SvedbergOpen
DISSEMINATION OF KNOWLEDGE
International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/
Research Paper Open Access

Automated Disease Detection And Classification Of Solanaceae Family Leaves: Focus On Tomato, Brinjal, And Pepper

A. Dhanalakshmi[1], Dr. D. Senthil Kumar[2], Dr.R.Mohan Kumar[3], Piskala Sathiyamurthy Kumaresh[4], Dr. R. Palani kumar[5], Elangovan Muniyandy[6]

[1]Assistant Professor, Postgraduate Department of Data Science, The American College. Madurai. Email id: Dhanuswathi020@gmail.com
[2]Assistant Professor, Department of Data Science, The American College Madurai. Email id : senthilkumar@americancollege.edu.in
[3]Professor, Department Bio medical Engineering, PSR Engineering College. Sivakasi, Sevalpatti, Appayanaickenpatti, Tamil Nadu 626140 email id : rmohankumar@psr.edu.in
[4]Department of Electronics and Communication Engineering, K.L.N. College of Engineering, Sivagangai, India, email: kumaresh.ps@klnce.edu
[5]Associate Professor,Department of Computer Science and Engineering ,P.S.R. ENGINEERING COLLEGE,,SIVAKASI-626 140., palanikumar@psr.edu.in
[6]Department of Biosciences, Saveetha School of Engineering. Saveetha Institute of Medical and Technical Sciences, Chennai - 602 105. , Chennai - 602 105.
Email id: muniyandy.e@gmail.com
Corresponding author mail id : Dhanuswathi020@gmail.com

Abstract

Plant diseases are a major concern for global agriculture, impacting a wide range of crops essential for food security. These diseases, which can be produced by bacteria, viruses, fungi, and environmental stress, lead to visible symptoms. Traditional disease detection methods lead to delayed intervention and crop damage. To address this challenge, there is a growing need for automated and efficient systems that can detect and classify plant diseases at an early stage. This research objectives to develop a Deep Learning (DL)-based finding and organization model for Solanaceae family leaves using a novel Stochastic Fractal Search-driven Intelligent Faster Region-based Convolutional Neural Network (SFS-InT-Faster R-CNN). The proposed model combines advanced techniques to optimize Faster R-CNN, ensuring improved accuracy with efficiency in detecting diseases. The dataset for this research includes high-resolution leaf images of tomato, brinjal, and pepper. The data pre-processing phase involves noise reduction using the Non-Local Means (NLM) filter and contrast enhancement via Adaptive Histogram Equalization (AHE), both crucial for improving image quality and highlighting subtle disease signs. Feature extraction is conducted through the Gray-Level Co-occurrence Matrix (GLCM), which captures critical texture and spatial features for accurate disease classification. The SFS-InT-Faster R-CNN model utilizes an SFS optimization algorithm to enhance the Faster R-CNN architecture's hyperparameters, significantly improving detection performance. The suggested model is implemented in Python software. The outcomes show that the model surpasses conventional techniques in precision of 95.2%, an accuracy of 98.5%, an F1-score of 98%, a recall of 98.36%, providing an efficient method to identify diseases in Solanaceae crops early. This research concludes that the proposed framework provides a reliable tool for plant disease monitoring, which can help optimize agricultural practices and mitigate crop losses.

Keywords: Disease Detection, Solanaceae Family, Leaves, Tomato, Brinjal, and Pepper.

1. Introduction

The Solanaceae family also known as the nightshade family, is notorious because many of its plants contain deadly alkaloids. Several species in this family thrive in shaded areas and bloom at night. There are around 102 genera and 2280 species in it (Matin et al. 2024). This family includes several valuable plants for food, medicine, crops, and ornamentation, including a variety of important crops, such as tomatoes, potatoes, eggplants, and peppers. The Solanaceae family comprises plants that are found all over the world (Houetohossou et al.2023). The majority of the plants are widespread and plentiful in both tropical and temperate climates. Mesophytes make up the majority of the plants. Some plants are cultivated for their culinary qualities and other purposes (Bütüner et al. 2024). This family's biggest genus, Solanum, has around 1700 species. Characteristics of Solanum xanthocarpum that are xerophytic and these crops are vital to global food security and agriculture (Tace et al.

2024). However, like all plants, members of the Solanaceae family are vulnerable to various kinds of diseases caused by fungi, viruses, bacteria, and stresses from the environment (Tammina et al. 2024). These diseases have a high potential for yield reduction not only in yields but also in the quality of yields. Such methods are visual assessment and selective sampling regularly miss early signals of diseases by the time diagnosis is made, the pathogens are well established (Polin et al. 2024). Conventional methods for early detection are viable but time-consuming and knowledge-intensive making them unsuitable for scaling up to cover large areas commonly found in agriculture (Martina et al. 2024). To address these challenges, there is a growing need for an automated and efficient system that can detect and classify diseases in plants (Kahya et al. 2024). Systems that accurately identify illnesses in Solanaceae plants are being developed using machine learning (Oh et al. 2024). These systems can be able to interpret figures of the diseases reflected on the leaves, categorize them, and produce feedback to farmers in almost real-time, hence most effective when treatment is promptly done (Thakare et al. 2024). The adoption of automated methods in disease detection for the Solanaceae family has benefits, such as quick results, scant need for expertise, and wide-area surveillance of fields (Lih et al. 2024). Such systems can improve decision-making, minimize the application of pesticides in the course of the growing season, and improve the general health of the crop. Crop diseases are becoming a major threat in the course of agriculture practices, hence the concern on improving the methods for disease detection and classification is an essential factor to the productivity of Solanaceae crops in the future (Putra et al. 2024). Optimized deep learning framework, combining SFS-InT-Faster R-CNN, for accurate early detection and classification of diseases in Solanaceae crops, enhancing agricultural productivity and mitigating crop losses effectively

Key contributions

NLM reduces image noise while retaining higher frequency information, which is beneficial, when classifying Solanaceae leaf images as the symptoms of diseases are often faint.

Local contrast enhancement through AHE enhances contrast of the specific regions where disease manifestations are subtle and helps visualize them significantly enhancing the capability of extraction of unique and discriminative features for exact disease identification.

GLCM extracts critical texture and spatial features from leaf images, enabling the identification of disease-specific patterns for more accurate classification of Solanaceae family leaf diseases.

SFS-InT-Faster R-CNN Optimizes hyperparameters of the Faster R-CNN model using the Stochastic Fractal Search algorithm, improving detection accuracy, efficiency, and performance for early disease detection in tomato, brinjal, and pepper crops.

The structure of the research is followed: Part 2 Literature Review, the methodology is established in Part 3, the performance evaluation is displayed in Part 4, the experimental setup illustrates in part 5 and the performance evaluation presented in Part 6 and Part 7 demonstrates the discussion and conclusion.

2. Literature Review

Gurusamy et al. (2024) suggested that using Convolutional Neural Networks (CNNs) to automatically recognize and detect leaf illnesses on potato crops. High precision, F1 scores, recall, and accuracy were attained by the CNN, suggesting that it could improve efforts to prevent disease and manage crops.

