**Crop Disease Detection Using Machine Learning**

**Research Paper**

**Session 2023-24**

**Pratham Agarwal (E22CSE1308)**

**Milind Kashyap (E22CSE1317)**

**Pranshu Saini (E22CSEU1300)**



**SUBMITTED TO**

**Mr. Kapil Juneja**

**SCHOOL OF COMPUTER SCIENCE ENGINEERING AND TECHNOLOGY, BENNETT UNIVERSITY**

**GREATER NOIDA, 201310, UTTAR PRADESH, INDIA**

***TABLE OF CONTENT***

Abstract………………………………………………………………………….…….3

1. INTRODUCTION…………………………………………………………….....4
2. LITERATURE REVIEW………………………………………………………..6
3. PROBLEM STATEMENT…………………………………………………..13
4. METHODOLOGY…………………………………………………………......17
5. ALGORITHM DESCRIPTION ……………………………………….…....19
6. LIMITATIONS……………………………………………………………………21
7. Result………………………………………………………………………………23
8. CONCLUSIONS………………………………………..……………………….25
9. REFERENCES……………………………………………………………..……26***ABSTRACT***

Crop diseases represent a significant threat to food security, but because to the absence of the necessary foundation, it is still difficult to quickly identify them in many regions of the world. Impressive achievements have been observed in the field of leaf-based image categorization with the emergence of accurate approaches. In order to distinguish between healthy and diseased leaves from the generated data sets, this study uses Random Forest. The implementation stages covered in our suggested study include feature extraction, dataset construction, classifier training, and classification. To categorize the photos of damaged and healthy leaves, a collective training process using Random Forest is applied to the collected datasets of infected and healthy leaves. The Histogram of an Oriented Gradient (HOG) is used to extract characteristics from an image. All in all, employing machine learning to educate the Large, publicly accessible data sets provide us with a clear means of identifying plant diseases on a massive scale.   
  
  
Keywords: Random forest, feature extraction, training, classification, diseased and healthy leaves.

# INTRODUCTION

Farmers in rural areas might believe that it's difficult to identify the diseases that could affect their crops. Going to the agricultural office to find out what the infection might be is not a moderate option. Our main goal is to identify the disease that has been introduced into a plant by observing its shape using image processing and machine learning. Food insecurity arises from reduced food production caused by pests and diseases that destroy crops or portions of plants. Additionally, fewer people in many less developed nations are knowledgeable about diseases and pest management or control. Factors such as toxic infections, inadequate disease control, and abrupt climate shifts are major causes of the decline in food output. Numerous contemporary technologies have surfaced to reduce postharvest processing in order to increase productivity, strengthen agricultural sustainability, and minimize waste. Numerous laboratory-based methods, including gas chromatography, thermography, mass spectrometry, polymerase chain reaction, and hyperspectral techniques, have been used to identify diseases. These methods require a lot of time and are not very cost-effective. The use of server-based and mobile-based approaches for illness identification has increased recently. The additional benefits of these technologies, which include a high-resolution camera, high-performance processing, and several built-in accessories, lead to automatic disease recognition.   
The use of contemporary methodologies like machine learning and deep learning algorithms has been implemented to enhance the recognition rate and precision of the outcomes. Numerous studies have been conducted in the topic of machine learning techniques such as random forests, artificial neural networks, support vector machines (SVM), fuzzy logic, K-means method, convolutional neural networks, etc. for the detection and diagnosis of plant diseases. Generally speaking, random forests are learning techniques for classification, regression, and other tasks that work by building a forest of decision trees during the training phase. In contrast to decision trees, random forests handle both numerical and categorical data and get around the issue of their training data set being overfitted. One element descriptor used in PC vision and image processing for object detection is the histogram of oriented gradients, or HOG. In this case, three component descriptors are being used:

1. Hu moments

2. Haralick texture

3. Color Histogram

Hu moments are essentially utilized to determine the leaf form. The distribution of colors in an image is represented by a color histogram, while the texture of the leaves is obtained using the Haralick texture.

