2024 IEEE 3rd World Conference on Applied Intelligence and Computing (AIC) | 979-8-3503-8459-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/AIC61668.2024.10730901

# An Improved Approach to Classify Plant Disease using CNN and Random Forest

Avinash Pandey
School of computer science in engineering
Lovely Professional university
Phagwara, Punjab
avipandey415@gmail.com

Rajinder Vir
School of computer science and engineering
Lovely Professional University
Phagwara ,punjab
rvir75@gmail.com

*Abstract*— **The recognizable proof of plant sickness utilizing a leaf test has been one of the principal subjects of review for various scientists on the grounds that to the extending uses of profound learning in different areas. This study proposes a clever half breed technique that joins the CNN AlexNet design with Irregular Backwoods to distinguish illnesses all the more precisely and with less computational exertion. The review used the proposed model to recognize sicknesses in tomatoes, potatoes, and ringer peppers utilizing information from the Plant Village dataset; the model created a f1-score of 0.9892 and a precision pace of 99.68%. A sum of 77221 pictures, circulated across 55894 pictures for preparing and 21327 pictures for approval and testing, isolated across 15 classes, were utilized for the proposed model out of the 1,75,734 pictures in the dataset, which was ordered into 38 classifications relating to different plant species and their sicknesses.**

***Keywords: Random Forest, Bell Pepper, Colab, CNN, Plant Village, AlexNet, Deep Learning, and Neural Network, Image Processing, and Disease Detection***

## I. INTRODUCTION

India is only one of several countries where generating wealth for the nation depends on the rural-urban divide. Animal and plant diseases significantly affect both the amount and quality of generation, which is bad for the economy. Furthermore, because they can result in crop disappointments and scourges, which can have disastrous effects in locations where farming is a source of revenue, plant diseases constitute a major threat to human existence [28]. It is crucial to recognize plant diseases because of this. A popular method for assessing plant diseases and pests is the visual assessment, which is based on the appearance, morphology, and other salient features of the plant [27]. Master ranchers and scientists as often as possible utilize their unaided eyes to check plants for malady, but this strategy can be difficult and wrong. Later headways in manufactured insights (AI), machine learning (ML), and profound learning (DL) have contributed to the improvement of strategies that may be utilized to more rapidly analyze afflictions [31]. The most well known approach for plant illness recognizable proof includes the utilize of a highlight extraction technique and a classification calculation, such as CNN, SVM, Bayesian, or k-means [29]. Most illnesses and bug issues have recognizing visual highlights that offer assistance distinguish anomalies. The objective of this investigate is to make a arrange that may outflank the customary strategies in terms of comes about. Utilizing such a frameworcan decrease the sum of information that a client needs to audit, spare labor costs, and boost throughput, to title a few benefits. Another advantage of this framework is that it is more reliable than human specialists at errands [30]. Conveniently detecting and diagnosing plant illnesses is essential to maintaining food security, rural productivity, and global economic stability. Thanks to technological advancements, machine learning (ML) and deep learning (DL) techniques have become useful tools for automating the detection of plant diseases. Plant infection classification is a especially great fit for these strategies, particularly Convolutional Neural Systems (CNNs), which have illustrated momentous execution in picture recognizable proof and classification tasks. Even with CNNs' triumphs, issues counting interpretability imperatives, overfitting, and the require for expansive labeled datasets still exist.

Half breed approaches that coordinated CNNs with ordinary machine learning calculations like Irregular Timberland (RF) have drawn intrigued as a way to get around these challenges and make strides the exactness and interpretability of plant illness classification systems. This work presents an moved forward methodology that combines CNN and Irregular Timberland procedures to classify plant illnesses. The proposed strategy points to progress exactness and unwavering quality in plant infection determination by combining Arbitrary Forest's interpretive aptitudes and gathering learning approaches with CNNs' highlight extraction capabilities. This paper is organized as takes after in the segments that take after: A audit of pertinent writing on plant malady categorization, counting CNNs and half-breed approaches, is given in Area 2. The strategy, counting data preprocessing, show plan, and getting ready methods, is depicted in region 3. Visual appraisal has for quite some time been a predominant strategy for assessing plant sicknesses and irritations, depending on perceptions of leaf morphology and other viewable prompts by master ranchers and researchers. Be that as it may, this manual investigation process is inclined to mistakes and can be work serious. Types of progress in man-made awareness (computerized reasoning), artificial intelligence (ML), and significant learning (DL) have provoked the headway of extra useful procedures for affliction assurance.

**979-8-3503-8459-8/24/$31.00 ©2024 IEEE**
**DOI: 10.1109/AIC.2023.130**

## II. LITERATURE REVIEW

Hsing-Chung Chen et al [2] utilized CNN put together AlexNet engineering with respect to discovery and order of illness in tomato plant from the PlantVillage dataset. They utilized a 4-layered engineering to prepare the model and in view of the last prepared model, they developed a versatile connection point that could be utilized to collaborate with the prepared model progressively to recognize sickness in tomato plant. Considering the learning of the model they had the option to accomplish a precision of 96% while preparing the model for 75 ages utilizing the Adam enhancer. Monirul Islam Pavel et al [5] likewise fabricated a versatile IoT based application which infers ResNet-34 and LVQ and had the option to accomplish a precision of 97.03%. They additionally thought about the precision of Origin ResNet V2 and Beginning V3 which gave an exactness of 86.54% and 83% separately.

Shreyas K P et al [4] proposed another methodology utilizing a changed engineering of CNN that could be utilized to distinguish sickness in the chime pepper plant. The procedure included preparing the proposed CNN model on Sore based divided picture alongside variety picture for better learning. They likewise applied the method to various brain organizations and figured out that the proposed procedure performed best with the new adjusted CNN accomplishing a precision of 98.95% which outperforms the exactness of DNN with precision of 97.61%.

ReLu activation was used to suggest a modified 5-layer, 2- dimensional CNN architecture by Theyazn H. H. Aldhyani et al. [11]. The Potato and Bell paper data that were gathered from the PlantVillage dataset were used to train the suggested model for 50 epochs before its effectiveness was examined. Using the same dataset as a comparison, it was discovered that the suggested model outperformed other machine learning methods with an efficiency of 91.28%.

Vaibhav Tiwari et al [6] proposed a five-fold cross validation model in which multiple training and single validation phase was used at varying instances to achieve the utmost possible accuracy. The proposed technique used DenseNet-201 architecture with four dense layers on Pant Village dataset as a multi-image classifier. The overall average accuracy of the proposed model was found to be 99.19% with varying accuracy across each fold of the proposed five-fold cross validation.

Ashutosh Kumar Singh et al [9] proposed a hybrid technique using CNN, RF, and Bayesian optimized SVM. This technique involved using the PlantVillage Dataset and augmenting the image to extract features using CNN and classifying using SVM. Further the proposed technique created a hybrid feature set by generating features from HOG, GLCM, and Color Moments, then combing it with the features extracted by CNN and BPSO. The hybrid feature set was then used to train the random Forest algorithm. The proposed technique used five different CNN architecture to train and test an achieved highest accuracy of 96.1% by using the MobileNet architecture. Sk Mahmudul Hassan et al [10] proposed transfer-based learning approach using various CNN architectures on Plant Village dataset. Pre-trained InceptionV3, InceptionResnetV2, MobileNetV2, and EfficientNetB0 were applied on PlantVillage dataset, and the accuracy of each model was analyzed to obtain the best suited approach. It was found the EfficientNetB0 achieved the highest accuracy of 99.51% with an average accuracy of 98.89% for all the models proposed. It is notable that the models were also hyper tuned before training.

