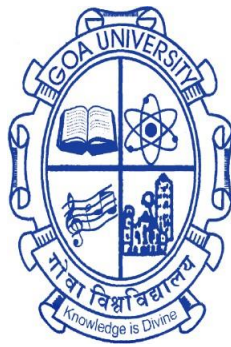


Early Detection Of Beetle Attack In Cashew Trees Using Machine Learning

A Thesis submitted in partial fulfillment for the Degree of
DOCTOR OF PHILOSOPHY

IN THE SCHOOL OF PHYSICAL AND APPLIED SCIENCES
GOA UNIVERSITY



By

Yogini Venkatraman Prabhu

Electronics Programme, School of Physical and Applied Sciences,
Goa University,
Goa

September 2023

DECLARATION

I, **Yogini Venkatraman Prabhu** hereby declare that this thesis represents work which has been carried out by me and that it has not been submitted, either in part or full, to any other University or Institution for the award of any research degree.

Place: Taleigao, Plateau.

Date: 16-09-2023

Yogini Venkatraman Prabhu

CERTIFICATE

I hereby certify that the work was carried out under my supervision and may be placed for evaluation.

Prof. Jivan S. Parab
Goa University

Acknowledgement

First and foremost, I would like to thank the Almighty GOD for showering upon me countless blessings and graces required by me during these years. I would like to express my deepest gratitude to my research guide, Dr. Jivan S. Parab, for his constant support and motivation throughout my research work. He has always been patient with me and inspired me to take up challenges during my research work.

I would like to convey my gratitude to Dr. M. Kunhanandan, Assistant professor, Department of Mathematics, School of Physical and Applied Sciences, Goa University, for his feedback and suggestions during the Department Research Committee (DRC) meetings which enabled me to improve in various facets of my research work.

I would like to acknowledge Dr. A. Rao Dessai, Etymologist and Dr. Pratap Singh, Director of ICAR-Ela Farm (year 2015), Old Goa for providing us the information on agricultural problems in Goa particularly the CSRB. The pioneering by Mr. V. K. Joseph led to opt, and constant surveillance by Dr. P. Samant accomplished the entire pursuance, of the research.

I acknowledge Media Lab Asia and Ministry of Electronics & Information Technology (MeitY) Government of India, towards support of fellowship under the Visveswaraya PhD Scheme for the period 2015-2020.

I extend a big thank you to, Prof. R. S. Gad, Vice Dean of School of Physical and Applied Sciences, Goa University, for his insightful comments. I am sincerely grateful to, Dr. G. M. Naik, School of Physical and Applied Sciences, Goa University, who has been kind to look after the arrangements, so as to improve focus on research.

My many thanks and appreciation go to Dr. N. G. Tavares, Dr. M. Sequeira, Dr. N. T. Vetrekar, Dr. C. Panem, Dr. S. S. Patil, and Dr. N. Lobo for their valuable contribution and help rendered during my research work. I would like to thank Mrs. Vilma M.T. Fernandes, Mr. A. Prabhu, Mr. A. Gaonkar, Dr. V. R. Gad, Mr. S. M. Patil, for the valuable scientific discussions we have had during my research work.

I am grateful to my colleagues, Dr. S. Gawali, and Mrs. S. Gawade, to retain the spirit for completions of PhD. I also thank the supporting staff of the Department of Electronics School of Physical and Applied Sciences, Goa University, Mr. W. D'Souza, Mr. V. Malik, Mrs. A. Velip, Mrs. P. Andrade, who have always extended their helping hand whenever needed. I also thank Mr. M. Lanjewar, Technical Officer, and Mr. A. Lopes, USIC, Goa University, for their moral encouragement.

Finally, I would like to thank each and every one, whose names might not have been mentioned here, but I have been forever indebted due to their inner real best wishes for my success. It has been a fruitful and enjoyable experience and the cooperation of all, meant a lot to me.

Contents

1 Introduction.....	1
1.1 Overview of worldwide cases of borers.....	3
1.2 Indian Statistics.....	4
1.3 In-depth Background.....	4
1.3.1 Insights to tunnelling.....	6
1.3.2 Insights to Physiological Changes.....	7
1.3.3 Pest Population Dynamics.....	9
1.3.4 Ecological Impacts.....	9
1.4 Actions at the impacted locations.....	11
1.4.1 Documented prevention methods.....	11
1.4.2 Documented treatment methods.....	12
1.4.3 Significance of scientific detection.....	13
2 Literature review.....	22
2.1 Techniques of pest detections.....	23
2.1.1 Remote Sensing.....	24
2.1.2 Proximal Sensing.....	28
2.1.3 Acoustics.....	28
2.1.4 Electronic Nose.....	30
2.1.5 Vis-NIR Spectroscopy.....	31
3 Pre-Processing and Machine Learning.....	50
3.1 Pre-processing.....	54
3.2 Dimensionality Reduction.....	59
3.3 Machine Learning.....	59
4 Methodology.....	69
4.1 Study location.....	69
4.2 Image Processing.....	69
4.3 Spectroscopy.....	73
4.3.1 Preliminary Studies.....	73

4.3.2 Refined Methodology.....	77
5 Results and Discussions.....	88
5.1 Leaf Spectral Analysis.....	89
5.2 Bark Spectral Analysis.....	93
5.3 Data Fusion (Leaf and Bark).....	97
5.3.1 Classification using BNNN.....	98
5.3.2 Classification Using SVM.....	101
6 Conclusion.....	105
 ANNEXURES	
ANNEXURE I.....	109
ANNEXURE II.....	115
ANNEXURE III.....	117
ANNEXURE IV.....	126

List of Tables

- 1.1 Health conditions of a cashew tree with detailed symptoms upon the CSRB attack
- 2.1 The non-destructive technologies for Agricultural and Food Products Quality Inspection, applicable to CSRB detection
- 2.2 Data-collection and processing techniques for studying metabolic processes at the different or moreover complex biological scenarios
- 2.3 Results achieved by the Machine Learning methods for the detection of diseases of plants, having target area as leaves
- 2.4 Results achieved by the Deep Learning methods for the detection of diseases of plants
- 4.1 Specifications of imaging device: Panasonic LUMIX DMC-SZ10
- 4.2 Chlorophyll values for each HC by applying the RGB equation to the pixels of images
- 4.3 Chlorophyll values from spectral data, for each health condition
- 4.4 Selected spectral ranges (wavelengths in nm) of leaf and trunk reflectance
- 5.1 Summary of results of the BPNN model for leaf spectral data
- 5.2 Summary of results of the BPNN model for bark spectral data
- 5.3 Summary of results of the BPNN model for leaf and bark spectral data, for 2 Classes
- 5.4 Summary of results of the BPNN model for leaf and bark spectral data, for 4 Classes.
- 5.5 Summary of results of SVM for leaf and bark spectral data, for 2 Class model.
- 5.6: Summary of results of SVM for leaf and bark spectral data, for 4 Class model.
- 5.7: Summary of results of all three modalities, for 4 class models.
- 6.1: Summary of results of all three modalities.

List of Figures

- 1.1.a Biotic Stressors of Plants
- 1.2.b Geographic distribution of the CSRB infestations.
- 1.2 On-site pictures of the CSRB infected plantation.
- 1.3 The life cycle of CSRB.
- 1.4 Sources (one of which is VOCs) and receptors of green-house gases.
- 2.1 Fluorescence technique: development process of images.
- 2.2 Picture of leaf and its thermograph.
- 2.3 Hyper-spectral technique.
- 2.4 Electronic Nose.
- 2.5 The Electromagnetic Spectrum.
- 2.6 A Preliminary assembly of a spectrophotometer.
- 3.1 Components of Dimensionality Reduction
- 3.2 Types of Dimensionality Reduction
- 3.3 Functioning of PCA
- 3.4 Workflow of project using Machine Learning algorithm
- 3.5 Tiers indicating the ascending capability of NNs in tracking severity of the HCs detection -epitome being the HC2, in terms of rapidness and accuracy.
- 4.1 Details of the study location along with map of Goa.
- 4.2 Pictures of the leaf samples for each study category.
- 4.3 An example - demo of positioning of location on the leaf area, for selection of pixels.
- 4.4 USB 2000+ Spectrophotometer setup.
- 4.5.a Study of leaf selection from different branches of cashew tree.
- 4.5.b The standard signature for chlorophyll in leaves.
- 4.6 A spectral recording setup
- 4.7 Study of height location at branch of ashewtree.
- 4.8 Jasco V-770 spectrophotometer, and integrating sphere with leaf

- 4.9 Viable spectral ranges (wavelengths in nm) for input definition to the models.
- 4.10.a Flowchart for the (PCA, NN based) individual of the leaf and bark spectra
- 4.10.b Flowchart for the (PCA, NN -SVM based) data fusion of the leaf and bark spectra
- 4.11 A neural network diagram for 2-class model.
- 5.1 The recorded spectra and the preprocessed spectra of leaf, for individual and averaged of each 10 trees in all the four health categories.
- 5.2 Score plot of leaf spectral data representing different health conditions.
- 5.3 PC loadings of first 2 components of the PCA to leaf data, with indication of minima and maxima of each.
- 5.4 Architecture of the BPNN model for leaf spectral data alone.
- 5.5 Confusion Matrix for leaf spectra.
- 5.6 Receiver Operating Characteristic Curve for leaf spectra alone
- 5.7 The recorded spectra and the preprocessed spectra of bark, for individual and averaged of each 10 trees in all the four health categories.
- 5.8 Score plot of leaf bark data representing different health conditions.
- 5.9 PC loadings of first 2 components of the PCA to bark data, with indication of minima and maxima of each.
- 5.10 Architecture of the BPNN model for bark spectral data alone
- 5.11 Confusion Matrix for bark spectra alone
- 5.12 Receiver Operating Characteristic Curve for bark spectra alone.
- 5.13 Architecture of the BPNN for leaf and bark spectral data together, for 2 class model.
- 5.14 Confusion Matrix for leaf and bark spectral data, for 2 class model using BPNN.
- 5.15 Architecture of the BPNN model for leaf and bark spectral data together
- 5.16 Confusion Matrix for leaf and bark together, using BPNN
- 5.17 Receiver Operating Characteristic Curve for leaf and bark together, using BPNN.
- 5.18 Confusion Matrix for leaf and bark spectral data, for 2 class model using SVM.
- 5.19 Confusion Matrix for leaf and bark together, using SVM.
- 5.20 Receiver Operating Characteristic Curve for leaf and bark together, using SVM.

PREFACE

This thesis is about the “Early detection of beetle attack in cashew trees Using Machine Learning”. The computational models for the three modalities – leaf spectra analysis, bark spectra analysis, data fusion was developed using MATLAB R2021a and Python 3.10.12. The data collection was done using two spectrophotometers- USB 2000+ and Jasco- V770.

Chapter 1 contains an introduction to the thesis. It describes the basis of research problem i.e., borers in global and Indian scenarios. It elaborates the process of infestation by CSRB i.e., the Cashew Stem and Root Borer, along with the actions at the impacted locations. It focuses on the problems associated with the manual detection, mentions the importance in developing a scientific and technological detection, accordingly defines the research statement.

Chapter 2 describes a literature review on various techniques of pest detections, plant diseases and general biological process (having stages of progression). It also reviewed on the application of Machine Learning and Deep learning for the detection of pest and plant diseases. It mentioned the importance of spectroscopy technique. The objectives for the research statement are defined.

Chapter 3 elaborates the basics of Preprocessing of spectral data, and details of features and parameters of Machine learning algorithms - BPNN and SVM. It also details the technique of PCA.

Chapter 4 discusses the methodology in designing a model for detection of borer pest at early stage. It presents two methodologies- 1) Image Processing, using images of leaves and a RGB equation for determining chlorophyll content, and 2) Spectroscopy, with two phases – preliminary analysis (related to confirmation of samples selection) and refined methodology

(focussed on Machine learning which consisted of steps of selection of wavelengths, and data fusion, performed on the entire recorded datasets of leaves and bark).

Chapter 5 elaborates on the description of steps conducted, model designs prepared, and corresponding results obtained from the spectroscopy- Machine Learning methodology. The calculated numerical metrics- Accuracy, Precision, Recall and F1-score, AUC, and plots of graphical metrics- Confusion Matrix and ROC (Receiver Operating Curve) are presented.

Chapter 6 delivers the comparison of all three modalities – leaf spectra analysis, bark spectra analysis, data fusion, by summarising the outcome of their results obtained, and concludes the objectives, to achieve the research statement.

Due to success of the computational model for the purpose of CSRB detection, there is establishment of a track based on the spectroscopy-ML method. The same track could be scaled up to the systems involving UAV's and satellite data which is widely known as field of remote sensing, so as encompass large geographical areas and thus large datasets containing great information, resulting in greater outcomes.

LIST OF ABBREVIATIONS

ANN: Artificial Neural Network

AUC: Area Under Curve

BPNN: Back-Propagation Neural Network

CABI: Centre For Agriculture And Biosciences International

CSRB: Cashew Stem and Root Borer

EW: Effective Wavelengths

LDA: Linear Discriminant Analysis

LS-SVM: Least Squares Support Vector Machine

LVQ: Learning Vector Quantization

MLP: Multi-Layer Perceptron

NIS: Non-Imaging Spectrometry

NN: Neural Network

PCA: Principal Component Analysis

PC: Principal Components

PLSR: Partial Least-Square Regression

PLS-DA: Partial Least Squares-Discrimination Analysis

PNN: Probabilistic Neural Network

RBF: Radial-Basis Function

RF: Random Forest

ROC: Receiver Operating Curve

Vis-NIR: Visible-Near Infra-Red

Journal Publications

1. Y. Prabhu, J. Parab, R. Gad and G. Naik, **Early Detection of Pest Attack in Trees Using Embedded System** in Journal of researches in dynamics and control systems, in Special Issue-03, Pages134-138,2018.
2. Y. Prabhu, J. Parab, R. Gad and G. Naik, **Farmer-friendly portable system for diagnosis of pest attack in cashew trees in CSI Transactions on ICT**, in Springer, Volume 6, Issue 2,Pages215-219, June 2018.

Conference Publications

1. Y. Prabhu, J. Parab, and G. Naik, **Back-propagation Neural Network (BP-NN) model for the detection of borer pest attack** in Journal of Physics: conference Series, IOP, March2021.
2. Y. Prabhu, J. Parab, and G. Naik, **Exploration of an Image processing model for the detection of borer pest attack** for AISC 1108, 484968_1_En, Chapter 67.
3. Y. Prabhu, J. Parab, R. Gad and G. Naik, **Design of Robust Meter for Monitoring Health of trees** for the 11th Annual Symposium of VLSI & Embedded system, 23rd March2018.
4. Y. Prabhu, J. Parab, R. Gad and G. Naik, **Development Of Embedded System for Early Detection Of Beetle Attack In Cashew Trees** for International Conference on Innovative Research in Science and Technology ICIRST-2017, November2017.
5. Y. Prabhu, J. Parab, R. Gad and G. Naik, **Early Detection of pest attack in trees using Embedded System** for the 10th Annual Symposium of VLSI & Embedded Systems, 28thApril2017.
6. Y. Prabhu, J. Parab, R. Gad and G. Naik, **Detection of defensive compounds in pest infested trees using Embedded System** for NCEMTFA 2016 –National Conference on Electronic Materials Thin Films & its applications, 9th&10thDecember 2016.
7. Y. Prabhu, J. Parab, R. Gad and G. Naik, **Development of Embedded system for detection of chemical compounds in pest infested trees** for Symposium on Plant pathology, October2016.
8. Y. Prabhu, J. Parab, R. Gad and G. Naik, **Detection of Stem and Root Borer pest in trees using embedded platform** for Evolving Trends and Challenges in Science Education and Research - an Exploration NC SER – 2016, on 29th& 30th January2016.

Motivation,

*Master Chirag Mukund S Khandeparkar,
the ebullient boy,
& my biological neural network - under training.
Mr. Mukund V S Khandeparkar,
the exuberant gent,
& my stacking ensemble - boosting all accuracies.*

Dedicated towards,

*my mother Laxmi V Prabhus' unconditional love and
boundless sacrifices.
my father V M Prabhus' aim to educate me as a doctor
of medicine, which is now materialized as a doctor of
philosophy, although with diagnosis and treatment of
a plant disease.
my husband Mukund V S Khandeparkars' herculean
hardships.*

Chapter 1

Introduction

Focused as a cash crop, the cashew has emerged in the cropping systems of many Asian countries, and is cultivated on insignificant and deteriorated lands. As a matter of surprising fact, regions that are left out either being unsuitable or somewhat less favorable for other crops, have been successfully producing the highly coveted food item. Accordingly, its only cashew that can yield income to the lesser fortunate farmers owning constrained land portions such as arid, marginal and sloping areas, as its an option for building a cultivation. Eight prime nuts of the world enumerated in the order of their production and consumption are almond, filbert, cashew, walnut, pecan, Brazil nut, pistachio and macadamia. Cashew, Brazil nut and macadamia are highly valued nut crops adapted to the tropics. The area under cashew in the world is 30.62 lakh ha. The world production of cashew is estimated to be around 20.83 lakh tons. [1] The cashew nut is an important item in the snack food market as well as in the confectionery industry. Moreover, cashew is also regarded as a means of improving food security and nutrition, world over. Besides being nutritious, it is high in durability and highly merchantable.

The cashew specialists strongly believe that the current efficiency of cashew growing enterprises is far below the inherent potential of the crop [2]. Productivity records reported from various locations leave much room to excel. There are wide disparities between actual yields obtained by the production sector and yields expected by researchers. The main reason attributed to this situation is the apparent neglect, low management levels applied to the crop. Efforts are therefore needed to narrow the gap between actual yields and potential yields with the application of modern agro-techniques [3] to protect against pest invasions. One factor which causes concern of management to any crop cultivation is stress. The stresses are of two types, abiotic and biotic. Abiotic are the ones which arise from non-living factors like lack of nutrients, harsh environmental conditions of low or high temperature, deficient or excessive water, high salinity, heavy metals, and ultraviolet radiation. Biotic stress in plants is caused by living organisms, especially viruses, bacteria, fungi, nematodes, insects, arachnids and weeds.

The cashew crop can survive under marginal conditions [4], which means the cashew is

less susceptible to the abiotic stresses. So, only issue which could cause loss is biotic stress. Biotic stress (as shown in Figure 1.1.a) gets even more intense in those cases, where a normal human monitoring cannot accurately predict the spread and strength of causing pests. The management of the plantation should be capable of detecting the biotic stress, and taking effective actions (section 1.4) in right time.

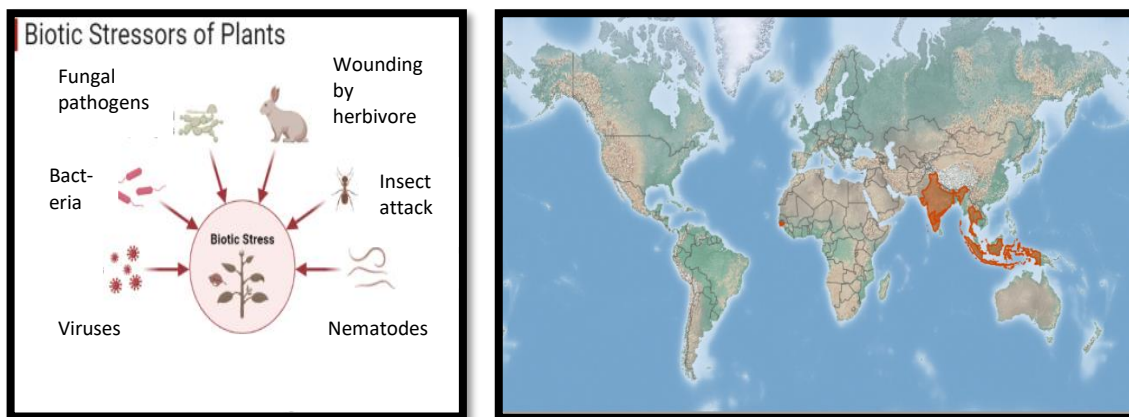


Figure 1.1: a) Biotic Stressors of Plants [5], b) Geographic distribution of the CSRB infestations [6].

Apart from Asian countries, in African country of Nigeria, work on cashew crop protection at the Cocoa Research Institute of Nigeria (CRIN), Ibadan [7], had commenced 1971 with the collection, identification and preservation of the insect pest and disease complex of the cashew plant. This was continued further to understand their symptoms and devise desirable practices to curb the menace of these pests. Thus, good yields can be truly obtained from well managed plantations which are devoid of any pests.

Cashew stem and root borers (CSRB) (*Plocaederus ferrugineus* L.) (Coleoptera: Cerambycidae) is the primary species, infesting cashew in all parts of India (as shown in Figure 1.b) and, two other species, viz., *P. obesus* G. and *Batocera Rufomaculata* De Geer were also reported in association with *P. ferrugineus*[8-10]. While in Ghana, the cashew bark borer *Mecocorynus* sp. (Coleoptera: Anthribidae) was identified [11]. The symptoms of damage include extrusion of frass, occurrence of gummosis, premature yellowing and shedding of leaves, drying of twigs and, finally, death of the tree. There is a need for accurate and prompt detection of CSRB as, 1) Incurability of infested trees, with damage of the trunk circumference 50% or with the leaf canopies yellowed. 2) Increased vulnerability of trees to stresses due to increased summer droughts and heat waves upon climate change.

1.1 Overview of worldwide cases of borers

Borers are a group of insects belonging to different families, including beetles (Coleoptera) and moths (Lepidoptera), which infest the inner tissues of plants, often causing substantial damage. Their larvae tunnel into the wood, stems, or other plant parts, disrupting nutrient and water transport, weakening the plant, and making it susceptible to diseases. Here's a brief overview of worldwide cases of borers:

1. **Agricultural Crops:** Borers can target a wide range of agricultural crops, including corn, sugarcane, coffee, and various fruits. For instance, the Asian corn borer (*Ostrinia furnacalis*) [12] is a significant pest of corn in Asia, causing yield losses due to stem damage. In sugarcane, the sugarcane borer (*Diatraea Saccharalis*) [13] can impact sugar production by feeding on the plant's internal tissues. the western corn rootworm (*Diabrotica virgifera virgifera*) [14], that develop resistance to previously effective pest-control measures and can-do considerable damage if not detected quickly. Whitestem borers (*Xylotrechus quadripes* & *Monochamus leuconotus* P.) are the most serious pests of coffee in India [15] and Zimbabwe [16] respectively.
2. **Fruit Trees:** Fruit trees including citrus (e.g. lemons [17]) are also vulnerable to borer infestations [18]. Apple trees may suffer from attacks by the codling moth (*Cydia Pomonella*) [19] and the flat-headed apple tree borer. Similarly, peach trees can be affected by the peach tree borer (*Synanthedon Exitiosa*) [20], causing injury to the trunk and branches.
3. **Palms:** Palms, both in natural habitats and urban environments, are not immune to borer infestations. The red palm weevil (*Rhynchophorus ferrugineus*) [21] is a notorious palm pest, tunnelling into the crowns of palm trees and causing extensive damage.
4. **Ornamental Plants:** Ornamental trees and shrubs in urban landscapes can also fall victim to borers. The bronze birch borer (*Agrilus Anxius*) [22] affects various species of birch trees, leading to their decline. D-shaped holes are left by the adult bronze birch borer beetle on limbs and trunks. The symptoms of clearwing and bronze birch borer (a flat-headed borer) include crown dieback [23]. The clearwing borer moth, such as ash clearwing borer resembles stinging wasps. Japanese maples and other popular ornamentals are also susceptible to borer damage.

5. Forest Trees: Borers pose a significant threat to forest trees. The emerald ash borer (*Agrilus Planipennis*) [24], for instance, has caused extensive mortality among ash trees in North America. This invasive beetle species bores into the bark and disrupts the water-conducting vessels, leading to the decline of ash tree populations.

The various borer cases in India over past century are Woolly apple aphid, San jose scale, Lantana bug, Cottony cushion scale, Potato tuber moth, Diamond back moth, Pine woolly aphid, Serpentine leaf miner, Coffee berry borer, Spiralling whitefly, Silver leaf whitefly, Papaya mealybug, Erythrina gall wasp, Lotus lily midge, Blue gum chalcid, Tomato pin worm, Invasive chilli thrips, Rugose spiralling whitefly, Fall army worm, Cassava mealybug and Woolly whitefly in chronological order [25].

1.2 Indian Statistics

Cashew, after its introduction from Brazil during the 16th Century, has developed very well in India. India's share in the world raw nut production accounts to about 25 per cent.

The process of attack by CSRB has been observed and documented by several researches. However, there has never been any research to record the stages of the infestation process by the CSRB, and develop a decision-making model. The only data available is about production of cashew. For year 2018-19, it is grown in an area of 11.05 lakh ha with production of 7.42 lakh tons of raw nuts [26]. To date, a total of 40 high-yielding varieties have been enlisted by the Directorate of Cashew Research, Puttur, and various Agricultural institutions.

Controlling this pest is a difficult job [9], as the borer remains in a concealed condition i.e. Green attack stage (when trees are yet to show distinct symptoms observable by the human eye) in the interface of bark and hard wood, and normally it escapes from the attack of any natural enemies.

Our research focus is to develop a farmer friendly technique to detect pest attack in cashew trees in initial stage.

1.3 In-depth Background

This section elaborates the infestation mechanism of CSRB, its primary and Stem borer attacks plant in mainly vital root and stem regions of plant. As the growth of

can occur [3, 4]. The occurrence, existence and proliferation of the infestation is totally unknown until the tree reaches the terminal stage. Figure 1.2 shows the on-site pictures of the issue. Few other pictures are documented by the Directorate of Cashew Research [27].



Figure 1.2: On-site pictures of the CSR B infected plantation. a) A tree after scraping off its lesions and two CSR B beetles' corpses, b) a lamented tree, c) frass, d) gum exuded on bark, inset: picture of beetles, e) Pre-mature leaf shedding in severely infested tree.

The three by-products of the infestation process are 1) Frass: It is the powder of trunk tissues out of burrowing by the CSR B, 2) Gum: It is the exudation by tree at the intrusion spots on trunk surface, 3) Lesions: the vascular portions at the lining of the hollow spaces of the created tunnel, which has collection of different generated molecules – defence compounds, stress compounds, plant-fungal interactive compounds, etc. In view of the biotic stress to cashew trees by the CSR B, four health conditions are identified. Table 1.1 illustrates the Health Conditions (HC) proceedings and their corresponding symptoms at the three locations on the tree - Foliage, Trunk, and Collar. The treatment procedures or remedial measures are known to be effective if applied at HC2 only [27], wherein the bark circumference damaged is under 50%. The HC4 will be a succumbing case and HC3 is known to deteriorate to HC4, hence both are irreversible to Healthy stage.

Accordingly, there has never been any research towards an ultimate instrument or a prototype model for design of a detection system, or even basic recording (data collection of HCs) compilations of CSR B.

Table 1.1: Health conditions of a cashew tree with detailed symptoms upon the CSRB attack.

Health Conditions	Symptoms		
	Foliage	Trunk	Collar
HC1 : Healthy	Green	Normal	Normal ground at base.
HC2 : Mild (Fig 1.2.d) (Green attack stage [2])	Green	Normal	Un-noticeable CSRB intrusions spots, Frass on ground at base.
HC3 : Moderate (Fig 1.2.c) (Grey attack stage)	Yellow-green	Gum or resin on the surface	Noticeable CSRB intrusions spots, Increased frass on ground at base.
HC4 : Severe (Fig 1.2.e) (Red attack stage)	Yellow	Lots of gum or resin	Copious frass on ground at base.