Pérez et al. (2024) presented qREAD-Raman, a non-coding program for evaluating and deciphering Raman spectrum data to identify tomato plant illnesses early. With its superior performance in identifying bacterial illnesses, the system preprocesses and classifies data using neural network techniques. With accuracy rates ranging from 71 to 83% and 58 to 67%, respectively, this research demonstrated respectable performance in differentiation.

Rajpal et al. (2024) introduced a plant disease that affects potatoes, a key food, which can result in large crop losses and jeopardize sustainable food production. Particularly in large-scale activities, conventional approaches were ineffective and susceptible to errors. To improve the classification of potato diseases, advanced Deep Learning (DL) models were employed, such as CNN and image transforms. With an accuracy of 85.06%, the hybrid model EfficientNetV2B3+ViT outperformed earlier techniques.

Sinamay (2024) analysed tomato detection of leaf diseases to increase the precision disorders, a DL model was created, and it achieved 97.44% classification accuracy. That strategy seeks to lower the incidence of these illnesses, which will support the expansion of the tomato business worldwide.

Khan et al. (2024) investigated the efficacy of CNN, Region-based Convolutional Neural Network (RCNN), dimensional plan models, in detecting fungal blight illnesses in potato crops using image processing. The CNN model ensures both economic expansion and food security by performing better than other models and approaches. An 80:20 ratio was used for training and testing, and the suggested model achieved an outstanding 97.44% classification accuracy.

Reis and Turk (2024) proposed a Deep Convolutional Neural Network (DCNN) based on images, the Multi-head Attention Mechanism Depthwise Separable Convolution Inception Reduction Network (MDSCIRNet) architecture, for the classification of potato leaf diseases. The research using the MDSCIRNet + Support Vector Machine (SVM) approach had the maximum success rate of 99.33%, while the integrated DL model and hard voting ensemble learning model had an accuracy rate of 99.11%.

Wagle et al. (2024) examined 4793 patent papers from Lens.org and examine international research endeavours, leading companies, and the patenting environment; it investigates the patented viewpoint of tomato plant disease using AI. The analysis examined the patent landscape, experts in the field, and research activity around the world. The various aspects of the intellectual property competitive landscape process pertaining to tomato plant disease are the main subject of this research.

Christakakis et al. (2024) detected Botrytis cinerea on Cucurbitaceae crops early and used the vision transformer encoders to enhance DL segmentation models. The model that performed the best, U-Net++ with MobileViTV2-125, had an IoU of 67.1% and a precision of 92%. On-site digital systems for practical uses could include their development in agricultural disease monitoring and management.

Owoeye et al. (2024) introduces a cloud-based software solution that uses neural networks to identify and cure plant illnesses in tomato leaves. With 98% accuracy during training and 80% accuracy in real-world scenarios, the system was trained on the plant village dataset and improved crop yields and agricultural sustainability.

Debnath et al. (2024) research offered a DL and EfficientNetV2B2-based disease identification system that analyses tomato leaf pictures. Accurate disease identification was possible because of the system's transfer learning training, which supports sustainable tomato farming methods and helps to manage the disease. On a dataset used for testing, the precision of the system was almost 100%.

Jung et al. (2024) improving agricultural productivity and quality requires the creation of a computerized system for early disease identification. The system has a high accuracy rate of 97.09% and employs a CNN algorithm and a stepwise illness diagnosis model. By using more crops as training datasets, the model could be extended to intelligent cultivation of the genus crops.

Esomonu et al. (2023) integrated a mobile system for tomato producers presented in combined ML methods, such as Universal Sentence Encoder (USE), and MobileNetV2, SVM, to accurately identify and classify diseases. A system of recommendations for early illness identification and prevention was provided by the system. This research find that the pretrained MobileNetV2 combined with SVM performs exceptionally well, beating other pre-trained models with 93.68% precision, 93.84% F-1 score, 96.81% accuracy, and 95.4% recall.

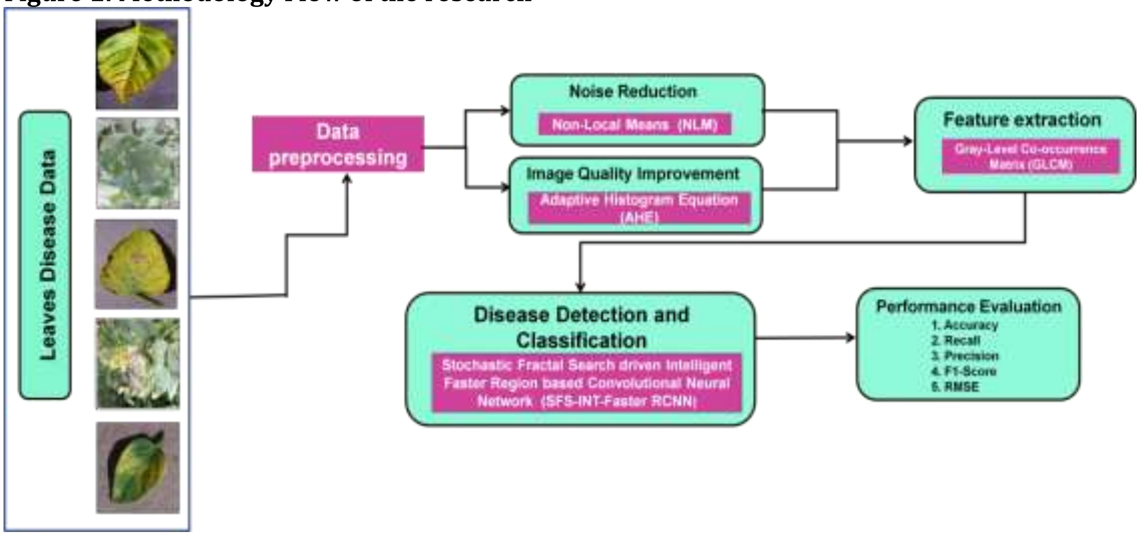
2.1 Research Gap

Despite the recent studies, some limitations exist with the current DL models used in the plant disease detection. Most of the research is done using specific plant species or certain disease pathogens, hence the models developed and restricted from diagnosis of diseases in other plant species. Further, models such as CNNs and EfficientNetV2B3 has been seen shown to perform well, but they come with its problems like computational slowness and rigid nature. It should be noted that important optimization methods necessary for the fine-tuning of the model and increasing the coefficient of correct real-time detection have not been integrated. Additionally, many of these models entail overfitting issues or, in other cases, exhibit performance that depends on the size of the dataset processed. To overcome these challenges, the proposed method named SFS-InT-Faster R-CNN integrates Stochastic Fractal Search (SFS) optimization technique, which increases their measure of the exploration-exploitation ratio and optimizes the relative error of the model for identification of various diseases that affect leaves. The integration of SFS enhances the performance of the model across crop-disease types, since it is less computationally complex, handles large and mixed data types effectively while increasing scalability.

3. Proposed Methodology

The high-resolution images of Solanaceae family leaves (tomato, brinjal, and pepper) diseased leaves were gathered for open source. Pre-processing involved noise reduction by using an NLM filters and contrast enhancement via AHE to improve image quality. Feature extraction was performed using the GLCM to capture critical texture features. The SFS-InT-Faster R-CNN model was implemented, leveraging the Stochastic Fractal Search algorithm for hyperparameter optimization. Figure 1 shows the proposed method flow.

