JOURNAL OF BASIC SCIENCE AND ENGINEERING

Prediction of Arecanut Disease Severity using Hyectral Imaging - Study in Channagiri Region

Kushalatha.M.R.[1], Dr. Prasantha.H.S [2]

1 Research Scholar, Visveswaraya Technological University, Assistant Professor, Department of Electronics and Communication Engineering, Nitte Meenakshi Institute of Technology, Bangalore, -560064, Karnataka, India

2 Professor, Department of Computer Science, Atria University, Bangalore- 560024, Karnataka, India

Keywords:

Hyperspectral Imaging
Deep learning
Machine Learning Algorithms
Disease severity determination

ABSTRACT

The Disease known as 'Hidimundige/ Crown -Choke Disease seen in Arecanut Farm has been spreading across the areas of Channagiri, Davanagere district of Karnataka, India over the past 5-6 years. The disease has caused a decrease in the yield of the area which is related to the downpour of the economic growth of the country as Arecanut is a cash crop. Karnataka's Channagiri district is "The Arecanut Hub" of India and is responsible for uplifting the economy and the export business of India. The paper focuses on the study of the Disease Severity determination of this area. Comparisons of the Disease Severity for the dataset collected in the month of February 2023 is calculated with that of the dataset collected during the year 2015-16. The comparison was again subdivided based on the age groups of the plants. The age group of below 6 months, less than 5 years, between 5 – 7 Years, above 7 and within 10 years, above 10 years to 15 years and 25 years were taken into consideration for the study. The features considered are spectral reflectance and vegetation indices. The classification models were trained with the samples taken where training and testing ratio were kept in the range of 80:20. Variation in the spectral reflectance was mainly seen in the 500 nm to 2000 nm range. Various classification algorithms were evaluated on the samples collected to get the better accuracy of classification. LightGBM (Light Gradient Boosting Machine) showed better classification with respect to the disease severity in terms of all classification parameters. The data taken of the site are Hyperspectral images of leaf samples from different sites.

Corresponding Author:

Kushalatha.M.R.
Department of Electronics and Communication Engineering, Nitte Meenakshi Institute of Technology, Bangalore – India.

1. INTRODUCTION

Cash crops grown in different parts of the world have been contributing to the economy of any county. Number of cash crops have been playing vital role in determining the economy status of the country. There are large number of cash crops that one can think of in this regard. Arecanut is one of the cash crops grown in many parts of the world. The contribution of this Arecanut crop in India is vast, many of the places in India the Arecanut crop is seen. Extensive usage of the crop is done by large set of people in India and considered to be linked to Indian religious practices.

India is one of the largest producers and consumers of Arecanut. Major states contributing to the large production are Karnataka, Kerala, Assam and Tamil Nadu, Meghalaya, and West Bengal. Among the Arecanut producing states of India, Karnataka's share is around 78%. Channagiri is a small village in Davanagere district of Karnataka is famously known as "Arecanut Hub". Davanagere is the district with Arecanut grown on 36,000 hectares - stood sixth in Arecanut production after Dakshina Kannada, Uttara Kannada, Shivmoga, Hassan / Chikmagalur and Kodagu of Karnataka state.

Ninety percent of the area covered by Arecanut crops is Channagiri. There are variations of this planation observed. The crop yields a profit that meets the farmer's daily necessities. Currently, however, several illnesses have harmed crops, which has impacted crop productivity. Preventing plant diseases is dependent on the assessment of the disease severity. Estimating different elements will contribute to the continued spread of these diseases. Farmers will be able to comprehend the fundamental cause of the diseases' spread with the aid of prompt therapies.

In agricultural Hyperspectral image analysis, remote sensing is clearly important. The goal of the study is to determine the index of diseases brought on by distinct factors. Twenty samples that were gathered from diverse locations are used in the study. The age groups of the samples vary.

Hyperspectral imaging is a powerful technique that can be utilized in the detection and monitoring of plant diseases, including those affecting Arecanut plants. It involves capturing images of an object or scene at numerous and contiguous spectral bands, providing detailed spectral information for each pixel in the image.

Hyperspectral imaging and deep learning algorithms can be used to monitor the health of Arecanut plants and identify and evaluate infections, dietary deficits, and other stressors. Here's how to use this combo strategy:

(i) Hyperspectral Imaging: Using a broad range of wavelengths, hyperspectral imaging provides precise spectral information on Arecanut plants. It is possible to discover disease signs or nutritional imbalances as well as minor changes in plant health by analysing the reflectance or absorption patterns in the hyperspectral data. Rich sources of information regarding the physiological and metabolic state of plants can be found in hyperspectral data.

(ii) Data Acquisition and Preprocessing: Specialized sensors or cameras that record the reflected light from the plants at many narrow spectral bands are used to obtain hyperspectral images of Arecanut plants. To ensure high-quality input for further analysis,

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

the acquired data is pre-processed to eliminate noise, adjust for atmospheric impacts, and improve the spectrum information.

(iii) Training Data Collection: Hyperspectral photos of Arecanut plants are combined with ground truth data about their health to create a labelled dataset. This dataset is essential for teaching deep learning algorithms to reliably identify and categorize various medical disorders.

(iv) Feature Extraction: Relevant characteristics can be automatically extracted from the Hyperspectral data using deep learning methods like convolutional neural networks (CNNs). These algorithms are trained to identify spectral signatures and patterns linked to various Arecanut plant health concerns. The labelled dataset is used to train CNNs so they can acquire discriminative features for additional health indicators and disease identification.

(v) Model Training and Validation: Using the hyperspectral data and associated health labels, the deep learning model is trained on the labelled dataset. The algorithm gains the ability to categorize Arecanut plants as healthy, diseased, or nutrient-deficient. To evaluate the trained model's accuracy and capacity for generalization, it is validated using an independent dataset.

(vi) Health Monitoring and Disease Detection: Hyperspectral photos of Arecanut plants can be used to build and verify a deep learning model that will be used to monitor health and identify diseases in the plants. To provide real-time information about the health status of plants and the presence of illnesses or other stress factors, the model can identify new, unseen photos.

This method allows for the precise and automated identification of diseases affecting Arecanut plants as well as other health problems by fusing Hyperspectral imagery with deep learning algorithms. It makes early diagnosis easier, which enables prompt intervention and management strategies to be put in place, reducing the negative effects of illnesses on plant yield.

2. LITERATURE SURVEY

In the paper [1] indicated how plant disease identification has helped in the quick development of Hperspectral imaging technologies and unmanned aerial vehicles (UAVs). Conventional manual inspection methods have superseded by UAV-borne Hyperspectral remote sensing (HRS) systems with high spectral, spatial, and temporal resolutions. This is because the technologies enable more precise and economical crop assessments and plant characteristics. The purpose of this work is to present a summary of the research on deep learning algorithms-based HRS for illness diagnosis. This study introduced the fundamentals of deep learning-based classifiers, Hyperspectral imaging, and aerial surveying with UAVs. As to investigate the viability of carrying out such study, generalizations on workflow and methodologies were established from prior research. Deep learning models outperform conventional machine learning algorithms in terms of accuracy, according to research findings. Lastly, some other difficulties and restrictions pertaining to this subject are discussed.

Data Acquisition: Flight planning and the construction of a Hyperspectral system based on UAVs are involved in data acquisition. Since most Hyperspectral camera systems require a microprocessor to function, UAV platforms must be able to transport payloads steadily and safely.