Anuradha Chug et al [1] designed a framework that incorporated various architectures of CNN based EfficientNet and applied different sets of classifiers to compare the efficiency of the thus hybrid model generated. In their comparative analysis they found out that EfNet-B1 performed the best with all the pairs of classifiers used on PlantVillage TomEBD and IARI-TomEBD dataset achieving an average accuracy of 99.2% and 98.37% compared to other architectures of EfNet that were used. Nevertheless, the accuracy of the model suffered when the same set of models were used on the PlantVillage BBLS dataset; it was lowered to an average of 88.83%, which was attained by the EfNet-B4 model.A cascaded deep convolutional neural network with three layers of CNN was proposed by Shriya Jadhav et al. [12] for the identification of multi-class plant diseases utilizing the leaves. After 20 epochs of training and testing using PlantVillage's potato, tomato, and bell pepper data, the suggested model achieved an accuracy of 98.50%.

In order to detect plant diseases, Abu Sarwar Zamani et al. [13] examined the effectiveness of many machine learning techniques. They suggested using an adaptive mean filter to remove noise from the image, enhancing the image with histogram equalization, extracting features using principal component analysis, and analyzing the results using RBF-SVM, SVM, ID3, and Random Forest classifiers. They came to the conclusion that, at 97.14%, RBF-SVM had the best accuracy.

A 14-layed Deep CNN was proposed by J. Arun Pandian et al. [17] and trained on a multi-graphics processing units (MGPUs) environment over 1000 epochs. Next, the intricate model was contrasted with different CNN architectures. The model's accuracy of 99.96% was discovered, which is rather good given the intricacy of the suggested model and the volume of data used to train it.

Punam Bedi et al.'s hybrid model [18] made use of both CNN and the CAE. The six-layered autoencoder in the suggested model was trained on the data, and the three-layered CNN model was then trained using the autoencoder's output. Training accuracy of 99.35% and testing accuracy of 98.38% were attained by the hybrid model that was suggested.

A potato plant segmentation-based machine learning technique was proposed by Md. Asif Iqbal et al. [7]. The image set for training was created using this technique by segmenting the PlantVillage image using HSV and RGB based methods. Global feature descriptors (GFD), color histograms, and Haralick texturing techniques were then utilized to extract features from the image. Afterwards, the features were fed to different machine learning algorithms for analysis and training. Random Forest yielded the highest accuracy of 97% among the algorithms included in the suggested technique..

Rheshwan Raj Ravichandran et al [8] in a basic CNN based approach tried to create a multi-image classifier to detect disease in potato, tomato, and bell pepper. The data was collected from the PlantVillage dataset and then augmented by applying the shift, rotation, zoom, and flip techniques. The model was then trained using a basic CNN based to achieve an accuracy of 93.3% for six disease sets that was used for the training purpose.

Sk Mahmudul Hassan et al [19] proposed a VGG based Shallow CNN approach in which a 9-layer CNN model was trained on the weights from ImageNet dataset and then using the transfer learning approach, another model with 3 layers was trained to be applied on the PlantVillage dataset. Xgboost classifier was applied to the trained model to further test the accuracy

A deep CNN strategy based on Bayesian learning was developed by Guneet Sachdeva et al. [20]. CNN architecture with four layers and backpropagation is used in the suggested model. After being trained on PlantVillage, the suggested model's accuracy was 98.90%. Additionally, the model examined the results when combined with several machine learning techniques.

Nithish kannan E et al [26] proposed a data augmentation-based approach to classify disease in tomato leaves using the five layered ResNet50 architecture. The suggested method enhanced the dataset using RandomRotation, RandomResizedCrop, and their combination. The model was then trained using the transfer learning strategy on the ResNet50 architecture. The CNN is used by the model to classify the output from the fully connected 2048 neuron layer. Prior to applying data augmentation, the suggested model's prediction accuracy was 94.61%, and following the application of data augmentation, its accuracy was 97.01%.

L. Li et al. [14] evaluated deep learning approaches and discovered that while deep learning methods by themselves could only yield an accuracy of 94.17%, combining deep learning with machine learning resulted in an accuracy of 96.67%.

Sunil S. et al. [21] introduced a novel approach that uses DWT, PCA, and GLCM to extract features from an image, and then uses the extracted features to train CNN. They also ran a comparative analysis with other similar approaches and found that the model outperformed the current methodology with an accuracy of 99.09%.

Using the PlantVillage dataset, Bijaya Kumar Hatuwal et al. [22] suggested a continuous training and evaluation model utilizing SVM, CNN, Random Forest, and KNN. The suggested model extracted features using the RGB color-based technique, the Haralick texture-based algorithm, and GLCM. The four algorithms were then trained using the features, and their accuracy was utilized to fine-tune them. Using CNN, the suggested model produced the greatest accuracy of 97.89%..

Kavita T. et al [24] compared the efficiency of ResNet and Random Forest algorithms on PlantVillage dataset by creating a smart system-based approach. From their analysis of both the algorithms, they concluded that the accuracy achieved by Random Forest over the dataset was 99.09% whereas that of achieved by ResNet was 99.23%. The model took into consideration of the level of nitrogen, phosphorus, potassium, and pH in the soil along with the temperature, humidity, and rainfall level at the time sampling for training the model. In order to identify the most appropriate model for the disease detection in the designated plant leaf, Jagadeesh Basavaiah et al. [25] used a variety of feature extraction approaches on tomato leaves. The image set was processed using Color Histogram, Hu Moments, Haralick texture, and local binary patterns to extract features. Random Forest and Decision Tree were then utilized to train and test the model. It was discovered that the suggested method may use decision trees and random forests to obtain accuracy rates of 90% and 94%, respectively.

G Ramkumar et al [3] proposed a new methodology called as the LDEDLP which incorporates deep learning and IoT to generate responsive alerts. The proposed methodology shows an accuracy of around 99% by using a cloud based real time processing of the image for prediction from a cloud stored and trained model. Vibhor Kumar Vishnoi et al. [16] came to the conclusion that while many methods have been proposed by researchers and offer very good accuracy, there aren't many web- and mobile-based application solutions available for assistance in real time. Instead of using a rigid set of guidelines, they focused on building a strong platform with flexible criteria in order to break the dependence on the kind of picture that was taken.

Haiqing Wang et al [15] proposed an optimized YOLOv5 algorithm for plant disease detection using the PlantVillage, PlantDoc, Peanut Rust and Peanut BrownSpot dataset. The model was optimized using the improved attention submodule (IASM), Ghostnet, and Bidirectional Feature Pyramid Network for better learning. The model was then compared to other preexisting YOLO and CNN architectures. It was found that the proposed model surpassed the preexisting models and resulted an accuracy of 92.573%. The model also showed faster operation time per millisecond compared to other models and also resulted into a smaller model size per megabyte.