Figure 1: Methodology Flow of the research



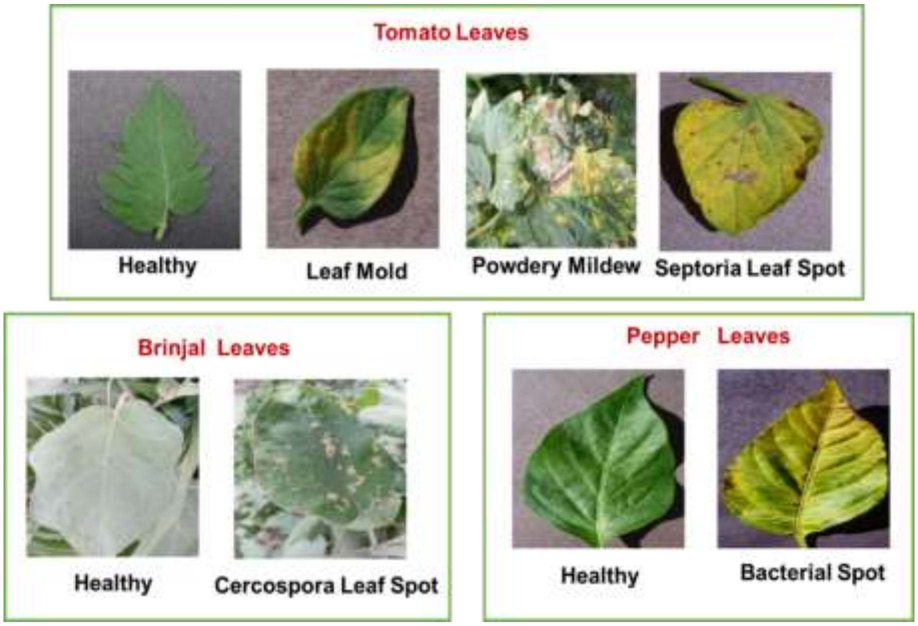
3.1 Dataset

The Solanaceae leaf dataset comprises high-resolution images of tomato, brinjal (eggplant), and pepper bell leaves, gathered to identify and categorize plant diseases [28]. Pictures are divided into categories of healthy and diseased leaves covering conditions like Leaf Mold, Powdery Mildew, and Septoria Leaf Spot (Tomato); Cercospora Leaf Spot (Brinjal); and Bacterial Spot (Pepper Bell). With 17,382 images, the dataset is organized into specific folders for each crop and disease type, facilitating deep-learning model development for automated plant health monitoring. Table 1 and Figure 2 present the data collection.

Table 1: Data Collection

Crop	Disease Type	Number of Images
Tomato	Healthy	3,051
	Leaf Mold	2,754
	Septoria Leaf Spot	2,882
	Powdery Mildew	1,004
Total Images		**9,691**
Brinjal	Healthy	1,177
	Cercospora Leaf Spot	1,039
Total Images		**2,216**
Pepper	Healthy	1,478
	Bacterial Spot	997
Total Images		**2,475**

Figure 2: Sample Data



3.2Data Pre-processing

To enhance image quality and prepare it for analysis. Noise is reduced using the Non-Local Means (NLM) filter, and contrast is improved through Adaptive Histogram Equalization (AHE). These steps ensure clarity and highlight critical disease features in plant leaf images.

Noise Reduction using Non-Local Means: The NLM filtering reduces noise in high-resolution leaf images, preserving essential details. This enhances image clarity, allowing the model to focus on subtle disease symptoms for accurate detection and classification of Solanaceae diseases and defective model assumptions using Equation (1).

$$h = e + \eta \quad (1)$$

Here, e represents the clean image, hthe noisy image, and $\eta$the Gaussian noise. H(j)denotes the filtered leaf image using Equations (2-4).

$$H(j) = \frac{\sum_{i\epsilon\Omega_j} \omega(j,i)h(i)}{\sum_{i\epsilon\Omega_j} \omega(j,i)} \quad (2)$$

$$\omega(j,i) = \exp\left(-\frac{c(j,i)}{h^2}\right) \quad (3)$$

$$c(j,i) = \|M(j) - M(i)\|_{2,\alpha}^2 \quad (4)$$

The method calculates the weighted Euclidean distance $c(j,i)$between image blocks $M(i)$and$M(j)$, with αas the Gaussian kernel's standard deviation. The normalization factor $g = \sigma$ensures weights sum to 1. To optimize speed, Ωlimits the search region size.

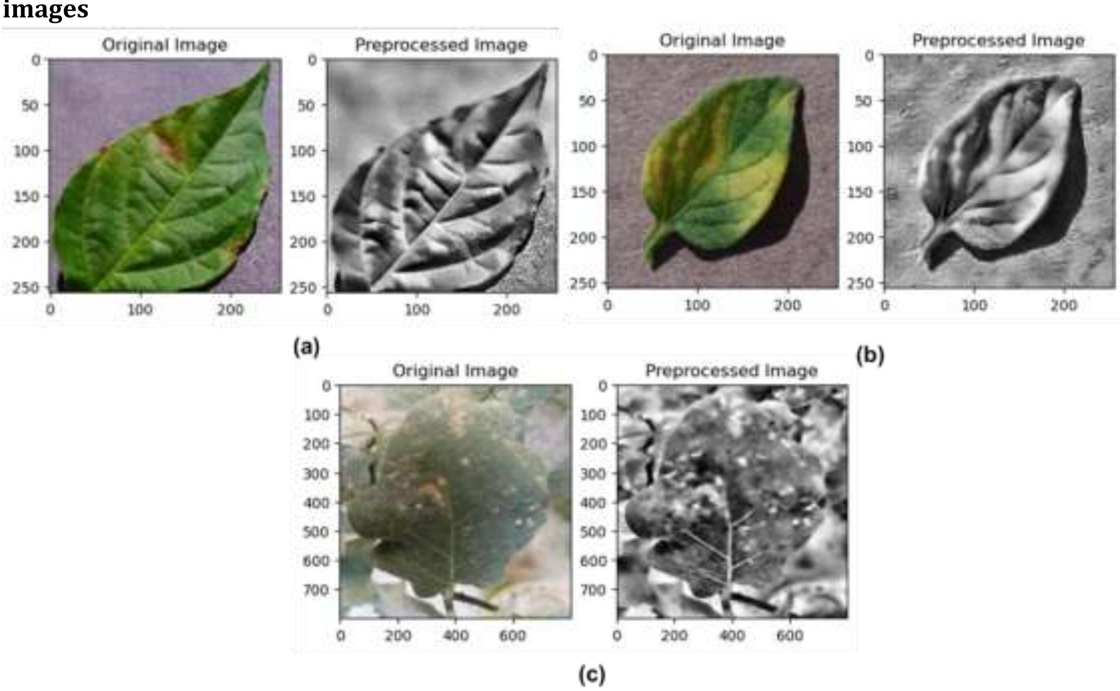
Image Quality Improvement using Adaptive Histogram Equalization (AHE): Itis an essential initial processing method that is used to improve the contrast of Solanaceae family leaf images, focusing on tomato, brinjal, and pepper. AHE works by dividing the image into localized regions, known as contextual regions, centred on each pixel. Histogram equalization is applied based on the intensity distribution in these regions, ensuring adaptive enhancement of contrast. This technique helps highlight subtle disease symptoms, such as discoloration or texture changes, which could be overlooked. By enhancing image quality AHE suggestive Ly advances the exactness of disease detection and classification using the Equations (5 and 6).

$$(s) = b[b, n^{--}(s) + (1 - a), n^{--}(s)] + (1 - b), (b, n^{--}(s) + (1 - a), n^{--}(s)] \quad (5)$$

$$b = (z - z^-)/(z^+ - z^-), a = (w - w^-)/(w^+ - w^-) \quad (6)$$

While AHE effectively enhances subtle disease symptoms, it can amplify noise in homogeneous regions, leading to over-enhancement. Figure 3 presents preprocesses image using NLM and AHE techniques.