Ground Survey: After the experimental site has been determined, a ground survey is still another essential step. Validation and matching between the diseased plant on the obtained photos and the diseased plant on the actual site are required. Using the sample's GPS location to mark its position is another workable solution. By analysing the sample morphology, the infection phases can be ascertained. The contaminated plants were identified by a ground survey, which also used to create a ground-truth dataset. Based on factors such as resin secretion, growth vigor, and needle colour, the PWD infection was classified into multiple stages.

Pre-processing: ArcGIS, Agisoft Metashape, PIX4Dmapper or ENVI are some examples of software that can be used for georeferencing in Pre-Processing stage.

Feature Extraction: Various colours may represent the sickness. Beyond the visible light bands, Hyperspectral scans show additional variations. Furthermore, spectral data can be used as input variables in a classifier model that is developed to identify sick areas. The model can discriminate between samples that are healthy and samples that are sick because it has a unique spectral signature.

Classification: By designating each band as a variable, a classifier model based on these differences can be constructed, and classifications can be made by classifying individual pixels at a time.

Band-Reduction Techniques: Decrease collinearity between variables and improve model performance by applying dimensionality reduction techniques like PCA, successive projection algorithm (SPA), stepwise discriminant analysis (SDA), linear discriminant analysis (LDA), minimum noise fraction (MNF) algorithm, and partial least squares discriminant analysis (PLS-DA).

Disease Detection Techniques: After the images have been divided, the separated portions can be sent into a classifier to identify whether the disease is present. The divided portions can then be merged into a single image with the diseased area highlighted. Depending on the study object, aerial photos are segmented differently and can be adjusted using a classifier to achieve optimal performance. A number of studies have used different methods, such as a segmentation procedure, to build the dataset by dividing the photos into 11 × 11, 13 × 13, 32 × 32, 64 × 64, 128 × 128, 224 × 224, 256 × 256, and 800 × 800 patches.

The research publication [2] showed how deep learning-based diagnosis of plant diseases is a perfect fit due to the extensive and sometimes redundant information contained in Hyperspectral data cubes. In this instance, scientists used a brand-new 3D deep convolutional neural network (DCNN) that absorbs the Hyperspectral data straight away. They also questioned the learned model to generate answers that made sense biologically. They concentrated on charcoal rot, a soil-borne fungal disease that has a significant economic impact and lowers soybean crop yields globally. Outcomes Their 3D DCNN exhibits an infected class F1 score of 0.87 and a classification accuracy of 95.73% based on hyperspectral imaging of inoculated and mock-inoculated stem images. They demonstrated how the model uses the geographical regions with obvious clinical symptoms for classification by visualizing the most sensitive pixel positions using the idea of a saliency map. The near infrared area (NIR), which is also a frequently used spectral range for assessing a plant's vegetative health, is where the researchers discovered that the model's most sensitive wavelengths for classification lie. The study concluded that the application of an explainable deep learning model yields high accuracy as well as physiological insight into model predictions, hence enhancing model prediction confidence. These clarified forecasts are amenable to ultimate application in automated phenotyping tools for research and precision agriculture.

A survey of the neural network algorithms currently in use for processing image data, with a focus on crop disease detection, conducted in the research paper [3]. An examination of the various image processing methods, deep learning models and architectures, and data acquisition sources used to manage the provided imaging data comes first. The study also emphasized the outcomes of evaluating several deep learning models that are already in use, and it concluded by discussing the potential applications of hyperspectral data analysis in the future. The purpose of this survey's preparation is to enable future study to discover deeper capabilities of deep learning while enhancing the accuracy and performance of the system's plant disease detection.

The survey conducted in [4] aimed to identify plant diseases using handmade features derived from deep learning-based models. Researchers have shown that deep learning-based techniques can reach appreciable accuracy rates on a given dataset; but, when the system is tested on different datasets or in field imaging conditions, the model's performance may suffer noticeably. Deep learning models with an inception layer, like GoogleNet and InceptionV3, are better at extracting features and yielding better performance outcomes than other models. They also discuss the difficulties that must be overcome to correctly identify plant diseases. Online resources include standard datasets on plant diseases, including those on Rice disease, Cotton disease, Hops disease, Cassava disease, and Plant Village disease. Thirteen different classes representing fourteen different plant species (fruits and vegetables) make up the Plant Village dataset. Five distinct types of diseases with nonuniform background circumstances make up the Hops dataset. Among the diseases are nutritional, pest, downy, powdery, and healthy diseases. Both healthy and sick cotton plants and leaves make up the Cotton dataset. Four distinct disease types that were addressed in the field are included in the rice disease dataset. The diseases brown spot, blast, bacterial blight, and tungro are included in the rice disease dataset.

A systematic approach to the investigation of different plant disease categorization models is described in the paper [5]. Researchers have conducted a systematic literature study in this paper on the applications of the most popular machine learning (ML) and deep learning (DL) algorithms for plant disease categorization. These algorithms include AlexNet, GoogLeNet, VGGNet, Support Vector Machine (SVM), Neural Network (NN), K-Nearest Neighbour (KNN), Naïve Bayes (NB), and other popular DL algorithms. The corresponding processing techniques for each algorithm—such as image segmentation and feature extraction—as well as the standard experimental setup metrics—such as the total number of training and testing datasets used, the number of diseases taken into account, the type of classifier used, and the percentage of classification accuracy—are used to characterize each algorithm. Researchers will find this study to be a useful resource for identifying plant diseases using data-driven methods.

Identify and categorize grapevines inoculated with the recently identified DNA virus grapevine vein-clearing virus (GVCV) during the early asymptomatic phases, the study [6] used hyperspectral photography at the plant level. Two grapevine groups—healthy and GVCV-infected—were used in an experiment at the South Farm Research Center test site in Columbia, Missouri, USA (38.92 N, 92.28 W), with other variables kept under control. An Oulu, Finland-based SPECIM IQ 400–1000 nm Hyperspectral sensor was used to take pictures of every vine. Only grapevine pixels were kept after pre-processing and calibration of hyperspectral photos. Distinguish between two reflectance spectrum patterns in healthy and GVCV vines, a statistical method was utilized. The researchers computed and investigated the significance of disease-centric vegetation indices (VIs) for the classification power. Within a framework including deep learning architectures and conventional machine learning, pixel-wise (spectral features) and image-wise (joint spatial–spectral features) classification were conducted concurrently. The findings demonstrated that: (1) the most discriminative indices were the following: (2) the normalized pheophytization index (NPQI), fluorescence ratio index 1 (FRI1), plant senescence reflectance index (PSRI), anthocyanin index (AntGitelson), and water stress and canopy temperature (WSCT) measures; (3) the support vector machine (SVM) was

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

effective in VI-wise classification with smaller feature spaces, while the RF classifier performed better in pixel-wise and image-wise classification with larger feature spaces; and (4) the automated 3D convolutional neural network (3D-CNN) feature extractor provided. When VI-based and pixel-based classification techniques were compared, both produced results that were similar in terms of classification success. In particular, the pixel-based model's 5-fold cross-validation accuracies varied from 85.10% to 95.30% and from 82.13% to 96.75%. According to their findings, the support vector machine (SVM) performed better in the VI-wise categorization when given little feature data. When employing pixels for classification, it did, in fact, perform as well as the random forest (RF) classifier for the simplest 2-feature model. However, the RF classifier proved to be the most effective classifier for reflectance data using Kernel-PCA and PCA, which covered a wider feature space.