1.3.1 Insights to tunneling

Instead of elaborating the background of the research for details about the life cycle of CSRB [28] i.e., sequence of egg-larvae(grub)-pupae-moth(insect) (as shown in Figure 1.3), the focus is onto only the grub stage as 1) it's resultant tunnelling is the root cause of the damage to the CSRB infested tree 2) the duration of grub stage was more than 40% of the total life duration [27]. The larval stage of CSRB consumes the bark portions of the stem and roots, thereby generating irregular tunnels which enlarge as the grubs grow in dimensions and numbers. Hence, the larval stage is of prime importance in this research. Also, it is a striking fact that neither shot borers nor their larvae feed mainly on the wood itself and their chief food consists of a fungus which grows on the walls of the burrows [18]. Hence it is concluded that the bygone extensive tunneling process is not only an outcome of the ingestion of the cell tissues. On the contrary, it is mainly for the creating space to implant the fungal colonies for food provision, so as to lead the prospective proliferation of the borer population.

The extensive tunneling destroys the vascular tissues supplying water and food through xylem and phloem respectively, and plant sap movements were affected and ultimately trees die.

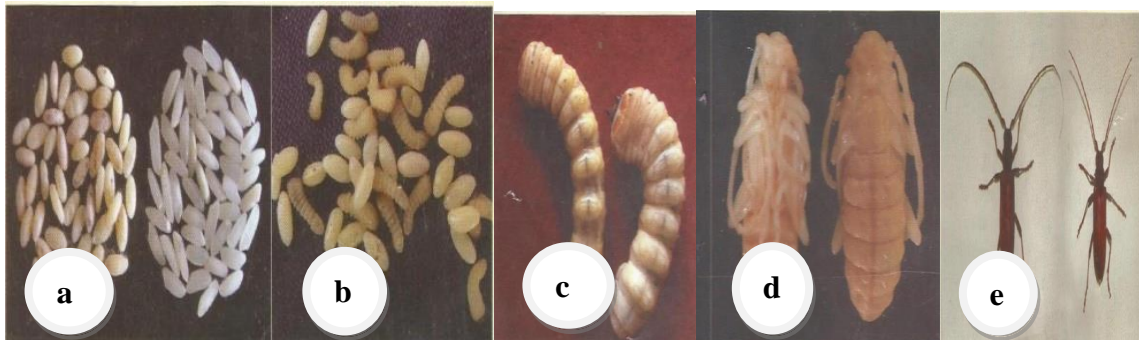


Figure 1.3: The life cycle of CSR: a) Eggs of CSR and Rice grains, b) nascent grubs of CSR, c) Mature grubs of CSR, d) Pupae of CSR, e) Adult beetles of CSR (Male, and Female).

1.3.2 Insights To Physiological Changes

Morphological change - Leaves Abscission

When a plant is under stress conditions, it needs to cope up with decreased supply of minerals and water. This is by an induced response of hormones, namely abscisic acid, ethylene and gibberellic acid in the plant, such that gradually the leaves abscission takes place [30,31]. It does so, by the closure of the stomata during water stress. In fact, in deciduous trees an abscission zone (also called a separation zone) is formed at the base of the petiole (i.e., joint area of leaf to stem) of leaf. It is composed of a top layer that has cells with weak walls, and a bottom layer that expands in the autumn or stress, breaking the weak walls of the cells in the top layer. Consequently, the leaf is shed.

Anatomical change – Plant-fungi interactions

Plant use of endophytes fungi in defense is a very common phenomenon. While some endophytes cause disease, others protect plants from herbivores (which in this research case is the Stem and root borer pest) and pathogenic microbes. Some cases of presence of fungus, which reside symbiotically within the trees, are as follows:

1. Fruit trees such as apple and plum, are apt to suffer from attacks of beetles, Bark beetle and shot borer. It is striking fact that neither Shot borers nor their larvae fed mainly on the wood itself and their chief food consists of a fungus which grows on the walls of the burrows. The fungus spores are carried by the beetles themselves when starting an attack on a new tree [18].

2. A tree-killing bark beetle, *Dendroctonus Frontalis* in the southeastern United States, was found to have three consistent fungal associates, *Entomocorticium sp. A*, *Ophiostoma ranaculosum* and *O. minus*, in 2003[32].
3. Fungal volatile organic compounds (FVOCs) are important drivers of interactions among bark beetle-associated fungi. The community of symbiotic fungi comprised of *Grosmannia clavigera*, *Ophiostoma montium*, and *Leptographium longiclavatum* associated with the Mountain pine beetle (*Dendroctonus ponderosae*) which has killed millions of hectares of pine forests in western North America [33].
4. In 2010, the spruce bark beetle *Ips Typographus* of Norway spruce *Picea abies* was found to have two fungal symbionts (*Grosmannia Penicillata* and *G. Europhioides*), which produced 2-methyl-3-buten-2-ol (MB), a beetles' aggregation pheromone [34].
5. Endophytes can help the plant by producing toxins harmful to other organisms that would attack the plant, such as alkaloid producing fungi which are common in grasses such as tall fescue (*Festuca Arundinacea*). Thus, the presence of alkaloids would be clear indication of the defensive status of the plant in case of infestation by pest [35].

Biochemical change - Defense Compounds

Plants are susceptible to the same kinds of diseases and organisms as animals, including viruses, bacteria and fungi, as well as physical invasion by insects and roundworms. In order to function and survive, plants produce a wide array of chemical compounds not found in other organisms. Because of immobility, plants must also defend themselves chemically from herbivores, insects [36], pathogens and competition from other plants. They do this by producing toxins and foul-tasting or smelling chemicals. Because the biology of plants differs from animals, their symptoms and responses are quite different [5].

Secondary Metabolites: The secondary metabolites deal in defenses against herbivores [37-39], even though they do not play any role in the normal growth, development or reproduction of organisms. A meta-analysis of recent relevant studies has suggested that they have a more complex involvement in defense [40]. Plants have evolved many secondary metabolites involved in plant defense, (which are collectively known as antiherbivore compounds). They can be classified into three sub-groups: nitrogen

compounds (including *alkaloids*, *cyanogenic glycosides*, *glycosylates* and *benzoxazinoids*), terpenoids, and phenolics[41]. In addition to the three larger groups of substances mentioned above, fatty acid derivatives, amino acids and even peptides[42] are also used as defense. The cholinergic toxin, cicutoxin of water hemlock, is a polyene derived from the fatty acid metabolism [43]. β -N-Oxalyl-L- α,β -diaminopropionic acid as simple amino acid is used by the sweet pea which leads also to intoxication in humans[44]. The synthesis of fluoroacetate in several plants is an example of the use of small molecules to disrupt the metabolism of herbivores, in this case the citric acid cycle[45].

1.3.3 Pest Population Dynamics

The female beetles initiate attacks. As they chew into the inner bark and phloem, pheromones are released, attracting male and female beetles to the same tree. The attacking beetles produce more pheromones resulting in a mass attack that overcomes the tree's defences, and results in attacks on adjacent trees. In order to tackle eruptive insect pests, their population dynamics rarely have been studied in a hypothesis-driven manner incorporating exogenous biotic variables. A conceptual framework to reveal the drivers of bark beetle populations has been prepared [46]. The variables that can be easily measured over large geographic and temporal scales are insect and tree host abundance, tree host connectivity, and abiotic climatic factors, while the other biotic factors are more difficult to measure at large scales. Also, in a study about the potential distribution of agricultural pests under projected climatic scenarios, with a focus on the factor of climate change, the results from both the BRT and GLM indicate that precipitation-related variables are more important in determining species range for the Coffee White Stem Borer than temperature related variables [47].

Regarding concerns about the migration of the beetle outbreak from non-intervention zones to adjacent managed forests, an analysis of spatiotemporal dynamics of large-scale disturbances was done by intervention and non-intervention type of management in the Tatra mountains (Central Europe), by collecting the time series of Landsat images [48].

1.3.4 Ecological Impacts

The borers have been reported worldwide to cause rampant damage to commercial plantations [1,3,7-17], gardens [18,19] as well as forests [20,23,28,29,40,41,49]. The result is loss monetarily and ecologically, respectively. The Japanese cypress bark beetle,

Phloeosinus rудis, *P. bicolor* and *P. Thujae* killed numerous shrubs and trees of Cupressaceae in The Netherlands, in the summer of 2004 [50]. Extensive beetle outbreaks by mountain pine beetle (MPB), on the lodgepole pine and ponderosa pine have destroyed over lakhs of km² of forest throughout British Columbia and the western United States [51]. Scientists warn that some forest ecosystems may never recover [52].

Following are harmful ecological turmoil due to bark beetle infestations: -

1. Green-house gases: emissions of monoterpenes, SOA (secondary organic aerosol) and VOCs (as depicted in Figure 1.4). The insect damage causes the carbon sink of forests to turn into carbon source [28, 44]. There has been a 4-fold increase in monoterpene emissions and up to a 40% increase in SOA (secondary organic aerosol) concentrations in some years (as observed in the lodgepole pine response), as result of Beetle infestations.

The Volatile organic compounds (VOCs) are emitted by plants and fungi [33] in response to the herbivore attacks. They provide functional information about the plant’s growth, defence, and health status. Besides pathogenic and insect damage, mechanical damage also arouses defense responses and causes changes to VOCs fingerprints. However, the mode of attack (mechanical or pest-attack) does not completely ascertain the composition of VOCs [53,54]. More importantly, the VOCs information from pest damage can be easily misguided by that from mechanical damage. Here in our study cashew is a crop that could be damaged by mechanical equipment used in the field. Therefore, detection of VOCs from mechanical damage to cashew tree is recommended to guarantee the reason of damage solely from pests.

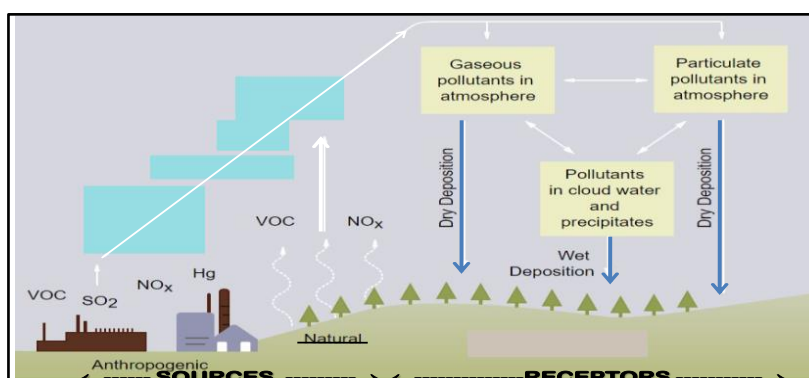


Figure 1.4: Sources (one of which is VOCs) and receptors of green-house gases.

2. Forest fires: The populations [55-57] as well as geographical locations [58,59] of bark-beetles are under alterations due to global warming and climate change. The subsequent rustling of dry leaves (upon the severe attacks of biotic and abiotic factors) to

warm breezes (upon increased atmospheric temperatures) augments the dead trees as inputs to fire [60].

A boon of the situation could be the findings of a study in Montana's Big Hole Valley foreexamining the genes of 'Supertrees' -those that survive beetle onslaughts [61]. This could include the deeper understanding of tree genes into a new kind of forest management.

1.4 Actions At The Impacted Locations

In order to mitigate biotic stress (Introduction), several tasks can be applied to the infested sites by the management authorities. Some of the following means can be attempted by prescribing them to the farmers[1,10, 27].

1.4.1 Documented prevention methods

1. Pest surveillance: All the trees in the plantation should be examined for CSRB at the tree base at monthly intervals, especially, during harvest (January-May). If external infestation is observed, CSRB grubs should be removed mechanically by carefully chipping the bark. The injured portion and base of the tree, and any exposed root should be swabbed with cotton plugs drenched in Chlorpyrifos 0.2% solution. [1,62-64]
2. Newly-planted grafts should be trained to bear branches at a height of 0.75m to 1.00m from ground level for facilitating inspection and pest management techniques effectively.[62]
3. For efficient management of another pest capable of destroying the cashew trees - tea mosquito bug (*Helopeltisantonii*), it is important to check the buildup of the pest population on the cashew crop as well as on the alternate hosts such as neem, drumstick, cocoa, guava etc. Tea mosquito bugs can be effectively controlled by three sprays at flushing, flowering and fruiting stages with endosulfan or monocrotophos (0.05 %) for the first and second sprays and carbaryl (0.15 %) for the third spray. In case of severe infestation, it may sometimes lead to die-back caused by a secondary infection of *Botry Diploida Theobromae*. In such instances it will require pruning of the diseased shoots and swabbing of the cut surfaces with 10 % Bordeaux paste and spraying the trees with a 1 % solution of Bordeaux mixture.
4. The effective control measure for CSRB is to remove immature stages of the pest and swabbing the trunk and exposed roots with carbaryl (0.2 %) or neem oil (5 %) and

application of Sevidol 8G (75g/tree) into the basin around the tree. Prophylactic treatment of swabbing the trunk up to one meter height with coal tar and kerosene in the ratio of 1:2 twice a year during March and November could also give effective control. The spray schedule indicated for tea mosquito bug will also be effective against the control of other foliage and inflorescence pests.

1.4.2 Documented Treatment methods

After detecting the presence of CSRB, appropriate treatment or curing measures need to be executed. In order to curb the spread and harm of CSRB and other similar pests, literatures specify the following instructions: -

1. *Mechanical: phytosanitation:* dead trees and trees beyond recovery (those with over 50% damaged bark circumference and / or showing yellowing of the canopy) should be uprooted and removed from the plantation immediately before and after the monsoon, since these serve as natural inoculum for multiplication and spread of the pest [63,64]. Plantation hygiene is to be ensured to prevent harbouring of the insect [62].

2. *Bio-Control Measures:* Entomopathogens like *Metarhizium anisopliae* and *Beauveria bassiana* are known to cause mycosis in grubs of CSRB [62,64]. Spores of *M. anisopliae* survive for three months under field conditions. Mixing the spawn with organic matter like farm yield manure, neem cake and cashew apple, can enhance spore load under field condition [65].

3. *Herbal:* a) Neem-oil swabbing: Application of Neem Oil 5% (50 ml of neem oil + 1,000 ml water +0.5 ml of teepol or 5g of bar soap) on collar portion upto 1 m height and on exposed roots keep the insects away for a period of 3 months.

b) Cashew Nut shell Liquid: Placing CNSL smeared – silk cotton logs in the cashew orchard did not give good attraction. Also, cashew trunks smeared with CNSL recorded more eggs, indicating attraction and increased infestation was noted in all treatments, at the end of the trails. However, the aim of reduction in infestation was obtained on trunks smeared with CNSL @ 2 &4 per block (6.5 and 2.9% respectively).

4. *Chemical:* a) In Brazil, progression of gummosis in cashew trunks could be mitigated by application of Benomyl either alone, or, in combination with copper oxychloride.

b) After pruning of trees, a standard practice is to smear all cut surfaces with Bordeaux mixture paste (10 %) to prevent fungal infections and die-back. A 1 % Bordeaux spray is also administered if the cut surfaces are small.

Out of all above, the Chlorpyrifos 20 EC @ 0.2% found better for the control of cashew stem and root borer [1], however, being a chemical method, it will indeed damage soil composition [66]. Hence, the bio-control measures scores over the former, however theoretically, as its sourcing, storage and application [67] need to be meticulously planned and executed. As of date, they are not under reach of farmers due to lack of training and awareness.

1.4.3 Significance of the scientific detection

A few of the above preventive and treatment measures (sections 1.4.1 and 1.4.2) were practiced at the study location and one of them was pursued premeditatedly with endurance, later found successful (hereafter, the arcane). However, the arcane is ultimately a chemical. Hence, its unrestrained usage will be having adverse effects on both the soil chemistry [68] and the air, to considerate extent. As per farmers expertise, the judicious application of the arcane compulsively be planned only for the specific critical trees, and not rampantly for all the trees of the plantation. The documentations states that the sure indicator of pest attack [27] is as follows: Normally, the pest incidence is noticed during month of Dec-May in the different cashew growing tracts of the country. Different stages of infestation are generally seen all round the year. However, the certain stages of the pest are noticed in certain months only. At the start of monsoon season, the healthy trees turn intense green. However, the infested trees which exhibit yellowish shades are sure indicators of the pest attack onto them. During the nut collection period, close observation reveals the initial infestation symptoms which can be treated suitably prior to onset of monsoon.

Eventually, the intense complexities involved in identifying the specific critical trees are enlisted as follows:

1. The several months taken by the CSRB HC2 (i.e., green attack stage) symptoms to appear in noticeable extent is known as the “latent period”. In the Latent Period, the symptoms are similar to cases of diseases not manifesting symptoms: asymptomatic.
2. Colour change of leaves may not be even at the whole canopy; only a few branches.

3. Other stressors - draught, soil composition, CO₂ and other gaseous pollutants, heat stress, heavy metal toxicity, exposure to UV radiation, herbicide treatments, salinity effects can disguise the same phenomenon.
4. Impracticality of spotting of frass on ground soil at collar for its laboriousness and costs.
5. Rescued trees may be re-infested year after year.

If the said issue is to be handled manually, it calls for exquisite observation by the farmer at every tree. This approach is undoubtedly prone to human errors of genuine indiligence, fostering unnoticed CSRB spread and destructive outbreak. The thriving of the beetle population is a potential disaster for the other plantation areas in entire region and catastrophe for farmer. Hence, an appropriate warning is to be sighted by resorting to timely information fusion. It is a key concept that ensures 1) immediate responses are taken to minimize damages and 2) hastening the prescribed recovery operations 3) opportunity to opt for the other direction of less-harsh management strategies.

Realistically, there is no alternative than resorting to a scientific and technical approach. The detection model solution should possess following characteristics:

1. Detecting the green attack stage for resorting appropriate corrective steps to save the affected trees in danger, just-in-time.
2. Furthermore critically, preventing the future outbreak devastating the entire plantation.

As per relevant science and technology, we planned the stated research objectives and obtained the results, interpreted the outcomes and conducted the early detection criterion.

References

1. A. K. Gupta, “Curative controls of cashew stem and root borer, *Plocaederus ferrugineus* L.”, *Journal of Entomology and Zoology Studies*, 8(4): 1583-1585, 2020. E-ISSN: 2320-7078 P-ISSN: 2349-6800 available at <https://www.entomoljournal.com/archives/2020/vol8issue4/PartY/8-3-228-232.pdf>
2. S. Banerjee and S. L. Shrivastava, “Economic Affairs”, 59(3): 59(3) : 429-437, 2014 doi: 10.5958/0976-4666.2014.00010.2
3. E.V.V. Bhaskara Rao, “Integrated Production Practices Of Cashew In India” Doi:<https://Www.Fao.Org/3/Ac451e/Ac451e04.Htm#Fn4>
4. M.A. Salam and, K.V. Peter, “Cashew, a Monograph”, ISBN 9789380012087, 2010, Studium Press (India).
5. S. C Bhatla, M. A. Lal, “Plant Physiology, Development and Metabolism. Springer Singapore”, 2018. <https://doi.org/10.1007/978-981-13-2023-1>
Biotic Stressors of Plants available at <https://app.biorender.com/biorender-templates/figures/5ffe36639a615d002847708d/t-620b0a94de4067009b2bf0db-biotic-stressors-of-plants>, Reprinted from "Biotic Stressors of Plants", by BioRender, December 2021, Copyright 2022 by BioRender. BioRender (2022).
6. Distribution maps showing the geographical occurrence of pest - cashew stem and root borer, *Plocaederus ferrugineus* L.(Reprinted from CABI, 2019, Infectious hypodermal and hematopoietic necrosis. In: *Invasive Species Compendium*. Wallingford, UK: CAB International)
7. E. U. Asogwa, L. A. Hammed, and T. C. N. Ndubuaku, “Integrated production and protection practices of cashew (*Anacardium occidentale*) in Nigeria”, *African Journal of Biotechnology* Vol. 7 (25), pp. 4868-4873, 2008 Available online at <http://w4868ww.academicjournals.org/AJB>, ISSN 1684–5315
8. P.S. Bhat, K. Vanitha, T.N. Raviprasad, K.K. Srikumar, “Arthropod Communities in Cashew: A Perennial Reservoir of Species Assemblages. In: Chakravarthy, A., Sridhara, S. (eds) *Economic and Ecological Significance of Arthropods in Diversified Ecosystems*” 2016, Springer, Singapore. https://doi.org/10.1007/978-981-10-1524-3_15
9. Dr. K. Vanitha, Dr. T.N. Raviprasad, R. Muthuraju and G.S. Mohan, “Cashew Stem & Root borer” available at <https://cashew.icar.gov.in/pestsite/cashew-stem-root-borer/>

10. Dr. K. Vanitha, Dr. T.N. Raviprasad, “Management of Cashew Stem and Root Borer (CSRB) – major pest of cashew”, ICAR-DCR Training manual on Cashew Production and Post Harvest Technologies. 2020 pp. 98-101 available at <https://krishi.icar.gov.in/jspui/bitstream/123456789/43498/1/e-Manual%20-%20CSRB.pdf>
11. E.A. Dwomoh, S.K. Ahadzie, G. Somuah, A.D. Amenga, “Preliminary studies on the damage symptoms and the spatial distribution of an emerging insect pest, *Mecocorynus* sp. (Coleoptera: Anthribidae) on cashew in Ghana”, 2011 Vol.5(7), pp. 144-151 , July 2011. <https://doi.org/10.5897/JCAB.9000121>
12. *Ostrinia furnacalis* [Online] available at <https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/ostrinia-furnacalis>
13. *Diatraea saccharalis* [Online] available at <https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/diatraea-saccharalis>
14. *Diabrotica virgifera virgifera* [Online] available at <https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/diabrotica-virgifera-virgifera>
15. S. Bhandarkar, R. Prasad, V. Agarwal, H. Ram, Y. Uma, Y. Raghuramulu, and G. R. Kasaragod, Ganesha Raj, “Deep Learning And Statistical Models For Detection Of White Stem Borer Disease In Arabica Coffee”. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2019, XLII-3/W6. 443-451. 10.5194/isprs-archives-XLII-3-W6-443-2019.
16. D. Kutuywayo, A. Chemura, W. Kusena, P. Chidoko, C. Mahoya, “The impact of climate change on the potential distribution of agricultural pests: the case of the coffee white stem borer (*Monochamus leuconotus* P.) in Zimbabwe”, PLoS One. 2013 Aug 27;8(8):e73432. doi: 10.1371/journal.pone.0073432. PMID: 24014222; PMCID: PMC3754943.
17. “Use of stethoscope in early detection of plant borer : A farmer’s offering to the agrarian community” , [Online] available at <https://abhinavkrishi.com/blog/2020/07/07/use-of-stethoscope-in-early-detection-of-plant-borer-a-farmers-offering-to-the-agrarian-community/>

18. J.C.F. Fryer, "Insect Pests of Fruit Crops", pp. 1999, Biotech Books, Daya Publishing House. ISBN: 81-7622-026-4
19. *Cydia Pomonella*[Online] available at <https://www.sciencedirect.com/topics/immunology-and-microbiology/cydia-pomonella>
20. *Synanthedon Exitios* [Online] available at https://entnemdept.ufl.edu/creatures/fruit/borers/peachtree_borer.htm
21. *Rhynchophorus ferrugineus* [Online] available at <https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/rhynchophorus-ferrugineus>
22. *Agrilus Anxius*[Online] available at <https://www.inaturalist.org/taxa/123961-Agrilus-anxius>
23. D. A. Potter and M. F. Potter, "Insect Borers of Trees and Shrubs" available at <https://entomology.ca.uky.edu/ent43>
24. *Agrilus Planipennis* [Online] available at <https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/agrilus-planipennis>
25. A. Vishwajeet, A. Kumar, N. Kumar, L. Rana & S. N. Singh, and A. K. Singh, "(2023). Modern approaches for the detection and prediction of invasive pests. Indian Farming 73 (04), pp. 25-28, 2023.
26. District-wise cashew production in India 2018-19 [Online] available at https://cashew.icar.gov.in/wp-content/uploads/2020/06/2018-19_districtwise-cashew-statistics.pdf.
27. NRCC Technical Bulletin No.:11, March 2012, Published by: The Director, Directorate of Cashew Research, Puttur-574202.
28. H. Abdullah et.al., "European spruce bark beetle (*Ips typographus*, L.) green attack affects foliar reflectance and biochemical properties", International Journal of Applied Earth Observation and Geoinformation, 64, pp 199-209, URL 10.1016/j.jag.2017.09.009, 2018.
29. K. Vanitha, and T.N. Raviprasad, "Biology and Morphometrics of Cashew Stem and Root Borers (CSRB) *Plocaederus Ferrugineus* and *Plocaederus Obesus* (Coleoptera: Cerambycidae) Reared on Cashew Bark", 2013.

30. Top 2 Plant Growth Inhibitors: Abscisic Acid and Ethylene [Online] Available at <https://www.biologydiscussion.com/plants/growth-of-plants/top-2-plant-growth-inhibitors-abscisic-acid-and-ethylene/15738>
31. W. Verelst, A. Skiryecz and D. Inzé, “Abscisic acid, ethylene and gibberellic acid act at different developmental stages to instruct the adaptation of young leaves to stress”, *Plant Signaling & Behavior*, Volume 5, Issue 4, pp. 473-475, 2010, <https://doi.org/10.4161/psb.5.4.11421>
32. R. W. Hofstetter, J. B. Mahfouz, K. D. Klepzig, And M. P. Ayres, “Effects Of Tree Phytochemistry On The Interactions Among Endophloedic Fungi Associated With The Southern Pine Beetle, *Journal of Chemical Ecology*, Vol. 31, No. 3, March 2005 (©2005) DOI: 10.1007/s10886-005-2035-4.
33. J.A. Cale, R.M. Colligno, J.G. Klutsch JG, S.S. Kanekar, A. Hussain A, “Fungal Volatiles Can Act as Carbon Sources and Semiochemicals to Mediate Interspecific Interactions Among Bark Beetle-Associated Fungal Symbionts”, 2016, PLOS ONE 11(9): e0162197. <https://doi.org/10.1371/journal.pone.0162197>.
34. T. Zhao, K. Axelsson, P. Krokene, A. Borg-Karlson, “Fungal Symbionts of the Spruce Bark Beetle Synthesize the Beetle Aggregation Pheromone 2-Methyl-3-buten-2-ol
35. P. Raven, H. Peter, E. F. Ray, S. Eichhorn, “Biology of Plants. New York: W. H. Freeman and Company. ISBN 0-7167-1007-2, 2005.
36. T. Will, A.C. Furch, & M. R. Zimmermann, “How phloem-feeding insects face the challenge of phloem-located defenses”, *Frontiers in plant science*, 4, 336, 2013, <https://doi.org/10.3389/fpls.2013.00336>
37. P.R. Ehrlich, P. H. Raven, "Butterflies and plants: a study of coevolution.". *Evolution. Society for the Study of Evolution.* 18 (4): 586–608. doi:10.2307/2406212, 1964,JSTOR 2406212.
38. G. Fraenkel, "The raison d'être of secondary plant substances". *Science.* 129 (3361): 1466–70. Bibcode:1959Sci...129.1466F, 1959, doi:10.1126/science.129.3361.1466. PMID 13658975.
39. R. H. Whittaker, Robert H., “Communities and ecosystems”, New York: Macmillan, 1975, ISBN 0-02-427390-2.
40. D. Carmona, M. J. Lajeunesse, and M. T.J. Johnson, "Plant traits that predict resistance to herbivores" (PDF). *Functional Ecology.* 25 (2): 358–367, 2011, doi:10.1111/j.1365-2435.2010.01794.x. Retrieved 26 June 2011.