Figure 3: Pre-processing Techniques(a) Tomato leaf images, (b) Brinjal leaf images, and (c) Pepper leaf images



3.3 Feature Extraction using Gray-Level Co-occurrence Matrix (GLCM)

GLCM captures the texture and spatial relationships by analysing the distribution of pixel intensity pairs in a given Solanaceae family leaf images (tomato, brinjal, and pepper). This technique helps identify subtle variations in texture that could indicate early disease symptoms. By extracting key features such as contrast, correlation, and homogeneity, GLCM enhances the replica's capability to organize and sense diseases effectively in plant leaves. Although there are other approaches to texture analysis, GLCM could be used to compare the Gray-level differences of any two neighbouring pixels in an identified axis or displacement of an image. Similarly, the GLCM of a picture is a function of the pixel spacing and angle relation inside the area. The GLCM of an image I of size M x M was defined by Equation (7).

$$KK(w,z) = \sum_{w=1}^{M}\sum_{z=1}^{M}\begin{cases}1, & J(t,u) = w \text{ and } J(t+\Delta t, u+\Delta u) = z \\ 0, & \text{otherwiwise}\end{cases} \qquad (7)$$

The collective probability of the co-existence of strengths j and i in $K(w,z)$ at any assumed point $\Delta t\&u$, where t and u are the exact locations in the image. The approach and reserve between the pixels are quantified by $\Delta t\&u$.

3.4 Leaf Diseases Detection using Stochastic Fractal Search-driven Intelligent Faster Region-based Convolutional Neural Network (SFS-InT-Faster R-CNN)

Plant leaf diseases significantly impact agricultural productivity, necessitating efficient detection systems. SFS-InT-Faster R-CNN for accurate disease detection in Solanaceae crops. The model leverages advanced optimization and deep learning techniques for enhanced accuracy and early diagnosis.

3.4.1 Faster Region-based Convolutional Neural Network(R-CNN)

Consequently, Solanaceae family leaves, especially tomato, brinjal, and pepper, diseases are detected and classified using Faster R-CNN. R-CNNs work most efficiently to detect areas of interest in the image, which makes them ideal for localized pathologies or what could be referred to as spotting of some diseases in plants, such as discoloration or lesions on the leaves due to pathogens. The R-CNN framework operates in three stages: the region suggestion, classification for pavement distress. First, proposal regions are obtained by selective search or other region proposal algorithms that search for regions in the image that could contain symptoms of the disease. These proposed regions, or regions of interest (RoIs), are taken through a CNN where its primary role is to extract features from each proposed RoIs. The CNN learns about the higher-order features and high-level textural and structural features that could enable the separation of the disease pretentious and healthy regions on the leaves.

Finally, the extracted structures are fed to a classifier that categorizes each RoI as diseased or healthy. Let J represent the image, where $J \in \mathbb{R}^{G\times X\times D}$is a 3D tensor with height I, width X, and channelD. B region proposal Q is extracted, and its features are obtained by passing through the CNN using Equations (8-10):

$$E_Q = CNN(Q) \quad (8)$$

Where E_Q denotes the feature vector of the region. The classifier D uses these features to classify the region as diseased or healthy:

$$O_Q = D(E_Q) \quad (9)$$

Where O_Q is the probability of the region Q being diseased. Bounding box regression is applied to refine the proposed regions, improving localization accuracy;

$$A_Q = Regressor(E_Q) \quad (10)$$

Such a regional method of early disease detection makes it possible to determine the presence of the disease on Solanaceae family leaves and to immediately take measures to prevent further damage, ultimately increasing the efficiency of crop protection measures.

3.4.2 Stochastic Fractal Search Optimization (SFSO)

In the disease detection of Solanaceae leaves, SFSO a balanced exploration-exploitation process, where the diffusion mechanism enhances the model's sensitivity to various disease signs, while the updating procedure fine-tunes disease classification. Incorporating SFS-R-CNN the test facilitates early diagnosis of diseases, hence optimizing crop health, and reducing losses. The SFS algorithm is an effective computational method of global optimization belonging to the metaheuristic category, based on natural processes, including fractal growth. Specifically in plant disease detection, SFS is used to fine-tune the detection model of Solanaceae family leaves, which includes (tomato, brinjal, and pepper leaves). Thus, SFS can achieve a better distribution of samples in the solution space given to copying the process of random fractals diffusion, which in turn makes it easier to identify and categorize plant diseases using the model. The SFS algorithm, two main processes occur: the diffusion process and the updating process. The dispersal process causes the elements to shift from their current positions, while the constriction process bounds the particles, confining them in a way that allows the algorithm to search within the area of interest without being trapped in a local minimum. This is important for recognizing the delicate patterns of the disease in the leaves of plants, as the model should eliminate primacy on well-defined forms while at the same time considering obscure forms as well. By contrast, the updating process updates the location of the particles in the context of all other particles in a specific group in the search for improved solutions using Equations (11 and 12).

$$HX_1 = Gaussian(\mu_{AO}, \sigma) + (\varepsilon \times AO - \varepsilon \times O_j) \quad (11)$$

$$HX_2 = Gaussian\ (\mu_O, \sigma) \quad (12)$$

Where μ_{AO}, μ_O, and σ are the Gaussian resources and normal aberration of, $\mu_{AO} = |AO|$and $\mu_O = |O_s|$ If g specifies the group amount, the normal deviation in the above equations is calculated using Equation (13).

$$\sigma = \left|(\log(f)/f) \times (O_j - AO)\right| \quad (13)$$

For a D-dimensional global optimization problem, a D-dimensional vector was used to create each potential solution. Each location in the initiation procedure is started at random using its minimum and maximum values. Given that VA and KA represent the vectors representing the issue's variables' upper and lower bounds, accordingly, the establishing equation for the j^{th} point, O_i, using Equation (14).

$$O_i = KA + \varepsilon \times (VA - KA) \quad (14)$$

By the dispersal procedure's exploitation property, all points stray from their existing position to utilize the issue search space. SFS employs two statistical approaches to improve space exploration because of the exploration attribute. Each vector index is subjected to the first process, and all locations are subjected to the second method. Initially, the first step ranks every point according to the fitness function's value. Next, a probability value o'_{bj}that follows a straightforward uniform distribution is assigned to each point j in the group using the Equations (15 and 16).