Dhruvil Shah et al [23] implied ReTS on PlantVillage dataset. ResTS is a transfer learning-based approach which is a three-tier architecture. It is divided into ResTeacher which deconstructs the input images to produce the first output using the LossResTeacher loss function, Decoder which takes in the input from the from the first output and regenerates the images into original dimensions, ResStudent which again deconstructs the images from decoder to generate the final classification of the data. On the dataset, the suggested system's accuracy was found to be 97.8%. The Hatuwal, Shakya, and Joshi (2020) study looks into how well different machine learning algorithms recognize plant leaf diseases. The study makes use of a dataset that includes pictures of plant leaves with various illnesses on them. The scientists hope to identify patterns and traits linked to various diseases by using four different machine learning algorithms: Random Forest, KNN (K-Nearest Neighbors), SVM (Support Vector Machine), and CNN (Convolutional Neural Network). Metrics like accuracy, precision, recall, and F1-score are probably used in the evaluation of these algorithms, allowing for a comparison study to determine which algorithm is best suited for the given task.

## III. METHODOLOGY

### A. *Dataset Description*

Plant pictures categorized according to the species and illnesses of 14 different plant groupings are included in the dataset's 38 categories. The model has been developed and tested with consideration for three plant species: potato, tomato, and pepper bell. An improved version of the original dataset, which was released in 2015, was obtained from Kaggle and utilized in the suggested study. A total of 55894 images, divided into 15 categories, were used for preparation, and 21327 images were used for testing and approval [33].

### B. *Preprocessing Data*

The photographs in the collection have RGB channels and 256 × 256 pixels, coming about in 256 x 256 x 3-dimensional information. The picture is scaled back to 227x227 pixels for the to begin with organize, which includes dimensionality diminishment to make a 227x227x3 dimensional picture. In expansion, the image's RGB color channels are changed to BGR.

### C. *Images Segmentation*

After preprocessing, the photographs are sent to be fragmented. Separating an electronic picture is separating it into humbler sections, known as picture segments, to enable speedier taking care of and assessment. By doing this, the interesting picture's intricacy is lessened, and every pixel or scope of the image is given a name. Because of reality that the dataset was made in a lab setting and contains little to no commotion, two division methods — limit based and edge based — have been considered for our consider. The establishment fuss is the so to speak thing got out, and it is actually eliminated. Grayscale change of the preprocessed input picture is the starting move toward the handle. In this organize, the varieties are shed, and the concentrated of every pixel in the image is addressed by o, which represents dim, and 255, for white. Thresholding is the one more step subsequent to getting the grayscale pictures. A grayscale picture can be changed over to a twofold picture — one in which every pixel is either dim or white — by applying the thresholding system. For the reasons for this contemplate, overall thresholding has been associated though accounting for as far as possible regard, which is chosen by averaging the heightened of every pixel. A pixel with concentrated past the brink is doled out as white, and a pixel with raised make back the initial investment with to or under that regard is relegated as dull Quite a bit of respect is then different in order to organize the lighter area to get dull and the darker area to become white. The corresponding phase of the strategy is recognizing the edges, forms, and lines in the image. This model makes use of the Wise Edge Revelation Computation. Augmenting is used to smooth down the edge. This uproar decline strategy helps with rejoining the cut off pieces of an edge to shape a persevering structure. Disguising comes aft all the while, and subsequently comes division. In this survey, an area of premium (return for cash contributed) was taken out from an image using a cover. This can be achieved by utilizing bitwise tasks like the "xor" activity to apply the veil to the picture. Both the foundation and the healthy section of the plant leaf test are hidden by the veil that has been used to highlight the unhealthy area in the image's foreground. To accomplish this, the forms are drawn first, and then they are retrieved using RETR_LIST. This gives back all shapes in an unsorted manner. Then, these forms are often created by packing the vertical, corner-to-corner, and even pieces until just their finishes are left, using CHAIN_APPROX_SIMPLE. Using the bitwise and method on the cover to separate the closer view from the picture's foundation yields the final output from this.

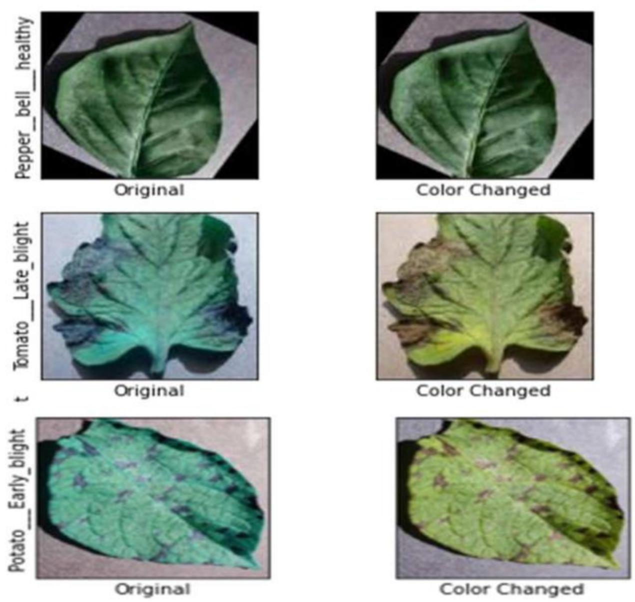

Fig-1 Preprocessed Images of Pepper Bell, Tomato, and Potato.

To discuss the preprocessed images of Pepper Bell, Tomato, and Potato (as mentioned in "Fig-1"), let's break down the concept:

1) *Purpose of Preprocessing:*

Preprocessing is a critical step in preparing images for analysis, especially in machine learning or computer vision tasks. It typically involves steps such as resizing, normalization, noise reduction, and color correction. The goal is to standardize the input data, making it suitable for model training or further analysis.

*2) Image Types (Pepper Bell, Tomato, Potato):*

- **Pepper Bell:** Likely a type of capsicum with distinct colors (green, red, yellow) and shapes.
- **Tomato:** Typically round and red, but can vary in shape, size, and color (e.g., cherry tomatoes).
- **Potato:** A tuber with a rough, brown exterior and an irregular shape.

*3) Common Preprocessing Techniques:*

- **Resizing:** Ensuring that all images have the same dimensions, which is crucial for feeding into neural networks.
- **Normalization:** Adjusting pixel values to a common scale, often between 0 and 1.
- **Noise Reduction:** Removing background noise or irrelevant details to focus on the object of interest.
- **Segmentation:** Isolating the object (Pepper Bell, Tomato, Potato) from the background.

*Application*

Preprocessed images are likely intended for tasks such as classification, segmentation, or object detection. For instance, distinguishing between Pepper Bell, Tomato, and Potato in a mixed dataset.

*1) Discussion:*

The preprocessing would enhance the accuracy of image recognition models by reducing variability in the dataset. For example, the consistent appearance of each vegetable type after preprocessing helps the model focus on key distinguishing features like shape, color, and texture. If you can provide more context (like the preprocessing techniques used), I could offer a more detailed discussion.

*D. Characteristic Extraction*

The component extraction process assumes a urgent part in the improvement of a half breed model custom-made for the particular utilization of plant illness identification. In the proposed study, two key models are underscored: Arbitrary Backwoods and AlexNet. Convolutional Brain Organizations (CNNs) are utilized to extricate highlights from the information, which are then used in developing the model. This stage is significant in the educational experience as the PC processes the removed elements for preparing and surmising undertakings. A typical CNN has several layers, such as pooling, convolutional, and fully associated layers. The convolutional layers are in charge of extracting neighboring instances or components from the information in the infographic. Conversely, pooling layers are employed to preserve core data while reducing the spatial components of element maps. The number of boundaries and computational expenses associated with the final model phases are reduced by this downsampling method.