According to research published in article [7], Hyperspectral imaging technology has made significant advancements in the gathering of agricultural data and the identification of the agricultural product's internal or external qualitative qualities. The analysis of hyperspectral images is now more performant thanks to deep learning algorithms. Deep learning architectures leverage both spatial and spectral information from Hyperspectral picture analysis, in contrast to typical machine learning methods. This study gives a systematic and complete evaluation of the current efforts in deep learning for Hyperspectral image analysis in agriculture, offering insights and suggesting future research possibilities. First, a summary of its uses in agriculture is given, which include the detection of plant diseases, the prediction of maturity and component content, and many categorization themes. Subsequently, the latest developments in hyperspectral image processing are examined from the perspectives of feature networks and deep learning models. Lastly, a summary of the current issues with deep learning-based hyperspectral image analysis is provided, along with an outlook for future research.

Establishing a mechanism for the presymptomatic identification of tobacco disease based on HSI was the primary goal of this investigation [8]. This ultimate aim was accomplished by fulfilling the subsequent particular goals: The process involves: (i) identifying the corresponding effective wavelengths (EWs) that exhibit the highest correlation between the spectral data and various disease stages; (ii) extracting texture features at the selected EWs using the grey-level co-occurrence matrix (GLCM); (iii) developing and comparing machine-learning models with spectral data, texture features, and data fusion, respectively, to quantitatively identify the tobacco disease; (iv) differentiating tobacco leaves infected with TMV from those that are not infected, and classifying three levels of disease degree during the infected period even before specific symptoms emerged. Random forests (RF), extreme learning machines (ELM), back propagation neural networks (BPNN), support vector machines (SVM), partial least squares-discrimination analysis (PLS-DA), and least squares support vector machines (LS-SVM). Provided are the created models' overall and individual (healthy, 2 DPI, 4 DPI, or 6 DPI) classification accuracy using the EWs. Most classification models did well; however, BPNN outperformed the others, with calibration and prediction classification accuracies of 95.00% and 93.33%, respectively. Overall, the results showed that using texture features to identify tobacco disease in relation to various developmental disease stages was both feasible and highly promising, even though the results of texture analysis based on GLCM are lower than those obtained by spectral reflectance.

Within the document [9], studies were conducted in several plant-pathogen systems. For microscopic studies, near-isogenic barley (Hordeum vulgare) lines cv. Ingrid wild type (WT) and Pallas with molo 3 and Mla1 resistance were employed. Evaluation of spectral changes on the leaf and in cells, a Hyperspectral microscope system was used. This configuration comprises of a foreoptic (Z6 APO, Leica, Wetzlar, Germany) with a magnification of up to 7.3x, mounted on a hyperspectral line scanner (spectral camera PFD V10E, Specim, Oulu, Finland) in the visible (400 – 700 nm) and near infrared (700 – 1000 nm) ranges. Time series imaging was used to track the emergence of symptoms and the early resistance responses in barley and sugar beet. The extraction, analysis, and classification of changes in spectral reflectance were done manually and by data mining techniques, respectively. The program ENVI 5.1 + IDL 8.3 (ITT Visual Information Solutions, Boulder, USA) was used to calculate reflectance in relation to a white reference and the dark current. Smooth hyperspectral images were subjected to the Savitzky-Golay filter (Savitzky & Golay, 1964) following normalization. These pre-processed images were utilized with Matlab, Python, or ENVI 5.1 + IDL 8.3 for additional analysis. The Simplex Volume Maximization (SiVM) technique was applied, which describes the data (hyperspectral images and signatures) in terms of a small number of extreme components. Additionally, K-means clustering was used on the as an unsupervised, data-driven method. Using the spectral angle mapper (SAM) technique, healthy and diseased pixels on the canopy scale may be classified spectrally. Every pixel in the hyperspectral photos was considered using the data analysis techniques that were demonstrated. Signatures in spectrums of seven cluster means. Pseudo-colour images were used to visualize the clustering results. Each cluster was given with a specific colour. Without human assistance, typical spectral clusters of green for healthy tissue, blue for leaf veins, and yellow to red for Cercospora leaf spots could be automatically detected.

In the study that was conducted [10], Numerous vegetation indices have been constructed to understand the data, either by biological reasoning or by presumption (for example, indices derived from satellite multispectral remote sensing data may only have had a restricted number of wavelengths available). These indices are referred regarded as "vegetation indices" when they are used on plant material. There are numerous vegetation indices available, and each one describes the physiological characteristics of vegetation using a unique set of wavelength measurements, focusing on either the general characteristics of the plant or growth metrics. Finding variations in the abrupt rise in reflectance at the red/near-infrared border is another often employed strategy. The term "red edge" refers to the small region in the electromagnetic spectrum (690–740 nm) where the near infrared and visible spectrums meet. Since chlorophyll extensively absorbs wavelengths up to around 700 nm, this section's green plant material exhibits a large change in spectral response (derivative). As a result, the material has low reflectance in this range but strongly reflects the infrared (from about 720 nm). Powdery mildew in wheat (Blumeria graminis f. sp. Tritici) has been identified using a disease index based on the red edge position; however, this method was not as accurate as Partial Least Squares Regression (PLSR), which used a statistical approach. In the study using avocado plants, the fungal disease Laurel wilt (Raffaelea lauricola) was investigated using the QDA method on plants grown in both a field and a glasshouse. 94% of the QDA classifications were accurate.

3. METHODOLOGY

The samples that were taken from the site were used to identify the disease. Disease identification in Arecanut requires a methodical strategy that combines technological advancements, laboratory analysis, and field observation. This is a general approach to identifying diseases in Arecanut.

1. Field Survey and Observation: This includes routine inspection, identifying symptoms, and observing seasonal trends.

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

2. Sample Collection: Samples of soil, water, roots, leaves, and other materials may be collected.
3. Symptomatology charts: These will show a correlation between potential illnesses and their severity levels.
4. Remote Sensing Technology: This method determines the location and severity of a disease by using Hyperspectral or drone imagery.
A high-level summary of the steps involved in applying hyperspectral imaging for disease detection is given in Figure 1. The precise Hyperspectral imaging technology, the features of the target crop (Arecanut), and the machine learning techniques selected for categorization may all influence the actual implementation.The following block diagram in Figure 1 represents the overall methodology adopted for the research work carried out.