$$o'_{bj} = rank(O_j)/M \quad (15)$$

$$O_j(i) = O_q(i) - \varepsilon \times \left(O_s(i) - O_j(i)\right) \quad (16)$$

The statistical technique is applied to the points' constituent parts. Once all of the points derived from the first statistical method are ranked before starting the second one using as in the first statistical procedure, the current location of O'_j is updated if the criterion $O'_j < \varepsilon$ is maintained for a new point O'_j; otherwise, no update takes place evaluating the Equations (17 and 18). Figure 4 displays the SFS's flowchart.

$$O_j'' = O_j' - \beta \times (O_s' - AO)|\beta \le 0.5, \quad (17)$$
$$O_j'' = O_j' - \beta \times (O_s' - O_q')|\beta > 0.5, \quad (18)$$

Figure 4: Flowchart of Stochastic Fractal Search optimization

Algorithm 1 combines Stochastic Fractal Search (SFS) optimization with R-CNN to improve Solanaceae leaf disease detection. R-CNN identifies regions of interest, extracts features, and classifies them, while SFS optimizes the R-CNN's hyperparameters, such as learning rate and anchor settings, ensuring superior detection accuracy and computational efficiency. By leveraging SFS, the model achieves enhanced performance in detecting diseases in tomato, brinjal, and pepper crops, making it a reliable tool for early plant disease intervention.

Algorithm 1: SFS-InT-Faster R-CNN

```
import tensorflow as tf
from SFS import StochasticFractalSearch
from FasterRCNN import FasterRCNN
from DataLoader import load_and_preprocess_data
def load_data_and_preprocess(path):
   return load_and_preprocess_data(path)
def setup_and_train_model(input_shape, num_classes, train_data, epochs = 50):
   model = FasterRCNN(input_shape, num_classes)
   optimizer = StochasticFractalSearch(). optimize(model, train_data)
model. compile(optimizer = optimizer['optimizer'], loss = 'categorical_crossentropy')
model. fit(train_data, epochs = epochs)
   return model
def evaluate_model(model, test_data, true_labels):
   predictions = model. predict(test_data)
   accuracy = accuracy_score(true_labels, predictions)
   print ("Accuracy: ", accuracy)
```

```
def main (dataset_path, input_shape, num_classes):
train_data, train_labels = load_data_and_preprocess(dataset_path + '/train')
test_data, test_labels = load_data_and_preprocess(dataset_path + '/test')
  model = setup_and_train_model(input_shape, num_classes, train_data)
evaluate_model(model, test_data, test_labels)
dataset_path = 'path_to_dataset'
input_shape = (224, 224, 3)
num_classes = 10
main (dataset_path, input_shape, num_classes)
```

4. Performance Evaluation

Accuracy: The model proved to be dependable and efficient in accurately identifying disease cases that are assessed using Equation (19) by achieving high accuracy in the detection and categorization of diseases in leaves of the Solanaceae family.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (19)$$

Precision: The proposed model measures the accuracy of the positive predictions, i.e., how many of the leaves identified as diseased are indeed diseased. For plant disease detection, high precision minimizes false positives, reducing unnecessary treatments and costs associated with incorrect disease classifications in Solanaceae crops using Equation (20).

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (20)$$

Recall: the model identifies all diseased leaves. In plant disease detection, high recall ensures that the model detects most of the diseased instances, preventing missed early-stage infections that could lead to crop loss in Solanaceae plants, ensuring timely interventions with Equation (21).

$$\text{Recall} = \frac{\text{TP}}{\text{TP=FN}} \quad (21)$$

F1-Score: The model's performance is measured in a balanced way by combining precision and recall. The reliability of early-stage interventions using Equation (22) is improved when the model retains both high sensitivity in recognizing diseases and accuracy in categorizing diseased leaves, as indicated by a high F1 score in Solanaceae leaf disease detection.

$$\text{F1} = 2 \cdot \frac{\text{Precision}\cdot\text{Recall}}{\text{Precision+Rcall}} \quad (22)$$

RMSE: the difference between predicted and actual disease classifications in Solanaceae leaves. A lower RMSE indicates better model performance, signifying more accurate predictions. In plant disease detection, it evaluates how effectively the model detects and classifies diseases, ensuring reliable early intervention and minimizing crop damage using Equation (23).

$$\text{RMSE} = \sqrt{\frac{1}{m}\sum_{j=1}^{m}(z_j - \hat{z}_j)^2} \quad (23)$$

5. Experimental setup

Python is the primary programming language used, with Scikit-learn for classical ML algorithms and TensorFlow/Keras for deep learning. Hyperparameter optimization is handled by Hyperopt. Table 2 presents the experimental setup.

Table 2: Experimental setup

Component	Specification/Tool
RAM	8 GB
Processor	Intel(R) Core (TM) i5-3320M CPU @ 2.60GHz 2.60 GHz
Disk	Hard Disk - 200 GB in C Drive
Operating System (OS)	Windows 10 64-bit configuration
Programming Language	Python 3.11.7
Frameworks/Libraries	TensorFlow, Keras, Scikit-Learn
Development Environment	Anaconda (Individual Edition)

IDE/Notebook	Jupyter Notebook
Anaconda URL	Anaconda Individual Edition
Jupyter Notebook URL	Jupyter Notebook

6. Performance Evaluation

The detection and classifying of diseases in the Solanaceae family leave with high accuracy and speed. The model, integrating advanced preprocessing techniques, feature extraction, and optimization algorithms, significantly outperforms traditional methods. By leveraging enhanced image quality, critical texture features, and optimized Faster R-CNN architecture, the method offers a dependable and effective way to detect diseases early, contributing to improve agricultural practices and reduced leaf diseases. The proposed SFS-InT-Faster R-CNN method was compared with existing techniques, including VGG19 (Islam et al. 2021), Xception (Islam et al. 2021), KNN (Alshammari et al. 2023), WoA-ANN (Alshammari et al. 2023), VGG Net (Umamageswari et al. 2022), and LSTM-SIFT (Umamageswari et al. 2022). The comparison evaluated performance metrics, such as accuracy, recall, F1-score, precision and RMSE to assess the effectiveness of each method for Solanaceae leaf disease detection. Table 3 presented the comparison of the methods. Figure (5-7) displays confusion matrices for three crops: (Figure 5) Tomato, (Figure 6) Brinjal, and (Figure 7) Pepper Bell, for healthy and diseased leaves with segments corresponding to model performance. For each crop, the matrix has different true labels plotted against predicted labels and highlights the classification accuracy, especially for diseases, such as Leaf Mold, Powdery Mildew, and Bacterial Spot. Within each matrix, the numbers are labelled as the amount of negative results and positive results. The numeral of misclassified diseases or healthy leaves; helps to understand which kind of disease is the model correctly recognizes to classify diseases and healthy leaves of all kinds as well.