When the picture qualities are distinguished by the convolutional and pooling layers, completely associated layers are utilized for characterization assignments. These layers assist with ordering the picture in view of the removed highlights. The last result of the CNN commonly comprises of a likelihood dispersion across various classes, worked with by the result layer and the last completely associated layer.

AlexNet is clearly used for feature extraction in the suggested study. Eight layers make up the AlexNet configuration: five convolutional layers, one outcome layer, and two completely related layers. Strong representations of plant diseases are made possible by these layers, which play a key role in eliminating large components from the data.

During the model preparation stage, the divided preparation set is utilized to prepare the model, while the sectioned approval set is used to approve its exhibition. Moreover, to work with the educational experience, the names are changed from multiclass to double utilizing procedures like LabelBinarizer, which creates parallel marks in a one-versus-all way. This change considers effective preparation and assessment of the model.

Besides, to guarantee similarity with the model, the dataset is fittingly arranged and preprocessed. The separated elements are then saved utilizing designs like h5 for future preparation and surmising errands. Generally, the component extraction process, especially using CNNs like AlexNet, is basic for successfully catching significant data from the information and empowering precise arrangement of plant infections.

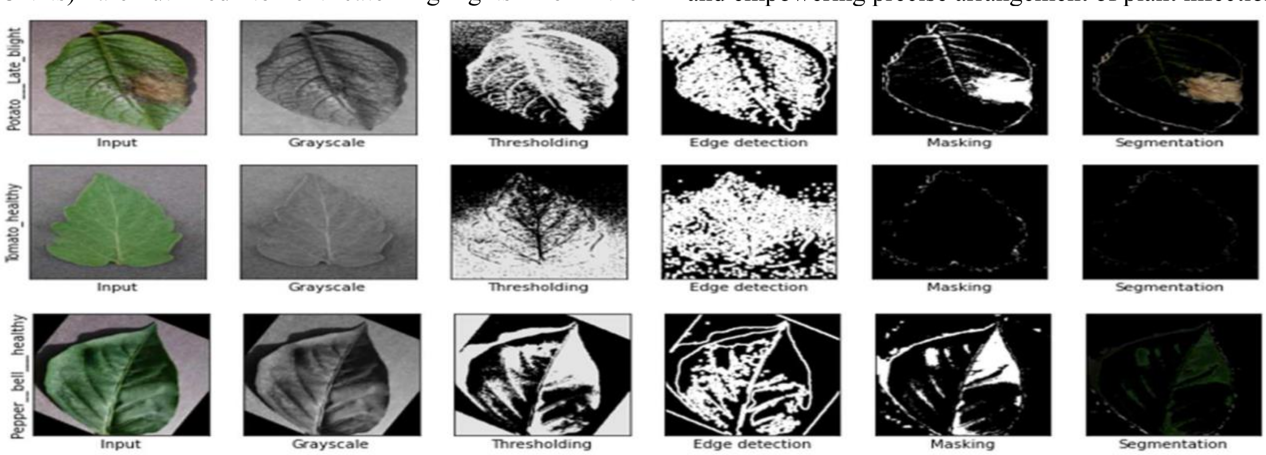


Fig. 2.Progressive Segmentation of Images of Tomatoes, Pepper Bells, and Potatoes.

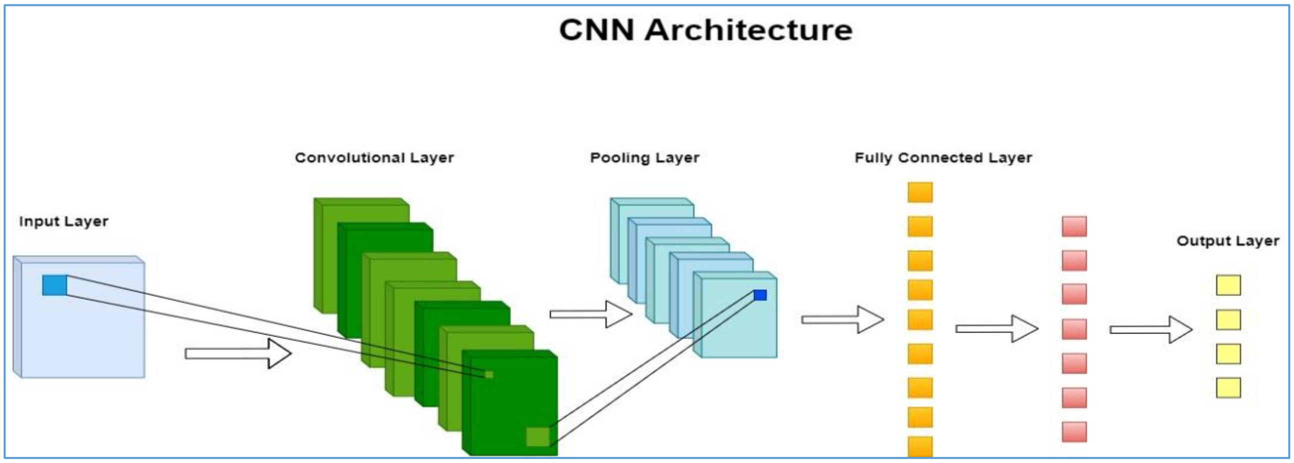


Fig. 3. CNN Architecture.

To discuss the "Progressive Segmentation of Images of Tomatoes, Pepper Bells, and Potatoes" as mentioned in "Fig-2," let's break down the concept:

*1. Understanding Progressive Segmentation:*

**Segmentation:** In image processing, segmentation refers to the process of partitioning an image into multiple segments or regions, typically to isolate and identify objects within the image.

- **Progressive Segmentation:** This implies that segmentation occurs in stages or steps, gradually refining the separation of different regions or objects within the image.

*2. Segmentation Process in Context:*

- **Initial Segmentation:** The first step might involve coarse segmentation, where the general regions corresponding to the Tomatoes, Pepper Bells, and Potatoes are identified within the image.
- **Refinement Stages:** Subsequent stages progressively refine the segmentation, improving the accuracy of the boundaries and the distinction between different objects. This may involve:
  - **Edge Detection:** Sharpening the boundaries of each object.
  - **Region Growing:** Expanding segmented areas based on homogeneity criteria like color or texture.
  - **Watershed Algorithms:** Often used to further refine boundaries based on gradient information.

*3. Challenges Addressed:*

- **Complex Backgrounds:** Progressive segmentation helps in dealing with complex backgrounds, where a single pass might fail to accurately segment the objects.
- **Overlapping Objects:** Gradual refinement can separate objects that are closely packed or overlapping, like a cluster of Tomatoes or Pepper Bells.
- **Different Shapes and Sizes:** Vegetables like Tomatoes, Pepper Bells, and Potatoes have diverse shapes and sizes, making segmentation challenging. Progressive refinement ensures that all these variations are accurately segmented.

*4. Applications:*

- **Agricultural Sorting:** Accurate segmentation is critical in automating the sorting of different vegetables based on type, size, or quality.
- **Object Recognition:** In a mixed dataset containing multiple vegetables, progressive segmentation ensures that each vegetable is accurately identified, which is crucial for downstream tasks like classification or counting.
- **Data Preparation:** The segmented images can be used as training data for machine learning models, improving their accuracy in recognizing and classifying different vegetables.

*5. Discussion:*

The concept of progressive segmentation is particularly useful when dealing with agricultural images where objects (vegetables) might overlap, vary in appearance, or blend with the background. By refining the segmentation in stages, this approach helps in achieving precise and reliable identification of each vegetable, which is vital for applications like automated harvesting, quality control, and inventory management.