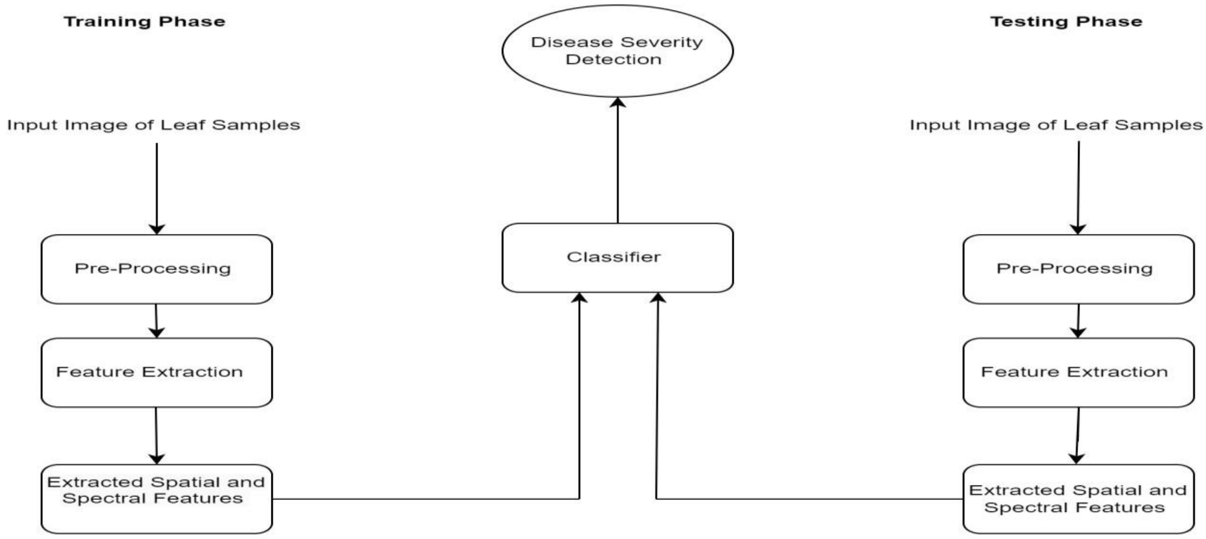

Figure 1 : Methodology of the Proposed System

Here is how Hyperspectral imaging can be applied in Arecanut plant disease detection.
3.1. Data acquisition: To obtain Hyperspectral images, specialized sensors or camer: s are used, which can record a broad range of wavelengths, usually from the visible to the near infrared. The excellent spectral resolution of these sensors allows for the detection of minute variations in the reflectance characteristics of plants.
3.2. Preprocessing: Reduce noise, account for atmospheric effects, and improve the spectrum information, the obtained Hyperspectral images go through several preprocessing stages. This encompasses atmospheric adjustment, geometric registration, and radiometric calibration.
Hyperspectral data offers a multitude of spectral information for every pixel in the image, making it ideal for feature extraction. To determine the pertinent spectral properties that can distinguish healthy plants from diseased ones, feature extraction techniques are used. These characteristics could include patterns of absorption or reflection linked to biochemical alterations brought on by illness in plants.
3.3. Classification: Machine learning methods or statistical models can be used for classification after the pertinent features have been extracted. These algorithms are trained using a tagged dataset, which consists of hyperspectral pictures labelled with the respective disease status (well or sick). Then, new, unseen hyperspectral photos may be classified using the trained models, and diseases in Arecanut plants can be detected.

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

3.4. Disease mapping and monitoring: These tasks can be completed by analysing Hyperspectral photos of Arecanut plants over time. It is possible to track changes in a plant's spectral signature, which makes disease early detection and spatial mapping possible. This data can help control the spread of diseases within Arecanut plants and facilitate the implementation of focused interventions.

It's important to remember that Hyperspectral imaging is a challenging and specialized discipline that calls for knowledge of machine learning, image processing, and remote sensing. It frequently works in tandem with additional diagnostic methods and field observations to offer a thorough comprehension of the dynamics of plant health and disease.

4. RESULTS AND DISCUSSIONS

4.1.Dataset Collection and the study area: The methodical gathering of data over several spectrum bands with meticulous attention paid to spectral, spatial, and temporal resolutions, is known as hyperspectral imaging. Applications utilizing hyperspectral imaging depend on precise data gathering, processing, and analysis.

4.1.1. The samples of several types of leaves—diseased and healthy are taken into consideration from the Channagiri, Davanagere, actual site. Additionally, the samples are separated according to age groups. Both the unhealthy and the healthy samples were recognized and categorized as the classification is divided into two categories: (i) healthy and (ii) diseased. Once again, the diseased category was split into three categories: diseased-low, diseased-mild, and diseased- severe. The age ranges that were considered were: 0.5 months to 10 years, 10-15 years, 15–25 years, and older than 25 years. In Karnataka, India, Channagiri is located at roughly 14.0291° N latitude and 75.9700° longitude. E.

The leaf samples collected were tested under spectroradiometer to get their respective reflectance values, which helped in further analysis. The spectroradiometer specifications are as below:

The ASD FieldSpec 4 standard res (FS4) spectroradiometer is a battery-operated, portable spectrometer. It is intended for measurements in the lab or during field campaigns. The wavelengths in the spectral range of 250–2500 nm is sampled at a rate of 0.2 seconds per spectrum. 251. Bands with a spectral resolution ranging from 3 nm at the very short wavelengths to 10 nm at the farther wavelengths.

Figure 2: Different sites visited (Different age groups and with and without disease conditions)

The following Figure 3 shows the laboratory arrangements made for testing the leaf samples.

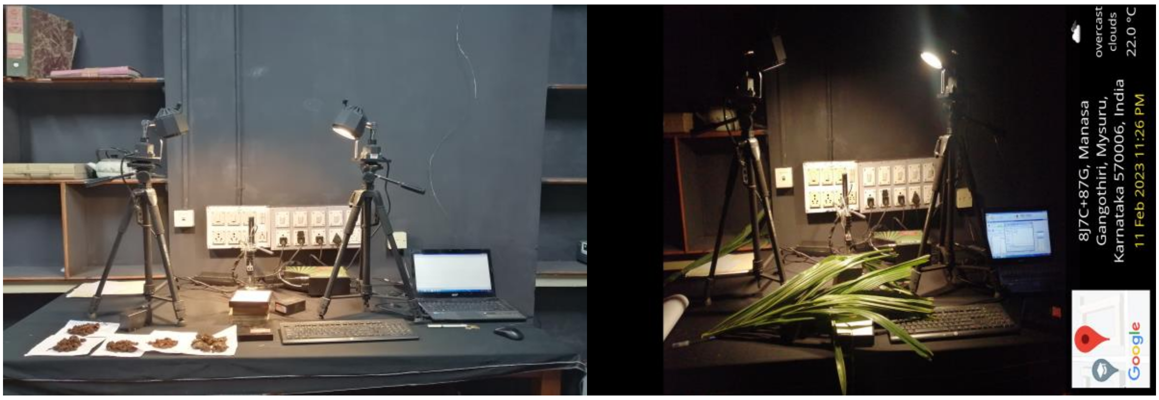

Figure 3 : Testing the samples under the Spectroradiometer.

4.1.Pre-Processing stage: Data collected was converted into its equivalent reflectance values, this stage is called as spectral library build-up. Once the spectral library is created. The pre-processing of the data was carried out using ENVI 5.6. Output of this stage contains only the required bands where useful information is present. The FLAASH (Fast Line of Sight Atmospheric Analysis of Spectral Hypercubes was performed along with MNF for dimensionality reduction. The process of PPI (Purest pixel Index) determines the ROI with only the purest pixels of the scene.