41. "Biochemical defenses: secondary metabolites". Plant Defense Systems & Medicinal Botany. Retrieved 2007-05-21.
42. J. W. Hylin, "Toxic peptides and amino acids in foods and feeds". Journal of Agricultural and Food Chemistry. 17 (3): 492–496, 1969, doi:10.1021/jf60163a003.
43. E. Anet, B. Lythgoe, M. H. Silk, and S. Trippett, "Oenanthotoxin and cicutoxin. Isolation and structures". Journal of the Chemical Society: 309–322, 1953, doi:10.1039/JR9530000309.
44. M. V. Barrow; C. F. Simpson, and E. J. Miller, "Lathyrism: A Review". The Quarterly Review of Biology. 49 (2): 101–128. doi:10.1086/408017, 1974, JSTOR 2820941. PMID 4601279.
45. D. A. Levin, and D. R. King, Dennis, ""The Impact of Fluoroacetate-Bearing Vegetation on Native Australian Fauna: A Review". Oikos. Blackwell Publishing. **61** (3): 412–430, 1991, doi:10.2307/3545249. JSTOR 3545249.
46. P. Biedermann, et.al "Bark Beetle Population Dynamics in the Anthropocene: Challenges and Solutions", Trends in Ecology & Evolution, 2019, 10.1016/j.tree.2019.06.002.
47. D. Kutuywayo, A. Chemura, W. Kusena, P. Chidoko, C. Mahoya, "The Impact of Climate Change on the Potential Distribution of Agricultural Pests: The Case of the Coffee White Stem Borer (*Monochamus leuconotus* P.) in Zimbabwe", PLOS ONE 8(8): e73432. <https://doi.org/10.1371/journal.pone.0073432>
48. M. Havašová, J. Ferencík, and R. Jakuš, "Interactions between windthrow, bark beetles and forest management in the Tatra national parks", Forest Ecology and Management, Volume 391, 2017, Pages 349-361, ISSN 0378-1127, <https://doi.org/10.1016/j.foreco.2017.01.009>. (<https://www.sciencedirect.com/science/article/pii/S0378112717300361>)
49. J.J. Couture, et.al., "Insect herbivory alters impact of atmospheric change on northern temperate forests. Nature Plants, 1(3), 2015, 15016. doi:10.1038/nplants.2015.16
50. L. G. Moraal, "Infestations of the cypress bark beetles *Phloeosinus rufus*, *P. bicolor* and *P. thujae* in The Netherlands (Coleoptera: Curculionidae: Scolytinae)", 140 entomologische berichten 70 (4) 2010 pg.140.
51. R. Berg, C. L. Heald, K. E. Huff Hartz, A. G. Hallar, A. J. H. Meddens, J. A. Hicke, J.-F. Lamarque, and S. Tilmes, "The impact of bark beetle infestations on

- monoterpene emissions and secondary organic aerosol formation in western North America”, *Atmos. Chem. Phys.*, 13, 3149–3161, 2013, www.atmos-chem-phys.net/13/3149/2013/, doi:10.5194/acp-13-3149-2013, © Author(s) 2013. CC Attribution 3.0 License
52. C. Katz, “Small Pests, Big Problems: The Global Spread of Bark Beetles”, 2017 available at <https://e360.yale.edu/features/small-pests-big-problems-the-global-spread-of-bark-beetles>.
 53. R. Ghaffari, F. Zhang; D.D. Iliescu, E.L. Hines, M.S. Leeson, R. Napier, “Detection of Diseases and Volatile Discrimination of Plants: An Electronic Nose and Self-Organizing Maps Approach”, *Intelligent Systems for Machine Olfaction: Tools and Methodologies*; IGI Global: Hershey, PA, USA, 2011; pp. 214–230.
 54. B. Zhou, J. Wang, “Discrimination of different types damage of rice plants by electronic nose”, *Biosyst. Eng.*, **2011**, 109, 250–257.
 55. P. Lehmann, et. al., “Complex responses of global insect pests to climate change”, doi: <https://doi.org/10.1101/425488> Now published in *Frontiers in Ecology and the Environment* doi: 10.1002/fee.2160
 56. C. L. Outhwaite, P. McCann and T. Newbold, “Agriculture and climate change are reshaping insect biodiversity worldwide”.
 57. C. A. Deutsch, J. J. Tewksbury, M. Tigchelaar, D. S. Battisti, S. C. Merrill, R. B. Huey, and R. L. Naylor, “Increase in crop losses to insect pests in a warming climate”, *SCIENCE*, Vol 361, Issue 6405, pp. 916-919, 2018, DOI: 10.1126/science.aat3466.
 58. H.C. Sharma, “Climate change vis-a-vis pest management. Proceedings in conference on national priorities in plant health management”, 2016, Tirupati pp 17-25.
 59. N. Masood et. al., "Insect Pest Management Under Climate Change", 2022, 10.1007/978-3-030-79408-8_15.
 60. <https://www.youtube.com/watch?v=d8TCT5UkjaU> : Do pine beetle fan the flames in western forest.mp4, Courtesy: U.S. Forest service Missoula technology and development center, landsat is joint program of NASA AND USGS: landsat.gsfc.nasa.gov, landsat.usgs.gov, nasa.gov/goddard (goddard space flight center).

61. Bark Beetles are decimating our forests. That might be actually a good thing, [Online] available at <https://www.motherjones.com/environment/2015/03/bark-pine-beetles-climate-change-diana-six/>
62. K. R. Sahu, “Survey Of Insect Pests Of Cashew (*Anacardium Occidentale* L.) And Management Of Key Pest *Plocaederus Ferrugineus* L. (Cashew Stem And Root Borer)”, 2009.
63. M.G. Bhat ,K.V. Nagaraja, And T.R. Rupa, “Cashew research in India. *Journal of Horticultural Sciences*”, 5(1), 1–16, 2010, <https://doi.org/10.24154/jhs.v5i1.483>
64. T. Raviprasad, T., and K. Vanitha, “Pests and Their Management in Cashew”, 2022, https://doi.org/10.1007/978-981-19-0343-4_59.
65. D. Sundararaju, and N. Bakthavatsalam, “Pests of cashew. In: *Advances in Horticulture*”, pp. 759-785,1994, Malhotra Publishing House, New Delhi- 10.
66. D. Sharma, K.R. Sahu, “Management of cashew stem and root borer, *Plocaederus ferrugineus* L. by microbial and plant products”, *Journal of Biopesticides*, pp. 121-123, doi : 1. 10.57182/jbiopestic.1.2., 2008.
67. V. Ambethgar, “Field assessment of delivery methods for fungal pathogens and insecticides against cashew stem and root borer, *plocaederus ferrugineus* L. (cerambycidae: Coleoptera). *Journal of Biopesticides*, 3(1), 121-125, 2010, Retrieved from <https://www.proquest.com/scholarly-journals/field-assessment-delivery-methods-fungal/docview/757064534/se-2>
68. M. D. Chodey, C. N. Shariff, “Pest detection via hybrid classification model with fuzzy C-means segmentation and proposed texture feature”, *Biomedical Signal Processing and Control*, Volume 84, 2023, 104710, ISSN 1746-8094, <https://doi.org/10.1016/j.bspc.2023.104710>.

Chapter 2

Literature Review

This chapter reviews the wide array of research studies conducted in the field of plant disease detection. The diseases can be originated from pathogens and insects, both of which are considered for the review despite the thesis based only on the beetle attack, due to the relevance in major stages of data collection and data processing. The integration of conventional methods, acoustic means, electronic nose, remote sensing, image processing, and Vis-NIR showcases the comprehensive nature of efforts aimed at enhancing plant disease diagnostics.

The numerous works and significant results over the challenge to detect multiple aspects-plant diseases, health and stresses have been compiled earlier in reviews [1-5]. The advanced areas broadly span across sensing by various means: ground(proximal) or aerial (e.g. UAVs [6]) or satellite remote sensing [7-9], invasive or non-invasive, and contact-related (laboratory based, conventional) or contactless.

The various conventional methods for crop disease detection [1] are categorized as

1. **Direct Methods:** Serological methods: Enzyme-Linked Immunosorbent Assay (ELISA), Immune-Fluorescence (IF), Flow-Cytometry. Molecular techniques: Polymerase Chain Reaction (PCR), Fluorescence In-Situ Hybridization (FISH), DNA arrays.
2. **Indirect Methods:** Biomarker-based disease detection: Gaseous metabolite profiling, Plant metabolite profiling.
3. **Sensors:** Biosensor Platforms Based on Nanomaterials, Affinity Bio-sensors, Antibody-Based Biosensors, DNA/RNA-Based Affinity Biosensors, Enzymatic Electrochemical Biosensors, Bacteriophage-Based Biosensors.

These methods encompass techniques that have been historically used to identify and diagnose plant diseases.

Numerous leaf disease-related detection could be tackled through the image-processing techniques as the disease was located on leaves itself.

While for the detection of borers, the point about the process - borer larva making winding beneath the bark, destroying phloem, cambium, outer xylem and girdling the trunk of young trees causes to opt for non-conventional methods. The effectiveness of

these methods has been established over time through extensive research and practical applications. They are detailed in below sections of this chapter.

In only one noteworthy implementation, the unique idea of using apparatus of stethoscope succeeded in preventing the damage and death of the mature bearing palms trees like Coconut, Arecanut etc. and fruit trees like lemons etc due to the Red Palm Weevil (*Ryncophorus ferrugienus*) and save the crop [10].

2.1 Techniques of pest detections

The non-conventional methods have played a significant role in early detection, patterns assessment and outbreaks estimation, which empowers efficient management of pests. Non-destructive methods have gained traction due to their ability to detect plant diseases without causing harm to the plant itself. Moreover, the integration of non-destructive methods with advanced data analysis approaches has led to more accurate disease identification and monitoring.

The literatures on pest detections were based on five primary technologies:

1. **Remote sensing** [7-9]: Imaging by means of Fluorescence, Thermal and Hyperspectral.
2. **Proximal sensing**: Images of pest bodies in caught mode by means of sticky traps [11].
3. **Acoustics-ultrasound**: The pest insects in flying mode were tracked using sound of wing beats [12], and resident mode using sound of boring.
4. **Electronic Nose**: The pest insects in resident mode were tracked using gases of Volatile Organic Compounds [13-15].

Technologies at Sr.No.1 & 2 are dependent on visual information and Sr.No.3 & 4 are hardware-based tracks. The quality of their output is relative to the quality of information and the device specifications respectively.

5. **Spectroscopy techniques** a. Vis-NIR spectroscopy b. Fluorescence spectroscopy.

The non-destructive techniques relevant to CSRB, is a meagre section of the realm of myriad agricultural products testing technologies [Table 2.1].

Table 2.1: The non-destructive technologies for Agricultural and Food Products Quality Inspection [16-18], applicable to CSRB detection.

Principals	Techniques	Measurand
For leaf		
Optics	Reflectance Image Analysis	Colour, Chemical Constituents Colour
For trunk		
Optics	Reflectance Image Analysis	Internal Defects, Chemical Constituents External Defects
Dynamics	Sonic Ultrasonic, X-ray image and CT	Density of Internal Cavity Internal Cavity And Structure Constitution Of Lesioned Tissues
Electro-magnetic Impedance	Olfactory Impedance MR/MRI	Internal Cavity Internal Defects And Structure

In a sole idea of using the X-ray computerized tomography (CT) and image processing for injury detection i.e., observation of the presence of Asian Long-horned beetle (*Anophophora glabripennis*) and damage at the holes inside the wood samples was implemented using a set of eight slices of Eucalyptus grandis trees in United states [19]. Simulation of the algorithm for the linear attenuation coefficient measurements on CT images, for the prepared set generated the density variation graphic. Thus, the results were in the form of images indicating correct identification of the location of residing beetle as well as damage evaluation with time.

The details of literatures on the aforementioned literatures are presented in next sections.

2.1.1 Remote sensing

One of these non-intrusive, non-destructive techniques id referred to as remote sensing. Basically, it was made to satisfy military requirements for thermal imaging and night vision.

A) Fluorescence Imaging Based on changes in the photosynthetic apparatus, and photosynthetic electron transport reactions, this technique measures the chlorophyll fluorescence on the leaves as a function of the incident light and uses the change in

fluorescence parameters to analyse the pathogen infections [20,21]. This method was used to precisely detect leaf rust and powdery mildew infections in wheat leaves by analysing the temporal and spatial variations of chlorophyll fluorescence at 470 nm using Pulse-Amplitude-Modulated (PAM)-Imaging system and monitoring the photochemical efficiency [19]. Although fluorescence measurement offers sensitive photosynthesis abnormality detection, there are few field applications for this technique [22-24].

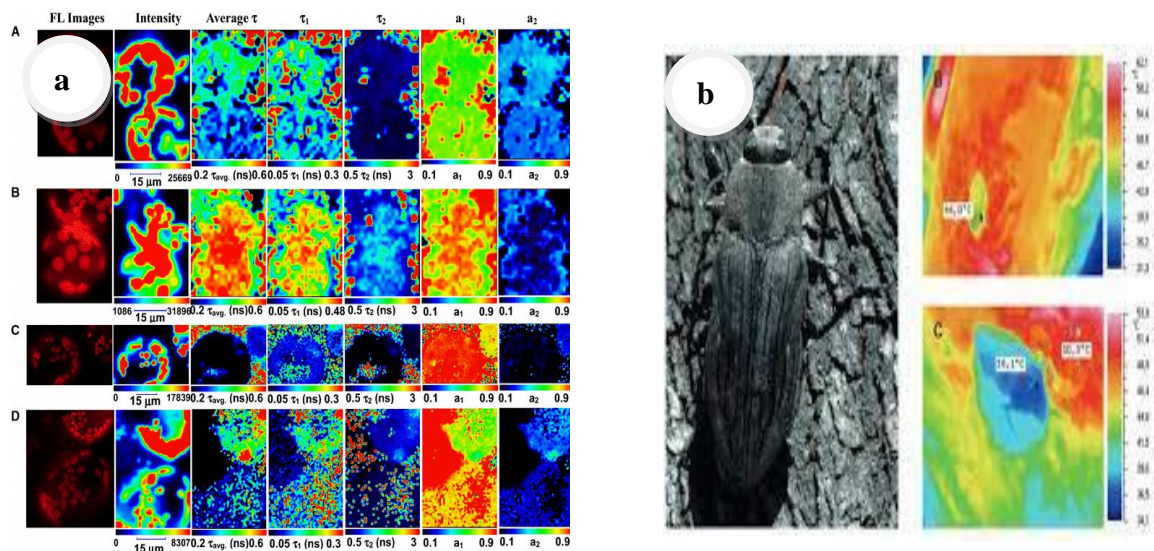


Figure 2.1: Fluorescence technique: development process of images for a) leaf and [25] b) borer, owing to variants [26]

b) Thermography makes it possible to visualize the variations in surface temperatures between plant leaves and canopies. Imaging has a unique ability to detect small temperature variations with and without light, which is used in many different fields today. Images of thermal sensors can be used to take pictures of plants, leaves, and plantations using a non-contact technique called thermal imaging. Thermograms are images that display temperatures or heat generated by objects. Thermograms can be examined and linked to a number of different parameters. The colour difference of the emitted infrared radiation can be examined using thermographic cameras. Previous studies have shown that phytopathogens would have an impact on the water loss in plants controlled by stomata and thus changes in temperatures [27-31]. The amount of water transpired can be measured without being influenced by the ambient temperature using thermographic imaging to monitor the resulting disease [27-31]. It is also possible to enhance the thermographic methods for direct disease detection in plants [32]. A promising method for tracking the diversity of soilborne pathogen infection is thermography [33]. However, because to its great sensitivity to changes in ambient circumstances during measurements, thermography's practical utility is restricted.

Additionally, because thermographic detection is not disease-specific, it cannot be utilized to determine the exact infection or condition. However, because to its great sensitivity to changes in ambient circumstances during measurements, thermography's practical utility for disease monitoring is restricted. Furthermore, because thermographic detection is not disease-specific, it cannot be utilized to determine the kind of infection or differentiate between diseases that have similar thermographic patterns.

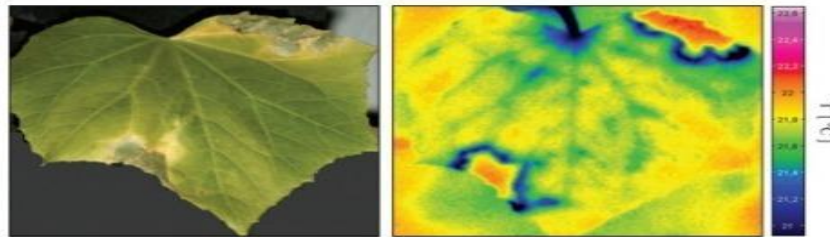


Figure 2.2: Picture of leaf and its thermograph [29].

When pathogens affect the physiological state of plant tissue, alterations in plant metabolism result in the production of heat. Three tomato plant species—Zheza205, Zheza207, and common type—that were grown in a controlled room at Zhejiang University were examined using this method. TMV-U1 was routinely injected into the plants. All plant species were subjected to thermal imaging. Once the plant became infected, its defense mechanism kicked in, producing or synthesizing phytoalexins that get deposited to strengthen the cell wall as well as pathogenesis-related proteins (PRs), of which salicylic acid (SA) acts as a signal in the plant's defense mechanism against pathogen infection. Different diseases have thermographic patterns that are similar to one other. The stomatal aperture is decreased by SA of this generation, which changes the leaf's surface temperature. These leaves' thermal images were recorded and examined with the MikroSpec2.7 program. When the temperature distribution was examined, it was discovered that the temperature of the infected plant tissue was between 0.5 and 1.3 degrees Celsius lower than the temperature of the healthy plant tissues.

c) Hyperspectral Spectroscopy serves as a tool for stress detection [34] and response assessment in plants, particularly in the context of diverse environments and stressors. Spectroscopic sensitivity has enabled the real-time tracking of phytochemical changes induced by various factors [34], including damage and herbivory. In forest ecosystems, light acts not only as a resource but also as a stressor and cue[35], intricately mediating the intricate relationship between diversity and productivity. This relationship is further

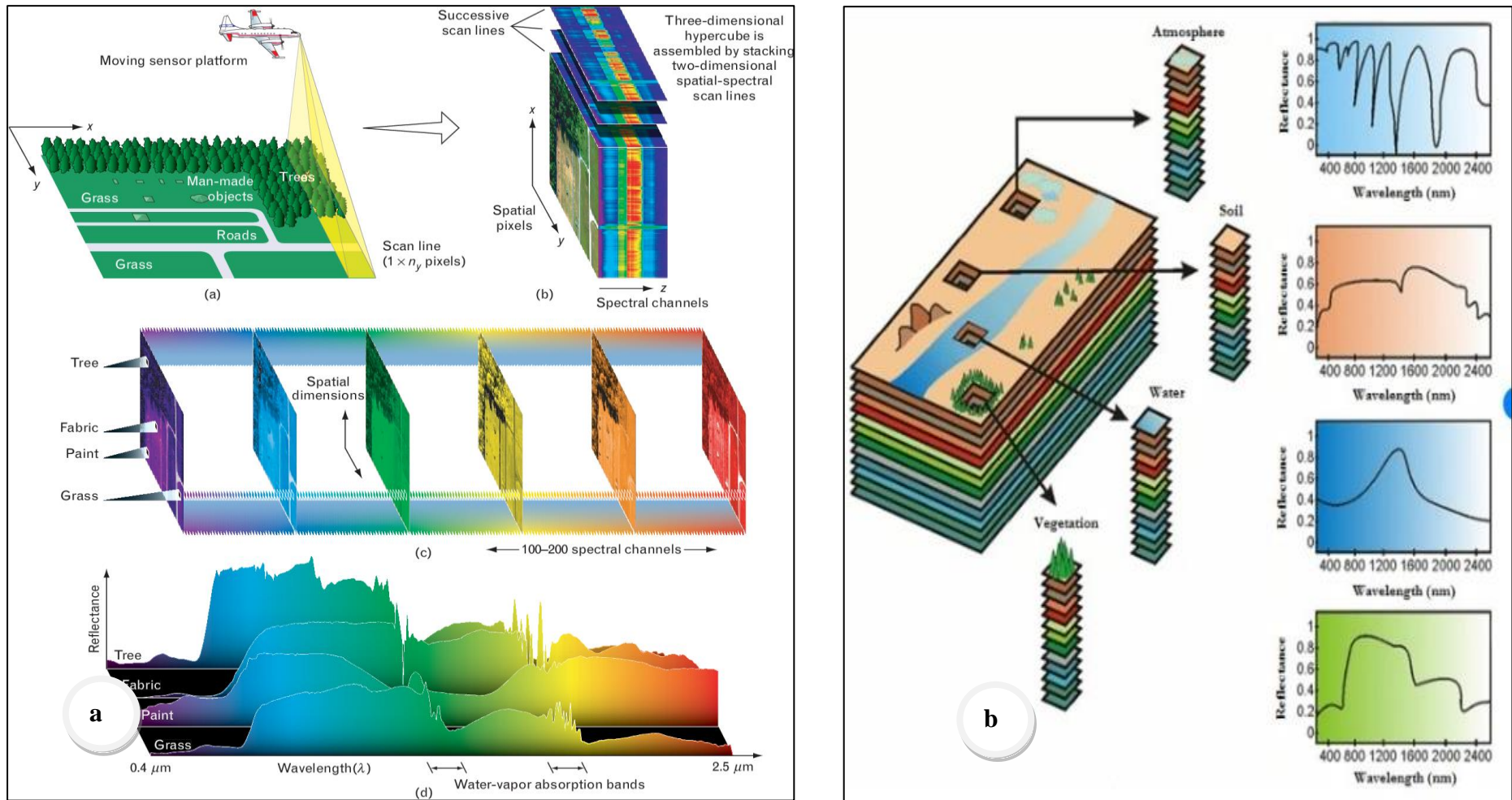


Figure 2.3: Hyper-spectral technique: a) pictorial representation of the entire demonstration: recording setup and fundamental principle for various non-agri components - Tree, Fabric, Paint and Grass [44], b) signature spectra of variants within the agri domain: Atmosphere, Soil, Water and Vegetation [45].

explored through hyperspectral analysis, which offers insights into the chemical and physiological responses of trees facing abiotic and biotic stressors via foliar means [36]. The potential of hyperspectral data extends to characterizing [36] and quantifying [34] these responses, offering a comprehensive view of the effects, even to extent of assessment of effect multiple agents in view of managing protected agro-systems [37]. Notably, hyperspectral techniques have even enabled the detection of plant chemical defences and responses to herbivory. Spectral profiles are positively related to variation in phenolic profiles [37]. Thus, the innovative approach finds applications, in abiotic-related cases such as understanding stress responses in Mediterranean plants [38], phenotyping leaf traits related to water status in maize [39], and in biotic-related cases such as investigating changes in tree canopy properties due to infestations by the Potted Lanternfly [40] and Late blight of tomato and potato caused by the oomycete pathogen *Phytophthora infestans*. By delving into the phylogenetic distribution of leaf spectra and optically derived functional traits, 1) hierarchical organization of plant diversity can be established and, 2) accurate classification of taxa than spectrally-derived traits (which by definition are of lower dimensionality) can be successfully achieved [41]. Satellite-derived vegetation indices, as a prime example, wield computational prowess in predicting occurrences of epiphyllous liverworts, highlighting how intricate mathematical analysis can provide insights into ecosystem dynamics [42]. Furthermore, the utilization of sentinel-2 spectral indices time series showcases the computational strength in identifying multiple stressors in regional agro-ecosystems, unravelling complex relationships through mathematical analysis [43].

2.1.2 Proximal sensing

Since the technique is making use of sticky traps, the beetles are caught in the same. In order to obtain images at the trap, the image capturing device need to be in close vicinity. Hence the type of recording is called proximal sensing [11-12].

2.1.3 Acoustics

The development of acoustic detection techniques has been crucial in the detection of wood- and grain-boring insect pests. The frequency range of 500–1000 Hz exhibited the strongest acoustic sound amplitude. Coefficients obtained through time-frequency wavelet analysis revealed distinct variations in coefficient values between infested trees

(0.3451 and 0.6588) and uninfested trees (0.0062). The identification of infested palms achieved an accuracy of 97%, and the likelihood of correctly identifying un-infested palms exceeded 92%. Performing a secondary assessment further improved the accuracy rate.[46].

To automatically detect RPW acoustic activity in offshoots and then implement it in a prototype setup, the mathematical method used was similar to those used in speech recognition, such as Vector quantization (VQ) or Gaussian mixture modeling (GMM). The algorithm was able to achieve detection ratios of up to 98.9%.[47]

The first work on the acoustic behaviour of cashew stem and root borer characterised the acoustic sounds of *Neoplocaederus ferrugineus* (Linnaeus) (Coleoptera: Cerambycidae), predicted the likelihood of infestation, and field evaluated the detection performance with 91% accuracy [48].

The other works in acoustics are for various pests and other parameters measurements.

The stepwise regression (SR) model was optimal for detecting the *Semanotus bifasciatus* larvae instar [coefficient of determination (R^2) = 0.71, root mean squared error of prediction ($RMSE_p$) = 0.42, and relative percent deviation (RPD) = 3.38] while the best model for predicting larval population size was the partial least squares regression (PLSR) model (R^2 = 0.97, $RMSE_p$ = 61.96, and RPD = 28.87). [49].

Moreover, a previous works on denoising acoustics for the same pest had designed the AI models based on RNN, for time domain denoising (TDD) and frequency domain denoising (FDD): the four variants of TDD-Nets were TDD-Nets using convolution (TDD-Nets-C), TDD-Nets using dilated convolution (TDD-Nets-D), TDD-Nets using dilated convolution with a shortcut (TDD-Nets-DS) and TDD-Nets using dilated convolution with a convolutional shortcut (TDD-Nets-DC) respectively. While the FDD had six kinds of recurrent layers viz. recurrent neural network (RNN), gated recurrent unit (GRU), long short-term memory (LSTM), bidirectional recurrent neural networks (BRNN), bidirectional gated recurrent unit (BiGRU) and bidirectional long short-term memory (BiLSTM). Thus, the corresponding FDD-Nets were called FDD-Nets using RNN (FDD-Nets-R), FDD-Nets using GRU (FDD-Nets-G), FDD-Nets using LSTM (FDD-Nets-L), FDD-Nets using BRNN (FDD-Nets-BR), FDD-Nets using BiGRU (FDD-

Nets-BG), FDD-Nets using BiLSTM (FDD-Nets-BL) respectively. Ultimately, the DD-Net-DC and FDD-Net-BL were selected and for each of them the denoising effects were compared using the metrics of a signal-to-noise ratio (SNR), a segment signal-noise ratio (SegSNR), and log spectral distance (LSD) as follows: The average SNR increment could achieve 17.53 dB and 11.10 dB, The average SegSNR increment achieved 18.59 dB and 12.04 dB, respectively, and the average LSD between pure feeding sounds and denoised feeding sounds were 0.85 dB and 0.84 dB, respectively [50].