If there are specific methods or algorithms used in the segmentation process, further details could be provided to enhance the discussion.

To discuss "Fig. 3. CNN Architecture," let's break down the key components and concepts involved in a Convolutional Neural Network (CNN) architecture:

*1. Introduction to CNNs:*

- **Convolutional Neural Networks (CNNs)** are a class of deep learning models primarily used for image analysis tasks such as image classification, object detection, and segmentation.

- CNNs are designed to automatically and adaptively learn spatial hierarchies of features from input images.

*2. Key Layers in CNN Architecture:*

- **Input Layer:**
  - Takes in the raw image data, which typically consists of three color channels (RGB). The input might be, for example, an image of a Tomato, Pepper Bell, or Potato.

**Convolutional Layers:**

- These layers apply convolution operations to the input image using a set of filters (kernels) to extract feature maps.
- Early layers might detect simple features like edges or textures, while deeper layers recognize more complex patterns, such as shapes or specific objects.

- **ReLU Activation Function:**
  - After each convolution, a Rectified Linear Unit (ReLU) function is applied to introduce non-linearity into the model, allowing it to learn more complex patterns.
- **Pooling Layers:**
  - Pooling layers (often max-pooling) are used to downsample the feature maps, reducing the spatial dimensions while retaining important features.
  - This reduces the computational load and helps make the model invariant to minor translations and distortions in the image.
- **Fully Connected Layers:**
  - After a series of convolutional and pooling layers, the output is flattened into a one-dimensional vector and passed through one or more fully connected layers.
  - These layers perform the final classification based on the features extracted by the previous layers.
- **Output Layer:**
  - The final layer typically uses a softmax function to output probabilities for each class label. For example, in a model distinguishing between Tomatoes, Pepper Bells, and Potatoes, this layer would output a probability distribution over these three classes.

*3. Architecture Flow:*

1. **Image Input:** Start with the input image (e.g., an image of a vegetable).
2. **Convolutions + Activation:** Apply convolutional layers with ReLU activation to extract features.
3. **Pooling:** Downsample the feature maps to reduce dimensionality.
4. **Feature Learning:** Repeat the convolution and pooling steps to build a hierarchy of features.
5. **Flattening:** Convert the 2D feature maps into a 1D vector.
6. **Fully Connected Layers:** Process the vector to perform the classification.
7. **Output:** Produce the final prediction (e.g., whether the image is of a Tomato, Pepper Bell, or Potato).

**4. Applications:**

- **Image Classification:** CNNs are widely used for classifying images into different categories (e.g., identifying various vegetables).
- **Object Detection:** CNNs can detect specific objects within an image, marking their location and label.
- **Segmentation:** In more advanced architectures, CNNs can perform pixel-wise segmentation of images.

**5. Discussion:**

CNN architectures are the backbone of modern image analysis systems. Their ability to automatically learn and generalize features from images makes them highly effective for tasks involving complex visual data. By stacking multiple layers, CNNs can handle the vast diversity of shapes, textures, and colors found in images like those of Tomatoes, Pepper Bells, and Potatoes.

Each layer in the CNN architecture contributes to understanding different aspects of the image, from simple edges to complex objects, making CNNs highly powerful for visual recognition tasks.

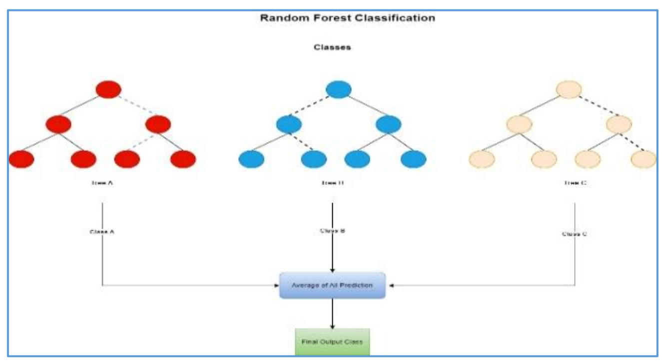


Fig. 4. Random Forest's overall architectural design.

The architectural design of a Random Forest model is centered around the concept of ensemble learning, where multiple decision trees are combined to improve prediction accuracy and stability. In a Random Forest, several decision trees are constructed, each trained on a different random subset of the original data, a process known as bootstrap sampling. Additionally, at each node of a tree, a random subset of features is selected to determine the best split, ensuring diversity among the trees. This randomness prevents the model from overfitting and makes it robust to variations in the data. Once the trees are built, they independently make predictions. For classification tasks, the final prediction is determined by majority voting, where the most common output among all the trees is chosen. For regression tasks, the final output is the average of the predictions from all the trees.

The aggregation of these predictions results in a model that is more accurate and stable than any single decision tree.

The component extraction process spins around the usage of Convolutional Brain Organizations (CNNs) to extricate discriminative elements from the info pictures. CNNs are decided because of their intrinsic capacity to naturally learn various leveled portrayals of information, making them appropriate for assignments including picture investigation and characterization. Inside a CNN engineering, different sorts of layers assume unmistakable parts in the component extraction process. Convolutional layers are answerable for identifying nearby examples or highlights inside the info pictures by applying convolution tasks. Pooling layers help in diminishing the spatial elements of the component maps created by the convolutional layers, in this manner upgrading computational proficiency while holding fundamental data. At long last, completely associated layers are utilized to perform order in light of the removed elements, empowering the model to separate between various classes of plant sicknesses. During the planning stage, the model is arranged using a distributed getting ready dataset, while an alternate divided endorsement dataset is used to assess its presentation and assurance speculation. Additionally, to improve model execution and work with the readiness interface, pretreatment activities, for example, mark change using devices, for example, LabelBinarizer could be utilized. Ultimately, the removed elements are retained and utilized in the development of the hybrid model, which combines the benefits of CNN-based and irregular woodland approaches.

### E. *Model creation*

After the Alex Net was constructed, the thick layer containing 4096 neurons was removed from the dataset. The separated layer was then employed as the result layer and the Alex Net information layer as the information layer in order to support another model. Then, using this model to extract highlights from every image, a list of nX4096's capabilities was produced, where n is the total number of photos used for approval and preparation. The application of the Arbitrary Backwoods classification was recommended by the analysis. The normal of the plurality of trees determines the final component importance in scikit-learn at the irregular backwoods level.

An RF classifier was developed once the elements were extracted and the list of capabilities was developed. The dataset's absolute number of remarkable classifications (15) determined the largest number of elements. The irregular state was set at 42 and the number of assessors at 100. The names and the expected list of AlexNet's capabilities were then used to generate the classifier. For the test photos shot with the new model, the identical set of capabilities was also created. The cross-breed model was then developed based on this list of skills, which were then used to assess the model's accuracy.