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

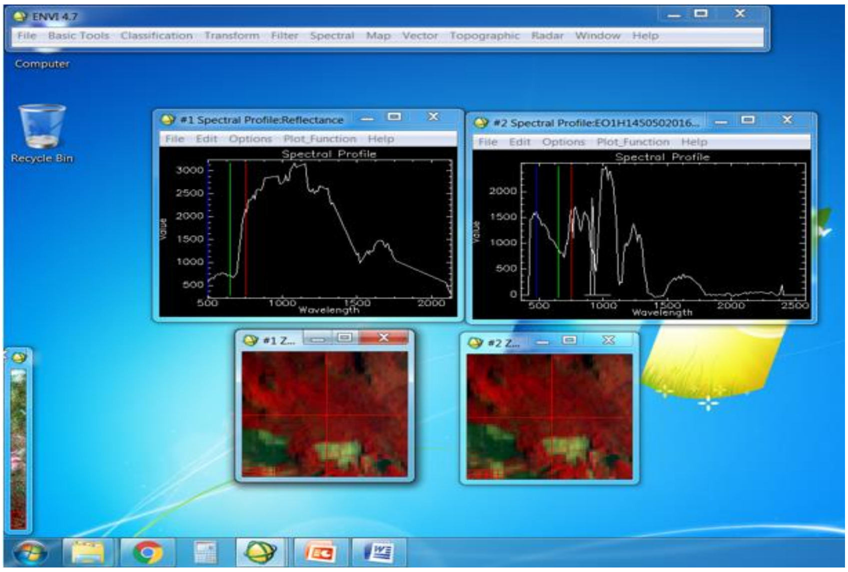

Figure 3 : Image1 with Atmospheric correction and Image 2 without Atmospheric Correction

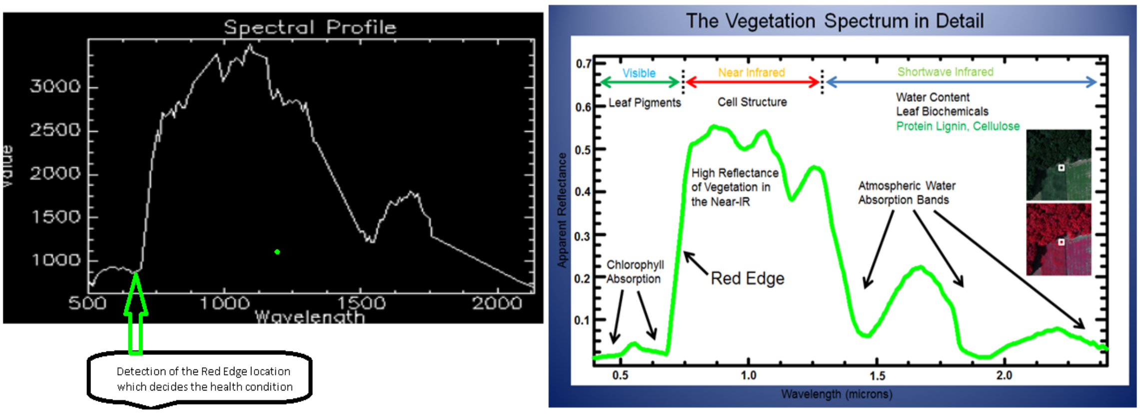

Figure 4: Radiance to Reflectance conversion Plot of vegetation :(i) Obtained (ii) Reference.

The spectral library created was used to give input to the machine learning algorithms models for further process of classification. The training and testing ratio was taken as 80:20. The plot of the reflectance graph in figure shows the spectral analysis of the healthy and diseased samples collected.

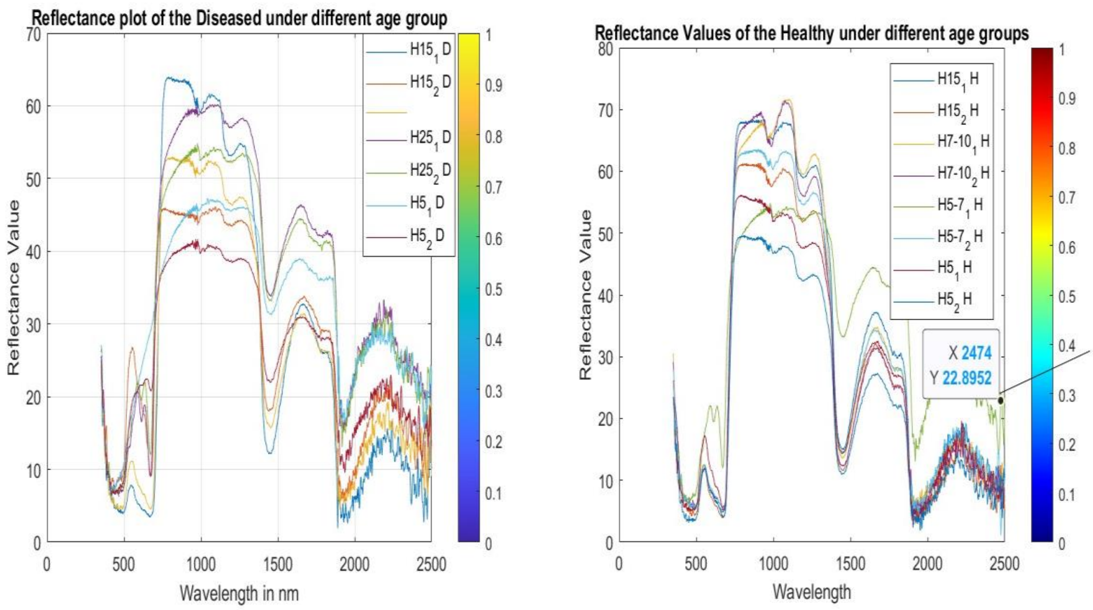

Figure 5: Image showing Reflectance spectral curve of the field data (i) Diseased (ii) Healthy for various age groups.

4.2.Classification: Detecting diseases in Arecanut crops can benefit from various classification algorithms in machine learning. The choice of algorithm depends on factors such as the complexity of the problem, the size of the dataset, and the features available. Here are some commonly used classification algorithms for disease detection in crops:

4.2.1.1.Decision Trees (DT): Decision trees are easy to interpret and can handle both numerical and categorical data. They are effective for feature selection and can be visualized to understand the decision-making process.

4.2.2. Random Forest (RF): Random Forest is an ensemble method that builds multiple decision trees and combines their prediction and often provides better accuracy than individual decision trees and is robust against overfitting.

4.2.3. K-Nearest Neighbours (KNN): KNN classifies data points based on the majority class of their k-nearest neighbour. It's a simple and effective algorithm, especially when there is spatial correlation in the data.

4.2.4. Naïve Bayes: Naive Bayes is based on Bayes' theorem and assumes independence between features. It is computationally efficient and works well for text classification and situations where the independence assumption is reasonable.

4.2.5. Logistic Regression: Logistic Regression is a simple yet powerful algorithm for binary classification and is easy to implement and interpret, making it a good choice for baseline models.

4.2.6. LightGBM (Light Gradient Boosting Machine): It is a gradient boosting framework that is designed for efficient and distributed training of large datasets. It is developed by Microsoft and is open source. LightGBM is particularly known for its speed and efficiency, making it suitable for handling large-scale datasets and tasks.

4.2.7. Extra Trees Algorithm: Extra Trees is an ensemble learning method, like Random Forests. It belongs to the family of tree-based ensemble algorithms.

4.2.8. XGBoost Algorithm: It stands for eXtreme Gradient Boosting, is a popular and powerful machine learning algorithm used for both classification and regression tasks. It is an implementation of gradient boosted decision trees designed for speed and performance. XGBoost is widely used in machine learning competitions and is considered a go-to algorithm for structured/tabular data problems.

4.2.9. Gradient Boosting: It includes the Gradient Boosting Classifier, is known for its effectiveness in a variety of tasks and often performs well on structured/tabular data. It is important to note that like XGBoost, the training of the model is sequential, and each new tree corrects errors made by the previous ones. It is a powerful algorithm but can be computationally expensive, especially for large dataset. Gradient Boosting is a machine learning technique that builds a series of weak learners (usually decision trees) and combines their predictions to create a stronger, more accurate model. In the context of scikit-learn, one of the popular libraries for machine learning in Python.