Classification between two species, *A. glabripennis* and *H. bajulus* has been achieved with rates of 92% and 96.5% using MLP and LVQ networks. Classification between the two species and noise has been achieved with rates of 88.7% and 95.6% using MLP and LVQ. Classification with an additional species *A. planipennis*, was at 88.8% using an LVQ network [51]

Acoustic analysis applied for detection of old house borer (*Hylotrupes bajulus L.*) [52] using the SVM module for various kernels (RBF, Polynomial and linear) methods of resulted in accuracy of detecting the beginning of the impulse between 84.5 to 88.6%.

2.1.4 Electronic Nose

A collection of studies revolves around the utilization of olfactory sensing [15] for various applications, particularly within the realm of agri-stress [53,54] and food safety [17,18]. These studies explore the use of olfactory receptors as a basis for biosensors and sensing systems that can detect specific substances, including alcohols [17] and volatile organic compounds (VOCs)[13] like acetic acid[14]. They delve into the biological foundations of olfactory systems and their mechanisms, extending their potential in improving food safety measures [17,18]. The studies collectively provide insights into the design and application of olfactory sensing technologies to address critical concerns in the agri and food.

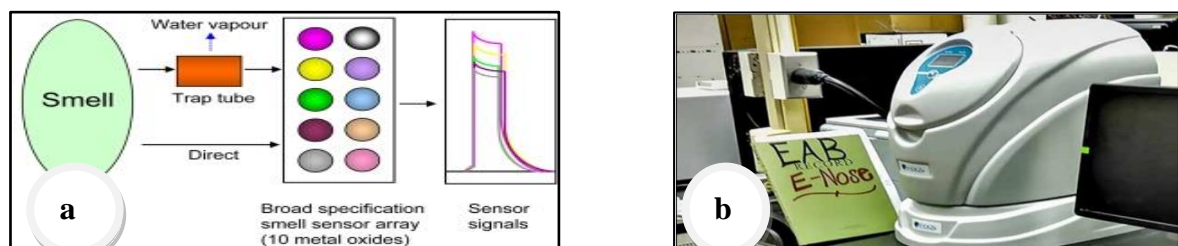


Figure 2.4: Electronic Nose: a) Fundamental Principle of functioning b) Instrument Image by United States Department

2.1.5 Vis-NIR Spectroscopy

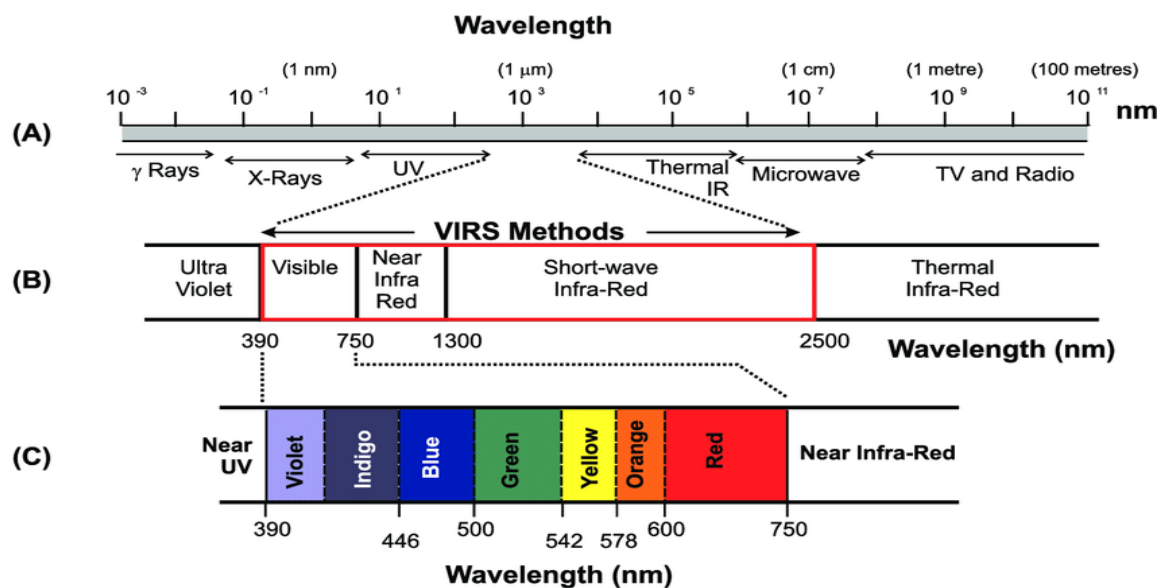


Figure 2.5: The Electromagnetic Spectrum

Spectral studies provide a powerful toolset for

1. understanding both general biological processes and specific processes to biotic stress caused by test they facilitated inside into the inner workings of organisms and environmental interaction with applications from agriculture to medicine.
2. detecting problems early and choose important (in terms of relevancy and effectiveness) management decisions, promptly detect test resistance, reduce chemical use, encourage non-chemical integrated pest management practices, and also assist in breeding plants with desirable traits. Further at a holistic scale, spectral profiling paves a way of food security. The former point is explored and presented in the next chapter, whereas the later point is explained in the next subsections.

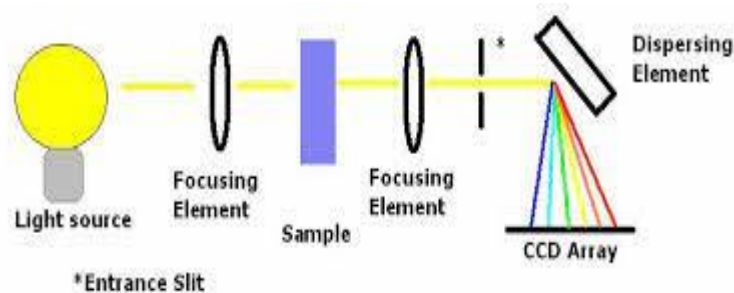


Figure 2.6: A Preliminary assembly of a spectrophotometer.

Figures 2.5 and 2.6 are about fundamental basics for Vis-NIR spectroscopy, former is the Electromagnetic Spectrum and later is the preliminary assembly of a spectrophotometer.

Spectroscopy For Cues to Plant Physiology

The Purdue researchers are using the standard spectroscopy approach: shine a light on a plant and break down the reflected light into its component wavelengths—collectively known as a spectral profile—which correspond to different molecules in the plant. The researchers are looking for certain compounds or combinations of compounds that the plant begins producing as soon as it faces an insect attack, infection with a virus, or some other stressor, such as drought. Since anything affecting a plant's physiology should also affect its spectral profile, this technique has the potential to predict multiple types of stress the plants are under. Early problem detection through spectral profiling has the potential to enhance pest and disease management, quickly spot pest resistance, reduce chemical use, support non-chemical integrated pest management strategies, and help in the breeding of plants with desirable traits. John Couture asserts that the issue is really one of general food security.

One study centres on the differentiating types of vegetation in a coastal wetland by analysing their distinct spectral characteristics [55]. Another study homes in on agricultural settings to detect both biotics dresses biotic stresses (related to living factors like pests and diseases) and abiotic stress (link to environmental factors) in crops. This detection is achieved using specialized methods, such as proximal optical sensing and self-organizing maps for the former [56], and hierarchical self-organization classifiers for both the stresses [57]. In essence, these investigations collectively contribute to an understanding of various stressors affecting vegetation and crops through the advanced sensing and classification approaches. The realm of spectral analysis enriches our comprehension of plant stress and adaptation mechanisms, positioning itself even at the forefront of studies on ecology.

Kooistra et al. (2004) applied the Fieldspec FR spectroradiometer (ASD Inc., Denver, Boulder, USA) to the Vis–NIR–SWIR region and defined the two spectral VIs, the red-edge position (REP) and the difference vegetation index (DVI). The results highlighted the potential of Vis–NIR–SWIR spectroscopy data, for studying the spatial distribution of soil contaminants in floodplains under natural grassland, via the spectral features of the vegetation.

Sahoo et al., (2015) discussed that spectral properties of vegetation are strongly determined by their physiological and biochemical attributes such as leaf area index, the

live and senesced biomass, the pigment and moisture content, and special arrangement of cells and structures.

Application To Pests and Diseases Detection

The previous section highlights that in the biological domain, the individual static characteristics - biochemical, physiological, structural, and morphology changes have been studied using Vis -NIR spectroscopy. However, in the complex biological scenarios, physiological changes during metabolic processes occur successively in sequence or a stepwise manner. and processing techniques applied towards the complex biological scenarios are stated in Table 2.2.

The stages during the process of CSRB infestation are termed as health conditions (details of each constituent stage are mentioned in table 1.1), hence the CSRB infestation is a complex biological process. The literatures on distinguishing the stages of crop diseases by classification of damage degree to have a consistent trend, marked by an increase in reflectance around the 580nm wavelength. This shared discovery highlights the potential significance of the specific reflectance pattern as a valuable indicator for gauging the impact of the biological stressors on plant health. The collection can be divided into two sets of studies. The first set is about European spruce bark beetle attack [58]. In contrast, the second set concentrates on fungal pathogens and includes studies on the detection of multi-tomato leaf diseases [59] and early visual detection of wheat stripe rust using hyperspectral imaging [60], early detection of ganoderma basal stem rot in oil palms [61], and optimum parameters for detecting laurel wilt disease in avocado [62].

The myriad other cases of applying Vis-Nir spectroscopy to biotic stress detection are: In avocado cultivation, researchers have assessed the effectiveness of spectral features and multivariate analysis tools in identifying the laurel wilt disease [63]. Similarly, in strawberries, spectroscopy has been applied in the field to detect anthracnose crown rot, offering a rapid and non-invasive approach to disease identification [64]. Additionally, studies focused on avocado trees have aimed to determine the optimal combination of spectral and geometric parameters that facilitate early detection of laurel wilt disease [65]. The utility of reflectance measurements and neural networks has extended to wheat crops, enabling automated recognition of 'yellow rust' infection [66]. These research destructive disease detection and monitoring across various agricultural contexts.

Even a step ahead, spectral profile will be able to sound alarm that certain insects have

developed resistance to a pest-control method. In this way, spectral profiling will be a part of a larger integrated pest management strategy.

With respect to the borer attack, the tunnelling damage to the vascular tissues would result in loss of nutrients supply to the trees. Corresponding physiological implications would be produced at the two thrust zones and indicated through spectral signatures. The spectral properties of the leaf are affected by several stress factors [31,32,34,35].

Machine learning methods have been used in recent years for crop disease prediction and these efforts have been proved worthwhile. Disease detection using images of crop leaves have been implemented using a pattern recognition branch of machine learning. Nikos Petrellis [67] detected plant diseases using machine-learning methods that achieved speed and accuracy. To identify and classify the diseases in turmeric leaf, C. Rajasekaran et al., [68] used Support Vector Machine, Decision Tree, Naive Bayes. The treatment for the diseases affected to the chilli plants is defined by using rule-based systems and Artificial Bee Colony, by applying machine learning techniques resulting in the best global-optimized solutions for recognizing the diseases in chilli plants.

This categorization provides a clear overview of the different techniques employed in the studies and their respective findings in the context of plant disease detection and classification.

Apart from the Machine Learning algorithms applied to spectroscopy data, remarkable results have been achieved by the Deep Learning methods for the detection of diseases of plants as mentioned in Table 2.4. Infact, the DL has resulted in a paradigm shift in image-based plant phenotyping. A wide range of DL architectures have been used in plant phenotyping [69], including DCNN, RCNN, GoogLeNet, ResNet, SegNet, AlexNet, VGGNet, ResNeXt, SqueezeNet, GAN, LeNet18 and ZFNet. DL architectures have performed well on a broad range of plant phenotyping tasks, such as plant identification based on leaf vein patterns, leaf counting, tassel counting in maize, plant stalk count and stalk width, panicle segmentation, root and shoot localization and feature detection, bloom detection, and plant recognition.

Spectrometry For Foliage: Changes In Chlorophyll (Leaf)

Because of the fundamental role of chlorophylls (Chls) in photosynthesis, researchers in the first half of last century, studied the effects of different light regimes, nutrients and other growth conditions on the efficiency of various photosynthetic reactions including O₂ evolution, CO₂ fixation, or carbohydrate biosynthesis. The method of obtaining spectra – selection of solvent and then equations for calculating the chlorophyll content have been evolved and critically updated over the past decade. Early studies were done by McKinney and compiled in 1941[70]. Arnon formulated Mc Kinney's work by 1949 [71].

R. J. Porra began exploration in 1989 so as to incorporate the corrections to the formulation of meaningful models of the photosynthetic apparatus [72]. Accordingly with his continuous works [73-74], and analysing the history on the development of accurate simultaneous equations by many research groups, he documented the grave flaws due to inaccurate spectrophotometric data in Arnon equations in 2002[75]. He brought out the difficulties and anomalies arising from their retention, and advises to abolish the usage as they are were slower, more difficult, and subsequent resultant of confusion upon continued retention. Even later in 2018, he stressed on need to apply the method to correct the data obtained by Arnon's procedure[76]. The disconcerting situation of citing and using the Arnon's assay of chlorophylls a and b extracted in aqueous 80% acetone, by even the high impact factor journals [77] is abolished by R.J. Porra [76].

Apart from all the above lineage, parallely individuals have tracked own methodologies. Lichtenthaler studied the pigments of photosynthetic bio-membranes: Chlorophylls and Carotenoids, for Acetone 100% (pure solvent), Acetone 80% (v/v), Methanol 100% (pure solvent)and Acetone 90% (v/v). Alan R. Wellburn formulated a collection of equationsets for various spectrometers of different resolutions (0.1-0.4 nm and 1-4 nm).The concentrations of chlorophylls and carotenoids were determined spectrophotometrically and calculated per unit fresh mass basis using the equations of Lichtenthaler and Wellburn [78]. The overtones of chlorophyll and carotenoids were brought out by R.Withnall and H. Schulz, while H. Schulz explored reasoning i.e. molecular fundamentals governing the carotenoids [79-80] signature. By the virtue to extensive works on the most suitable solvent for chlorophyll [81], Raymond Ritchie and Michael M. Blanke recommended the determination of chlorophyll using DMSO. J.D. Barnes illustrated use of DMSO for determination of chlorophylls a and b in lichens and

higher plants. Extraction of Chlorophyll using Dimethyl Sulfoxide and Acetone was experimented by Kurtis Allain. Bhagat Chand *et al* documented effect of solvent polarity by analysing the extraction efficiencies of acetone (80%), ethanol, DMSO, and DMF for chlorophyll extraction from mango, pine, spinach, and hibiscus.

Spectrometry For Trunk: Changes In Cell Components

The application of spectroscopy to estimate concentrations of trunk composition has not been attempted earlier. It's bizarre that only meagre literatures for other purposes are available, as listed: -

1. Wood spectra[82]: 70 years of NIR band assignment literature for wood and wood components had been reviewed, to establish the band assignments for purpose of practical considerations important for the application of NIR to wood conclusions to be drawn on the chemistry and physio-chemical properties of samples.
2. Saw-dust spectra[83]: importance of selecting a suitable reference wavelength in a two-wavelength device. In cases where it cannot be assured that the size of sample will be uniformly presented to the sensor, the reference wavelength needs to be closer to the absorption wavelength.

Moreover, they are based on only woods, thus the intense point is that the substance is a product - derived case.

Hence, none other than for the plant foliage, the improvements in the sensitivity and portability of spectrometers as well as advances in computation power and chemometric modelling methods has been harnessed immensely to estimate specific chemical constituents [55-57].

Table 2.2: Data-collection and processing techniques for studying metabolic processes at the different or moreover complex biological scenarios.

Aim	Data collected	Results	Conclusions	Technology
Evaluating metabolic changes during ripening of citrus fruit.	FT-IR spectral and HPLC data of fruit peels and flesh of Citrus fruit (Citrus unshiu Marc. cv. Miyagawa)	During citrus fruit ripening: decrease in the amide region (1500–1700 cm ⁻¹), increase in the carbohydrate region (1000–1200 cm ⁻¹). During citrus fruit development: increase in β -cryptoxanthin and decrease in lutein. or each of pigments of the citrus fruit, antheraxanthin, β -carotene, β -cryptoxanthin and violaxanthin, the R2 was 0.9117, 0.8816, 0.8856 and 0.7305. content.	FT-IR in combination with multivariate analysis enables discrimination of metabolic variation during fruit development, and prediction of carotenoid content from citrus fruit.	PLS-discriminant analysis (PLS-DA) of FT-IR spectral data and PLS regression modelling of carotenoid content.
Real-time monitoring of alcalase hydrolysis of egg white protein.	NIR data of Egg-white, Reflectance of egg white -freeze-dried, ground into power as natural EWP.	$R_p^2 = 0.9525$.	The relative error less than 10% for 10 independent samples confirmed the stability and reliability	Efficient variables selection algorithms – multi-variate calibrations; the partial least squares-uninformative variables elimination and ant colony optimization.

Table 2.3: Results achieved by the Machine Learning methods for the detection of diseases of plants.

Sr. No.	Crop, Pest Target zone (samples), Data collected,	Assessment aim	Results	Algorithm
1	Greenhouse Pepper (Capsicum Annuum), Two-spotted spider mites (TSSM; Tetranychusurticae Koch) Leaves, Reflectance data transformed into vegetation indices.	Spectral detection of the destruction of chloroplast-containing cells,	Early identification by multispectral means.	ANOVA (One-way analysis of variance).
2	Strawberries (Fragaria spp.), Two-spotted Spider Mite (Acari: Tetranychidae), Leaflets, Spectral regions.	Classes of density (mites/leaflet): Low infestation = 0-20, Moderate infestation = 20-50, High infestation =50	Separation between leaf damage levels, in the reflectance. Mite density class (categorical): validation accuracy = 81%. Mite density(continuous): High validation efficiency, R2 = 0.85, Root-mean-squared error (RMSE)= 12.2 mites per leaf	Partial Least Squares (PLS) regression
3	Citrus species, citrus HLB-associated bacteria (Candidatus Liberibacter asiaticus, CLas) Leaves, Spectra from MicroNIR 1700 Spectrometer,	Characterization of the chemical differences between HLB- infected and healthy leaves rapidly and non-destructively.	Correct rejection rate of CLas-negative leaves by Principal Component Analyses (PCA) model: more than 80%, & PLS-DA model: more than 83%.	Soft Independent Modeling of Class Analogy(SIMCA) and Partial Least Squares Discriminant Analysis (PLS-DA) & RT-qPCR for verification.

2.2 Objectives

Following are the objectives to attain the goal of designing a model to detect the CSRB infestation at early stage:

1. Performing the Spectral Analysis of the bark of healthy as well as infested cashew plant, and analyzing the results.
2. Chlorophyll content analysis of healthy as well as infested cashew plant, and analyzing the results.
3. Frequency response of healthy as well as infested cashew plant, and analyzing the results.
4. Developing the best model for early detection by combining bark and leaf spectral information.
5. Designing a classification model for deciding the status of a productive plant, with respect to presence of infestation.

Hence, this chapter concludes to select the spectroscopy technology for further research as it is a fast, low-cost, accurate, multi-analytical technique based on the electromagnetic absorption of organic compounds at wavelength ranges [31,32]. This technique comprises both qualitative and quantitative analyses of multiple component samples by a single measurement. Also, as no reagents are employed, and no hazardous waste is produced, it is environmentally friendly. The Vis-NIR technique has been applied in a wide range of fields in the last decades, and several of these applications are currently used for routine analyses, including online monitoring.

The next chapter explains the description of the theories and fundamentals of technologies utilized to achieve the objectives.

Table 2.4: Results achieved by the Deep Learning methods for the detection of plant diseases

Authors	Crop	Aim	Algorithm	Results
Nurul Nabilah Ahmad Loti, Mohamad Roff Mohd Noor	Chili: 5 types of diseases, 2 types of pest infestations, and 1 healthy type.	Integrated analysis of machine learning and deep learning for the identification task of 974 chili leaf images	Three classifiers: 1) Support vector machine (SVM), 2) Random forest (RF), 3) Artificial neural network (ANN)	Deep learning feature-based approaches performed better than the traditional feature-based approaches. SVM classifier gave best accuracy 92.10%.
Huiqun Hong, Jinfa Lin, et.al.	Grape Leaf: Black Rot, Black Measles, Leaf Blight And Phylloxera.	Disease Detection and Classification Using Machine Learning. Adjust 1) different learning rates, 2) appropriate training parameters, and test those networks	Transfer learning technique based on three pre-trained machine learning models (VGG16, MobileNet, and AlexNet).	Comparative evaluation results show better accuracy and performance improvement compared with pre-trained models. An ensemble model improves final detection and classification accuracy.
Ally Qin, Jingshu Lu, Aparna Menon, Jerry Gao, et.al.	Tomato Leaves: 9 types of leaves diseases.	appropriate training parameters, and test those networks	Five Deep Learning networks were applied to perform the feature extraction	Best recognition accuracy 1) Densenet_Xception, with most parameters = 97.10%, 2) ShuffleNet, with least parameters is = 83.68%.

References

1. Y. Fang, R.P. Ramasamy, “Current and Prospective Methods for Plant Disease Detection”, *Biosensors (Basel)*. 2015 Aug 6;5(3):537-61. doi: 10.3390/bios5030537. PMID: 26287253; PMCID: PMC4600171.
2. F. Martinelli et.al. “Advanced methods of plant disease detection. A review”, *Agronomy for Sustainable Development*, Springer Verlag/EDP Sciences/INRA, 2015, 35 (1), pp.1-25. 10.1007/s13593-014-0246-1.hal-01284270
3. A.K. Mahlein et.al. “Recent advances in sensing plant diseases for precision crop protection”, *Eur J Plant Pathol* 133, 197–209 (2012). <https://doi.org/10.1007/s10658-011-9878-z>
4. O.W. Liew, et.al, “Signature Optical Cues: Emerging Technologies for Monitoring Plant Health”, *Sensors*. 8. 10.3390/s8053205, 2008.
5. A. Galieni et. al, “Past and Future of Plant Stress Detection: An Overview From Remote Sensing to Positron Emission Tomography”, *Front Plant Sci.*, 2021
6. J. G. A. Barbedo, “A Review on the Use of Unmanned Aerial Vehicles and Imaging Sensors for Monitoring and Assessing Plant Stresses”, *Agricultural Informatics*, 2019.
7. A. Meddens et.al, “Evaluating methods to detect bark beetle- caused tree mortality using single-date and multi-date Landsat imagery”, *Remote Sensing of Environment*, 2013.
8. S. Ortiz, J. Breidenbach, and G.Kändler, “Early detection of Bark beetle green attack using TerraSAR-X and Rapid Eye Data”, *Remote Sensing*, 2013.
9. R. M. Hoffer, “Biological and physical considerations in applying computer-aided analysis techniques to remote sensor data, in *Remote Sensing: The Quantitative Approach*”, P.H. Swain and S.M. Davis (Eds), McGraw-Hill Book Company, 227-289, 1978.
10. Use of stethoscope in early detection of plant borer: A farmer’s offering to the agrarian community available at <https://abhinavakrishi.com/blog/2020/07/07/use->

of-stethoscope-in-early-detection-of-plant-borer-a-farmers-offering-to-the-agrarian-community/

11. K. Espinoza, et.al “Combination of image processing and artificial neural networks as a novel approach for the identification of Bemisiatabaci and Frankliniella occidentalis on sticky traps in green house agriculture”, *ComputersandElectronicsinAgriculture*127(C)pp495-505.
12. Y. Chen et. al. “Flying Insect Detection and Classification with Inexpensive Sensors”, *Journal of Visualized Experiments* 92(e52111) URL<https://doi.org/10.3791/52111>,2014.
13. M. Valencia-Ortiz et.al, “VOCs Emission Profiles Associated with Plant-Pest Interaction for Phenotyping Applications. *Sensors*. 2022; 22(13):4870. <https://doi.org/10.3390/s22134870>
14. S. Panigrahi, et.al. “Olfactory receptor-based polypeptide sensor for acetic acid VOC detection”, *Materials Science and Engineering: C*, Volume 32, Issue 6, 2012,Pages 1307-1313, ISSN 0928-4931, <https://doi.org/10.1016/j.msec.2011.11.003>
15. S. Sankaran, et.al. “Biology and applications of olfactory sensing system: A review, *Sensors and Actuators B: Chemical*, Volumes 171–172, 2012, Pages 1-17, ISSN 0925-4005, <https://doi.org/10.1016/j.snb.2012.03.029>.
16. H. Elmesiry, et.al., “Applications of Non-destructive Technologies for Agricultural and Food Products Quality Inspection. *Sensors*”,19, 846, 2019, 10.3390/s19040846.
17. S. Sankaran, et.al., “Olfactory receptor based piezoelectric biosensors for detection of alcohols related to food safety applications, *Sensors and Actuators B: Chemical*, Volume 155, Issue 1, 2011, Pages 8-18, ISSN 0925-4005, <https://doi.org/10.1016/j.snb.2010.08.003>.
<https://www.sciencedirect.com/science/article/pii/S0925400510006647>
18. S. Sankaran, et.al. “Olfactory Sensing: Overview of the Biology, Mechanism and Applications for Food Safety”, <https://doi.org/10.13031/2013.22374>, 2006,

- Olfaction Sensing Mechanism, Food Safety, Sensors. URL - <https://elibrary.asabe.org/abstract.asp?aid=22374&t=8>.
19. P. Cruvinel, et.al., “Detection of beetle damage in forests by X-ray CT image processing”, *Revista Árvore*. 27. 10.1590/S0100-67622003000500017.
 20. L. Chaerle, et. al, “Multi-sensor imaging of plant stresses: Towards the development of a stress-catalogue.
 21. J. Kuckenberg, I. Tartachnyk, and G. Noga, “Temporal and spatial changes of chlorophyll fluorescence as a basis for early and precise detection of leaf rust and powdery mildew infections in wheat leaves. *Precision Agric* 10, 34–44 (2009). <https://doi.org/10.1007/s11119-008-9082-0>
 22. L. Chaerle et.al., “multi-sensor imaging of plant stresses: Towards the development of a stress-catalogue. 2009
 23. L. Cséfalvay, et.al., “pre-symptomatic detection of *Plasmopara viticola* infection in grapevine leaves using chlorophyll fluorescence imaging. *European Journal of Plant Pathology*, 125, 2009, 291-302. 10.1007/s10658-009-9482-7.
 24. J. Scholes, and R. Stephen, “Chlorophyll fluorescence imaging as tool for understanding the impact of fungal diseases on plant performance: A phenomics perspective. *Functional Plant Biology*, 2009, 10.1071/FP09145.
 25. R. Lei, et. al., “Chlorophyll fluorescence lifetime imaging provides new insight into the chlorosis induced by plant virus infection”, *Plant Cell Reports*, 36, 2017,10.1007/s00299-016-2083-y.
 26. J. Al-doski, “Thermal imaging for pests detecting-a review”, <https://api.semanticscholar.org/CorpusID:11593688>, 2016.
 27. L. Chaerle, I. Leinonen, H.G. Jones, and D.V. D. Straeten, “Monitoring and screening plant populations with combined thermal and chlorophyll fluorescence imaging. *J. Exp. Bot.* 2007, 58, 773–784.
 28. M. Lindenthal, U. Steiner, H.W. Dehne, E.C. Oerke, “Effect of downy mildew development on transpiration of cucumber leaves visualized by digital infrared thermography”, *Phytopathology* 2005, 95, 233–240.