***2. LITERATURE REVIEW***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S No. | Topic | Methodology | Advantages | Disadvantages |
| 1. | Crop Disease Prediction Using Deep Learning Techniques. | This study highlights recent research on the use of deep learning to detect crop illness. Various deep learning methods and configurations are discussed, including modified tight dense networks, Faster RCNN, CNN configurations, and Inception-ResNet-v2.  The authors provide an overview of the techniques employed in each research, outlining the key procedures and approaches for identifying crop sickness.  The data utilized to train and evaluate the deep learning setups, as well as the methods for determining the efficacy of individual studies, are discussed in the paper. | Early and accurate detection of crop illnesses is crucial for prompt intervention and disease management, and deep learning provides this capability.  Deep learning reduces the time and costs associated with diagnosing diseases by reducing the need for human and expert inspections.  Deep learning configurations can handle large amounts of agricultural data with ease, enabling the examination of intricate indicators and connections for disease detection.  This article provides a thorough analysis of recent developments in the sector and offers insightful advice for academics and agricultural workers. | For training, deep learning requires a large amount of marked data, which might be difficult to get, particularly for novel or uncommon crop diseases.  The quality and scope of the data greatly influence how effectively deep learning performs, and this can introduce biases and limitations for diagnosing illness.  To create and train models, deep learning frequently requires a large amount of computational power and expertise, which can be challenging in locations with limited resources.  Knowing why deep learning detects an illness in a particular way might be challenging due to its complexity. |
| 2. | Deep-Learning based Framework for Accurate Prediction of Diseases of Crops in Smart Agriculture | In order to accurately detect crop diseases using computer vision and deep learning, the research study presents deep learning (dCrop). The process included a number of crucial steps:  Gathering of datasets: utilizing the 54,306 photos of healthy and diseased plants from 14 crops in the Plant Village collection, which covers 38 distinct illnesses.  Descriptions of the datasets: Images of certain creatures are included in each folder containing data, making explanation and training simple.  Models for training: Deep convolutional neural networks (CNNs) can be trained on data using adaptive learning, particularly with sophisticated training techniques like ResNet 50, ResNet 34, and AlexNet. Using photos of leaves, the model is trained to recognize and categorize pests. | High Accuracy: The suggested method predicts diseases with 99.24% accuracy, increasing farmer profitability.  Quick Prediction: The dCrop app makes it simple to identify diseases in crop photos instantaneously, providing farmers with timely information to safeguard their harvests.  Offline functionality: Farmers in rural or under connected areas can use the application because it operates without an internet connection.  Simple to use: Farmers can record and evaluate crop photographs even in the absence of an internet connection thanks to the dCrop app's user-friendly UI. | Restricted Coverage: Despite the framework's extensive coverage of agricultural diseases, some diseases or conditions might still be unaddressed by the necessary standard.  Dependency on image quality: When the illumination is good or the picture is not good, the accuracy of disease prediction may be impacted by the quality of the photographs that were taken.  Model Upkeep: To guarantee that the training model continues to be effective against evolving or newly discovered crop diseases, regular updating and maintenance are necessary. |
| 3. | Application of machine learning techniques in forecasting crop disease | In order to forecast the severity of potato late blight in Sardinia, this study employs two machine learning techniques: feedforward neural networks and support vector machine classification.  The ARPAS weather station provides meteorological data that is used to model machine learning.  Based on meteorological factors, the SimCast prediction model is also utilized to forecast the intensity of late blight.  Three categories for late blight disease were created using support vector machine classification: low, medium, and high.  explains the input, hidden layer, activation, and output of an electronic neural network. | With the use of machine learning technology, crop disease predictions may be made using data, leading to more precise and reliable agricultural management choices.  By combining machine learning with decision-making systems (DSS) like DSS LANDS, farmers may save production costs, predict crop danger, and enhance business management. This study shows that feedforward neural networks and support vector machines can accurately predict potato late blight with 96% and 98% accuracy, respectively. It also shows how effective these learning machines are at doing so. | The learning machine's performance will be impacted by an uneven distribution of classes in the dataset, particularly if some classes are significant.  It's possible that additional variables like soil, crop management, or genetics that affect disease transmission will be unaccounted for if disease prediction relies solely on climate data.  Some farmers and ranchers may not be able to utilize machine learning models due to their complexity, which will require specialist professionals to apply and analyse. |
| 4. | Crop Prediction Using Deep Learning Techniques | Accepted Research Article makes use of a variety of artificial neural networks, such as short-term (LSTM) networks, hybrid networks, convolutional neural networks, deep neural networks, and neural networks (ANNs). These methods are used to forecast crop yield in relation to variables including crop shape, soil area, and environmental conditions. Machine learning algorithms are also used in the paper for reproduction and categorization. Using extensive data sets that include details on crop development, environmental influences, soil characteristics, and other elements, these models are trained using this method. | Enhanced precision: When compared to conventional techniques, deep learning yields higher precision and offers noteworthy advantages in the field of crop forecasting.  Real-time monitoring: To enable prompt intervention, deep learning can be used to track the health and growth of crops in real-time.  Automatic decision-making: By enhancing agricultural decision-making processes, deep learning models raise productivity and efficiency. | Dependency on data: Training deep learning models requires a lot of recorded data, which can be costly and time-consuming to collect.  Complexity: Adoption of deep learning techniques in agriculture may face difficulties due to the need for data science and computing resource skills. There are finite resources in space.  Overfitting: Deep learning models have a tendency to overfit; as a result, they perform well on training data but poorly on unknown data. |