4.2.10. Support Vector Machines: Support Vector Machines (SVM) is a supervised machine learning algorithm used for classification and regression tasks. SVM works by finding a hyperplane that best separates different classes in the feature space. The hyperplane is chosen in such a way that it maximizes the margin, which is the distance between the hyperplane and the nearest data points of each class, also known as support vector.

4.2.11. Linear Discriminant Analysis (LDA): LDA is a dimensionality reduction and classification technique. In the context of linear discriminant analysis, LDA is used to find the linear combinations of features that best separate two or more classes. It aims to maximize the distance between the means of different classes while minimizing the spread (variance) within each class.

4.2.12. Quadratic Discriminant Analysis: It is a classification algorithm closely related to Linear Discriminant Analysis (LDA). Like LDA, QDA is a supervised learning algorithm used for classification tasks. However, unlike LDA, QDA assumes that the covariance of each class is different, allowing for a more flexible decision boundary.

4.2.13. AdaBoost Algorithm: “Ada" typically refers to AdaBoost (Adaptive Boosting), which is an ensemble learning algorithm used for classification and regression tasks. AdaBoost is particularly powerful and effective in improving the performance of weak learners (e.g., decision trees) by combining them to create a strong learner. The algorithm gives more weight to the misclassified instances in each iteration, allowing subsequent weak learners to focus on the mistakes made by the previous one.

The various models discussed above were considered for the classification purpose. The following parameters given in Table 1 are obtained after the classification. Model analysis was done based on ACC (Accuracy), AUC, Recall and PREC (Precision), F1 Score, Kappa Coefficient, MCC (Mathew’s Correlation Coefficient and TT (Time Taken) .

Table 1: Performance Analysis of Various Algorithms

Model	ACC	AUC	Recall	PREC	F1	Kappa	MCC	TT(Sec)
lightgbm	**0.8942**	0.9926	0.8826	0.8942	0.8941	0.8636	0.8637	0.6480
ET	**0.8930**	0.9663	0.8772	0.8931	0.8930	0.8620	0.8620	0.3380

KNN	**0.8876**	0.9866	0.8743	0.8877	0.8875	0.8550	0.8550	0.1270
RF	**0.8863**	0.9901	0.8713	0.8866	0.8863	0.8534	0.8534	0.5540
XGBoost	**0.8861**	0.9914	0.8749	0.8863	0.,8861	0.8532	0.8533	2.5150
DT	**0.8768**	0.9247	0.8593	0.8766	0.8765	0.8410	0.8411	0.0300
GBC	**0.8281**	0.9809	0.8325	0.8347	0.8286	0.7799	0.7811	3.4579
LR	**0.6311**	0.9156	0.5952	0.6579	0.6309	0.5235	0.5283	2.8340
Ridge	0.6289	0.0000	0.595	0.652	0.633	0.5251	0.5282	0.0140
SVM	0.5911	0.0000	0.5135	0.5662	0.5632	0.4653	0.4746	0.0740
LDA	0.5872	0.9066	0.6784	0.8501	0.6058	0.5050	0.5536	0.02420
NB	0.4702	0.8973	0.5991	0.5414	0.4231	0.3800	0.4477	0.0170
Ada	0.3782	0.7057	0.3054	0.4108	0.3109	0.1720	0.2132	0.2380
QDA	0.2948	0.000	0.2000	0.0869	0.1342	0.000	0.000	0.0260
Dummy	0.2948	0.5000	0.2000	0.0869	0.1342	0.000	0.000	0.0120

From table 1 it is evident that, Lightbgm Algorithm performed better in all the parameters on the datasets taken. It is considered as the best classification algorithm for disease detection at various levels. The following graphs indicated the spectral reflectance curve for the broad category of healthy and diseased (mild, severe, and light)

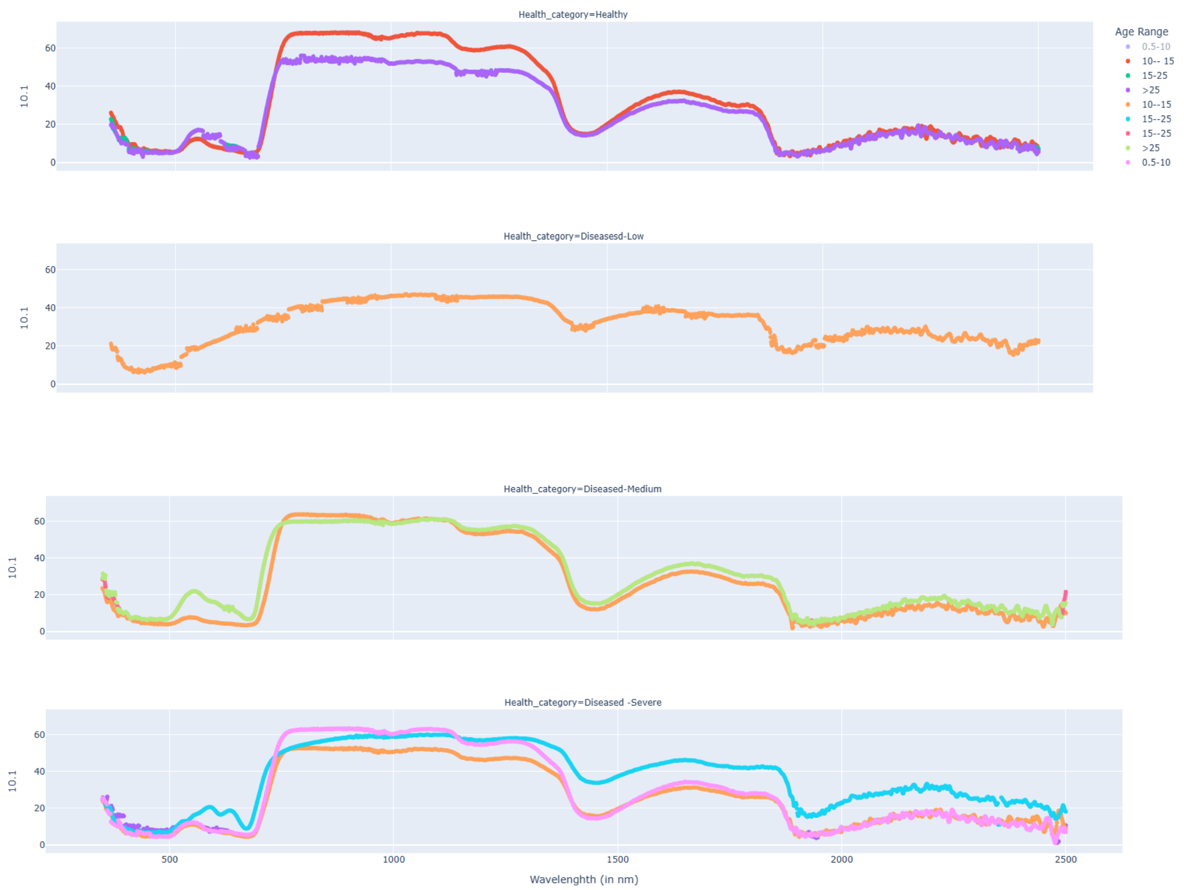

Figure 6: Spectral Prediction Plots of (i) Healthy and (ii) Diseased –Low (iii) Diseased – Medium and (iv) Diseased-Severe

Figure 6 shows the spectral reflectance curve taken between 200 nm to 2500 nm which is the range of spectroradiometer. We can observe that prominent wavelength lies between 700 nm to 2350nm range. Hence for evaluation purpose the plot of reflectance in this range is only considered which is shown below in figure 8. The red edge transition between 500 nm to 700 nm is the point to be considered in the analysis of disease. Red edge point gives whether the samples taken can be considered under healthy category or non – healthy category.