29. E. Oerke, U. Steiner, H.W. Dehne, and M. Lindenthal “Thermal imaging of cucumber leaves affected by downy mildew and environmental conditions”, *J. Exp. Bot.*, 2006, 57(9), 2121–2132, <https://doi.org/10.1093/jxb/erj170>.
30. E.C. Oerke, P. Fröhling, and U. Steiner, “Thermographic assessment of scab disease on apple leaves, *Precis. Agric.* 2011, 12, 699–715.
31. M. Stoll, H.R. Schultz, G. Baecker, and B. Berkelmann-Loehnertz, “Early pathogen detection under different water status and the assessment of spray application in vineyards through the use of thermal imagery”, *Precis. Agric.*, 2008, 9, 407–417.
32. M. Lindenthal, “Visualisierung der Krankheitsentwicklung von Falschem Mehltau an Gurken durch *Pseudoperonospora cubensis* mittels Thermografie”, Dissertation, University of Bonn, Germany, 2005.
33. C. Hillnhütter, A.K. Mahlein, R.A. Sikora, and E.-C. Oerke, “Remote sensing to detect plant stress induced by *Heterodera schachtii* and *Rhizoctonia solani* in sugar beet fields”, *Field Crops Res.*, 2011, 122, 70–77.
34. J. J. Couture, “Spectroscopic sensitivity of real-time, rapidly induced phytochemical change in response to damage” *New Phytol.* 2013 Apr;198(1):311-319. doi: 10.1111/nph.12159. Epub 2013 Feb 5. PMID: 23384059.
35. L. Williams, “Light as a Resource, Stressor and Cue in Mediating Diversity-Productivity Relationships in Forests”, vol. 2021, 2021.
36. S. Park, et.al., “Characterizing Tree Chemical and Physiological Responses to Abiotic and Biotic Stress using Hyperspectral Information”, AGU Fall Meeting Abstracts, vol. 2022, pp. B52B-05. 2022.
37. C. Lorenzo, and J. J. Couture, “Hyperspectral assessment of plant responses to multi-stress environments: Prospects for managing protected agrosystems”, *Plants, People, Planet* 2, no. 3 (2020): 244-258.
38. C. Lorenzo, et al., “Spectroscopy: an efficacious approach for better understanding of stress responses of Mediterranean plants”, Book of abstracts, 2nd CAPERmed meeting, p. 11. 2016.

39. C. Lorenzo, et al., "Spectral phenotyping of physiological and anatomical leaf traits related with maize water status", *Plantphysiology* 184, no. 3 (2020): 1363-1377.
40. E.G. Joll, et al., "Changes In Tree Canopychemical And Spectral Properties In Response To Spotted Lanternfly (*Lycormadelicatula*) Infestations", PhD diss., Purdue University Graduate School, 2023.
41. J. Cavender-Bares, et.al., "Phylogenetic Distribution of Leaf Spectra and Optically Derived Functional Traits in the American Oaks", *AGU Fall Meeting Abstracts*, vol. 2015, pp. B51K-08. 2015.
42. A.de Castro, Ana I., Yeyin Shi, Joe Mari Maja, and Jose M. Peña. 2021. "UAVs for Vegetation Monitoring: Overview and Recent Scientific Contributions" *Remote Sensing* 13, no. 11: 2139. <https://doi.org/10.3390/rs13112139>.
43. M. Liu, X. Liu, Y. Meng, A. K. Skidmore and T. Wang, "Identifying multiple stressors in regional agro-ecosystems based on sentinel-2 spectral indices time series," *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, Valencia, Spain, 2018, pp. 9335-9338, doi: 10.1109/IGARSS.2018.8651434.
44. N. Sulaiman N, et.al., "The Application of Hyperspectral Remote Sensing Imagery (HRSI) for Weed Detection Analysis in Rice Fields: A Review", *Applied Sciences*. 2022; 12(5):2570. <https://doi.org/10.3390/app12052570>
45. J. Plaza, A. Plaza, and C. Barra, "Multi-Channel Morphological Profiles for Classification of Hyperspectral Images Using Support Vector Machines. *Sensors* (Basel, Switzerland). 9. 196-218. [10.3390/s90100196](https://doi.org/10.3390/s90100196).
46. K.A.P. Siriwardena, L.C.P. Fernando, N. Nanayakkara, K.F.G. Perera, A.D.N.T. Kumara, T. Nanayakkara, Portable acoustic device for detection of coconut palms infested by *Rynchophorus ferrugineus* (Coleoptera: Curculionidae), *Crop Protection*, Volume 29, Issue 1, 2010, Pages 25-29, ISSN 0261-2194, <https://doi.org/10.1016/j.cropro.2009.09.002>.
47. J. Pinhas, V. Soroker, A. Hetzroni, A. Mizrach, M. Teicher, and J. Goldberger, "Automatic acoustic detection of the red palm weevil", *Computers and Electronics in Agriculture*, 63, 131-139, 2008, [10.1016/j.compag.2008.02.004](https://doi.org/10.1016/j.compag.2008.02.004).

48. R. Maruthadurai, T. Veerakumar, ChannabasavaVeershetty, A.N. S. Chakaravarthi, “Acoustic detection of stem and root borer *Neoplocaederus ferrugineus* (Coleoptera: Cerambycidae) in cashew, *Journal of Asia-Pacific Entomology*, Volume 25, Issue 3, 2022, 101968, ISSN 1226-8615, <https://doi.org/10.1016/j.aspen.2022.101968>.
49. Q. Jiang, Y. Liu, L. Ren, Y. Sun, and Y. Luo, “Acoustic detection of the wood borer, *Semanotus bifasciatus*, as an early monitoring technology”, *Pest Manag Sci*, 78, 4689-4699, 2022, <https://doi.org/10.1002/ps.7089>.
50. X. Liu, H. Zhang, Q. Jiang, L. Ren, Z. Chen, Y. Luo, and J. Li, “Acoustic Denoising Using Artificial Intelligence for Wood-Boring Pests *Semanotus bifasciatus* Larvae Early Monitoring”, *Sensors*, 2022, 22(10):3861, <https://doi.org/10.3390/s22103861>.
51. J.Schofield, “Real-time Acoustic Identification of Invasive Wood-boring Beetles” [Online] Available at <https://core.ac.uk/download/pdf/1145982.pdf>. Bilski, Piotr & Bobiński, Piotr & Witomski, Piotr. (2017).
52. P. Bilski, P. Bobiński, A. Krajewski, and P. Witomski, “Detection of Wood Boring Insects’ Larvae Based on the Acoustic Signal Analysis and the Artificial Intelligence Algorithm”, *Archives of Acoustics*, 42,1, pp. 61-70, 2017, 10.1515/aoa-2017-0007.
53. S. Cui, P. Ling, H. Zhu, and H. M. Keener, "Plant Pest Detection Using an Artificial Nose System: A Review", *Sensors* 18, 2018, no. 2: 378, <https://doi.org/10.3390/s18020378>.
54. S. Cui, E. A. A. Inocente, N. Acosta, H. M. Keener, H. Zhu, and PP Ling, “Development of Fast E-nose System for Early-Stage Diagnosis of Aphid-Stressed Tomato Plants”, *Sensors (Basel)*, 2019, 19(16), 3480, doi: 10.3390/s19163480. PMID: 31395823, PMCID: PMC6721161.
55. K.S. Schmidt, A.K. Skidmore, “Spectral discrimination of vegetation types in a coastal wetland”, *Remote Sensing of Environment*, Volume 85, Issue 1, 2003, Pages 92-108, ISSN 0034-4257, [https://doi.org/10.1016/S0034-4257\(02\)00196-7](https://doi.org/10.1016/S0034-4257(02)00196-7).
56. D. Moshou, C. Bravo, S. Wahlen, et. al., “Simultaneous identification of plant stresses and diseases in arable crops using proximal optical sensing and self-

- organising maps”, *Precision Agric* 7, 149–164, 2006, <https://doi.org/10.1007/s11119-006-9002-0>.
57. X.E. Pantazi, et. al., “Detection of biotic and abiotic stresses in crops by using hierarchical self-organizing classifiers” *Precision Agric* 18, 383–393, 2017, <https://doi.org/10.1007/s11119-017-9507-8>.
58. H. Abdullah, R. Darvishzadeh, A. Skidmore, T. Groen, and M. Heurich, “European spruce bark beetle (*Ips typographus*, L.) green attack affects foliar reflectance and biochemical properties”, *International Journal of Applied Earth Observation and Geoinformation*, 64, 199-209, 2018, [10.1016/j.jag.2017.09.009](https://doi.org/10.1016/j.jag.2017.09.009).
59. J. Lu, R. Ehsani, Y. Shi, A. De Castro, and S. Wang, “Detection of multi-tomato leaf diseases (late blight, target and bacterial spots) in different stages by using a spectral-based sensor”, *Scientific Reports*, 8, 2018, [10.1038/s41598-018-21191-6](https://doi.org/10.1038/s41598-018-21191-6).
60. Z. Yao, Y. Lei, and D. He, “Early Visual Detection of Wheat Stripe Rust Using Visible/Near-Infrared Hyperspectral Imaging”, *Sensors*, 2019, 19(4):952. <https://doi.org/10.3390/s19040952>.
61. P. Ahmadi, F. M. Muharam, K. Ahmad, S. Mansor, and I. A. Seman, “Early Detection of Ganoderma Basal Stem Rot of Oil Palms Using Artificial Neural Network Spectral Analysis”, *Plant Disease* 2017 101:6, 1009-1016.
62. A. I. de Castro, R. Ehsani, R. Ploetz, C. Jonathan, and J. Abdulridha, “Optimum spectral and geometric parameters for early detection of laurel wilt disease in avocado”, *Remote Sensing of Environment*, 171, 33-44, 2015, [10.1016/j.rse.2015.09.011](https://doi.org/10.1016/j.rse.2015.09.011).
63. J. Abdulridha, Y. Ampatzidis, R. Ehsani, and A. I. de Castro, “Evaluating the performance of spectral features and multivariate analysis tools to detect laurel wilt disease and nutritional deficiency in avocado”, *Computers and Electronics in Agriculture*, Volume 155, 2018, Pages 203-211, ISSN 0168-1699, <https://doi.org/10.1016/j.compag.2018.10.016>.
64. J. Lu, R. Ehsani, Yeyin Shi, J. Abdulridha, A. I. de Castro, Y. Xu, “Field detection of anthracnose crown rot in strawberry using spectroscopy technology”, *Computers*

- and Electronics in Agriculture, Volume 135, 2017, Pages 289-299, ISSN 0168-1699, <https://doi.org/10.1016/j.compag.2017.01.017>.
65. A.I. De Castro, R. Ehsani, R. Ploetz, J.H. Crane, and J. Abdulridha, “Optimum spectral and geometric parameters for early detection of laurel wilt disease in avocado”, *Remote Sensing of Environment*, Volume 171, 2015, Pages 33-44, ISSN 0034-4257, <https://doi.org/10.1016/j.rse.2015.09.011>.
 66. D. Moshou, C. Bravo, J. West, S. Wahlen, A. McCartney, and H. Ramon,” Automatic detection of ‘yellow rust’ in wheat using reflectance measurements and neural networks”, *Computers and Electronics in Agriculture*, Volume 44, Issue 3, 2004, Pages 173-188, ISSN 0168-1699, <https://doi.org/10.1016/j.compag.2004.04.003>.
 67. N. Petrellis, “Exploring Feature Correlation in Plant Disease Diagnosis -A Smart Phone Application for Tomato Disorder Classification”, [Online] Available at <https://dl.acm.org/doi/pdf/10.1145/3503823.3503825>.
 68. C. Rajasekaran, S. Arul, S. Devi, G. Gowtham, and S. Jeyaram, “Turmeric Plant Diseases Detection and Classification using Artificial Intelligence”, 1335-1339. 10.1109/ICCSP48568.2020.9182255.
 69. A.K. Singh, B. Ganapathysubramanian, S. Sarkar, A. Singh, “Deep Learning for Plant Stress Phenotyping: Trends and Future Perspectives”, *Trends Plant Sci.* 2018 Oct;23(10):883-898. doi: 10.1016/j.tplants.2018.07.004.
 70. B. G. Mackinney, “Absorption Of Light By Chlorophyll Solutions.” *Journal of Biological Chemistry* 140 (1941): 315-322.
 71. D. I. Arnon, “Copper Enzymes in Isolated Chloroplasts. Polyphenoloxidase in *Beta Vulgaris*”, *Plant Physiology*, Volume 24, Issue 1, January 1949, Pages 1–15, <https://doi.org/10.1104/pp.24.1.1>
 72. R.J .Porra, W.A. Thompson, and P.E. Kriedemann, “Determination of accurate extinction coefficients and simultaneous equations for assaying chlorophylls a and b extracted with four different solvents: verification of the concentration of chlorophyll standards by atomic absorption spectrometry”, *Biochim. biophys. Acta* 975: 384-394, 1989.

73. R.J. Porra, "A simple method for extracting chlorophylls from recalcitrant alga, *Nannochloris atomus*, without formation of spectroscopically-different magnesium-rhodochlorin derivatives. – *Biochim. biophys. Acta* 1019: 137-141, 1990.
74. R.J. Porra, "Recent advances and re-assessments in chlorophyll extraction and assay procedures for terrestrial, aquatic and marine organisms, including recalcitrant algae. – In: Scheer, H. (ed.): *Chlorophylls*", Pp. 31-57. CRC Press, Boca Raton – Ann Arbor – Boston – London 1991.
75. R.J. Porra, "The chequered history of the development and use of simultaneous equations for the accurate determination of chlorophylls a and b". *Photosynth. Res.* 73: 149-156, 2002.
76. R.J. Porra, "Towards a more accurate future for chlorophyll a and b determinations: the inaccuracies of Daniel Arnon's assay", *Photosynthesis Research.* 140, 2018, 10.1007/s11120-018-0579-8.
77. R. Esteban, et.al., "On the recalcitrant use of Arnon's method for chlorophyll determination", *New Phytol.* 217: 474-476, 2018 <https://doi.org/10.1111/nph.14932>.
78. H. K. Lichtenthaler and A. R. Wellburn, *Biochem. Soc. Trans.* 11, 591 (1983).
79. H. Schulz, M. Baranska, and R. Baranski, *Biopolymers* 77, 212 (2005). – carotenoids.
80. R. Withnall, B. Z. Chowdhru, J. Silver, H. G. M. Edwards, and L. F. C. de Oliveira, *Spectrochim Acta, Part A* 59, 2207 (2003). overtone of the carotenoids.
81. H. Schulz and M. Baranska, *Vib. Spectrosc.* 43, 13 (2007). the overtone of chlorophylls.
82. M. Schwanninger, J. C. Rodrigues, and K. Fackler, "A Review of Band Assignments in near Infrared Spectra of Wood and Wood Components. *Journal of near Infrared Spectroscopy.* 19. 287-308, 2011, 10.1255/jnirs.955.
83. C.R. Bull, "Wavelength selection for near-infrared reflectance moisture meters", *Journal of Agricultural Engineering Research*, Volume 49, 1991, Pages 113-125, ISSN 0021-8634, [https://doi.org/10.1016/0021-8634\(91\)80032-A](https://doi.org/10.1016/0021-8634(91)80032-A).

Chapter 3

Pre-Processing and Machine Learning

This chapter explains the basics of the selected theories and technologies, to plan a methodology towards achieving the objectives of the research.

3.1 Pre-Processing

The preprocessing methods in spectroscopy are invented for the cater to the critical requirement of enhancing Data Quality and Interpretation [1,2]. Spectroscopy, a powerful analytical technique, involves the interaction of electromagnetic radiation with matter to provide insights into its composition, structure, and properties. However, raw spectral data often contain noise, baseline shifts, and other artifacts that can hinder accurate analysis and interpretation. Preprocessing methods are essential to mitigate these issues and extract meaningful information from spectra. In this section, we will explore various preprocessing techniques commonly used in spectroscopy, along with their corresponding equations, advantages, and disadvantages.

1. Baseline Correction:

Baseline correction is crucial for removing the continuous, non-sample-related signal that can obscure true spectral features. Polynomial fitting and spline interpolation are common methods for baseline correction. While it enhances peak identification, inaccurately estimated baselines can lead to misinterpretation.

$$\text{Baseline-corrected spectrum (S')} = S - B \dots\dots\dots (1)$$

where $S \rightarrow$ Original spectrum, $B \rightarrow$ Baseline

It enhances the visibility of spectral features by removing baseline interference and improves accuracy in peak detection and quantification. However, its effectiveness depends on the accuracy of baseline estimation.

2. Smoothing:

Smoothing methods involve averaging nearby data points to suppress high-frequency noise. Thus, it reduces noise and fluctuations, improving signal-to-noise ratio. It enhances

the detection of subtle spectral features. While it improves data quality and peak detection, oversmoothing can lead to the loss of small but important features causing loss of fine details.

$$\text{Smoothened intensity at point } i (I') = \frac{I(i-k)+I(i-k+1)+ \dots + I(i+k-1)+I(i+k)}{2k+1} \dots\dots(2)$$

3. Normalization:

Normalization scales spectra to a common intensity range, facilitating comparison between samples. However, it can amplify noise if noise levels are high relative to signal strength.

$$\text{Normalized intensity at point } i (I') = \frac{I(i)-\min(I)}{(\max(I) - \min(I))} \dots\dots\dots(3)$$

It reduces variations in intensity, making spectra comparable and improves consistency in analysis across different samples. However, it may amplify noise itself, in cases of significant noise components in data.

4. Derivative :

$$\text{First derivative at point } i (D1') = I(i + 1) - I(i - 1) \dots\dots\dots(4)$$

Derivative spectroscopy accentuates spectral features by focusing on their rate of change. While it enhances peak detection, noise amplification and potential loss of information due to noise are drawbacks.

It enhances fine details and small spectral features, and improves peak identification and differentiation. It amplifies noise and might require additional smoothing.

5. Multiplicative Scatter Correction (MSC):

$$\text{MSC - corrected intensity at point } i (I_{msc}) = \frac{I(i)-\text{mean}(I)}{(\text{reference}(i)-\text{mean}(\text{reference}))} \dots\dots\dots(5)$$

MSC mitigates baseline shifts caused by scattering effects, making spectra more comparable. However, the accuracy of the reference spectrum is crucial, and artifacts may arise if it's not representative.

It corrects for scattering effects, improving spectral comparability, and reduces variations due to sample preparation differences. However, it assumes that the reference spectrum is scatter-free and can introduce artifacts.

6. Fourier Transform:

$$F(\omega) = \int f(t) * e^{-i\omega t} dt \dots\dots\dots(6)$$

Fourier Transform is pivotal for frequency-based spectroscopic techniques. It uncovers spectral components hidden in the time-domain data. However, its accuracy depends on precise instrument calibration and a solid grasp of frequency-domain data interpretation.

It converts time-domain data into frequency-domain data, revealing complex spectra. It is

essential for techniques like Fourier Transform Infrared (FTIR) spectroscopy. It depends on precise instrument calibration and understanding of frequency-domain data.

7. Noise Reduction:

Noise reduction methods like wavelet denoising and median filtering improve the signal-to-noise ratio. Yet, excessive noise reduction might eliminate genuine spectral features, necessitating a balance between noise reduction and feature preservation.

It reduces noise components, enhancing the quality of spectral data, and improves signal detection and quantification. It is overaggressive denoising may lead to loss of important information.

8. Peak Detection and Integration:

$$\text{Peak Area} = \int I(x)dx \text{ (Integration over the peak region) } \dots\dots\dots (7)$$

Peak detection and integration techniques facilitate the identification and quantification of specific features. Manual intervention might be necessary for complex spectra or overlapping peaks.

It enables identification and quantification of specific spectral features, and provides insights into concentration and composition changes. But, it depends on accurate peak models and may require manual intervention.

9. Standard Normal Variate (SNV):

$$\text{SNV} - \text{corrected intensity at point } i \text{ } (I_{snv}) = \frac{(I(i) - \text{mean}(I))}{\text{std}(I)} \dots\dots\dots (8)$$

SNV scales spectra to a standard deviation of one, correcting for multiplicative effects. However, its effectiveness is reduced when noise dominates the signal.

It corrects for multiplicative effects, such as baseline shifts and scattering, and improves comparability of spectra by scaling them to zero mean and unit variance. Nonetheless, it may amplify noise if applied to noisy data.

10. Mean-Centering:

$$\text{Mean} - \text{centered intensity at point } i \text{ } (I_{mc}) = I(i) - \text{mean}(I) \dots\dots\dots (9)$$

Mean-centering removes constant offsets, emphasizing relative changes in spectral data. However, it does not address multiplicative variations like baseline shifts.

It removes constant offsets and highlights spectral variations, and simplifies data interpretation by focusing on relative changes. Nevertheless, it doesn't account for multiplicative variations, such as baseline shifts.

11. Savitzky-Golay Filter for Data Smoothing

The Savitzky-Golay filter is a widely used method for smoothing noisy data while

preserving important features of the original signal [3]. It is particularly valuable in scenarios where noise reduction is required without significant distortion of underlying trends. Developed by Abraham Savitzky and Marcel J. E. Golay in the 1960s, this filter employs polynomial fitting to local segments of data, providing a versatile tool for signal processing and data analysis.

Given a set of data points (x_i, y_i) , where $i=1,2,\dots,N$, the goal is to smooth the data by estimating the underlying trend while reducing noise. The Savitzky-Golay filter achieves this by fitting a polynomial of degree p to a window of $2m+1$ neighboring points centered at each data point (hence, m is the “half width” of the approximation interval). The polynomial is fitted using the method of least squares regression.

$$p(n) = \sum_{k=0}^N a_k n^k \dots\dots\dots (10)$$

The S-G filter of order N , for each time sample n , accepts an input set $x[n]$ and produces a single output sample $y[n]$ defined to be the polynomial $p[n]$ evaluated at $n=0$; i.e,

$$y[n]=y[0]=a_0 \dots\dots\dots(11)$$

Incidentally as the window slides along the input data length, at every time $n=d$ a new output sample $y[d]$ is computed according to the window of samples.

The coefficients c_j are computed by solving a least squares problem. The goal is to minimize the sum of squared differences between the polynomial approximation and the original data points within the window. The solution involves finding the coefficients that minimize the following equation, by differentiating ϵ_N in (11) with respect to each of the $N+1$ unknown coefficients and setting the corresponding derivative equal to zero:

$$\epsilon_N = \sum_{n=-M}^M ((p(n) - x(n))^2 = \sum_{n=-M}^M (\sum_{k=0}^N a_k n^k - x[n])^2 \dots\dots (12)$$

Where p_j are the coefficients of the polynomial $P(x)$ fitted to the data points within the window.

The solution for the coefficients c_j involves matrix operations and can be expressed as:

$$c = (X^T X)^{-1} X^T y \dots\dots\dots (13)$$

Where \mathbf{c} is a vector of the coefficients c_j , \mathbf{X} is a matrix containing powers of x for the data points in the window and \mathbf{y} is a vector containing the original data values in the window.

The Savitzky-Golay filter offers several advantages, including noise reduction while maintaining important features of the data. The choice of polynomial degree p and window size $2m+1$ allows for customization based on the characteristics of the data.

However, it's important to note that Savitzky-Golay filtering might not perform optimally when dealing with data containing rapid changes or when the noise characteristics are not

well-behaved. Careful parameter selection is required to achieve desired results.

The Savitzky-Golay filter is a valuable tool for data smoothing, widely used in signal processing and data analysis. Its mathematical foundation, involving polynomial fitting and least squares regression, allows for effective noise reduction while preserving essential features of the original signal. Proper parameter tuning is essential for achieving the desired balance between smoothing and maintaining data integrity.

Ultimately, the preprocessing methods are essential for enhancing the quality and interpretability of spectroscopic data. Each technique has its advantages and limitations, and their selection should be based on the specific goals of analysis and the characteristics of the data. Experimentation with different preprocessing steps is often necessary to determine the most suitable approach for a particular spectroscopic study.

3.2 Dimensionality Reduction

Dimensionality reduction is a technique used in machine learning and data analysis to reduce the number of features (variables or dimensions) in a dataset while preserving the essential information. This process helps simplify data, reduce noise, and improve the efficiency of machine learning algorithms. There are two main approaches to dimensionality reduction: feature selection and feature extraction.

Curse Of Dimensionality is a common problem in machine learning, where the performance of the model deteriorates as the number of features increases. This is because the complexity of the model increases with the number of features, and it becomes more difficult to find a good solution. In addition, high-dimensional data (with a large number of features or variables) can also lead to overfitting, where the model fits the training data too closely and does not generalize well to new data.

Feature Selection: Feature selection is a fundamental technique in dimensionality reduction. Its primary objective is to choose a subset of the most relevant features from the original dataset while discarding less important or redundant ones. This process offers several advantages, including reduced computational complexity, improved model efficiency, and enhanced interpretability. Feature selection methods range from simple statistical tests to more advanced algorithms. They can help identify which features have the most significant impact on the target variable, making it easier to understand the

underlying patterns in the data. However, feature selection may not perform optimally when there are intricate relationships between features, and it may struggle to capture nonlinear dependencies in the data.

Feature Extraction: Feature extraction, on the other hand, goes beyond simple feature selection by transforming the original feature space into a lower-dimensional representation. This transformation is achieved through mathematical techniques such as linear algebra or manifold learning. Feature extraction methods aim to capture complex relationships and patterns within the data, often revealing hidden structures that may not be apparent in the original high-dimensional space. Principal Component Analysis (PCA) is a prominent example of a feature extraction technique. It identifies linear combinations of the original features, known as principal components that maximize the variance in the data. Feature extraction is valuable when dealing with high-dimensional datasets or when the relationships between features are nonlinear and intricate. However, the resulting features may lack direct interpretability.

In summary, feature selection and feature extraction are essential dimensionality reduction techniques. Feature selection simplifies the dataset by retaining only the most relevant features, while feature extraction transforms the data into a lower-dimensional space, capturing complex relationships. PCA, a well-known feature extraction method, identifies linear combinations of features to reduce dimensionality effectively. Each approach has its advantages and limitations, and the choice between them depends on the nature of the data and the specific goals of the analysis.

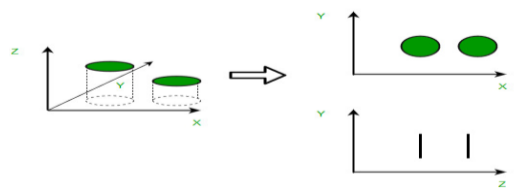


Fig 3.1 Components of Dimensionality Reduction

Advantages of Dimensionality Reduction

1. It helps in data compression, and hence reduced storage space.
2. It reduces computation time.
3. It also helps remove redundant features, if any.
4. Improved Visualization: High dimensional data is difficult to visualize, and dimensionality reduction techniques can help in visualizing the data in 2D or 3D, which can help in better understanding and analysis.

5. **Overfitting Prevention:** High dimensional data may lead to overfitting in machine learning models, which can lead to poor generalization performance. Dimensionality reduction can help in reducing the complexity of the data, and hence prevent overfitting.
6. **Feature Extraction:** Dimensionality reduction can help in extracting important features from high dimensional data, which can be useful in feature selection for machine learning models.
7. **Data Preprocessing:** Dimensionality reduction can be used as a preprocessing step before applying machine learning algorithms to reduce the dimensionality of the data and hence improve the performance of the model.
8. **Improved Performance:** Dimensionality reduction can help in improving the performance of machine learning models by reducing the complexity of the data, and hence reducing the noise and irrelevant information in the data.