| 5. | Plant Disease Detection Using Machine LearningSeismocardiography (23) | Feature extraction: To extract features from photos, use the Histogram of Oriented Gradients (HOG) technology. There are three descriptors used: color histogram, Haralick texture, and search time.  Training classifier: Utilizing the attributes that were retrieved, train a random forest classifier.  Classification: New photos are classified as either healthy or diseased leaves using a classifier. | Accuracy identification: To diagnose illnesses and health, use HOG feature extraction technology along with random forest distribution.  Scalability: The capacity to identify plant illnesses scalable is made possible by machine learning techniques, which is crucial for large-scale agriculture.  Cost-Effectiveness: Compared to conventional testing methods, this approach is less time- and money-consuming. | Dependency on dataset quality: The diversity and quality of the dataset used for training determines the classification accuracy.  Complexity: utilizing machine learning algorithms such as Random Forests, etc. Recognize its limitations and make appropriate adjustments.  Resource-intensive: Training learning models can be computationally demanding and demand a substantial amount of resources, particularly when done on big data sets. |
| 6. | Machine Learning Classification Techniques for Plant Disease Detection | Plant disease detection is achieved through data comparison using different machine learning techniques, including support vector machine (SVM), artificial neural network (ANN), K-nearest neighbour (KNN), fuzzy classifier, and convolutional neural network (CNN).  Facts are presented, specialized applications are discussed, and all classification techniques are applied in the study of various crop diseases. | An overview of machine learning classification methods for plant disease diagnosis.  Compare and contrast various technologies so that readers can grasp the benefits and drawbacks of each.  It is understood that this technology can be used in practical settings, such as identifying agricultural illnesses. | It might not contain comprehensive information about other facets of plant diseases, such as image detecting technology or engineering tools, because its primary focus is the classification process.  The difficulties or restrictions related to applying this technology in agriculture are not covered in this article. |
| 7. | Plant Disease Detection Techniques | With an emphasis on imaging techniques, this research article explores several applications for identifying plant diseases. There are four primary stages to it:  Image capture: Takes pictures of web pages from the Internet or digital media.  Segmenting images: Simplifying image representation with k-means clustering, Otsu algorithm, and other techniques  Feature extraction involves taking segmented images and extracting features (such as color, form, and texture) by combining grey levels. spatial grayscale level dependency matrix (GLCM), color co-construction technique, and histogram-based feature extraction. | Efficiency: Compared to farmer monitoring, automatic detection is quicker and more precise.  Accuracy: The article discusses a number of techniques that yield an accuracy of 82% to 99.9%.  Robustness: Methods like Support Vector Machines (SVM) yield good results and are simple to use.  Save Time: Automated technologies have the potential to expedite the audit process when compared to traditional systems. | Complexity: Using imaging techniques to provide automated analysis can be challenging and require specialist knowledge.  Hard Environment: Accurately detecting diseases can be challenging when capturing photos in harsh settings and outdoor illumination.  Restricted application: Certain techniques might only work with particular crops or illnesses, making them unsuitable for usage in a broader setting. |
| 8. | A Model for Prediction of Paddy Crop Disease Using CNN | 1. Data gathered: - Images illustrating different crop diseases. These numbers ought to include a broad spectrum of illnesses and levels of severity.  - Has pictures of wholesome crops for comparison and accurate categorization.  2. Pre-processing: - To guarantee uniformity of pixel values and sizes, normalize and normalize gathered photos.  - Enlarge the configuration file to make it more flexible and large. To generate more training models, one might employ methods like translation, scaling, and rotation | Increasing decision-making skills: By using CNN to anticipate crops and classify diseases, one may make better decisions. Making - Early Disease Detection: The suggested approach makes it easier to identify crop diseases, allowing for prompt intervention to lower crop losses.  Boosting productivity: The approach helps avert significant production losses and hence raises overall efficiency by identifying diseases in their early stages. | input dependency: Both the quantity and quality of the input affect how well the CNN model performs. Estimates that are off can result from incomplete or biased data.  Model complexity: CNN models have the potential to be computationally demanding and resource-intensive during training and inference, which might lead to issues in certain situations.  Interpretability: CNNs show how accurate they are in classifying data, but they are frequently difficult to interpret, which makes it challenging to comprehend the model that guides the decision-making process. |
| 9. | Development of Two Robust Classifiers for Crop Disease Prediction Using Deep Learning and Weather Data | 1. Literature research: To comprehend current techniques for yield and crop disease prediction, a literature research was done.   2. Classifier Development: - CNN-CA-I and CNN-CA-W, two potent classifiers, were created for the prediction of agricultural diseases.  - CNN-CA-W uses meteorological data for the same purpose as CNN-CA-I, which processes photos to predict crop diseases.  - To increase prediction accuracy, a set of six cellular automata was created using convolutional neural networks (CNN).  3. Data Collection: 12,45,678 photos of healthy leaves and ill plants that CNN-CA-I photographed under ideal circumstances are included in the collection of public data.  - CNN-CA-W used 8,52,624 files of environmental data that were gathered by ECMWF as weather data. | 1. Better disease identification: CNN-CA-I's usage of image segmentation technology improves illness identification accuracy, which results in more precise forecasts.  2. Dual method: This work offers an integrated approach to the crop of disease prediction by building two classifiers, one based on image analysis and the other on weather data.  3. High accuracy: The plan's efficacy was demonstrated by the two categories' respective high accuracy rates of 92.6% and 90.1% for the sickness and profit predictions, respectively. | 1. Dependency on ideal conditions: The classifier's application in an agricultural setting may be limited if recorded data does not accurately reflect real-world conditions. Use by all.  2. Data Availability: The accuracy and dependability of forecasts can be impacted by the quantity and quality of data, particularly when it comes to climate change.  3. Complexity: Managing and implementing deep learning-based classification and meteorological data processing will demand resources and experience, which will hinder adoption, particularly in situations where resources are limited. |