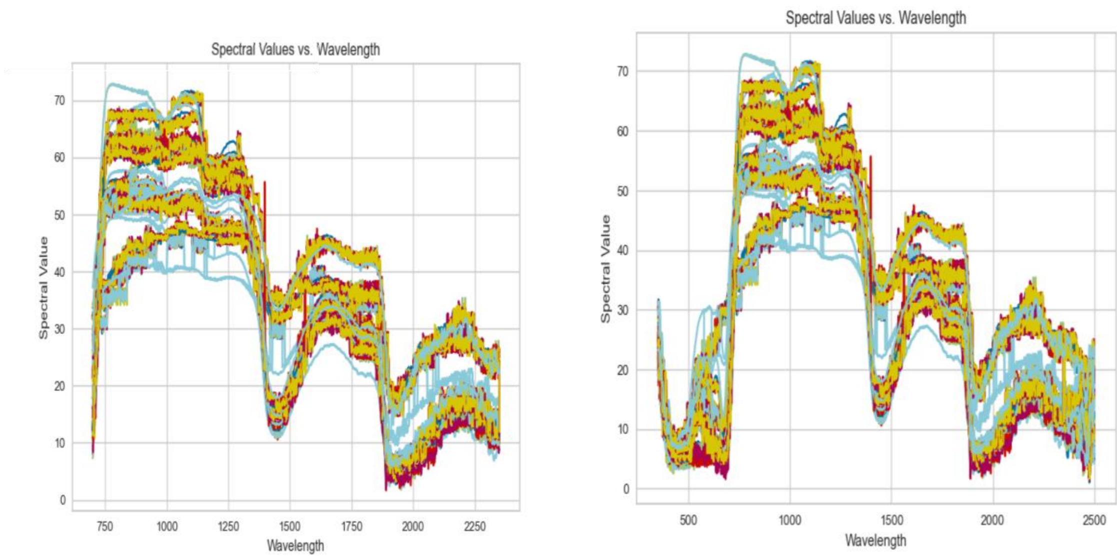

Figure 7: Plot of the (i) Entire Spectral Reflectance Versus Wavelength between the range of 200 to 2350 nm (ii) spectral reflectance Versus wavelength between the range of 700 to 2350 nm for various Status categories

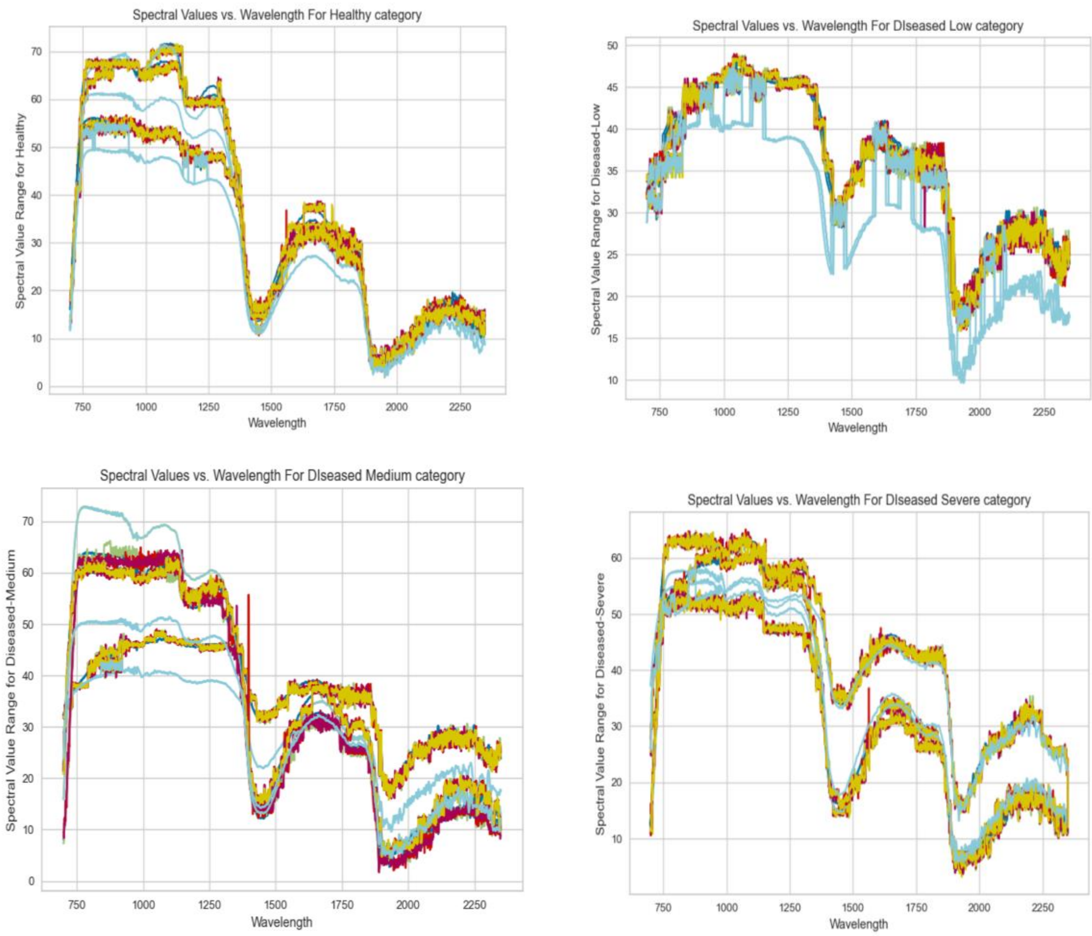

Figure 8: Plot for spectral reflectance Versus wavelength between the range of 700 to 2350 nm for (i) Healthy Status (ii) Diseased – Low Status (iii) Diseased – Medium Status (iv) Diseased – Severe Status

4.4. Feature Importance Metrics:

This section gives an overview of the important features that should be considered as to understand the important wavelengths which helps one to determine the health status of the Arecanut samples.