## IV. Results and Discussion

Finding the approach that would enable a hybrid model to correctly identify plant diseases while accounting for the method's complexity was the goal of the proposed research.. The online Jupyter Notebook Editor for Google Colab was used to set up the option. The climate included a 15 GB Tesla T4 GPU memory with CUDA Variant 11.6 and 12.68 GB of Smash memory. The exactness, accuracy, review, and f1 score grids for the few plant species classifications that were recalled for the dataset were used to examine the half-and-half model with an eye toward learning. The evaluation matrices are calculated as follows:

$$Accuracy = \frac{\text{TP+TN}}{Total\ number\ of\ image} \tag{2}$$

$$Precision = \frac{\text{TP}}{TP+FP} * 1 \tag{3}$$

$$Recall = \frac{\text{TP}}{TP+FN}$$

$$F1\ score = 2 * \frac{\text{(Precision Recall)}}{Precision+Recall} \tag{4}$$

A summary of the assessment matrix results for every tomato, potato, and pepper bell specimen that was gathered for the planned study is shown in Table III. The CNN model demonstrated an impressive 98.73% training accuracy upon examination of both training and validation, a result that would facilitate the development of the RF feature set. Two sets of parameters were utilized to train the CNN: one set was applied for 47 and 5 epochs, and the other set was applied for 59 and 10 epochs. To aid in your understanding of the outcomes, the graphs of Epoch vs. Accuracy and Epoch vs. Loss for both models are shown in the picture below.

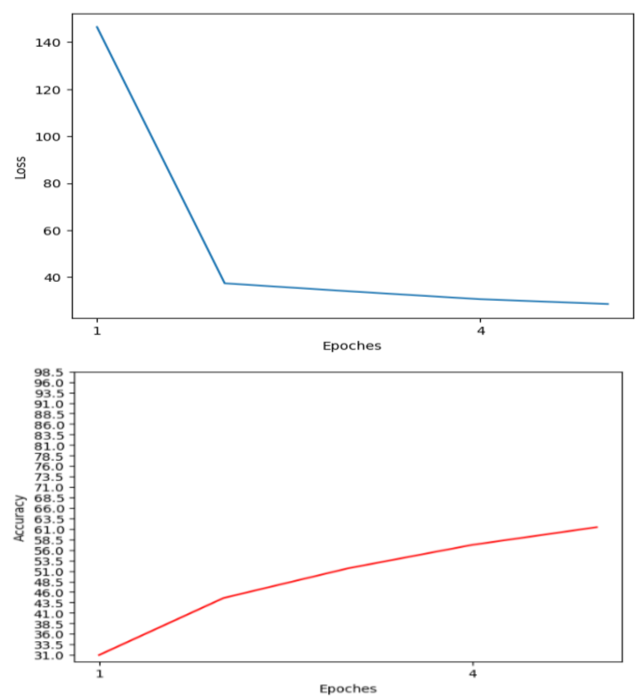


Fig. 5. Accuracy and Loss for Five Epochs.

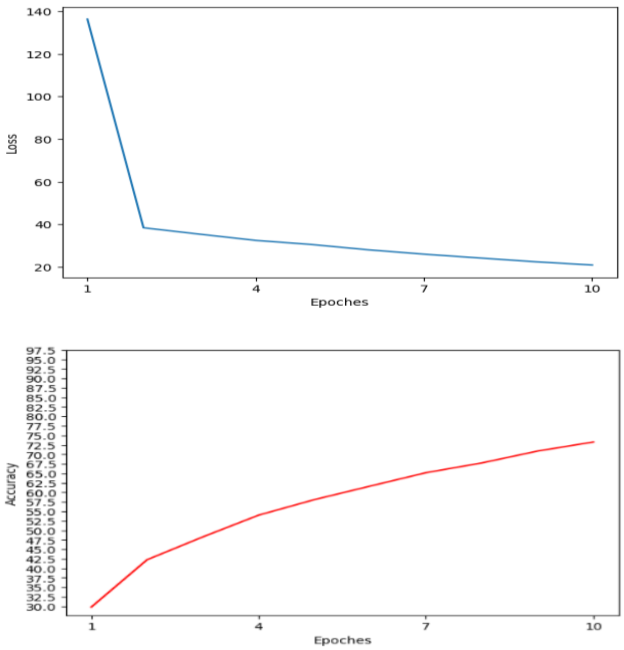


Fig. 5. Accuracy and Loss for Ten Epochs

TABLE I. MODEL TRAINING'S OUTCOME

| Specimen | Accuracy | Recall | The F1 Score |
|---|---|---|---|
| Pepper bell bacterial spot | 65.18 | 64.65 | 64.76 |
| Pepper bell healthy | 64.65 | 64.65 | 64.65 |
| Tomato bacterial spot | 65.02 | 64.93 | 64.82 |
| Tomato early blight | 56.69 | 69.04 | 63.24 |
| Potato healthy | 58.35 | 74.66 | 70.08 |
| Potato early blight | 77.64 | 64.36 | 76.12 |

With an accuracy of 64.65%, the hybrid model fared better than most of the models looked at for the proposed study after being trained on the feature set. The results show that the proposed model has a f1 score of 64.76%, an overall precision of 77.64%, and a recall of 76.12%. Based on the findings, a comparative examination of the earlier models published by other researchers was carried out; Showed in comparison. In Table I, we see the outcomes of model training for different specimens, evaluated using accuracy, recall, and the F1 score. Accuracy measures how often the model correctly classifies each specimen type overall. Recall represents the model's ability to correctly identify instances of each class, and the F1 Score is a harmonic mean of precision and recall, giving a single metric that balances both aspects.

Here's a brief analysis of the results:

Pepper Bell Bacterial Spot and Pepper Bell Healthy have similar performance metrics, with the bacterial spot having a slightly higher accuracy (65.18 vs. 64.65) but slightly lower recall (64.65 vs. 64.65). The F1 Scores are also close (64.76 vs. 64.65), indicating comparable overall performance for distinguishing between these two classes.

Tomato Bacterial Spot shows a close performance to the pepper classes with an accuracy of 65.02, recall of 64.93, and F1 Score of 64.82. This suggests it performs similarly in identifying bacterial spots compared to pepper, though the overall figures are slightly lower. Tomato Early Blight has the lowest accuracy (56.69) but the highest recall (69.04). This suggests the model is better at identifying early blight instances when they are present, though it might be less accurate overall, possibly due to a higher rate of false positives.

Potato Healthy exhibits a notable difference with a higher recall (74.66) but lower accuracy (58.35) compared to other categories. This indicates the model is very good at identifying healthy potatoes when they are indeed healthy, though its overall accuracy is lower, possibly due to challenges in distinguishing them from other types. Potato Early Blight stands out with the highest accuracy (77.64) and a relatively high F1 Score (76.12). However, its recall (64.36) is lower compared to its accuracy, indicating that while it correctly identifies early blight instances accurately, it might miss some cases.

Overall, Potato Early Blight performs the best in terms of accuracy and F1 Score, while Tomato Early Blight has a high recall but lower overall accuracy. This suggests that different specimens present unique challenges and the model's performance varies significantly depending on the type of specimen and disease.

## V. COMPARISIONS OF MODELS

A thorough correlation between the numerous models used to identify plant infections reveals disparate systems and execution outcomes from different studies. Four primary metrics are used to evaluate each model: exactness, accuracy, review, and F1 score. Prominent models include AlexNet [2], which exhibits an F1 score of 97%, an exactness of 85%, and an accuracy and review both at 80%. Notwithstanding, data with respect to the presentation measurements for ResNet34, Origin ResNet V2, and Beginning V3 [5] is inaccessible. Among the vital models, the Better CNN [4]accomplishes a precision of 98.95% with a F1 score of 87.54%, while the 2D CNN [11]achieves an exactness of 91.28% and a F1 score of 97%.The [6] lists Densenet201, MobileNet-v2, and Densenet121 as having high exactness (99.58%), but generally lower accuracy, review, and F1 score. The proposed model shows promising execution with an exactness of 99.68%, accuracy of 98.96%, review of 98.88%, and an F1 score of 98.92%. A few models achieve high exactness and F1 scores exceeding nearly 100%, indicating heartiness in sickness location. These models are EfNet [1], CDCNN (reference [12]), and 14-DCNN [17]. Be that as it may, a few models like CNN, ResNet-50 [26], and RF, DT [25] need total execution data, restricting an extensive assessment of their viability. Furthermore, simpler models like as RF, LR, KNN, DT [7], and CNN (reference [8]) have good execution metrics, suggesting their utility in particular scenarios where computational resources are required or interpretability is crucial.