| 10. | Crop Disease Prediction and Solutions using Machine Learning Techniques. | 1. Pre-record: Images that have been pre-recorded to enhance quality and eliminate noise. Techniques like resizing and normalization can be applied.   2. Feature Extraction: Using methods like Convolutional Neural Networks (CNN) or manual extraction approaches, remove pertinent features from the front image.   3. Model Design: Create machine learning models to forecast agricultural illnesses from extracted data, potentially based on CNN or other deep learning techniques.   4. Model training: Utilizing historical data, train the model. In order to attain high accuracy, the test model must be developed in this step. | 1. Early Disease Detection: Early disease detection is made possible by machine learning, which lowers crop losses by allowing for prompt intervention.   2. Prediction accuracy: The model achieves accuracy in crop disease prediction through the use of sophisticated techniques and extraction procedures.   3. Provide solutions: This reduces the need for manual intervention by automatically generating recommendations or solutions based on predictions from machine learning models.   4. Efficiency: By streamlining the process of predicting diseases and addressing problems, machine learning improves efficiency and cuts down on the amount of time needed to make decisions. | 1. Data dependency: Both the quantity and quality of training data affect the model's performance. Estimates that are off can result from incomplete or biased data.   2. Complexity: Small farms and places with low resources may find it particularly difficult to use and manage machine learning models, which calls for specific knowledge and resources.   3. Interpretation: It might be challenging to comprehend the fundamental ideas underpinning prediction-making since deep learning models, like CNNs, are frequently difficult to interpret. User adoption of solutions and trust may be impacted by this.   4. General difficulties: Missing data or newly discovered disease variations might cause general concerns for a learning model based on specific data when it comes to continuous modelling and adjusting to changing agriculture. |
| 11. | Integration of Deep Learning and Weather Data for Crop Disease Prediction and Management | 1. Data collection: Thorough information was gathered, along with images of crops that were both healthy and afflicted with different diseases. Weather information was also gathered on crop growth and the emergence of diseases.   2. Pre-processing: \* Enhancing, eliminating noise, and guaranteeing uniformity by pre-processing of photos and meteorological data. It is possible to employ methods like feature scaling and normalization.   3. Custom extraction: Convolutional neural networks (CNNs) for picture data and engineering techniques for meteorological data are two examples of deep learning tools that may be used to extract features from historical data. | 1. Integration: By combining deep learning with climate data, integrated methods for managing and predicting crop diseases that take into account both biotic (found in organisms) and abiotic (found in the environment) aspects can be developed.   2. Boost accuracy: To improve crop disease management techniques, employ a range of data, including photos and meteorological information, to boost crop disease accuracy and dependability.   3. Prompt response: By anticipating crop illnesses based on the environment, the planning process lowers the risk of crop loss and improves food safety by enabling prompt intervention and management techniques. | 1. Data dependency: The fusion model's effectiveness is reliant on the quantity and calibre of training data, such as weather and picture data. Inadequate or prejudiced data can result in poor management tactics and erroneous forecasts.  2. Complexity: Proficiency in deep learning, data analysis, and agricultural research is necessary for the application and maintenance of hybrid models. Adoption may be hampered by its complexity, particularly in settings with limited resources.   3. Interpretation: It might be challenging to comprehend the underlying ideas behind predictions made by deep learning models, like CNNs, since they are frequently difficult to interpret. User acceptability of management tactics and user trust may be impacted by this. |
| 12. | A Convolutional Neural Network Model for Wheat Crop Disease Prediction | 1. Data collection: Images of crops with various diseases as well as healthy crops are included in the data.   2. Pre-processing: Enhance quality, eliminate noise, and guarantee consistency prior to image collection. Techniques like resizing and normalization can be applied.   3. Feature extraction: Convolutional neural networks (CNNs), a deep learning technique tailored for image data, are used to extract pertinent features from early photos.   4. Model design: Using attributes taken from picture data, a CNN model specifically created for crop disease prediction was created. | 1. High accuracy: The model achieves high accuracy in crop disease prediction, prompt intervention, and management through the application of deep learning technology and a customized CNN model.   2. Automatic detection: CNN's architecture may facilitate quick determination of crop health clearance, minimize the need for manual inspection, and expedite the process of inspecting crops for diseases.   3. Scalability: The CNN model can be developed to fit larger datasets and can be applied to other crops; this makes it ideal for many agricultural situations. | 1. Performance Materials: The availability and calibre of training data determine how well CNN models operate. Inadequate or prejudiced data can result in poor management tactics and erroneous forecasts.   2. Model Interpretability: It might be challenging to comprehend the fundamental ideas behind prediction-making since deep learning models, like CNNs, are frequently difficult to read. This may affect users' acceptance and confidence in predictive models.   3. Requirements: Deep learning knowledge and computational capacity for training and reasoning are necessary for the implementation and management of CNN models. Adoption may be hampered by its complexity, particularly in settings with limited resources. |