Figure 5: Confusion Matrix of Tomato,

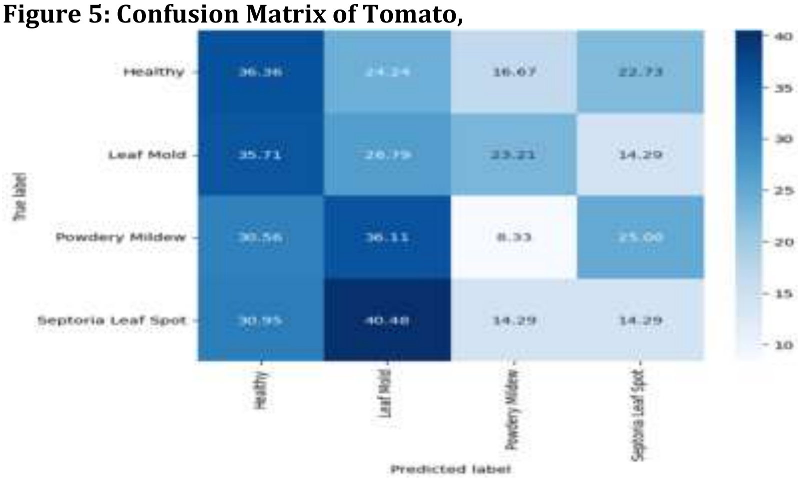


Figure 6: Confusion Matrix analysis of Brinjal

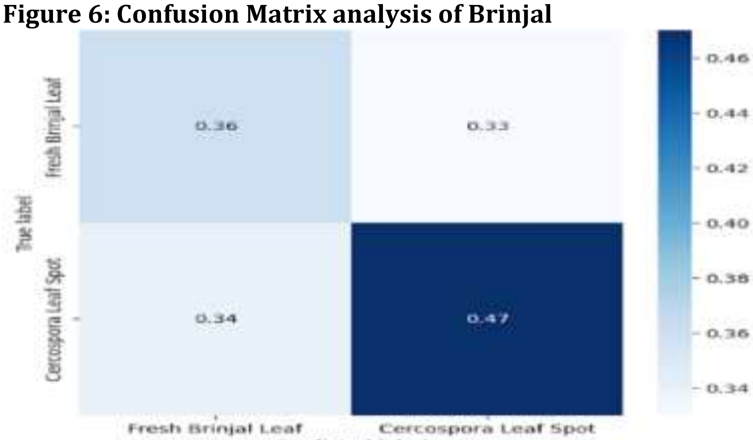
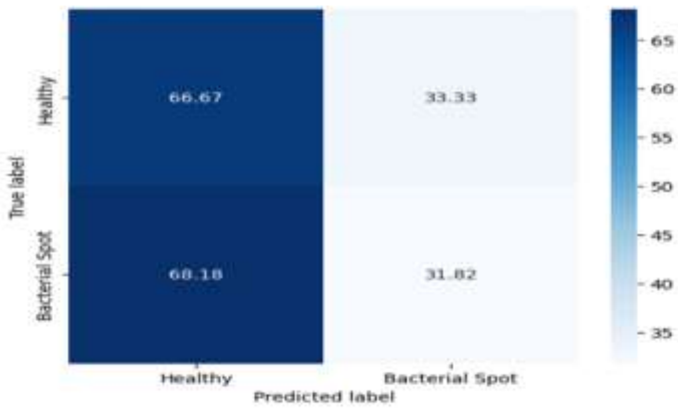


Figure 7: Confusion Matrix Pepper Bell



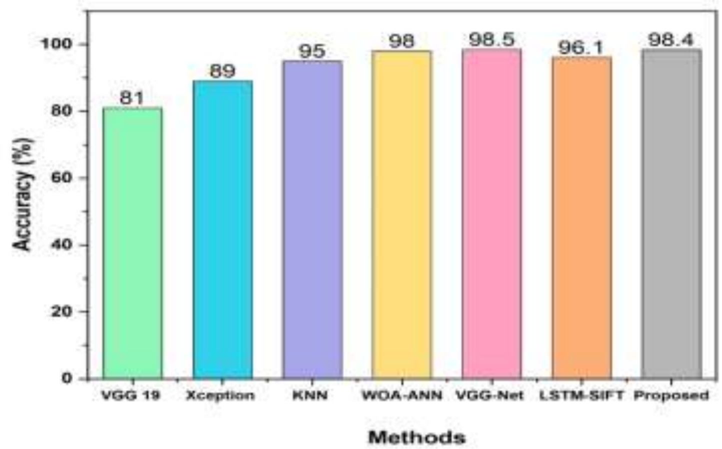
The identification of tomato, brinjal, and pepper plant diseases has been presented in Table 3, form with the classification performance. But diseases, such as Tomato (Powdery Mildew) got 98.4% and Pepper (Bacterial Spot) got 94.3% accuracy. Analysing Root Mean Square Error (RMSE) values, a similar level of precision is shown, thereby implying the efficiency of the model with respect to disease and healthy leaf discrimination.

Table 3: Plant Diseases Classification

Plant Diseases classification			
Plant	**Disease**	**Accuracy (%)**	**RMSE**
Tomato	Leaf Mold	96.2	0.2
	Septoria Leaf Spot	97.8	0.18
	Powdery Mildew	98.4	0.37
	Healthy	97.1	0.26
Brinjal	Cercospora Leaf Spot	97.5	0.35
	Healthy	95.9	0.52
Pepper	Bacterial Spot	94.3	0.37
	Healthy	97.2	0.32

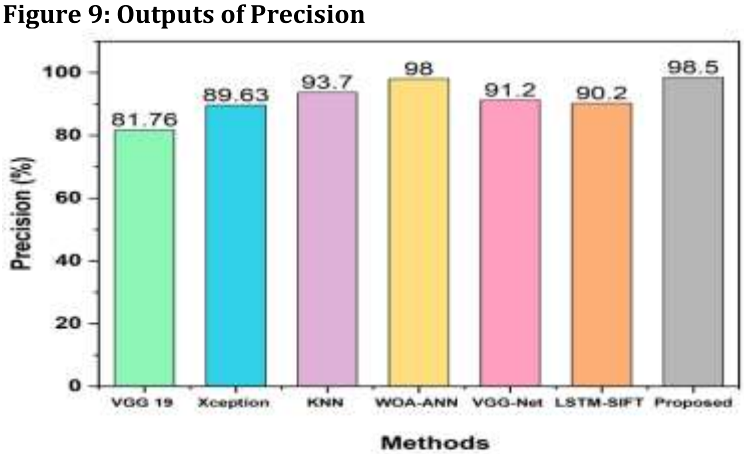
Accuracy: Figure 8 and table 4 presented the comparison of accuracy, VGG 19 achieved 81% accuracy, Xception improved to 89%, and KNN reached 95%. VGG Net scored 96.1%, WoA-ANN demonstrated 98%, while LSTM-SIFT scored 98.4%. The proposed SFS-InT-Faster R-CNN model excelled with 98.5%, showcasing the highest accuracy in detecting Solanaceae crop diseases.