Disadvantages of Dimensionality Reduction

1. It may lead to some amount of data loss.
2. PCA tends to find linear correlations between variables, which is sometimes undesirable.
3. PCA fails in cases where mean and covariance are not enough to define datasets.
4. We may not know how many principal components to keep- in practice, some thumb rules are applied.
5. **Interpretability:** The reduced dimensions may not be easily interpretable, and it may be difficult to understand the relationship between the original features and the reduced dimensions.
6. **Overfitting:** In some cases, dimensionality reduction may lead to overfitting, especially when the number of components is chosen based on the training data.
7. **Sensitivity to outliers:** can result in a biased representation of the data.
8. **Computational complexity:** Some dimensionality reduction techniques, such as manifold learning, can be computationally intensive, especially when dealing with large datasets.

Important points: Dimensionality reduction is performed during pre-processing stage before building a model to improve the performance. It is important to note that dimensionality reduction can also discard useful information, so care must be taken when applying these techniques.

It is commonly used in the fields that deal with high-dimensional data, such as speech recognition, signal processing, bioinformatics, etc. It can also be used for data visualization, noise reduction, cluster analysis, etc.

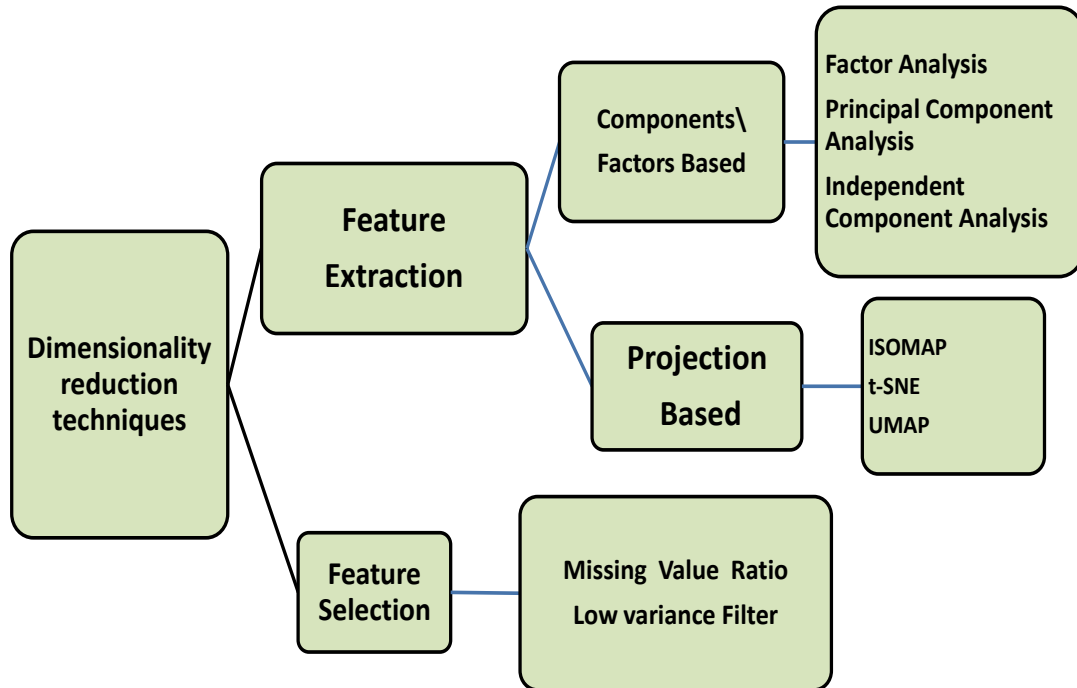


Fig. 3.2 Types of Dimensionality Reduction

Principal Component Analysis (PCA)

PCA is the most widely used tool in exploratory data analysis and in machine learning for predictive models [4]. Principal Component Analysis (PCA) technique was introduced by the mathematician Karl Pearson in 1901.

Principal Component Analysis (PCA) is an unsupervised learning algorithm technique used to examine the interrelations among a set of variables. It is also known as a general factor analysis where regression determines a line of best fit.

The main goal of Principal Component Analysis (PCA) is to reduce the dimensionality of a dataset while preserving the most important patterns or relationships between the variables without any prior knowledge of the target variables. It does so, by finding a new set of variables, smaller than the original set of variables, retaining most of the sample's information, and useful for the regression and classification of data. It works on the basis that the information is carried in the variance of the features, that is, the higher the variation in a feature, the more information that features carries.

It works on the condition that the data in a higher dimensional space is mapped to data in a lower dimension space, such that

1. The variance of the data in the lower dimensional space should be maximum.
2. The principal components are linear combinations of the original variables in the dataset and are ordered in decreasing order of importance. The total variance captured by all the principal components is equal to the total variance in the original dataset.
3. The correlated variables are converted to a set of uncorrelated variables.

The first principal component captures the most variation in the data, but the second principal component captures the maximum variance that is orthogonal to the first principal component, and so on.

Overall, PCA is a powerful tool for data analysis and can help to simplify complex datasets, making them easier to understand and work with.

Principal Component Analysis can be used for a variety of purposes, including data visualization, feature selection, and data compression. In data visualization, PCA can be used to plot high-dimensional data in two or three dimensions, making it easier to interpret. In feature selection, PCA can be used to identify the most important variables in a dataset, and to reduce the size of a dataset without losing important information. Fig 3.3 indicates factors (f_1 , f_2) and errors (e_1 , e_2) of the PCA.

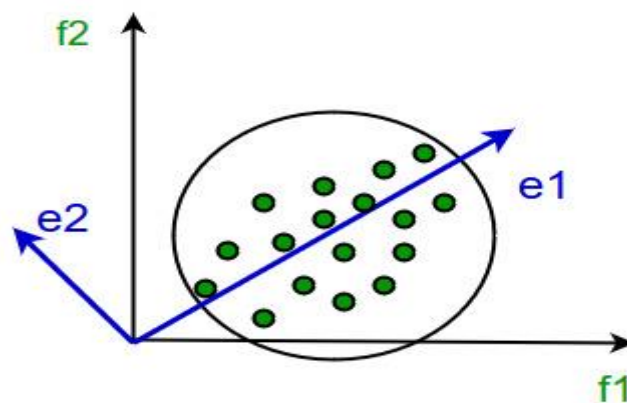


Fig 3.3 Functioning of PCA

It involves the following steps:

1. Construct the covariance matrix of the data.
2. Compute the eigenvectors of this matrix.
3. Eigenvectors corresponding to the largest eigenvalues are used to reconstruct a large fraction of variance of the original data.

Hence, we are left with a lesser number of eigenvectors, and there might have been some data loss in the process. But, the most important variances should be retained by these

eigenvectors.

3.3 Machine Learning

Machine learning, a subset of artificial intelligence, has emerged as a transformative force in our technologically advanced world. It is the field that empowers computers to learn from data, discern patterns, and make intelligent decisions without explicit programming. Machine learning is not merely a technological innovation; it represents a paradigm shift in how the complex problems are approached and the vast amounts of available data are leveraged.

The roots of machine learning trace back to the mid-20th century when early pioneers like Alan Turing and Arthur Samuel began exploring the concept of teaching machines to learn from data. However, it wasn't until the digital age and the availability of vast datasets that machine learning truly began to flourish. One of the key breakthroughs came with the development of neural networks, inspired by the structure of the human brain. Neural networks, particularly deep learning models, have fuelled remarkable advancements in image recognition, natural language processing, and speech recognition.

While machine learning holds immense promise, it also raises important ethical and societal questions. Concerns regarding data privacy, algorithmic bias, and job displacement must be addressed as the technology continues to evolve. Additionally, the black-box nature of some machine learning models can make it challenging to understand their decision-making processes fully. Furthermore, machine learning's effectiveness heavily relies on the quality and quantity of data available. Biased or incomplete data can lead to biased or inaccurate results, potentially perpetuating discrimination or misinformation.

Machine learning is not merely a technological tool; it represents a transformative shift in how we harness the power of data to drive innovation and make informed decisions. Its applications span across diverse domains, from healthcare to agriculture, revolutionizing industries and improving our quality of life. A machine has the ability to learn if it can improve its performance by gaining more data.

Fig 3.4 shows block diagram explains the functioning of a project utilising Machine Learning algorithms

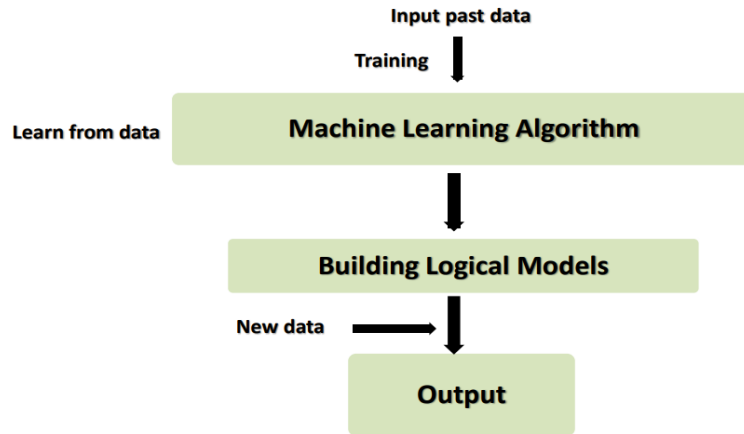


Fig.3.4 workflow of project using Machine Learning algorithm

Features of Machine Learning:

1. Machine learning uses data to detect various patterns in a given dataset.
2. It can learn from past data and improve automatically.
3. It is a data-driven technology.
4. Machine learning is much similar to data mining as it also deals with the huge amount of the data.

Need for Machine Learning

The importance of machine learning can be easily understood by its use cases, Currently, machine learning is used in self-driving cars, cyber fraud detection, face recognition, and friend suggestion by Facebook, etc. Various sectors like healthcare, agriculture, e-commerce have machine learning models that are using a vast amount of data.

Importance of Machine Learning:

1. Rapid increment in the production of data
2. Solving complex problems, which are difficult for a human
3. Decision making in various sector including finance
4. Finding hidden patterns and extracting useful information from data.

Classification: Classification is the process of categorizing data into predefined classes or labels based on input features. In Agriculture: Machine learning classification is used to identify and categorize plant diseases. For instance, it can distinguish between fungal,

bacterial, or viral infections by analysing images of diseased leaves. It's also applied to differentiate various pest species well as levels of infestations aiding in targeted control measures [5].

Regression: Regression is a technique used to predict continuous numerical values based on input features. In Agriculture: Regression models are utilized to forecast crop yields by considering factors like weather conditions, soil quality, and irrigation practices. These models can also predict future pest population levels based on historical data and environmental variables [6].

Supervised Learning: Supervised learning involves training models on labelled data to map input features to predefined output labels. In Agriculture: Supervised learning is applied to detect plant diseases. Models are trained on labelled images of both healthy and diseased plants, allowing them to identify and classify diseases accurately. It's also used for optimizing crop yield by learning from historical data on crop performance and agricultural practices [7].

Unsupervised Learning: Unsupervised learning deals with unlabelled data to discover patterns, structures, or clusters within the data. It can be further classified into two categories of algorithms: Clustering and Association. In Agriculture: Unsupervised learning is utilized in crop management by clustering fields with similar soil properties and historical crop data for optimal crop rotation strategies. It's also valuable for detecting anomalies in pest populations or unexpected changes in crop health, facilitating early intervention [8,9].

Reinforcement Learning: Reinforcement learning enables agents to make sequential decisions through interaction with an environment, receiving rewards or punishments. In Agriculture: Reinforcement learning powers autonomous farming robots that navigate fields, identify weeds, and apply herbicides efficiently, adapting their actions over time for maximum efficiency. Drones equipped with reinforcement learning scan fields for pest infestations and deploy targeted interventions, such as releasing natural predators or applying pesticides selectively [10].

These machine learning paradigms, when applied in agriculture, play crucial roles in plant-pest management, disease detection, and optimizing agricultural processes, ultimately

contributing to more efficient and sustainable farming practices.

Currently, the most advanced technological domain indecision-making is Artificial Intelligence-Machine Learning. It has been employed in agriculture generally [11,12] and plant stress studies [13], and plant disease specifically [14]. Disease detection using images of crop leaves have been implemented using a pattern recognition branch of machine learning. Machine learning algorithms effectively take into consideration all the possible factors, historic data as well as satellite/sensor data of fields to provide valuable disease classifiers. These methods deal well with noisy and multi-faceted data.

One of the widely used algorithms is different Neural Network (NN) techniques: Single-Layer Perceptron (SLP), Multi-Layer Perception (MLP), Support Vector Machine (SVM), Radial-Basis Function (RBF) networks, Learning Vector Quantization (LVQ), Kohonen's Self-Organizing Map (SOM) networks, Convolutional Neural Network (CNN) and Probabilistic Neural Network (PNN). As per a review [14], these have been exerted thoroughly for the specific application of various plant disease detection. Amongst the various approaches and machine learning techniques, the NN classifiers have the best classification with the fastest speed and best accuracy for classification work.

Generally, NNs have been used for data mining purposes only, but their various applications with hyperspectral data are now showing significant promise for disease detection. More often than not, like many other technologies, researchers have been confronted with emerging challenges in NN applications. For example, the detection of three different categories of disease manifestation, viz., pre-symptomatic, symptomatic, and asymptomatic diseases from a single plant requires the best trainer sets for accurate classification. NNs have shown incredible capabilities in adapting to new challenges of disease detection using spectral data. NNs have been used for a variety of purposes, such as reduction of data dimensionality, training of image pixels or spectra as the input sets, generalization of the input sets, and classification of wavebands or SDIs.

Specifically, the BP-NN algorithms employ a supervised learning paradigm, which minimizes errors between the desired outputs and the calculated outputs driven from the inputs and network learning.

Higher degree of reliability owing to unique capabilities of learning, generalization, and

imagination due to the learning function that adjusts all the weights and biases to the assigned layer.

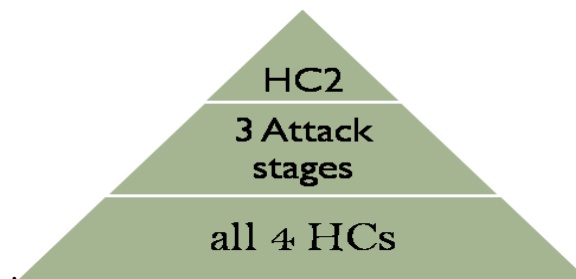


Fig 3.5: Tiers indicating the ascending capability of NNs in tracking severity of the HCs detection -epitome being the HC2, in terms of rapidness and accuracy.

The Back Propagation succeeds over other NNs for plant disease detection, for its following striking features [14]: -

1. Higher degree of diagnosis due to incredible capabilities of adapting to new challenges of disease detection using spectral data.
2. The different Health Conditions are traceable in specificity indicated through tiers (Fig.3.4) in terms of rapidness and accuracy.
3. BPNN has the highest correction and discriminating capabilities of reflectance wavelength between 460nm and 1130nm at 10nm increments.

The **Scaled Conjugate Gradient (SCG)** [6] is an optimization algorithm commonly used in machine learning and numerical optimization. It is primarily employed for training artificial neural networks, particularly in the context of supervised learning tasks like regression and classification. SCG is used to find the optimal parameters (weights and biases) of a machine learning model, typically a neural network, by minimizing a defined cost or error function. The primary objective is to find the model's parameter values that result in the lowest possible error or cost on a given dataset. SCG is a variant of the Conjugate Gradient (CG) method, which is an iterative optimization technique for solving unconstrained optimization problems. SCG has been applied in various fields, including pattern recognition, image processing, natural language processing, and more, wherever optimization of machine learning models is required.

Scaled Approach: The "scaled" aspect of SCG involves dynamically adjusting the step size

(learning rate) for each parameter during optimization. This scaling helps speed up convergence and stabilize training.

Advantages: SCG is known for its efficiency and fast convergence, making it suitable for training large and complex neural networks. It can handle high-dimensional parameter spaces efficiently.

Usage in Neural Networks: SCG is popular for training neural networks because it combines the advantages of the conjugate gradient method with adaptive learning rates.

Adaptive Learning Rates: SCG adjusts the learning rate individually for each parameter based on their historical gradient information, allowing for faster progress when the gradient direction remains consistent and smaller steps when it changes direction.

Avoiding Overfitting: The adaptive learning rates can help prevent overfitting, a common problem in machine learning, by effectively controlling the size of parameter updates.

Alternatives: Other optimization methods, such as stochastic gradient descent (SGD), Adam, and RMSprop, are also commonly used in machine learning, each with its strengths and weaknesses.

In summary, the Scaled Conjugate Gradient algorithm is a widely used optimization technique in machine learning, particularly in the context of training neural networks. Its adaptive learning rates and fast convergence properties make it efficient for finding optimal model parameters, ultimately leading to better machine learning model performance.

Support Vector Machines (SVMs) [15] are a type of supervised machine learning algorithm used for classification and regression tasks. SVMs work by finding a hyperplane that best separates data points into different classes or predicts a continuous target variable. SVM functions in these steps:

Data Preparation: SVMs require labelled training data. Each data point should be represented by a feature vector and assigned to a class label (for classification) or have a corresponding target value (for regression).

Feature Space: The feature vectors represent the data points in a high-dimensional feature space. Each feature corresponds to a specific attribute or characteristic of the data.

Choosing a Kernel: SVMs use a kernel function to transform the data into a higher-dimensional space. Common kernel functions include Linear, Polynomial, Radial Basis Function (RBF), and Sigmoid kernels. The choice of kernel depends on the nature of the data and the problem at hand.

Finding the Hyperplane: The goal of SVM is to find the hyperplane that best separates the data into different classes while maximizing the margin between the two classes. The margin is defined as the distance between the hyperplane and the nearest data points (support vectors) from each class.

Here are important features of SVM:

1. Support vectors are data points that lie closest to the hyperplane and play a crucial role in determining the hyperplane's position. These are the data points that, if removed, would change the position of the hyperplane.
2. Margin Optimization: SVM aims to maximize the margin while minimizing classification errors. It uses a mathematical optimization technique to find the optimal hyperplane. This involves solving a constrained optimization problem.
3. Handling Non-Linear Data: In cases where the data is not linearly separable, SVMs can still work by using a kernel trick. The kernel function implicitly maps the data into a higher-dimensional space, making it linearly separable in that space.
4. Regularization: SVMs often include a regularization parameter (C) that controls the trade-off between maximizing the margin and minimizing classification errors. A smaller C value results in a larger margin but may allow some misclassification, while a larger C value reduces the margin but minimizes misclassification.

Tuning Hyperparameters: The choice of kernel, regularization parameter, and other hyperparameters can significantly impact SVM performance. Hyperparameter tuning is essential to optimize model accuracy.

SVM models are evaluated using appropriate performance metrics such as accuracy, precision, recall, F1-score (for classification), or mean squared error (for regression) on a separate validation or test dataset.

In summary, Support Vector Machines work by finding the optimal hyperplane that maximizes the margin between classes or fits the target values for regression. They are powerful for handling both linearly separable and non-linearly separable data, making them a versatile tool in machine learning. Proper hyperparameter tuning is crucial for achieving good performance with SVMs.

References

1. P. Mishra, A. Biancolillo, J. M. Roger, F. Marini, D. N. Rutledge, “New data preprocessing trends based on ensemble of multiple preprocessing techniques”, *Trends in Analytical Chemistry*, Volume 132, 2020, 116045, ISSN 0165-9936, <https://doi.org/10.1016/j.trac.2020.116045>.
2. L. Dave L, S. Mukhopadhyay, L. Lin L and J. Harnly, “A Comparison of Analytical and Data Preprocessing Methods for Spectral Fingerprinting”, *Applied spectroscopy*, 65, pp 250-259. URL 10.1366/10-06109.
3. W. H. Press and S. A. Teukolsky, “Savitzky-Golay Smoothing Filters”, 1990, *Computers in Physics* 4:669 URL 10.1063/1.4822961
4. I.T. Jolliffe and J. Cadima, “Principal component analysis: a review and recent developments”, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374, 2065, 20150202, 2016, 10.1098/rsta.2015.0202.
5. D. Wei, J. Chen, T. Luo, T. Long, H. Wang, “Classification of crop pests based on multi-scale feature fusion”, *Computers and Electronics in Agriculture*, Volume 194, 2022, 106736, ISSN 0168-1699, <https://doi.org/10.1016/j.compag.2022.106736>.
6. A. A. Sarangdhar and V. R. Pawar, "Machine learning regression technique for cotton leaf disease detection and controlling using IoT", *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, India, 2017, pp. 449-454, doi: 10.1109/ICECA.2017.8212855.
7. S. S. Harakannavar, J. M. Rudagi, V. I. Puranikmath, A. Siddiqua, R. Pramodhini, Plant leaf disease detection using computer vision and machine learning algorithms, *Global Transitions Proceedings*, Volume 3, Issue 1, 2022, Pages 305-310, ISSN 2666-285X, <https://doi.org/10.1016/j.gltip.2022.03.016>.
8. D. Moshou, C. Bravo, J. West, S. Wahlen, A. McCartney, H. Ramon, “Automatic detection of yellow rust in wheat using reflectance measurements and neural networks”, *Computers and Electronics in Agriculture*, Volume 44, Issue 3, 2004, Pages 173-188, ISSN 0168-1699, <https://doi.org/10.1016/j.compag.2004.04.003>.
9. E. C. Tetila, B. B. Machado, G. Astolfi, N. A. de Souza Belete, W. P. Amorim, A. R. Roel, H. Pistori, “Detection and classification of soybean pests using deep learning with UAV images”, *Computers and Electronics in Agriculture*, Volume 179, 2020, 105836, ISSN 0168-1699, <https://doi.org/10.1016/j.compag.2020.105836>.

10. L. Eduard, G. Aitor, A. Ander, S. Loreto, T. Carlos, R. Ivan, L. Libor, 2019, “A Benchmarking of Learning Strategies for Pest Detection and Identification on Tomato Plants for Autonomous Scouting Robots Using Internal Databases, 5219471, 2019, ISSN - 1687-725X, <https://doi.org/10.1155/2019/5219471>, 10.1155/2019/5219471, Journal of Sensors, Hindawi
11. A. Kamilaris and F.X. Prenafeta-Boldú, “Deep Learning in Agriculture: A Survey”, *Computers and Electronics in Agriculture*. 147 pp 70-90 URL [10.1016/j.compag.2018.02.016](https://doi.org/10.1016/j.compag.2018.02.016), 2018.
12. K. G. Liakos, P. Busato, D. Moshou, S. Pearson and D. Bochtis, “Machine Learning in Agriculture: A Review Sensors”, 2018, 18(8) p 2674 URL <https://doi.org/10.3390/s18082674>
13. A. K. Singh, B. Ganapathy Subramanian, S. Sarkar and A. Singh, “Deep Learning for Plant Stress Phenotyping: Trends and Future Perspectives”, *Trends in Plant Science*, 2018, 23(10) pp 883–898 URL <https://doi.org/10.1016/j.tplants.2018.07.004>
14. K. Golhani, S. Balasundram, G. Vadamalai and B. Pradhan, “A Review of Neural Networks in Plant Disease Detection using Hyperspectral Data Information Processing in Agriculture”, 2018, 5(3) pp 354-371 URL [10.1016/j.inpa.2018.05.002](https://doi.org/10.1016/j.inpa.2018.05.002)
15. M.F. Moller, “A Scaled Conjugate Gradient Algorithm For Fast Supervised Learning Neural Networks”, 1993, pp 525-533 URL [10.1016/S0893-6080\(05\)80056-5](https://doi.org/10.1016/S0893-6080(05)80056-5).

Chapter 4

Methodology

Two approaches were pursued for designing a solution to the research problem of early detection of CSRB, viz., Image Processing and Spectroscopy. Both methods were centered on the subject of the stress, i.e., cashew trees, instead of tracking the beetle up front. Hence, they are more effect-oriented cases than cause-oriented types. The fundamental constituents in the functioning of Image Processing and Spectroscopy are the pixel data (RGB components) and the spectral signature (response to certain wavelength bands), respectively.

4.1 Study location

The study location is Mollem in the Western ghats of Sanguem Taluka of Goa, which covers an area of 28000 m² between 74.23° E (longitude) and 15.38° N (latitude). The elevation of the plantation is nearly 16m. This region is categorized into a tropical climate with a total annual precipitation 2570 mm. Mean annual temperature that varies between 25 and 30 °C. These details are populated in figure 4.1, along with map of the State of Goa. The area is divided into two kinds of soil profiles ecological zones: laterite and loose soil. Around 75% of the tree stands in laterite soils are Cashew (*Anacardium occidentale*) 12% are timber and the remaining 13% are covered by other commercial trees, like mango, jamuns and kokums. In the valley bottoms, 83% of the trees are areca nut and coconut, and the rest is a mix of banana trees and pepper creepers.

A storm in 1990s and a channel of low-pressure area in 2018 might have aggravated the borer pest attack on the cashew trees.

4.2 Image Processing

The procedure revolves around gathering images from healthy and unhealthy plants of insect attacks. Then incorporate the RGB model to help figure out the intensity of damage and estimate any scope for recovery of HC2 trees using prescribed remedies. Images of leaves of cashew were used to identify chlorophyll content and trying to estimate the early pest attack.

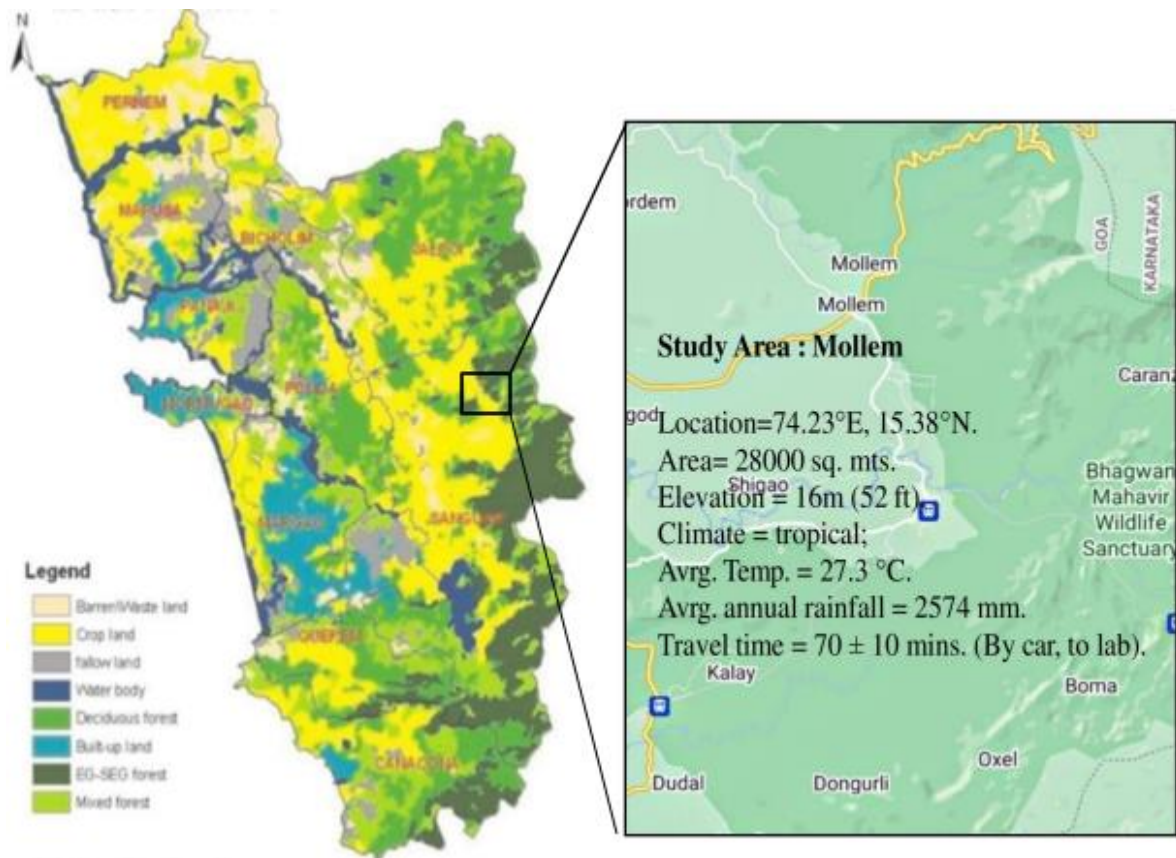


Figure 4.1: Details of the study location along with map of Goa.