# 3. PROBLEM STATEMENT:

Crop disease detection and management provide major issues for farmers in rural and impoverished areas because of:   
  
**3.1-Access to diagnostic labs and agricultural extension services is restricted:**  
Geographical Barriers: Agricultural extension services and diagnostic laboratories are usually located in urban centers, yet many farmers are located in distant areas. Farmers find it expensive and challenging to get these services because of the distance.   
Inadequate Infrastructure: Access to essential agricultural support services is further hampered by the frequently inadequate roads and transportation networks found in rural areas.   
Professional Scarcity: In rural areas, there is a dearth of extension officers and other qualified agricultural experts, which restricts farmers' access to professional advice and diagnostic services.

**3.2-Traditional laboratory-based methods for disease identification are expensive and time-consuming**:

* Cost of Laboratory Tests: Many smallholder farmers cannot afford the high equipment and material costs associated with techniques like polymerase chain reaction (PCR), gas chromatography, mass spectrometry, thermography, and hyperspectral imaging.
* Time Delays: Sending samples to far-off laboratories is a common practice in traditional diagnostic methods, which causes major delays in receiving results. In the event that this delay persists, infections may spread quickly before the proper action is done.
* Logistical Difficulties: Gathering, storing, and bringing samples to labs can be labor-intensive and logistically challenging, which exacerbates the issue.

# 3.3-Lack of knowledge and resources to implement effective pest and disease management strategies:

# Educational Gaps: Many farmers lack formal education and training in modern agricultural practices and disease management strategies. This knowledge gap makes it challenging for them to identify and manage crop diseases effectively.

# Limited Access to Information: In rural and underdeveloped regions, access to up-to-date information on pest and disease management is often limited. Farmers may not have access to the internet or other resources that provide the latest agricultural research and recommendations.

# Resource Constraints: Even when farmers have some knowledge of disease management, they may lack the necessary resources, such as appropriate pesticides, fungicides, and disease-resistant crop varieties, to implement effective strategies.

**3.4-The quick spread of pests and illnesses, which causes large crop losses and a decline in food production:**

* High Disease Prevalence: Rapid pathogen transmission can destroy crops in high-disease prevalence areas, resulting in significant losses. This problem is made worse by a lack of prompt intervention and early discovery.
* Impacts of Climate Change: New illnesses and pests have emerged and spread as a result of climate change. Crop infestation risk can rise due to the favorable conditions that unpredictable weather patterns can produce for disease propagation.
* The economic impact of crop losses resulting from pests and diseases is felt directly by farmers, who experience a reduction in income and a disruption to their way of life. In rural areas, this economic strain may result in a vicious cycle of food insecurity and poverty.
* Food Security Danger: Widespread agricultural diseases have reduced food output, posing a hazard to food security both locally and nationally. Malnutrition and other health problems may result from this in areas where agriculture is the primary industry.
* The creation of an affordable, effective, and easily available agricultural disease prediction system can help to substantially improve disease control, raise crop yields, and guarantee food security for farmers in rural and impoverished areas.
* Create a crop disease prediction system based on images that uses contemporary technology, like image processing and machine learning, to deliver precise and rapid disease detection. Farmers will find it simple to utilize the system in the field because it is made to be available via mobile devices.

**The following are the suggested system's salient features:**

1.Image Capture and Preprocessing: Using their mobile devices, farmers can take pictures of sick plants. These photos will undergo preprocessing by the system to improve quality and extract pertinent features.

2.Feature extraction: Use feature descriptors like color histograms for color distribution analysis, Hu Moments for shape analysis, and Haralick Texture for texture analysis.

3.Machine Learning Model: To categorize and forecast the kind of plant disease, apply machine learning methods such Random Forest, Support Vector Machine (SVM), and Convolutional Neural Networks (CNN).

4.  
User Interface: Give farmers an easy-to-use interface via which they can upload photos, get diagnosis findings, and get details on how to manage and control diseases.

# 4.Methodology

There are measures that must be taken in order to determine whether the leaf is infected or healthy. For example, feature extraction, preprocessing, classifier training, and classification Reducing each image's size to a consistent, smaller size is known as preprocessing. Next, with the aid of HOG, features from a preprocessed image are extracted. A feature descriptor called HoG [6] is employed in object detection. This feature descriptor uses gradients of intensity to describe how an object looks and how the image is outlined. The fact that HoG feature extraction works with the generated cells is one of its benefits. This remains unaffected by any alterations. Three feature descriptors were used in this instance.

Hu moments: Picture moments that contain significant details from the image pixels aid in the description of the items. Hu moments are useful in characterizing the general shape of a leaf in this context. Hu moments are only computed for one channel. Hu moments are computed after converting RGB to grayscale in the first phase. An array of shape descriptors is provided in this stage. Haralick Texture: Typically, the textures of sick and healthy leaves differ. Here, we differentiate between the textures of healthy and diseased leaves using the Haralick texture feature. The adjacency matrix, which contains the position of (I,J), serves as its foundation. Based on the frequency of the pixel I occupying the texture [7], position next to pixel J. To calculate Haralick texture it is required that the image be converted to gray scale.



Fig.1. RGB to Gray scale conversion of a leaf.

Color Histogram: A color histogram provides an image's color representation. The histogram is computed for RGB after it has been transformed to HSV color space. The RGB image must be converted to HSV since the latter closely resembles how the human eye perceives color in an image. The description of the number of pixels available in the specified color ranges is given by the histogram plot [8].