Figure 8: Accuracy comparison across models



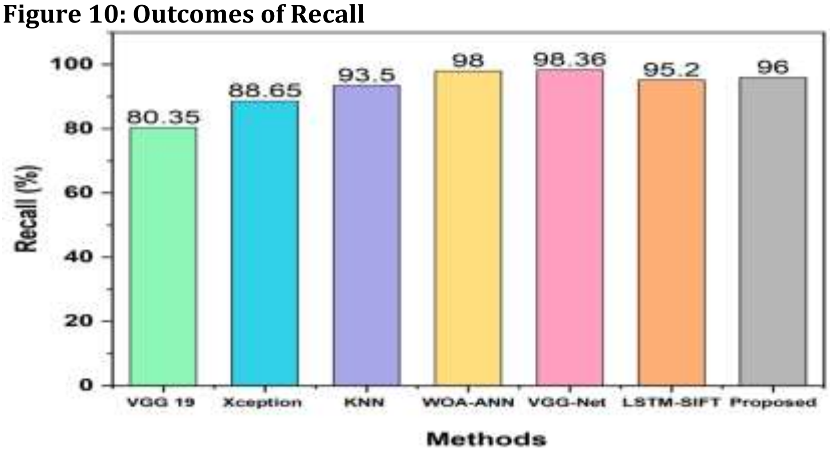
Precision: VGG 19 recorded 81.76% precision, Xception achieved 89.36%, and KNN reached 93.7%. WoA-ANN maintained 98% precision, VGG Net scored 91.2% and LSTM-SIFT scored 90.2%. The proposed model excelled with 95.2%, reducing misclassification and offering superior precision. Figure 9 and table 4 demonstrates the outcomes of precision.

Figure 9: Outputs of Precision



Recall: Figure 10 and table 4 presented the recall, VGG 19 achieved 80.35% recall, Xception reached 88.65%, and KNN scored 93.5%. VGG Net scored 95.2%, WoA-ANN and LSTM-SIFT achieved 98% and 96%, respectively. The proposed model outperformed with a recall of 98.36%, ensuring comprehensive disease detection.

Figure 10: Outcomes of Recall



F1 Score: VGG 19's F1-score was 80.41%, Xception scored 88.23%, and KNN achieved 93.78%. WoA-ANN reached 98%, VGG Net scored 92.88% while LSTM-SIFT scored 92.31%. The proposed model excelled with an F1-score of 98, showing exceptional balance in precision and recall. Figure 11 and table 4 illustrates F-1score.

Figure 11: Comparison of F-1 score

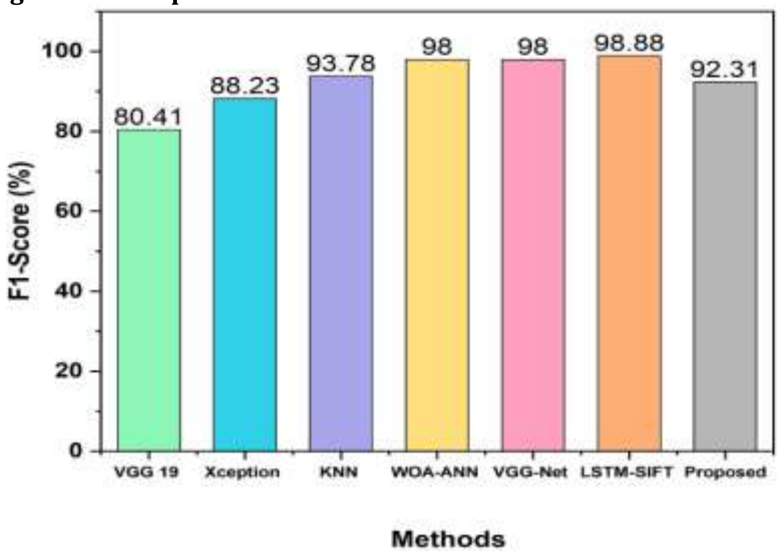


Table 4: Comparison of the Methods

Methods	Accuracy	Precision	Recall	F-1 score
VGG 19 (Islam et al. 2021)	81	81.76	80.35	80.41
Xception (Islam et al. 2021)	89	89.36	88.65	88.23
KNN (Alshammari et al. 2023)	95	93.7	93.5	93.78
WoA-ANN (Alshammari et al. 2023)	98	98	98	98
VGG Net (Umamageswari et al. 2022)	96.1	91.2	95.2	92.88
LSTM-SIFT (Umamageswari et al. 2022)	98.4	90.2	96	92.31
SFS-InT-Faster R-CNN (Proposed)	**98.5**	**95.2**	**98.36**	**98**

7. Discussion and Conclusion

The proposed SFS-InT-Faster R-CNN model achieves high accuracy in the Solanaceae leaf diseases detection; the model suffers from the problem of high computational time. VGG19 and Xception, which are conventional approaches, require significantly less time for training and performing than deep feature-based & CNN Combinations while they provide higher error rates due to less complex structures. While approaching the problem with KNN or LSTM-type models provides sufficient accuracy, their usage in the analysis of severe diseases registered in high-resolution images allows observing their weaker performance in terms of intricate pattern identification. The proposed SFS-InT-Faster R-CNN has eliminated these problems by applying SFS optimization technique. The hyperparameters are adjusted by the SFS algorithm, making the improvement in the Faster R-CNN. Indeed, integrating exploitation with exploration to minimize the model training causes fast convergence but high accuracy, as demonstrated by SFS. Additionally, the optimization reduces the computational resources required, as the model complexity is adapted through iterative learning, making the method more applicable to real-time applications in farming. Both in terms of speed and accuracy, this combination of optimization and deep learning results in a boost in efficiency for early-stage plant diseases.

Research presents an advanced framework for the automated discovery and organization of diseases in Solanaceae family leaves using optimized Faster R-CNN architecture. In this research, techniques including NLM, AHE, and GLCM are integrated into the proposed model for feature extraction. The SFS-InT-Faster R-CNN, resulting in a precision of 95.2%, an accuracy of 98.5%, an F1-score of 98%, a recall of 98.36%, These numerical results validate the model's efficiency and accuracy, providing a transformative approach to crop disease monitoring and prevention. It is limited to being based on a specific dataset of images of tomato, brinjal, and pepper leaves, which could limit its generalization to other crops. Furthermore, future classical technologies could not be viable for implementation on devices with limited resources due to its computational complexity. The dataset should be expanded to include more varied crops and climatic situations in future researches. It should optimize the model for real-time mobile applications. Furthermore, it integrates well with IoT-based systems for large-scale agricultural monitoring.