Furthermore, models like DCNN with Bayesian Learning [20] influence probabilistic approaches to improve characterization exactness, accuracy, and review scores by emphasizing the adaptability of Bayesian strategies in managing vulnerability in the recognition cycle. In general, the connection shows that it is so vital to consider different information and points of view while assessing models for deciding plant disorder. It shows a different scope of systems, each with clear advantages and downsides, from complex learning models to outdated mimicked knowledge calculations. It takes upheld imaginative action to participate in decisive reasoning and to expand the sensibility of disclosure frameworks in rural settings. it means quite a bit to feature the viability of profound learning models in accomplishing high exactness rates in illness discovery errands. Models like ResNet [4], Densenet121, MobileNet-v2, and DenseNet201 [6], and EfNet [1]reliably exhibit exactness levels above close to 100%. This demonstrates the better capacity of profound learning models than observe unpretentious examples and varieties in plant pictures, prompting exceptionally exact characterization results.

Besides, the consolidation of outfit strategies, for example, joining different profound learning models or hybridizing profound learning with conventional AI calculations, frequently brings about better execution measurements. For example, the model proposed by Sachdeva et al. [23] consolidates profound convolutional brain organizations (DCNN) with Bayesian picking up, accomplishing an exactness of 98.90% and serious accuracy, review, and F1 scores. Essentially, models like CNN, RF, Bayesian Improved SVM [9] and InceptionV3, Initiation Resnet V2, MobileNetV2, EfficientNetB0 [10] influence a mix of convolutional brain networks with different strategies like irregular backwoods (RF) and backing vector machines (SVM), prompting upgraded characterization execution.

Then again, less complex models like RF, LR, KNN, DT [7] and CNN [8]offer decent execution measurements, yet with marginally lower exactness and F1 scores contrasted with additional mind boggling designs. These models are in many cases leaned toward in situations where computational assets are restricted or where interpretability is focused on over crude prescient power. Regardless of their straightforwardness, these models give a reasonable and interpretable answer for plant illness discovery errands.

Also, the presence of deficient data in specific models, as seen in CNN, ResNet-50 [26] and RF, DT [25], highlights the significance of far reaching announcing in research studies. Without admittance to finish execution measurements, it becomes testing to figure out the capacities and constraints of these models, restricting their relevance in commonsense situations completely.

In rundown, the examination uncovers a rich scene of approaches in plant sickness location, going from complex profound learning designs to more straightforward group strategies and customary AI calculations. Each approach has its assets and shortcomings, and the choice of a proper model relies upon different factors, for example, dataset qualities, computational assets, and interpretability prerequisites. Proceeded with innovative work are fundamental to additionally refine these models and address the developing difficulties in horticultural settings.

## VI. Conclusion and Future Scope

In outline, the review has successfully represented the viability of a crossover model that consolidates CNN and Irregular Woods calculations to recognize plant sicknesses. In light of three plant species and fifteen particular sickness classifications, the model has a surprising precision pace of 99.68%, making it a solid device for overseeing rural diseases. The flexibility and straightforwardness of purpose of the proposed model are among its principal benefits. Without beginning once again without any preparation, the framework can be effectively redone to new establish species or infection classes by utilizing put away model records. This works with move learning and velocities up the adaption interaction of the model, empowering further developed execution on new datasets. Also, the audit includes the drawbacks of depending just upon benchmark datasets like Plant Village, which couldn't unequivocally address the intricacies of authentic agricultural settings. The prescribed model gives a helpful philosophy to infection recognizing verification that can be conveniently evaluated and adjusted to various datasets and division approaches by utilizing direct yet capable techniques. The conceivable usage of this exposure is exceptionally uplifting. If there is adequate picture data open for setting up, the model can be reached out to integrate a greater grouping of plant species and illnesses. Furthermore, endless headway of division techniques and computation limits could really chip away at the model's accuracy and strength considerably more. Basically, this work gives serious areas of strength for a to the production of adaptable and useful answers for overseeing rural infections. It additionally gives quick data on the capability of half and half demonstrating ways to deal with address true issues connected with crop insurance and food security.

## REFERENCES


[1.] S. Dixit and N. Kaur, "An Improved Approach To Classify Plant Disease Using CNN And Random Forest," 2023 International Conference on Artificial Intelligence and Applications (ICAIA) Alliance Technology Conference (ATCON-1), Bangalore, India, 2023, pp. 1-8, doi: 10.1109/ICAIA57370.2023.10169830.

[2.] A. Chug, A. Bhatia, A. P. Singh, and D. Singh, "A novel framework for image-based plant disease detection using hybrid deep learning approach," Soft Computing,Jun. 2022, doi: https://doi.org/10.1007/s00500-022-07177-7.

[3.] H.C. Chen, A.M. Widodo, A. Wisnujati; M. Rahaman; J.C-W Lin; L. Chen; C.E. Weng, "AlexNet Convolutional Neural Network for Disease Detection and Classification of Tomato Leaf,"