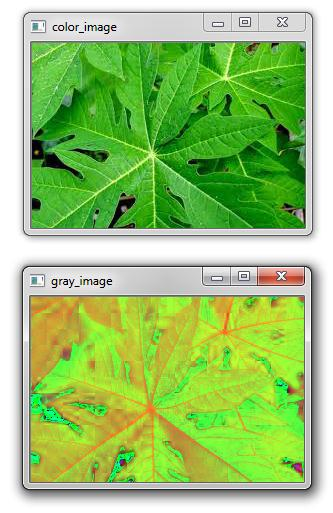


Fig.2. RGB to HSV conversion of leaf

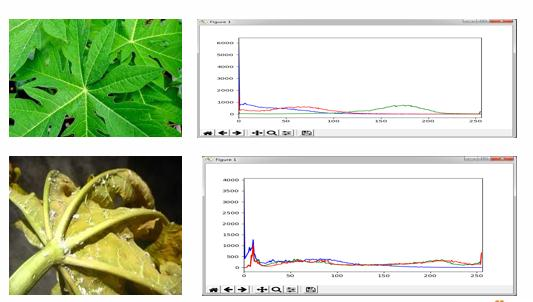


Fig.3. Histogram plot for healthy and diseased leaf.

***6. ALGORITHM DESCRIPTION***

Here, the random forests classifier is used to implement the algorithm. They are adaptable and suitable for both regression and classification methods. Random forests provided higher accuracy with less picture data sets than other machine learning approaches including SVM, Gaussian Naïve Bayes, logistic regression, and linear discriminant analysis. The architecture of our suggested method is depicted in the following image.

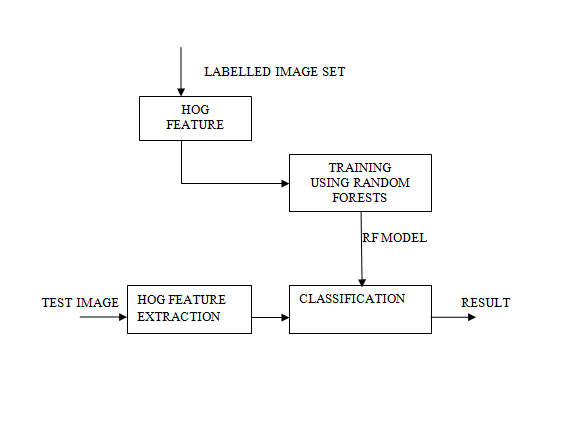


Fig.4. Architecture of the proposed model

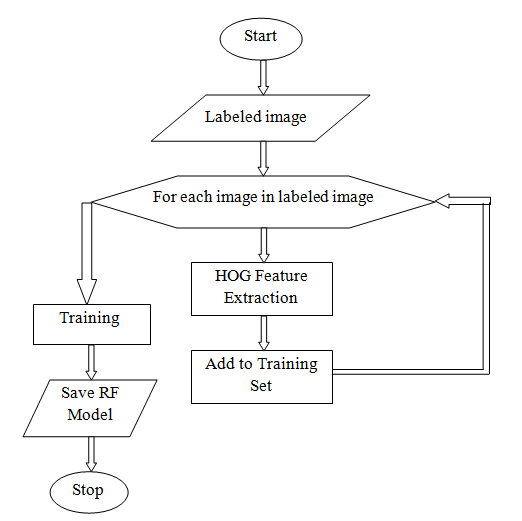


Fig.5. Flow chart for training.

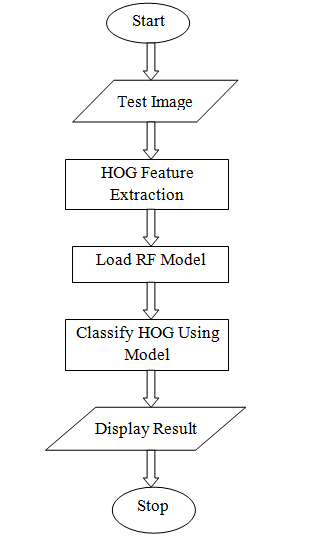


Fig.6. Flow chart for classification

Training and testing data are separated from the labeled datasets. HoG feature extraction is used to create the feature vector for the training dataset. The Random Forest classifier is used to train the resulting feature vector. Additionally, the trained classifier receives the feature vector for the testing data produced by HoG feature extraction for prediction, as shown in "Fig.4". Labeled training datasets are transformed into the corresponding feature vectors via HoG feature extraction, as illustrated in "Fig. 5." The training datasets contain the extracted feature vectors that have been preserved. Additionally, the Random Forest classifier is used to train the trained feature vectors [9, 10]. HoG feature extraction is used to extract the feature vectors for the test image, as shown in "Fig. 6." These produced vectors of features are provided

the saved and trained classifier for predicting the results.

# 7. Limitations

The Crop Disease Prediction System's Limitations

7.1-Image Consistency and Quality:   
  
Dependency on Image Quality: The quality of the photographs that the farmers take is a major factor in how accurate the disease prediction system is. The system's capacity to accurately diagnose illnesses can be seriously impacted by poor lighting, fuzzy images, or uncorrected viewing angles.   
Device Variability: differing mobile device kinds with differing camera resolutions and capabilities may be used by farmers, resulting in unpredictable image quality and influencing the diagnosis's trustworthiness.