References

1. Matin, M.M.H., Ava, M., Anjum, A. and Uddin, M.S., 2024. A Dataset for Accurately Identifying Eggplant Diseases. Golam Moazzam and Uddin, Mohammad Shorif, A Dataset for Accurately Identifying Eggplant Diseases. https://dx.doi.org/10.2139/ssrn.4630716
2. Houetohossou, S.C.A., Houndji, V.R., Hounmenou, C.G., Sikirou, R. and Kakaï, R.L.G., 2023. Deep learning methods for biotic and abiotic stresses detection and classification in fruits and vegetables: State of the art and perspectives. Artificial Intelligence in Agriculture. https://doi.org/10.1016/j.aiia.2023.08.001
3. Bütüner, A.K., Şahin, Y.S., Erdinç, A., Erdoğan, H. and Lewıs, E., 2024. Enhancing Pest Detection: Assessing Tutaabsoluta (Lepidoptera: Gelechiidae) Damage Intensity in Field Images through Advanced Machine Learning. Journal of Agricultural Sciences, 30(1), pp.99-107. https://doi.org/10.1007/s11540-024-09735-y
4. Tace, Y., Tabaa, M., Elfilali, S., Bensag, H. and Leghris, C., 2024, May. A novel approach for detecting Bacterial spot combining Transfer Learning and Large Language Models. In 2024 IEEE 12th International Symposium on Signal, Image, Video and Communications (ISIVC) (pp. 1-6). IEEE. https://doi.org/10.1109/ISIVC61350.2024.10577820
5. Tammina, M.R., Sumana, K., Singh, P.P., Lakshmi, T.V. and Pande, S.D., 2024. Prediction of Plant Disease Using Artificial Intelligence. In Microbial Data Intelligence and Computational Techniques for Sustainable Computing (pp. 25-48). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-99-9621-6_2
6. Polin, J.A., Hasan, N., Habib, M.T., Rahman, A., Vasha, Z.N. and Sharma, B., 2024. Tomato pest recognition using convolutional neural network in Bangladesh. Bulletin of Electrical Engineering and Informatics, 13(1), pp.619-627. https://doi.org/10.3389/fpls.2024.1278760
7. Martina, M., De Rosa, V., Magon, G., Acquadro, A., Barchi, L., Barcaccia, G., De Paoli, E., Vannozzi, A. and Portis, E., 2024. Revitalizing agriculture: next-generation genotyping and-omics technologies enabling molecular prediction of resilient traits in the Solanaceae family. Frontiers in Plant Science, 15, p.1278760. https://doi.org/10.5539/jas.v16n2p9
8. Kahya, E., Ozduven, F.F. and Aslan, Y., 2024. YOLOv5 Model Application in Real-Time Robotic Eggplant Harvesting. Journal of Agricultural Science, 16(2). https://doi.org/10.6084/m9.figshare.26421439
9. Oh, S., Kim, M.S., Kang, H.J., Kim, T., Kong, J. and Choi, D., 2024. Conserved effector families render Phytophthora species vulnerable to recognition by NLR receptors in nonhost plants. Nature Communications, 15(1), p.10070. https://doi.org/10.3389/fpls.2024.1482275
10. Thakare, A.P., Della Lucia, M.C., Mulagala, C., Bertoldo, G., Cagnin, M. and Stevanato, P., 2024. Omics-based approaches to decipher the leaf ionome and transcriptome changes in Solanum lycopersicum L. upon Tomato Brown Rugose Fruit Virus (ToBRFV) infection. PloS one, 19(11), p.e0313335. https://doi.org/10.1371/journal.pone.0313335
11. Lih, H.T., Airemwen, C.O. and Halilu, E.M., 2024. Phytochemical studies and evaluation of silver nanoparticles synthesized from Solanum elaeagnifolium leaves extract for antioxidant and antibacterial activities. Tropical Journal of Natural Product Research, 8(2), pp.6440-6445.
12. Putra, I.G.A.M., Dewi, P.J.N., Yuarini, D.A.A., Al Hakim, M.E. and Mahatmananda, I.B.G., 2024. Characteristics of Takokak (Solanum torvum Sw.) Leaves Herbal Tea Based on Drying Temperature. International Journal on Food, Agriculture and Natural Resources, 5(1), pp.44-48. https://doi.org/10.33366/bs.v21i1.2662
13. Gurusamy, S., Natarajan, B., Bhuvaneswari, R. and Arvindhan, M., 2024. Potato plant leaf diseases detection and identification using convolutional neural networks. In Artificial Intelligence, Blockchain, Computing and Security Volume 1 (pp. 160-165). CRC Press.
14. Pérez, M.R.V., Denis, J.J.C., Ley, A.C., Herrera, J.A.S., Ortiz, J.C.D., Vázquez, Á.G.R. and Contreras, H.R.N., 2024. Early Plant Disease Detection by Raman Spectroscopy: An Open-Source Software Designed for the Automation of Preprocessing and Analysis of Spectral Dataset. Crop Protection, p.107003. https://doi.org/10.1016/j.cropro.2024.107003

15. Rajpal, A., Mishra, R., Rajpal, S., Kavita, Bhatia, V. and Kumar, N., 2024. Explaining deep learning-based leaf disease identification. Soft Computing, pp.1-24.
16. Sinamay, J.H., 2024. Improving Potato Disease Image Classification Through Advanced Deep Learning Models (Master's thesis, Oslo Metropolitan University). https://doi.org/10.1177/00368504241275371
17. Khan, M., Gulan, F., Arshad, M., Zaman, A. and Riaz, A., 2024. Early and late blight disease identification in tomato plants using a neural network-based model to augment agricultural productivity. Science Progress, 107(3), p.00368504241275371. https://dx.doi.org/10.2139/ssrn.4962883
18. Reis, H.C. and Turk, V., 2024. Potato leaf disease detection with a novel deep learning model based on depthwise separable convolution and transformer networks. Engineering Applications of Artificial Intelligence, 133, p.108307. https://doi.org/10.1016/j.engappai.2024.108307
19. Wagle, S.A., Volkova, E. and Tiwari, A.K., 2024. Patinformatics Studies on Tomato Plant Disease. Recent Patents on Engineering, 18(3), pp.84-101. https://doi.org/10.2174/1872212118666230330110559
20. Christakakis, P., Giakoumoglou, N., Kapetas, D., Tzovaras, D. and Pechlivani, E.M., 2024. Vision Transformers in Optimization of AI-Based Early Detection of Botrytis cinerea. AI, 5(3), pp.1301-1323. https://doi.org/10.3390/ai5030063
21. Owoeye, S., Durodola, F., Adekunle, M., Oyelami, A. and Bisiriyu, B., 2024. Development of a Tomato Leaf Disease Detection System Using Convolutional Neural Networks. Journal of Engineering Research, 29(3), pp.22-31. https://doi.org/10.1109/TST.2017.7830891
22. Debnath, A., Hasan, M.M., Raihan, M., Samrat, N., Alsulami, M.M., Masud, M. and Bairagi, A.K., 2023. A smartphone-based detection system for tomato leaf disease using efficientNetV2B2 and its explainability with artificial intelligence (AI). Sensors, 23(21), p.8685. https://doi.org/10.3390/s23218685
23. Jung, M., Song, J.S., Shin, A.Y., Choi, B., Go, S., Kwon, S.Y., Park, J., Park, S.G. and Kim, Y.M., 2023. Construction of deep learning-based disease detection model in plants. Scientific reports, 13(1), p.7331. https://doi.org/10.1038/s41598-023-34549-2
24. Esomonu, N.F., Eze, U.F., John-Otumu, A.M. and Ayogu, I.G., 2023. A Cascaded Intelligent Recommender Model for Multi-Class Tomato Leaf Diseases Identification and Control. https://doi.org/10.5281/zenodo.10053690
25. Islam, M.A., Shuvo, M.N.R., Shamsojjaman, M., Hasan, S., Hossain, M.S. and Khatun, T., 2021. An automated convolutional neural network-based approach for paddy leaf disease detection. International Journal of Advanced Computer Science and Applications, 12(1).
26. Alshammari, H.H., Taloba, A.I. and Shahin, O.R., 2023. Identification of olive leaf disease through optimized deep learning approach. Alexandria Engineering Journal, 72, pp.213-224. https://doi.org/10.1016/j.aej.2023.03.081
27. Umamageswari, A., Deepa, S. and Raja, K., 2022. An enhanced approach for leaf disease identification and classification using deep learning techniques. Measurement: Sensors, 24, p.100568. https://doi.org/10.1016/j.measen.2022.100568
28. Data source: https://www.kaggle.com/datasets/ziya07/solanaceae-family-leaves-dataset/data