Data acquisition: The database was obtained for four different combinations: Two cases of illumination scenarios, such as a day-light and incandescent lamp, and two image-capturing devices, such as a digital camera (Panasonic LUMIX DMC-SZ10) and a cell phone camera (Obi Boa S503) were used. Figure 4.2 shows the pictures of the leaf samples for study category, comprising of two Health conditions and four illumination scenarios. Table 4.1 gives the details of imaging device: Panasonic LUMIX DMC-SZ10.

The data collection included selecting six pixels in the leaf area for every leaf (approx. region of tissues other than the veins, as shown in Figure 4.3), and their RGB values were extracted using MATLAB R2018b Image Processing functions. These values are substituted in Equation 1 to estimate chlorophyll [1]. The equation 4.1 describes the calculation of the chlorophyll content using the R, G, B values at a selected pixel.

$$Chl = \frac{\text{logsig}\left(G - \frac{R}{3} - \frac{B}{3}\right)}{255} \dots\dots\dots(1)$$

Table 4.1 Specifications of imaging device: Panasonic LUMIX DMC-SZ10.

Property	Features	Details
	Camera	
Pixels	Effective Pixels	16 Megapixels
	Sensor Size /	
	Total Pixels /	1/2.33-type CCD Sensor / Total Pixel Number 16.6
Sensor	Filter	Megapixels / Primary Colour Filter
Lens	Aperture	F3.1 - 6.3 / 2-step (F3.1 / 7.8 (W), F6.3 / 16.3 (T))
	Optical Zoom	12x
	Focal Length	f = 4.3 - 51.6mm (24 - 288mm in 35mm equiv.)/(26 - 314mm in 35mm equiv. in 16:9 video recording)
	Extra Optical Zoom (EZ)	16.9x (4:3 / 8M), 27.0x (4:3 / under 3M)
	Intelligent Zoom	24x
	Lens	8 elements in 8 groups/(4 Aspherical Lenses / 8 Aspherical surfaces / 1 ED Lenses)
	Optical Image Stabilizer/Five Axis Correction	O.I.S. (On / Off) / No
	Digital Zoom	Max. 4x
Focus	Focusing Area	Normal / Intelligent Auto / Motion Picture: Wide 3 cm - infinity / Tele 150 cm - infinity
	Focus	Normal/Quick AF (Always On), Continuous AF (only for motion picture)
	AF Metering	Face / 1-area / 9-area
Shutter	Shutter Speed	Approx. 8 - 1/2000 sec/Starry Sky Mode: 15, 30, 60 sec
Recording Modes	Mode Dial / Mode Button	Intelligent Auto, Normal Picture, Panorama Shot, Scene, Creative Control
	Still Image Scene Mode	Portrait, Soft Skin, Scenery, Sports, Night Portrait, Night Scenery, Food, Baby1*, Baby2*, Pet*, Sunset, High Sensitivity, Glass Through, HDR, Starry Sky * Birthday only.
Flash	Built-in-Flash	Auto, Auto/Red-eye Reduction, Forced On, Slow Sync. /Red-eye Reduction, Forced Off/0.6 - 5.2m (Wide / ISO Auto), 1.5 - 2.6m (Tele / ISO Auto)

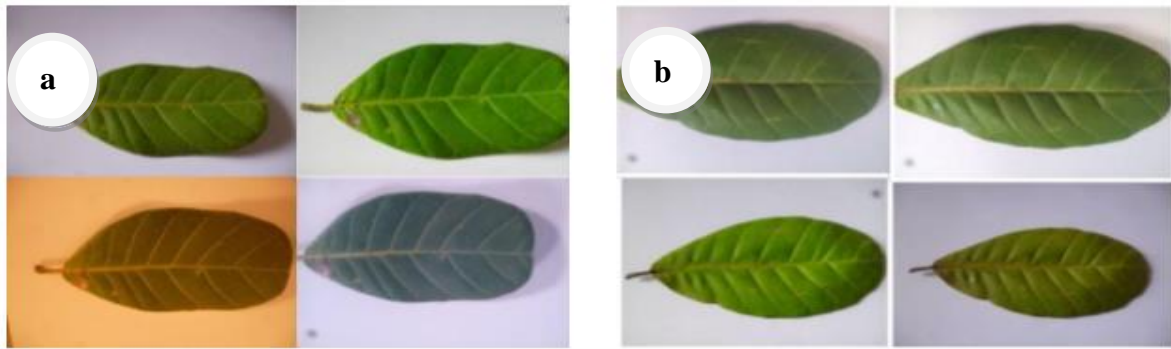


Figure 4.2: Pictures of the leaf samples for each study category (in clockwise order, starting from top left) a) healthy condition: (NI-DC, LI-DC, LI-CC, NI-CC) b) severe condition: LI-CC, NI-CC, NI-DC, LI-DC) where LI is Lamp Illumination, NI is Normal Illumination, CC is cell phone camera and DC is Digital Camera.

For results tabulation, the six output chlorophyll values for a given leaf were averaged to obtain single values for overall analysis. This was done for ten leaves in each Health Condition: Healthy, Mild, Moderate, and Severe infestations.

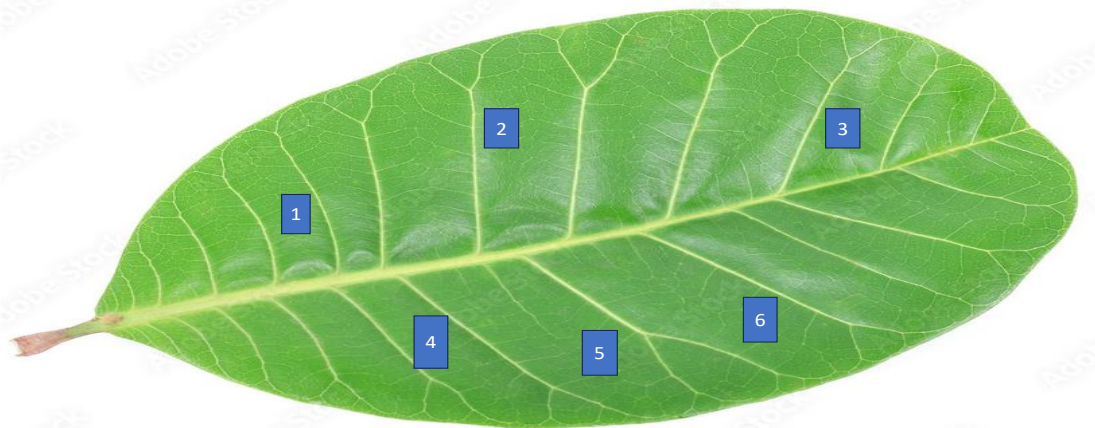


Figure 4.3: An example - demo of positioning of location on the leaf area, for selection of pixels

Table 4.2: Chlorophyll values for each HC by applying the RGB equation to the pixels of images.

Illumination scenario and device	Healthy	Mild	Moderate	Severe
Day light and digital camera- LUMIX DMC SZ10	0.5567	0.559	0.5505	0.5511
Day Light and cell phone	0.583	0.5861	0.5692	0.5756
Incandescent lamp and digital camera- LUMIX DMC SZ10	0.5386	0.5406	0.5358	0.5452
Incandescent lamp and cell phone	0.5357	0.539	0.5353	0.5503

From the Table 4.2, a certain chlorophyll content value does not assert a specific

corresponding infestation stage. This will create an uncertainty in the decision-making about the status of CSRB attack. Hence, the results obtained with computational model of RGB do not achieve recognition and classification of the infestation stages categories or levels of CSRB infestation.

Image processing was attempted on leaves to explore any possibility of detection of the borer pest. In a borer attack, the bark is also affected along with the leaves. This technique was based solely on leaf; hence it could have been a simple option to tackle beetle detection, without having to include any other perspective or factor (i.e., bark). It was potentially capable to detect beetle attack only if successful. Ultimately, the considered RGB image processing model is not found to be relevant for the purpose of timely of detection of beetle CSRB pest attack in cashew trees.

4.3 Spectroscopy

The procedure revolves around spectral signatures from healthy and unhealthy plants of insect attacks. Then the chlorophyll estimation equation or machine learning model is incorporated to help figure out the intensity of damage and estimate any scope for recovery of the HC2 tress using prescribed remedies.

4.3.1 Preliminary Studies

This section is about confirmation of substances to be utilized for spectra recording. Also, considering that the target specimen is a tree having large structure the location of region needs to be spotted on; the concise aim is to establish a rule for material or substance selection.

These studies will also augment the nullification of disturbances arising from natural variations 1) within tree (as different branches are selected), 3) between species and growing conditions.

Preliminary studies are conducted to finalize the collection spots (or area) out of the entire thrust zone. For foliage, the developed leaves (matured and fully-matured) are considered while the extremities of the branch (tip and young leaves) are overlooked. For the bark, the intrusion (and gum spots) on bark is pieced into, to find hollow cavity and vertically navigated to discover the tunnel created by the borers. These mined regions are called lesions and chunks of tissues at the lining of tunnel constitute raw material.

Leaf node selection:

Since the developed portion of the cashew tree branch has two kinds of leaves - matured and fully-matured, there was ambiguity towards the response in spectral data produced due to each. Hence, this study is to analyze the effect of type of leaf and decide about selection of leaf on tree branch.

Five number of the two kinds of leaves - matured and fully-matured, for each of forty trees (ten in each Health category) were plucked, transported and analysed in the laboratory.

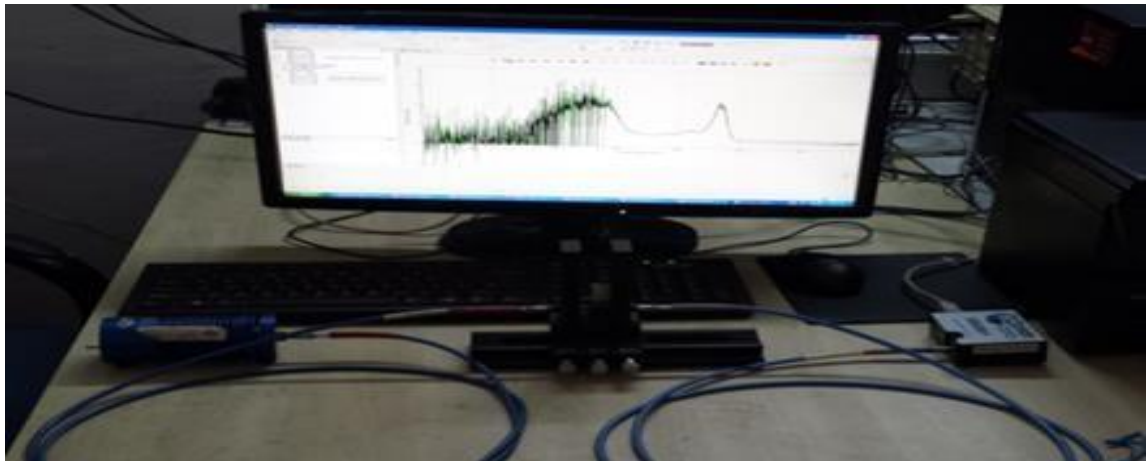


Figure 4.4: USB 2000+ Spectrophotometer setup

Matured Leaves:

- Healthy: $10 \times 5 = 50$
- Mild: $10 \times 5 = 50$
- Moderate: $10 \times 5 = 50$
- Severe: $10 \times 5 = 50$

Total = 200

Fully matured leaves:

- Healthy: $10 \times 5 = 50$
- Mild: $10 \times 5 = 50$
- Moderate: $10 \times 5 = 50$
- Severe: $10 \times 5 = 50$

Total = 200

Input preparation consisted of crushing the 0.5g of leaf portions other than veins, in 3.5 g of DMSO [1] in a crucible. The glass cuvettes of the extractant were stored in cold storage for few minutes [1]. Absorbance spectra were obtained using Ocean Optics USB 2000+ spectrophotometer setup, as shown in figure 4.4. The spectrophotometer supports the spectral range of 200 to 800 nm.

The figure 4.5a shows recorded spectra of fully-matured and matured leaves of four Health condition (HCs). As per the literature reported signature of chlorophyll (figure 4.5b) the two peaks are 662 nm and 642 nm for chlorophyll A, and chlorophyll B respectively, and second highest peak is at 620 nm. The recorded spectra show peaks of Chlorophyll A i.e., 620 and 666 nm. However, the response at each of them due to maturity cases differ. At

the former peak, the fully matured and matured response in coinciding. Whereas at the main wavelength of 666 nm, the maturity cases are having differences.

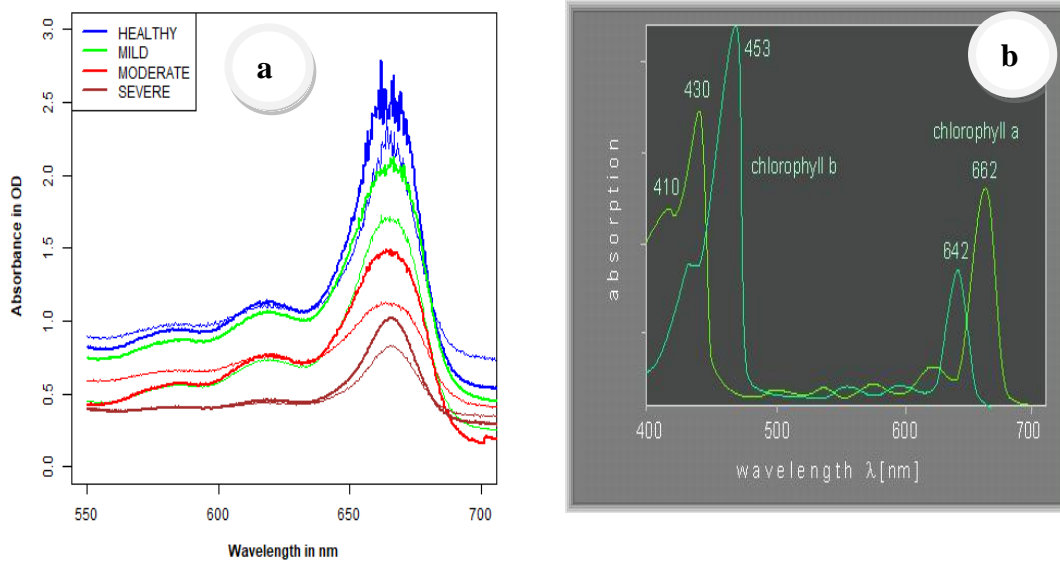


Figure 4.5: a) Study of leaf selection from different branches of cashew tree: matured and fully-matured. b) the standard signature for chlorophyll in leaves [2].

Hence, if there is random selection i.e., without considering any maturity case, the amplitudes at 666nm will be close by and not significant gaps. The selection has to be group specific – either only matured or only fully-matured. For our study, we have finalized fully-matured leaves (its tissues other than veins hereafter referred as substance 1), which are present at farther nodes of a branch.

Bark area selection:

Since the borer attack is located at different heights on the bark region of the cashew tree, there was ambiguity towards the response in spectral data produced due to them. Hence, this study is to analyze the effect of progression of the vertical tunneling by the beetle on the spectral response, to decide the sample collection.

The study material consisted of the inner scrapings of the tissues at the lesions for three different height levels -ground (collar region), 1 mtr and 1.5 mtr., of a given tree were selected and processed in the laboratory. Altogether 60 samples were used, fifteen for each of the four Health Condition of the tree. The spectral data is recorded in reflectance mode using the setup shown in Figure 4.6 (Jasco V-770 spectrophotometer).

Input preparation

- Healthy: $5 \times 3 = 15$
- Mild: $5 \times 3 = 15$
- Moderate: $5 \times 3 = 15$
- Severe: $10 \times 5 = 15$

Total = 60



Figure 4.6: A spectral recording setup.

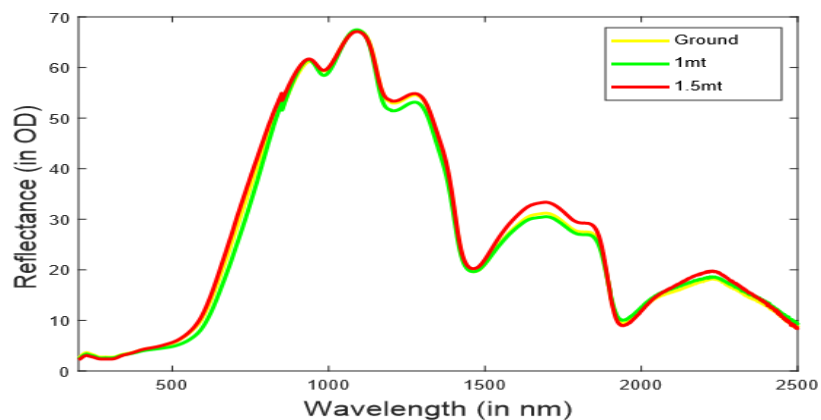


Figure 4.7: Study of height location at branch of cashew tree: ground, 1mtr and 1.5 mtr (Average Spectra)

Mass basis: by weighing the material on electronic balance. 0.5g of fresh cashew bark sample is crushed in 5ml of hot water in a crucible.

Figure 4.7 shows recorded spectra at different heights – ground level, 1mt and 1.5 mt of

CSRB attacked cashew tree. Since there are no significant differences, we can ascertain that location of truck sampling will not affect the spectral data. Samples can be taken from any height of tree, out the exposed cavities (the tissues at the lesions) within the tunnels.

Absorbance Processing (Leaf): Chlorophyll Calculations

For each HC, absorbance spectra were noted for 4 samples. The concentrations (in $\mu\text{g/ml}$) of *chlorophyll a*, *chlorophyll b* and thus total *chlorophyll* was determined using following equations [2], and tabulated in table 4.3.

$$chl_a = 14.85A_{664.9} - 5.14A_{648.2} \quad \text{-----(2)}$$

$$chl_b = 25.48A_{648.2} - 7.36A_{664.9} \quad \text{-----(3)}$$

$$chl_{a+b} = 7.49A_{664.9} + 20.34A_{648.2} \quad \text{-----(4)}$$

Table 4.3: Chlorophyll values from spectral data, for each health condition

HC	Sample 1	Sample 2	Sample 3	Sample 4	Avg. Chlorophyll ($\mu\text{g/ml}$)
1(Healthy)	47.22	41.751	45.742	43.426	44.53 \pm 2.43
2(Mild)	39.516	35.183	32.587	32.837	35.03 \pm 3.23
3(Middle)	18.703	20.873	29.303	32.328	25.30 \pm 6.53
4(Severe)	14.238	19.296	14.238	14.238	15.50 \pm 2.53

The results of Spectro-photometric method of chlorophyll reflectance measurement of cashew leaves show that there is considerable change in the magnitudes for different HCs, which means tree is under attack by pest.

4.3.2 Refined Methodology

After establishing the rule for study material selection using the preliminary studies on the two thrust zones, we could proceed to develop the solution - computational model of machine learning.

Since the symptoms of the CSRB attack disaster encompasses both the leaf and bark conditions of a tree, the detection system should conglomerate the data obtained at both the tree locations. Only a certain kind of Data Fusion technique would augment the functionality and complete the totality, and possibly improve accuracy. SVM and BPNN model was applied to Data fusion of leaf and bark. The flow-chart same is depicted in

Figure 4.10.a. Then the performance was evaluated. For the leaf and bark individually, BPNN model was applied. The flow-chart for same is shown in figure 4.10.b.

The spectral data for the four models- leaf alone (with BPNN), bark alone (with BPNN), leaf and bark with BPNN and leaf and bark with SVM, is recorded in reflectance mode using integrating sphere attachment in Jasco V-770 spectrophotometer, as shown in figure 4.8.

For the refined methodology, 40 cashew trees were selected from the plantation at Mollem, Goa. Ten trees were identified from each category. The leaf and bark samples were selected from the same trees. Total database for this study is 400 leaf spectra and 80 bark samples from these identified trees of different health condition.



Figure 4.8: Jasco V-770 spectrophotometer, and integrating sphere with leaf

Number of leaves

samples:

- Healthy: $10 \times 10 = 100$
- Mild: $10 \times 10 = 100$
- Moderate: $10 \times 10 = 100$
- Severe: $10 \times 10 = 100$

Total = 400

Number of bark

samples:

- Healthy: $10 \times 2 = 20$
- Mild: $10 \times 2 = 20$
- Moderate: $10 \times 2 = 20$
- Severe: $10 \times 2 = 20$

Total = 80

Spectral range Selection:

For the reflectance spectra (after pre-processed with Savitzky-Golay smoothing [4] with size=15 and polynomial order=2) of foliage and peduncle, the changing trend is not consistent throughout the recorded range of wavelengths. Even though

the fact is visible in the diagram, series of trials to figure best suitable portions out of the entire recordings yielded the selections, as shown in table 4.4. The selected portions are indicated by blue regions in Figure 4.9.a and 4.9.b.

Table 4.4: Selected spectral ranges (wavelengths in nm) of leaf and bark reflectance.

Spectra Ranges (in nm)	
Leaf spectra	550-670 & 750-1000
Bark spectra	705-1400 & 1500-2000

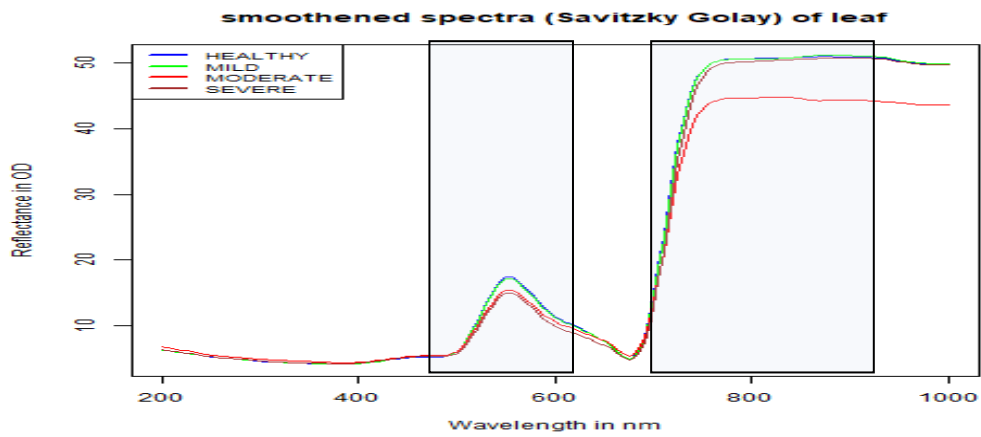


Figure 4.9.a: Viable spectral ranges (wavelengths in nm) of leaf reflectance for input definition to the models.

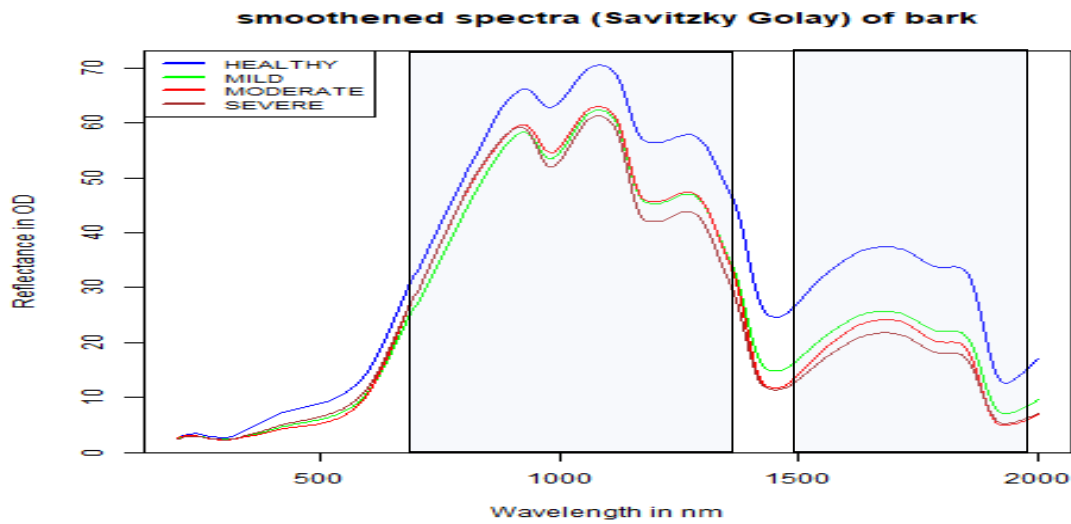


Figure 4.9.b: Viable spectral ranges (wavelengths in nm) of bark reflectance for input definition to the models.

Significance of Principal Component Analysis (PCA)

In case of spectral signatures of a diseased plant, Parametric approaches such as simple or multiple regression and functional statistics, do not analyze correctly. Whereas Non-parametric approaches: Principal Component Analysis(PCA), Fuzzy logic, Support Vector Machine (SVM), Cluster Analysis (CA), Partial Least Square

(PLS), Neural Networks (NNs) are more suitable[3].Hence, PCA is chosen for processing the data of spectral recordings.

For a given dataset,

1. PC coefficients (loadings), each column corresponds to each principal component, and columns are in descending order of component variance.
2. PC scores are the representations of data matrix X in the principal component space. Rows of score correspond to observations, and columns correspond to components.
3. PC variances are Eigen-values of the covariance matrix of X.
4. By default, PCA centers the data and uses the singular value decomposition (SVD) algorithm.

Implementation of PCA for CSRB detection

Full spectra have all the spectral information, including large data sets, information redundancy and a number of correlated variables. A full spectrum would thus make a model even more complex, slow down the calculation speed and affect its accuracy. Thus, efficient wavelength bands need to be identified that have a major role in the modelling. Unsupervised approach principal component analysis (PCA) is widely used to reduce the dimension of multivariate datasets. The principal components (PCs) score plot is used to show the characteristics of the variable distribution, and the PC loadings plot can show how important certain variables are. Typically, the ideal wavelength selection is done using the loadings of the first several PCs.

For the individual study of leaf and bark, PCA was used to identify various suitable wavelengths at the chosen spectral range of leaf data, i.e., 550-670 & 750-1000, which might simplify the model structure and enhance model accuracy. The wavelengths located at local maximal values (peaks) or lowest values (valleys) of the loadings curve have a more significant contribution to the PC score since each PC score is a linear combination of each spectral point multiplied by the associated loading. As a result, the effective wavebands were chosen based on the loadings curve of the first two PCs. The wavelengths that correspond to these loadings were

chosen as the key wavelengths for the classification models as loadings at local peaks and valleys can be regarded as characteristic loadings.

For the data fusion models of leaf and bark data together, PCA is used for dimensionality reductions i.e. obtain the features in the form PCA scores.

