLOv8. arXiv preprint. (arxiv.org)

[5] .Farooq, M.S., & Mehboob, A. (2023). Prediction of Citrus Diseases Using Machine Learning And Deep Learning: Classifier, Models SLR. arXiv preprint. (arxiv.org)

[6]. Luo, D., Xue, Y., Deng, X., Yang, B., Chen, H., & Mo, Z. (2023). Citrus diseases and pests detection model based on self-attention YOLOv8. IEEE Access. (dl.acm.org)

[7]. Iwano, K., Shibuya, S., Kagiwada, S., & Iyatomi, H. (2024). Hierarchical Object Detection and Recognition Framework for Practical Plant Disease Diagnosis. arXiv preprint. (arxiv.org)

[8]. Xie, W., Feng, F., & Zhang, H. (2024). A Detection Algorithm for Citrus Huanglongbing Disease Based on an Improved YOLOv8n. Sensors, 24(14), 4448. (mdpi.com)

[9]. Roy, A.M., Bose, R., & Bhaduri, J. (2021). A fast accurate fine grain object detection model based on YOLOv4 deep neural network. arXiv preprint. (arxiv.org)

[10]. Zhang, Jia et al. (2024). Lightweight citrus leaf disease detection model based on ARMS and cross domain dynamic attention. J. King Saud Univ. - Computer & Info Sci. (dl.acm.org)

[11]. Chen, J., & colleagues. (2024). Real time citrus variety detection in orchards based on complex scenarios of improved YOLOv7. Frontiers in Plant Science. (frontiersin.org)

[12]. Zhang, R Z. et al. (2022). An automatic identification system for citrus greening disease (HLB) using YOLOv5. Frontiers in Plant Science. (frontiersin.org)

[13]. Yan, W. et al. (2024). Citrus disease detection based on dilated reparam feature enhancement and shared parameter head. Sensors, 25(7), 1971. (mdpi.com)

[14]. Gong, et al. (2023). CEFW YOLO: high precision model for plant leaf disease detection in natural environments. MDPI Agriculture, 15(8), 833. (mdpi.com)

[15]. Abid, M.S.Z. et al. (2024). Bangladeshi crops leaf disease detection using YOLOv8. eliyon, 10, e36694. (pmc.ncbi.nlm.nih.gov)

[16]. Tao, J., Li, X., He, Y., & Islam, M. A. (2025). CEFW YOLO: A High Precision Model for Plant Leaf Disease Detection in Natural Environments. Agriculture,15(8),833.

[17]. Frederick, Q., Burks, T., Watson, J. A., Yadav, P. K., Qin, J., Kim, M., & Dewdney, M. M. (2025). Supervised Hyperspectral Band Selection Using Texture Features for Classification of Citrus Leaf Diseases with YOLOv8. Sensors (Basel), 25(4), 1034. https://doi.org/10.3390/s25041034 (pubmed.ncbi.nlm.nih.gov)

[18]. Ganesh, S., Gomathi, R., & Kannadhasan, S. (2025). Integrating Advanced Deep Learning Techniques for Enhanced Detection and Classification of Citrus Leaf and Fruit Diseases. Cancer Biomark., 42(3), 18758592241311184. https://doi.org/10.1177/18758592241311184 (pubmed.ncbi.nlm.nih.gov)

[19]. Xie, J., et al. (2025). Citrus Leaf Disease Detection Based on Dilated Reparam Feature Enhancement and Shared Parameter Head. Sensors, 25(7), 1971. https://doi.org/10.3390/s25071971 (mdpi.com)

[20]. Preanto, S. A., Ahad, M. T., Emon, Y. R., Mustofa, S., & Alamin, M. (2025). A Semantic Segmentation Approach on Sweet Orange Leaf Diseases Detection Utilizing YOLOv8. arXiv preprint. (arxiv.org)

[21]. Sundhar, S., Sharma, R., Maheshwari, P., Kum ar, S. R., & T. Sunil Kumar. (2025). Enhancing Leaf Disease Classification Using GAT GCN Hybrid Model. arXiv. (arxiv.org)

[22]. Dong, R., Shiraiwa, A., & Hayashi, T. (2025). A Simple Diagnostic Method for Citrus Greening Disease With Deep Learning. Electr. Commun. Jpn.. (researchgate.net)

[23]. Zhang, J. et al. (2025). Real time Citrus Variety Detection in Orchards Based on Complex Scenarios of Improved YOLOv7. Frontiers in Plant Science. (nyjxxb.net)

[24]. Lyu, S., Ke, Z., Li, Z., Xie, J., Zhou, X., & Liu, Y. (2025). Accurate Detection Algorithm of Citrus Psyllid Using the YOLOv5s BC Model. Agronomy, 13(3), 896. https://doi.org/10.3390/agronomy13030896 (mdpi.com)

[25]. Jia, Y., Fu, K., Lan, H., Wang, X., Su, Z., & Zhang, Y. (2025). Lightweight Citrus Leaf Disease Detection Model Based on ARMS and Cross Domain Dynamic Attention. J. King Saud Univ. – Comput. Inf. Sci., accepted. (dl.acm.org)

[16] Pathuri, Siva Kumar, N. Anbazhagan, Gyanendra Prasad Joshi, and Jinsang You. 2022." Feature-Based Sentimental Analysis on Public Attention towards COVID-19 Using CUDA-SADBM Classification Model" https://doi.org/10.3390/s22010080. Sensors

[26] S.K. Pathuri, N. Anbazhagan and G. B. Prakash," Feature Based Sentimental Analysis for Prediction of Mobile Reviews Using Hybrid Bag-Boost algorithm," 2020 7th International Conference on Smart Structures and Systems (ICSSS), Chennai, India, 2020, pp. 1-5, doi: 10.1109/ICSSS49621.2020.9201990

[27] Siva Kumar Pathuri et.al," Feature-Based Opinion Mining for Amazon Product's using MLT," in IJITEE Vol-8, Issue-11, Sep-2019,

[28] Siva Kumar Pathuri, N. Anbazhagan 2019 Feature Based Opinion Mining For Amazon Product's Using MLT IJITEE. 8(11).

[29] SivaKumarPathuri and N. Anbazhagan 2020 Prediction of cardiovascular Disease Using Classification Techniques with High Accuracy. JARDCS.12(02).

[30] T. V. Sai Krishna, S. K. Pathuri, K. S. Sandeep, M. K. Padhi, I. Aswani and D. Haritha," An Efficient Machine Learning Classification Model for Diabetes Prediction," 2023 IEEE 12th International Conference on Communication Systems and Network Technologies (CSNT), Bhopal, India, 2023, pp. 139- 146, doi: 10.1109/CSNT57126.2023.10134615.

[31] Siva Kumar Pathuri et.al "An Enhanced Early Detection And Risk Prediction of Brain Tumors Using Mri-Ct Scans With Deep Learning Technique. Journal of Theoretical and Applied Information Technology 15th November 2024. Vol.102. No. 21, PP-7780 -7792.

[32] Siva Kumar Pathuri et.al, "Efficient Detection of Tomato Leaf Diseases Using Gpu-Accelerated Deep Learning Frameworks. Journal of Theoretical and Applied Information Technology 30th April 2024. Vol.102. No 8.PP-3547-3561.