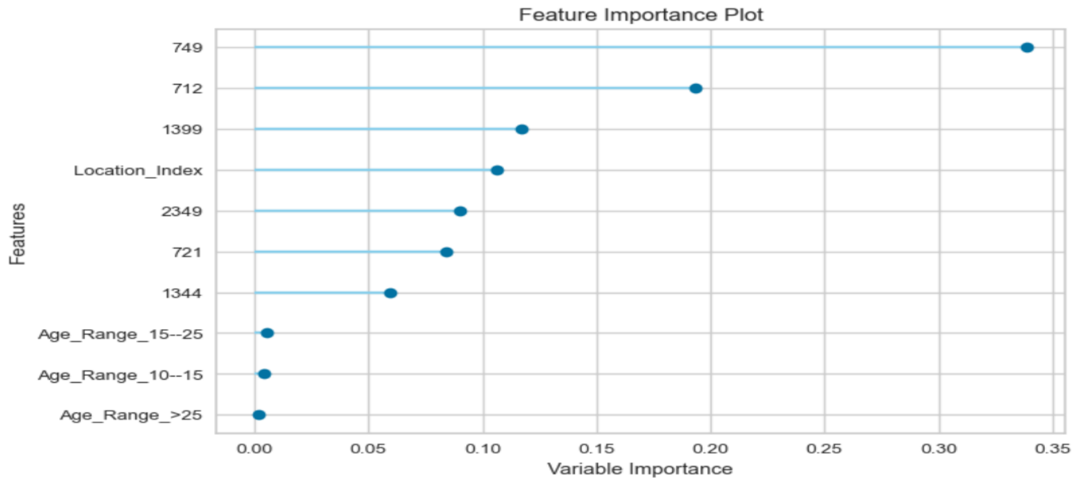

Figure 9: Feature importance Metrics

The above graph shows that 749nm (Red edge transition value), 712nm, 1399nm are the prominent wavelengths that decide the health status of the Arecanut plot samples that are taken into consideration. Along with these, the location index is the next parameter that must be given with importance. As the maximum variation is seen with respect to 749nm, it must be given major importance.

4.3.Location wise impact of Health Status: The following figure gives the relation between the various locations (Age based) and the impact of health status in those locations. The graph is a clear indication of the amount of healthy Arecanut plantation present in different locations.

The Arecanut plants of 0.5-10 years are seen more in the 5th to 10th sample places taken and are healthy. Location-wise these samples are taken as healthy samples. Hence healthy samples among the total samples are present in these places. Diseased – Low Arecanut plants are present in 12th-14th sample locations. Diseased-Medium can be seen in 8th - 13th sample locations. Diseased- Severe and found in 7th – 12th Sample locations. This way we could identify the locations of the Diseased and healthy category Arecanut from the samples taken using spectroradiometer.

CONCLUSION

The overall research carried out in various locations/ sites of Channagiri, Davanagere on Arecanut plantations have given us details of the health status of the various samples. The study is helpful in determining the area as healthy, diseased etc. so that careful measures can be taken beforehand to find out the reason for the effected diseases. The analysis carried out is helpful for the farmers of that area to know the status of their crop as to apply the corrective measures. The research done on these real time samples also will be helpful in predicting the yield and improvement of the same in the future.

The various machine learning and deep learning approaches helped us to perform the classification of the samples based on their disease severity measures.

The analysis of soil and water can give the better solution for the farmers to judge the deficiencies which caused the diseases. Work can be further progressed by taking more samples into consideration and comparing health status year wise/ seasonwise.

REFERENCES

[1] L. W. Kuswidiyanto, et.al, “Plant Disease Diagnosis Using Deep Learning Based on Aerial Hyperspectral Images: A Review”, published in Remote Sensing, https://doi.org/10.3390/rs14236031 in 2022.

[2] Koushik Nagasubramanian et.al, “Plant disease identification using explainable 3D deep learning on Hperspectral images”, published in plant Methods, https://doi.org/10.1186/s13007-019-0479-8 in 2019.

[3] M. Nagaraju et.al, “Systematic review of deep learning techniques in plant disease detection”, published in Int. J. Syst. Assur. Eng. Manag., https://doi.org/10.1007/s13198-020-00972-1 in 2020.

[4] S. M. Hassan et.al, “ A Survey on Different Plant Diseases Detection Using Machine Learning Techniques”, published in Electronics, https://doi.org/10.3390/electronics11172641 in 2022.

[5] Majji V Applalanaidu et.al, “A Review of Machine Learning Approaches in Plant Leaf Disease Detection and Classification”, published in 021 Third International Conference on Intelligent Communication Technologies , https://doi.org/10.1109/ICICV50876.2021.9388488 in 2021.

[6] Canh Nguyen et.al, “ Early Detection of Plant Viral Disease Using Hyperspectral Imaging and Deep Learning”, published in Italian National Conference on Sensors, https://doi.org/10.3390/s21030742, in 2021.

[7] Chunying Wang et.al, “A review of deep learning used in the hyperspectral image analysis for agriculture”, published in Artif. Intell. Rev, https://doi.org/10.1007/S10462-021-10018-Y, in 2021.

[8] Hongyan Zhu et.al, “Hyperspectral Imaging for Presymptomatic Detection of Tobacco Disease with Successive Projections Algorithm and Machine-learning Classifiers”, published in Scientific Reports, https://doi.org/10.1038/s41598-017-04501-2 in 2017.

[9] A-K. Mahlein1 et.al, “ Plant disease detection by Hyperspectral imaging: from the lab to the field”, https://doi.org/10.1017/S2040470017001248, in 2017.

[10] Amy Lowe et.al, “Hyperspectral image analysis techniques for the detection and classification of the early onset of plant disease and stress”, published in Plant Methods, https://doi.org/10.1186/s13007-017-0233-z in 2017.

[11] Ankita Chatterjee, JayasreeSaha, Jayanta Mukherjee, Subhas Aikat and Arundhati Misra, “Unsupervised land cover classification of hybrid and dual-polarized images using deep convolutional neural network”, IEEE Geoscience and Remote Sensing Letters, vol. 18, no. 6-Pg -66.

[12] . Yue Zhu, Christian Gei, Emily So and Ying Jin, “Multitemporal relearning with convolutional LSTM models for land use classification”, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 3251-3265, 2021.

[13]. Ziheng Sun, Liping Di and Hui Fang, “Using long short-term memory recurrent neural network in land cover classification on landsat and cropland data layer time series”, International Journal of Remote Sensing, vol. 40, no. 2, pp. 593-614, 2018.

[14] Anju Unnikrishnan, Sowmya V and Soman K. P, “Deep learning architectures for land cover classification using red and near-infrared satellite images”, Multimedia Tools and Applications, vol. 78, no. 13, pp. 18379–18394, 2019.

[15] Noureldin Laban, Bassam Abdellatif, Hala M Ebeid, howida A Shedeed and Mohamed Flba, “Sparse pixel training of convolutional neural networks for land cover classification”, IEEE Access, vol. 9, pp. 52067-52078, 2021.

[16] Chermprayong P, Hongkarnjanakul N, Rouquette D, Schwob C and Mezeix L, Convolutional Neural Network for Thailand’s Eastern Economic Corridor (EEC) land cover classification using overlapping process on satellite images”, Remote Sensing Applications:Society and Environment, vol. 23, pp. 1-14, 2021.

[17] Su Wit Yi Aung, SoeSoeKhaing and ShweThinzar Aung, “Multi-label land cover indices classification of satellite images using deep learning”, Springer, Singapore, 1st Edition, ISBN: 978-981-13-0868-0, 2019.

[18] Ioannis Papoutsis, Nikolaos IoannisBountos, AngelosZavras, DimitriosMichail and Christos Tryfonopoulos, “Benchmarking and scaling of deep learning models for land cover image classification”, ISPRS Journal of Photogrammetry and Remote Sensing, vol. 195, pp. 250- 268, 2023.

[19] Ozlem Akar and EsraTuncGormus, “Land use/land cover mapping from airborn Hyperspectral images with machine learning algorithms and contextual information”, Geocarto International, vol. 37, no. 14, pp. 3963-3990, 2021.

[20] Rajat Garg, Anil Kumar, Manish Prateek, and Shashi Kumar, “Land cover classificatio spaceborne multifrequency SAR and optical multispectral data using machine learning”, Advances in Space Research, vol. 69, no. 4, pp. 1726-1742, 2022.