7.2-Limited Disease Database: ⎫ Database Coverage: The extent to which the disease database it uses is extensive determines how accurate the system can be. The method might not correctly detect some diseases or variations of diseases if the database does not contain them.   
Regional Variability: Illnesses unique to particular areas or weather patterns might not be adequately reflected in the database, which could result in incorrect diagnoses or illnesses going unnoticed.   
7.3-Environmental Factors: Symptom Variability The system's ability to correctly diagnose illnesses based only on morphology is challenged by environmental factors that might affect the presentation of disease signs, including weather, soil type, and crop management practices.   
Similar Symptoms: A variety of illnesses might exhibit comparable symptoms, which could cause misunderstanding and incorrect diagnosis. The system could have trouble distinguishing between various illnesses with overlapping symptom profiles.

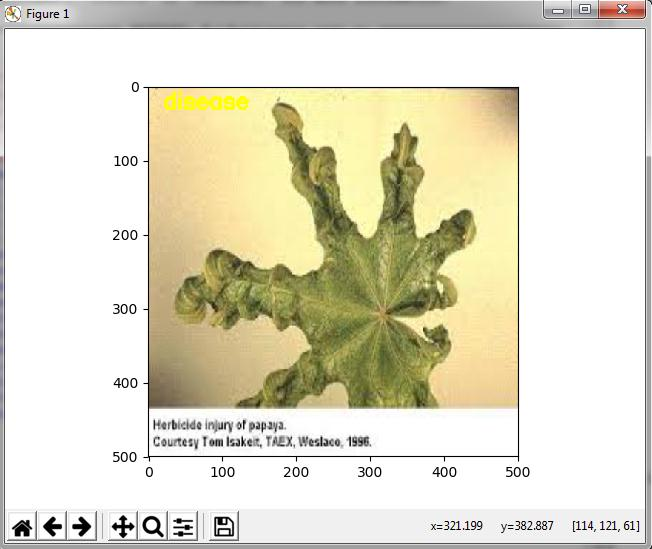
7.4-Limitations of Machine Learning Models: ⎫ Training Data Bias: The caliber and breadth of the training data influence how well the machine learning models operate. The model's predictions could be less accurate if the training data is biased or does not adequately represent certain medical problems.   
  
Interpretability of the Model: Complex machine learning models, like deep learning networks, can be challenging to understand. It can be difficult to understand why the model arrived to a specific diagnosis, which could reduce user confidence and system usefulness.   
7.5-Restrictions on Resources:   
  
Computational Requirements: Complex image processing and machine learning algorithms can be resource-intensive to run on mobile devices, which may cause problems for devices with lower end hardware.   
Internet connectivity: Access to reliable internet connectivity may be restricted in rural and impoverished areas. As a result, the system's ability to upload images for processing on cloud-based servers or to access updated disease databases.

7.6: User Proficiency and Accessibility: technical Literacy: The degree of technical proficiency among farmers may differ. The farmers' ability to utilize the mobile application correctly, take excellent pictures, and analyze the data will determine how beneficial the system is.   
  
Language Barriers: In order to make the system interface and illness information accessible to all users, they must be provided in many languages and dialects. Linguistic obstacles may prevent the system from being widely used and from working as intended.   
7.7-Expense and Upkeep:   
Even though the system is meant to be economical, there can be upfront expenses related to getting internet access and suitable mobile devices.   
Updating and Maintenance: To keep the system accurate and functional, regular maintenance and upgrades are needed for the mobile application, machine learning models, and illness database.This ongoing requirement may pose challenges in terms of funding and resource allocation.

7.8-Privacy and Ethics Issues:  
  
  
Data privacy: Gathering and analyzing crop image data may cause farmers to worry about their privacy. Gaining the trust of farmers requires making sure that data is handled securely and with consideration for user privacy.   
Bias & Fairness: To avoid any unforeseen negative outcomes, it is crucial to ensure that the system is impartial and fair in its forecasts across various areas, crop types, and socio-economic factors.The crop disease prediction system can be enhanced to give farmers in rural and impoverished areas more dependable and easily available solutions by identifying and resolving these constraints.

**8. The outcome**

Every image must first be converted from RGB to grayscale. This is only done because Haralick features and the Hu moments shape descriptor can only be computed over a single channel. Therefore, before calculating Hu moments and Haralick features, RGB data must be converted to grayscale. as illustrated in picture 4. In order to calculate the histogram, the image must first be transformed to HSV (hue, saturation, and value). Figure 5 illustrates this process of transforming an RGB image to an HSV image. Lastly, the primary goal of our project is to use a Random Forest classifier, as shown in "Fig.7," to determine if a leaf is infected or healthy.