Electronics,vol. 11, no. 6, p. 951, Mar. 2022, doi: https://doi.org/10.3390/electronics11060951.
[4.] G. Ramkumar, T.M. Amrithalakshmi, R. T. Prabu, and A. Sabarivani, “An Effectual Plant Leaf Disease Detection using Deep Learning Network with IoT Strategies,” Annals of the Romanian Society for Cell Biology, pp. 8876–8885, Apr. 2021.
[5.] K. P. Shreyas and G. V. S. Narayana, "Bell Pepper Crop Disease Detection Using Convolution Neural Network," 2022 Second International Conference on Computer Science, Engineering and Applications (ICCSEA), Gunupur, India, 2022, pp. 1-6, doi: https://doi.org/10.1109/ICCSEA54677.2022.9936171.
[6.] M.I. Pavel, R.I. Rumi, F. Fairooz, S. Jahan, and M. A. Hossain, “Deep Residual Learning Approach for Plant Disease Recognition,” EAI/Springer Innovations in Communication and Computing, pp. 511– 521, Jan. 2020, doi: https://doi.org/10.1007/978-3-030-49795-8_49.
[7.] V. Tiwari, R. C. Joshi, and M. K. Dutta, “Dense convolutional neural networks based multiclass plant disease detection and classification using leaf images,” Ecological Informatics, p. 101289, Mar. 2021, doi: https://doi.org/10.1016/j.ecoinf.2021.101289.
[8.] M. A. Iqbal and K. H. Talukder, "Detection of Potato Disease Using Image Segmentation and Machine Learning," 2020 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET), Chennai, India, 2020, pp. 43-47, doi: https://doi.org/10.1109/WiSPNET48689.2020.9198563.
[9.] R. R. Ravichandran, N. Yusof, & N.K. Ibrahim, “Detection using Convolution Neural Network for Plant Disease Identification in Tomato”, Proceedings of the 3rd International Conference on Intelligent and Interactive Computing 2021 (UTeM Press), pp. 14-18.
[10.] N. Venkatakrishnan and M. Natarajan, “Comparative Study of Various Machine Learning Algorithms with MDLDPTS for Plant Leaf Disease Analysis,” Lecture notes in networks and systems, pp. 543–561, Jan. 2023, doi: https://doi.org/10.1007/978-981-19-7402-1_39.
[11.] S. M. Hassan, A. K. Maji, M. Jasiński, Z. Leonowicz, and E. Jasińska, “Identification of Plant-Leaf Diseases Using CNN and Transfer-Learning Approach,” Electronics, vol. 10, no. 12, p. 1388, Jun. 2021, doi: https://doi.org/10.3390/electronics10121388.
[12.] T. H. H. Aldhyani, H. Alkahtani, R. J. Eunice and D. J. Hemanth, "Leaf Pathology Detection in Potato and Pepper Bell Plant using Convolutional Neural Networks," 2022 7th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2022, pp.1289-1294, doi: https://doi.org/10.1109/ICCES54183.2022.9835735.
[13.] S. Jadhav and A. M. Lal, “Multi-Class Plant Leaf Disease Detection Using a Deep Convolutional Neural Network,” International Journal of Information System Modeling and Design, vol. 13, no. 1, pp. 1–14, Jan. 2022, doi: https://doi.org/10.4018/ijismd.315126.
[14.] A. S. Zamani, L. Anand, K.P. Rane, P. Prabhu, A.M. Buttar, H. Pallathadka, A. Raghuvanshi and B.N. Dugbakie, “Performance of Machine Learning and Image Processing in Plant Leaf Disease Detection,” Journal of Food Quality, vol. 2022, pp. 1–7, Apr. 2022, doi: https://doi.org/10.1155/2022/1598796.
[15.] L. Li, S. Zhang and B. Wang, "Plant Disease Detection and Classification by Deep Learning—A Review," in IEEE Access, vol. 9,pp.5668356698,2021,doi:17.https://doi.org/10.1109/ACCESS.2021.3069646.
[16.] H. Wang, S. Shang, D. Wang, X. He, K. Feng, and H. Zhu, “Plant Disease Detection and Classification Method Based on the Optimized Lightweight YOLOv5 Model,” Agriculture, vol. 12, no. 7,p.931, Jul. 2022, doi: https://doi.org/10.3390/agriculture12070931.
[17.] V. K. Vishnoi, K. Kumar, and B. Kumar, “Plant disease detection using computational intelligence and image processing,” Journal of Plant Diseases and Protection, J Plant Dis Prot 128, 19–53, 2021. doi: https://doi.org/10.1007/s41348-020-00368-0.
[18.] J. A. Pandian, V. D. Kumar, O. Geman, M. Hnatiuc, M. Arif, and K. Kanchanadevi, “Plant Disease Detection Using Deep Convolutional Neural Network,” Applied Sciences, vol. 12, no. 14, p. 6982, Jul. 2022, doi: https://doi.org/10.3390/app12146982.
[19.] P. Bedi and P. Gole, “Plant disease detection using hybrid model based on convolutional autoencoder and convolutional neural network,” Artificial Intelligence in Agriculture, vol. 5, pp. 90–101, 2021, doi: https://doi.org/10.1016/j.aiia.2021.05.002.
[20.] S. M. Hassan, M. Jasinski, Z. Leonowicz, E. Jasinska, and A. K. Maji, “Plant Disease Identification Using Shallow Convolutional Neural Network,” Agronomy, vol. 11, no. 12, p. 2388, Nov. 2021, doi: https://doi.org/10.3390/agronomy11122388.
[21.] G. Sachdeva, P. Singh, and P. Kaur, “Plant leaf disease classification using deep Convolutional neural network with Bayesian learning,” Materials Today: Proceedings, vol. 45, no. 6, p. 5584-5590, 2021, doi: https://doi.org/10.1016/j.matpr.2021.02.312.
[22.] S. S. Harakannanavar, J. M. Rudagi, V. I. Puranikmath, A. Siddiqua, and R. Pramodhini, “Plant Leaf Disease Detection using Computer Vision and Machine Learning Algorithms,” Global Transitions Proceedings, vol. 3, no. 1, p. 305-310, 2022, doi: https://doi.org/10.1016/j.gltp.2022.03.016.
[23.] B. K. Hatuwal, A. Shakya, & B. Joshi, “Plant Leaf Disease Recognition Using Random Forest, KNN, SVM and CNN”. Vol. 62, pp. 13-19, Polibits, 2020.
[24.] D. Shah, V. Trivedi, V. Sheth, A. Shah, and U. Chauhan, “ResTS: Residual Deep interpretable architecture for plant disease detection,” Information Processing in Agriculture, vol. 9, no. 2, pp. 212-223, 2021, doi: https://doi.org/10.1016/j.inpa.2021.06.001.
[25.] T. Kavitha, S. Deepika, K. Nattaraj, P. Shanthini and M. Puranaraja, "Smart System for Crop and Diseases Prediction using Random Forest and Resnet Architecture," 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2022, pp.1513-1519, doi: https://doi.org/10.1109/ICSCDS53736.2022.9760813.
[26.] J. Basavaiah and A. Arlene Anthony, “Tomato Leaf Disease Classification using Multiple Feature Extraction Techniques,” Wireless Personal Communications, vol. 115, pp. 633-651, 2020, doi: https://doi.org/10.1007/s11277-020-07590-x.
[27.] P. Tm, A. Pranathi, K. SaiAshritha, N. B. Chittaragi and S. G. Koolagudi, "Tomato Leaf Disease Detection Using Convolutional Neural Networks," 2018 Eleventh International Conference on Contemporary Computing (IC3), Noida, India, 2018, pp. 1-5, doi: https://doi.org/10.1109/IC3.2018.8530532.
[28.] J. N. Reddy, K. Vinod, and A. S. Remya Ajai, “Analysis of Classification Algorithms for Plant Leaf Disease Detection,” IEEE Xplore, Feb. 01, 2019.
[29.] M. B. Ahmad Supian, H. Madzin and E. Albahari, "Plant Disease Detection and Classification Using Image Processing Techniques: a review," 2019 2nd International Conference on Applied Engineering (ICAE),2019,pp.14,doi:https://doi.org/10.1109/ICAE47758.2019.9221712.
[30.] M.H. Saleem, J. Potgieter, and K.M. Arif, “Plant Disease Detection and Classification by Deep Learning,” Plants, vol. 8, no. 11, p. 468, Oct. 2019, doi: https://doi.org/10.3390/plants8110468
[31.] M. Sardogan, A. Tuncer and Y. Ozen, "Plant Leaf Disease Detection and Classification Based on CNN with LVQ Algorithm," 2018 3rd International Conference on Computer Science and Engineering (UBMK), Sarajevo, Bosnia and Herzegovina, 2018, pp. 382-385, doi: https://doi.org/10.1109/UBMK.2018.8566635.
[32.] P. Vasavi, A. Punitha and T.V. Narayana Rao, "Crop leaf disease detection and classification using machine learning and deep learning," International Journal of Electrical and Computer Engineering (IJECE) Vol. 12, No. 2, April 2022, pp. 2079-2086, doi: https://doi.org/10.11591/ijece.v12i2.pp2079-2086 .
[33.] https://catalyst.earth/catalyst-system-files/help/concepts/focus_c/oa_classif_intro_rt.html