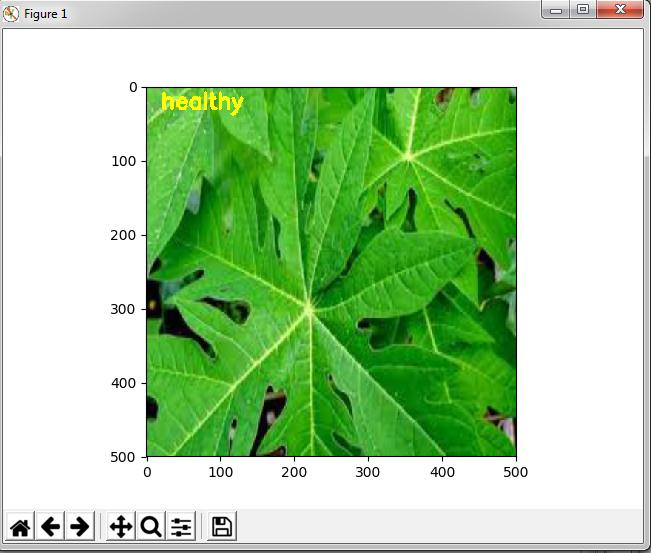


Fig.7. Final output of the classifier.

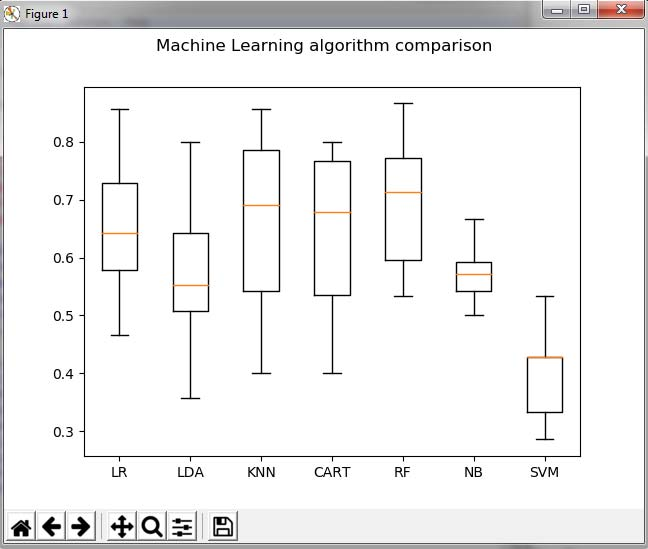


Fig.8. Comparison between different machine learning models.

| **Machine Learning Model** | **Accuracy (percent)** |
| --- | --- |
| Logistic Regression | 65.33 |
| Support Vector Machine | 40.33 |
| k-Nearest Neighbor | 66.76 |
| CART | 64.66 |
| Random Forests | 70.14 |
| Naïve Bayes | 57.61 |

Fig .9. Table showing the comparison.

***9.conclusion***

This program aims to identify anomalies that arise on plants in their natural habitat or in greenhouses. To avoid occlusion, the picture is typically taken against a plain background. The accuracy of the algorithm was compared with different machine learning models. 160 photos of papaya leaves were used to train the model using the Random Forest classifier. The model's classification accuracy was about 70%. Increased training volume and the use of additional local features in addition to the global features—such as SURF (Speed Up Robust Features), DENSE, and SIFT (Scale Invariant Feature Transform)—along with BOVW (Bag Of Visual Word) can both improve accuracy.

# 10.Reference

[1] S. S. Sannakki and V. S. Rajpurohit,” Classification of Pomegranate

Diseases Based on Back Propagation Neural Network,” International

Research Journal of Engineering and Technology (IRJET), Vol2 Issue:

02 | May-2015

[2] P. R. Rothe and R. V. Kshirsagar,” Cotton Leaf Disease Identification

using Pattern Recognition Techniques”, International Conference on

Pervasive Computing (ICPC),2015.

[3] Aakanksha Rastogi, Ritika Arora and Shanu Sharma,” Leaf Disease

Detection and Grading using Computer Vision Technology &Fuzzy

Logic” 2nd International Conference on Signal Processing and

Integrated Networks (SPIN)2015.

[4] Godliver Owomugisha, John A. Quinn, Ernest Mwebaze and James

Lwasa,” Automated Vision-Based Diagnosis of Banana Bacterial Wilt

Disease and Black Sigatoka Disease “, Preceding of the 1’st

international conference on the use of mobile ICT in Africa ,2014.

[5] uan Tian, Chunjiang Zhao, Shenglian Lu and Xinyu Guo,” SVM-based

Multiple Classifier System for Recognition of Wheat Leaf Diseases,”

Proceedings of 2010 Conference on Dependable Computing

(CDC’2010), November 20-22, 2010.

[6] S. Yun, W. Xianfeng, Z. Shanwen, and Z. Chuanlei, “Pnn based crop

disease recognition with leaf image features and meteorological data,”

International Journal of Agricultural and Biological Engineering, vol. 8,

no. 4, p. 60, 2015.

[7] J. G. A. Barbedo, “Digital image processing techniques for detecting,

quantifying and classifying plant diseases,” Springer Plus, vol. 2,

no.660, pp. 1–12, 2013.

[8] Caglayan, A., Guclu, O., & Can, A. B. (2013, September).

“A plant recognition approach using

shape and color

features in leaf images.” In International Conference on

Image Analysis and Processing (pp. 161-170). Springer,

Berlin, Heidelberg.