1. The PCA applied to datasets: data matrix X of spectral data, rows are the number of samples (trees) and columns are the observations (individual wavelength over reordered ranges).
2. The PC variances values are used for choosing number of components - n , with a criterion as $\sum(\text{PCA variances}) \approx 99.99$

Value of 'n' is the restriction to recompute the PC scores within the n count. This is done separately for leaves and bark, to obtain n_1 and n_2 .

3. PC selected are $n_1=4$ for leaves and $n_2=6$ for bark which are concatenated together to form the input dataset for the computational models.

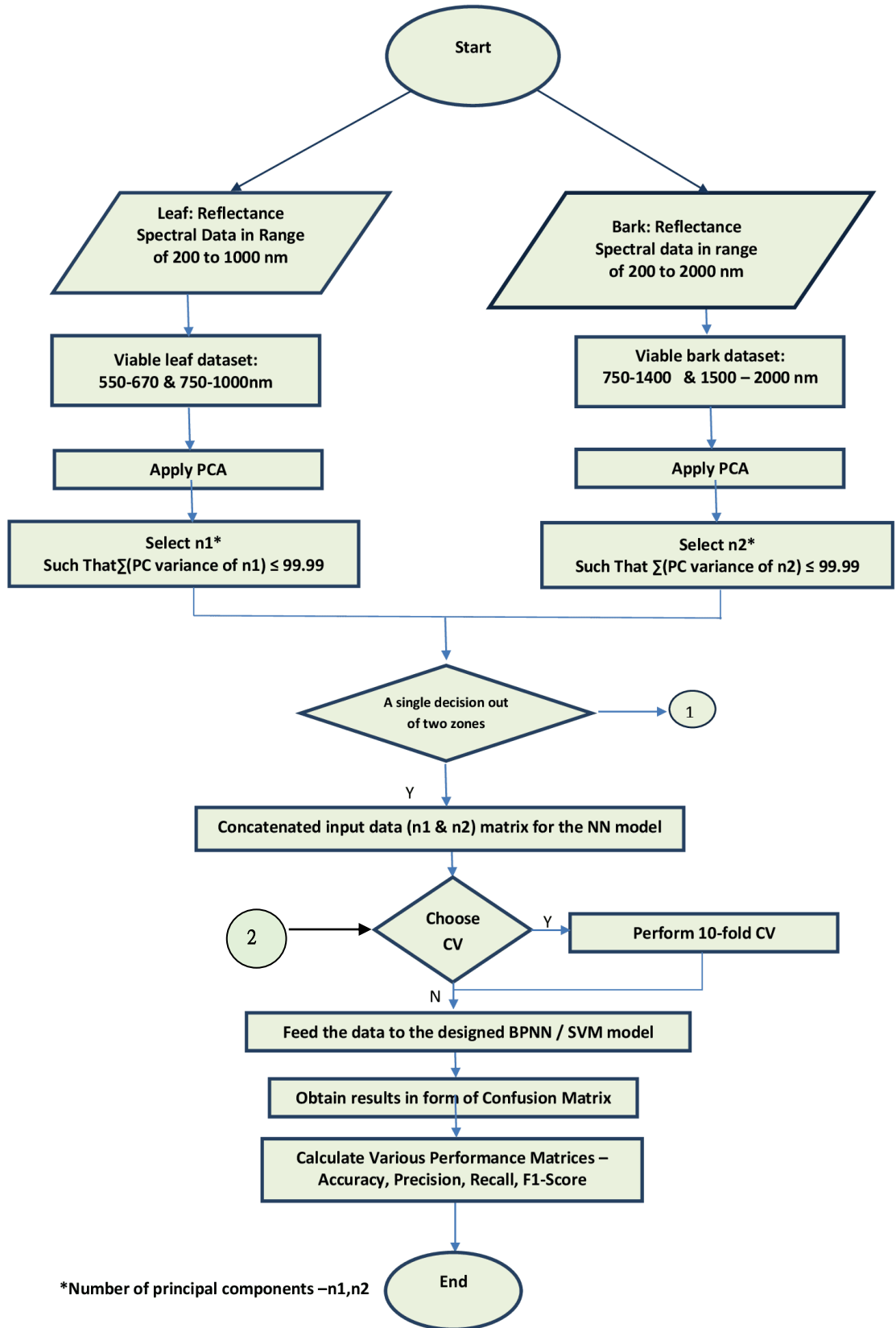


Figure 4.10.a. Flowchart for the (PCA, NN -SVM based) data fusion of the leaf and bark spectra

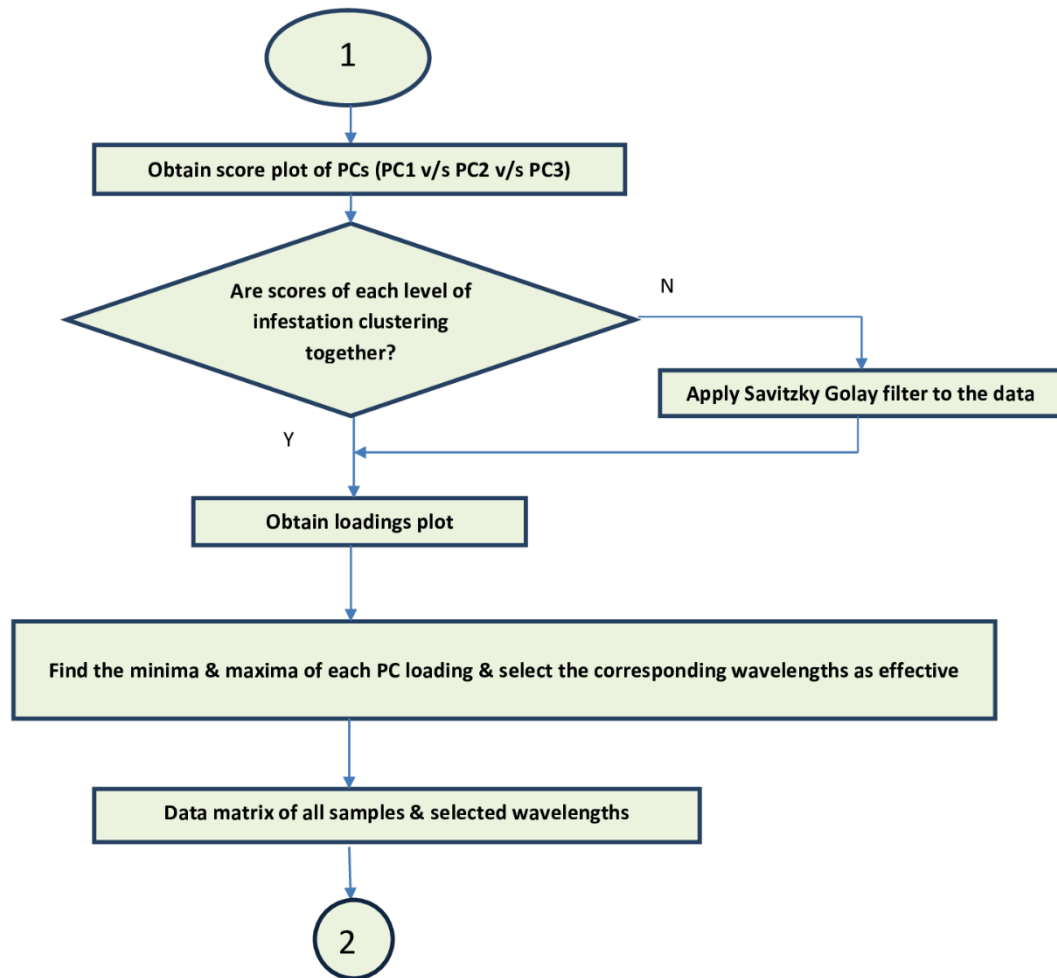


Figure 4.10.b. Flowchart for the (PCA, NN based) individual of the leaf and bark spectra

Classification using BPNN

The advantages of the neural network method over traditional classifiers are its nonparametric nature, arbitrary decision boundary capabilities, easy adaptation to different types of data and input structures, fuzzy output values that can enhance classification, and good generalization for use with multiple images. The disadvantages of the method are slow training time, inconsistent results due to random initial weights, and the requirement of obscure initialization values (e.g., learning rate and hidden layer size) [5].

Specifications of the Back-Propagation Neural Network Model

1. Trained with scaled conjugate gradient algorithm[6].
2. Input: PCA scores or selected wavelengths values (as per the case – Fusion or individual- leaf and bark)

3. Number neurons of hidden layer: 10 (for leaf alone) & 15 (individual bark and fusion of leaf and bark)
4. Output: four stages or levels of infestations.
5. Number of training iterations: 10.

Figure 4.11 is an example of four-class model.

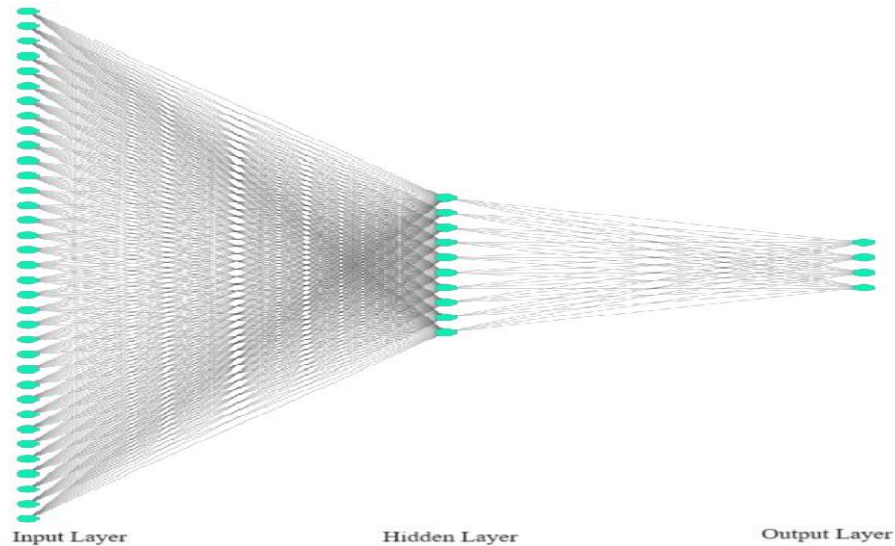


Figure 4.11: A neural network diagram for a four-class model.

Since the considered research problem has four HCs, the number of classes for Machine Learning are four, and each class has equal number of samples (i.e. 10), the method is Balanced Classification.

Cross-Validation

To validate the model a 10-fold cross validation [7] is applied with Stratification to ensure that the test sets include observations from all classes. Each fold has the same proportion of observations with a given class.

Performance metrics

The performance metrics [7] are defined in order to conduct the evaluation of designed models. In case of classification, the outputs are termed as: False Positive (FP), True positive (TP), False Negatives (FN), and True Negative (TN).

Conventionally, the mensuration of Machine learning outcomes is indicated using the defined performance matrices, in the form of

1. Values: Accuracy, Precision, Recall, F1-score [8,9,10].

The Accuracy is an exclusive per class matrix. However, the matrices viz. Precision,

Recall, F1-score per class as mentioned in Equations 4.5, 4.6, 4.7 and 4.8, requires a one-vs-rest manner.

Accuracy is a measure of how many predictions a model got correct out of the total predictions it made. It represents the ratio of correctly predicted instances to the total instances in the dataset. It is a good metric when the class distribution in the dataset is approximately balanced. However, it may not be suitable when dealing with imbalanced datasets, where one class is significantly more frequent than the other.

$$\text{Accuracy} = \frac{(TP+TN)}{TP+TN+FP+FN} \dots\dots\dots(5)$$

Precision is a measure of how many of the predicted positive instances were actually positive. It focuses on the accuracy of the positive predictions. It is important when the cost of false positives is high. For example, in medical diagnosis, you want to ensure that when you predict a disease, it is actually present.

$$\text{Precision} = \frac{TP}{TP+FP} \dots\dots\dots(6)$$

Recall is a measure of how many of the actual positive instances were correctly predicted as positive. It focuses on capturing all positive instances. It is important when the cost of false negatives is high. For example, in spam email detection, you want to ensure that all spam emails are correctly identified to avoid missing important emails.

$$\text{Recall} = \frac{TP}{TP+FN} \dots\dots\dots(7)$$

F1-Score is the harmonic mean of precision and recall. It provides a balance between precision and recall, which can be especially useful when you need to consider both false positives and false negatives. It is a good metric when you want to find a balance between precision and recall. It is particularly useful in situations where there is an uneven class distribution or when both false positives and false negatives have significant consequences.

$$F1 - score = \frac{2*Precision*Recall}{Precision+Recall} \dots\dots\dots(8)$$

These metrics help assess the performance of classification models and can guide model selection and parameter tuning to achieve the desired trade-offs between precision and recall, depending on the specific problem and its requirements.

2. Graphics: Confusion Matrix and ROC curves- AUC values [11,12].

The confusion matrix compares the actual values with the predicted values. The confusion matrix indicates values of False Positive (FP), True positive (TP), False Negatives (FN),

and True Negative (TN), in its different cells. It is used to track the classification model's usefulness and types of errors.

The ROC curves and their corresponding AUC values provide a visual and quantitative assessment of the model's performance for each class. Analysts and data scientists use ROC analysis to understand how well a classification model distinguishes between different classes by examining the trade-off between true positive and false positive rates. In multi-class problems, multiple ROC curves are generated to assess the model's performance for each class independently, and evaluate its discrimination capabilities. The diagonal dashed line represents the ROC curve for a random classifier, indicating the baseline performance. Any ROC curve above this line indicates better-than-random performance. The AUC values, shown in the legend, quantify the quality of each ROC curve. A higher AUC value indicates better classification performance. A perfect classifier would have an AUC of 1.0, while a random classifier would have an AUC of 0.5.

References

1. M. Ali M, A. Al-Ani, D. Eamus, and D. Tan, “A New Image Processing Based Technique to Determine Chlorophyll inPlants”,*American-EurasianJournalonAgriculturalandEnvironmentalScience*,12(10),pp1323–1328, 2012,URL10.5829/idosi.aejaes.2012.12.10.1917
2. J.D. Barnes, L. Balaguer, E. Manrique, S. Elvira, A.W. Davison, “A reappraisal of the use of DMSO for the extraction anddetermination of chlorophylls a and b in lichens and higher plants”, *Environmental and Experimental Botany*, Volume 32, Issue2,1992,Pages 85-100,ISSN0098-8472,https://doi.org/10.1016/0098-8472(92)90034-Y.
3. Absorption Spectra of Chlorophyll a (Light Green) and Chlorophyll b (Turquoise) © Peter v. Sengbusch, [Online] Available athttps://www1.biologie.uni-hamburg.de/b-online/e24/3.hAtm
4. P.W.Teukolsky,“Savitzky-GolaySmoothingFilters”,1990,ComputersinPhysics4669URL10.1063/1.4822961
5. K. Golhani, S. Balasundram, G. Vadamalai, and B. Pradhan, “A Review of Neural Networks in Plant Disease DetectionusingHyperspectralDataInformationProcessinginAgriculture”,2018,5(3)p p354-371URL10.1016/j.inpa.2018.05.002
6. M.F. Moller, “A Scaled Conjugate Gradient Algorithm For Fast Supervised Learning Neural Networks”, 1993, pp 525-533URL10.1016/S0893-6080(05)80056-5.
7. G. S. Handelman, H. K. Kok, R. V. Chandra, A. H. Razavi, S. Huang, M. Brooks, M. J. Lee, and H. Asadi, “Peering into the Black Box of Artificial Intelligence: Evaluation Metrics of Machine Learning Methods”, *American Journal of Roentgenology*, volume 212, number 1, pp. 38-43, 2019, doi 10.2214/AJR.18.20224, note PMID: 30332290.
8. H. Mohammad, and M.N. Sulaiman, “A Review on Evaluation Metrics for Data Classification Evaluations”, *International Journal of Data Mining & Knowledge Management Process*, 5, 01-11, 2015, 10.5121/ijdkp.2015.5201.
9. J. Dessain, “Machine learning models predicting returns: Why most popular performance metrics are misleading and proposal for an efficient metric”, *Expert Systems with Applications*, Volume 199, 2022, 116970, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2022.116970.

Chapter 5

Results and Discussions

The main goal of this study is the detection of the pest attack in cashew tree at early stage. To achieve it, we have used three different modalities. Initially the analysis was done on the leaf and bark individually. Later on, by combining both the leaf and bark spectral data to enhance the detection accuracy. According to the prior preliminary analysis, the fully matured leaves from the tree branch, and bark samples from any height of the tree are decided to record the spectral data set. The range of spectra was chosen according to the standard spectral signatures which has regions for indications of chlorophyll (visible band with reflectance peak around 550 nm), water content (700-1400 nm), cell structure (1300-2000nm).

For the first two individual analysis of leaf and bark, PCA is applied on the original spectra and score plot is observed to decide the tuning parameters of Savitzky-Golay i.e., window size and polynomial order. A score plot helps to observe patterns or trends in the data points with their alignment along the axes. These patterns can provide insights into the relationships between the principal components and the categories or classes. The locations of spectral data points in the 3D space indicate the distribution of spectral data across the principal components. The data points of each health category convey their relation to each other: overlap or separate different health categories. The overlap indicates similarities between categories, whereas the cluster of data points of the same category together indicates similar characteristics or values within the specific principal components. Thus, the Savitzky-Golay filter was tuned till the clustering of four health categories can be visualized in a (2D or 3D) plot.

Similarly, the loadings plot of PCA is used for variable selection and interpretation of a spectra, as it helps to identify the variables which are most relevant in explaining the variance in a spectrum. The identified variables are the ones which contribute most to each principal component. Each wavelength is represented as a vector that points in the direction of maximum variance for that wavelength in the PCA space. The direction of the loading plot indicates whether the variable has a positive or negative relationship with the principal component. A wavelength with a positive loading contributes positively to the

principal component's values, while a wavelength with a negative loading contributes negatively. Variables with higher vectors in the loadings plot are more important in explaining the variance in the data along that principal component. Thus, they have a strong influence on the principal component's representation of the data. Conversely, variables with lower vectors have less influence. Hence, the minima and maxima of the loadings plot of leaf and bark spectra, was used to figure out the important wavelengths, during their individual analysis.

The next sections will give the performance analysis of all three different methodologies i.e Leaf Spectral analysis, Bark Analysis and Fusion of leaf and bark analysis.

5.1 Leaf Spectral Analysis

After the preliminary analysis, fully matured leaves from the tree branches are selected to record the spectral data set. The total number of leaves selected are 400. This database is created by plucking the leaves from different branches of a given tree at each of the 10 trees for each of the four health categories. The selected wavelength range is 550 nm to 670 nm & 750 nm to 1000 nm having 76 wavelength points. To preprocess the spectra, a processing window of 15 and all the polynomial order 2 are selected for Savitzky-Golay smoothing function.

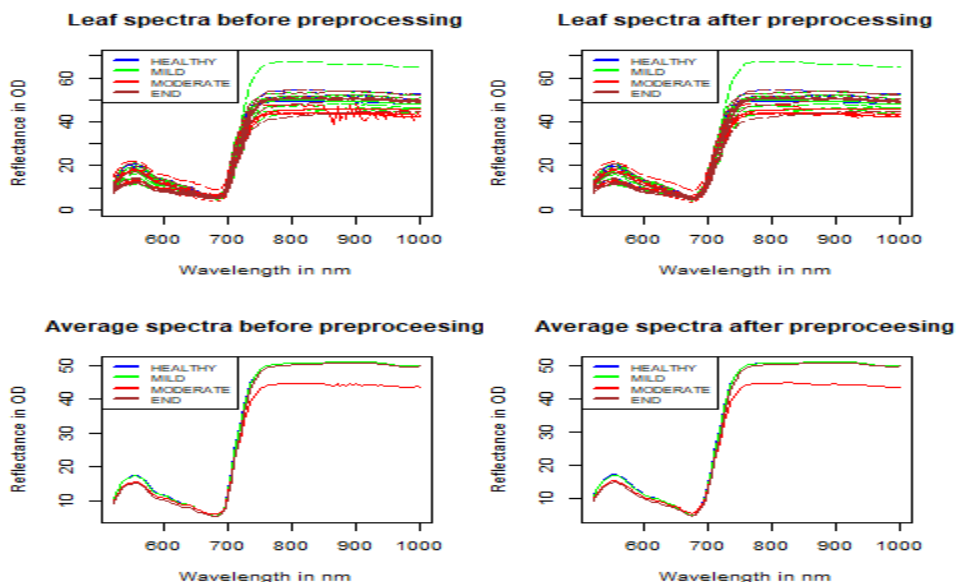


Figure 5.1: The recorded spectra and the preprocessed spectra of leaf, for individual and averaged of each 10 trees in all the four health categories.

Figure 5.1 shows the 40 recorded leaves spectra (Averaged per tree) and the preprocessed spectra, and then averaged for each 10 trees in all the four health categories - Severe, Moderate, Mild, and Healthy. Whereas the figure 5.2 shows the score plot (program code 3-D space. It has axis of principal components PC1, PC2, and PC3 out of seven Principal

components. Individual variances of first seven principal components were 88.24, 9.72, 1.38, 0.53, 0.05, 0.02, 0.01, summing to 99.99% total variance.

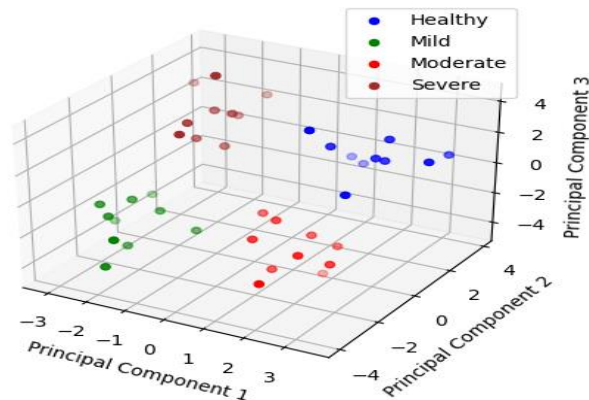


Figure 5.2: The score plot of leaf spectral data representing different health conditions. There are no outliers for the leaf data as there are no unusual data points noticed far away from the main clusters of data. These outliers could indicate exceptions or anomalies within the dataset.

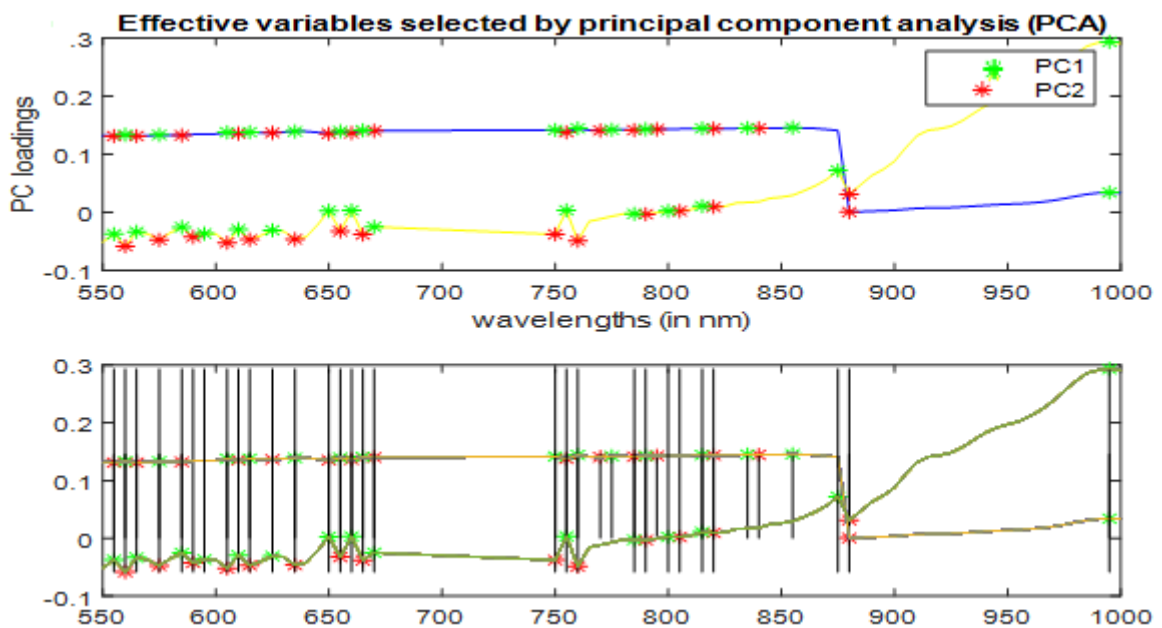


Figure 5.3: PC loadings of first 2 components of the PCA with minima and maxima.

Here only the PC1 and PC2 are selected which accounts for approx. 98% variance (program code as per Annexure I) in order to simplify the model. The loadings plot (program code as per Annexure I) of leaf spectra shown in fig. 5.3 which helps to identify the wavelengths with maximum variance (555, 560, 565, 575, 585, 590, 595, 605, 610, 615, 625, 635, 650, 655, 660, 665, 750, 755, 760, 770, 775, 785, 790, 795, 800, 805, 815, 820, 835, 840, 855, 875, 880, 995 nm). These 35 numbers of wavelengths were the selected variables among original 76 wavelength points, and the corresponding spectra of 40

sample trees constituted the input matrix. These 35 selected variables are given as an input to the designed BPNN model for classification of the 4 health conditions of cashew tree. This would help in simplifying the models and improving the rate of operation greatly, due to minimal number of variables.

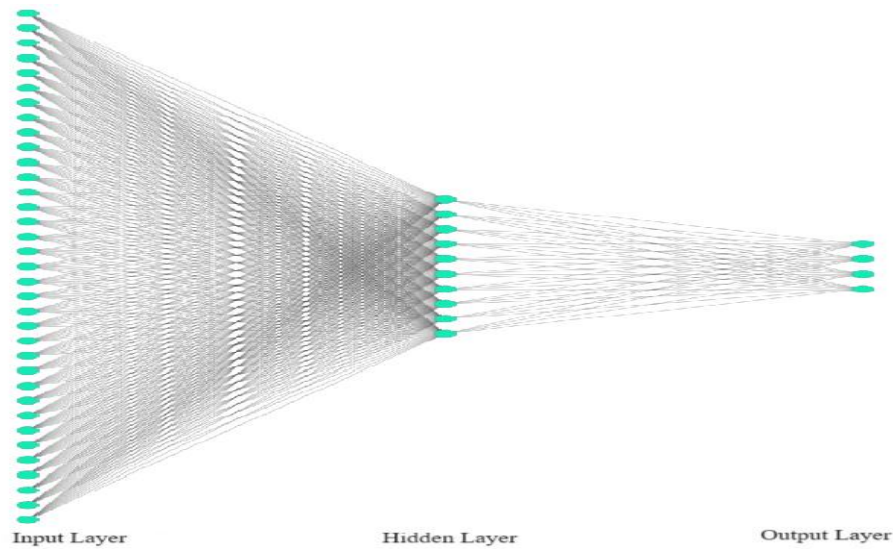


Figure 5.4: Architecture of the BPNN model for leaf spectral data alone.

The BPNN model of leaf alone (architecture as depicted in figure. 5.4) was designed having the inputs layer, hidden layer and output layer with 35,10 and 4 neurons respectively. After 10-fold cross validations, the classifications results are depicted in form of confusion matrix as shown in figure 5.5.

	Healthy	Mild	Moderate	Severe
Prediction	Healthy	Mild	Moderate	Severe
	10	0	0	0
	0	8	1	0
	0	2	8	1
	0	0	1	9
	Healthy	Mild	Moderate	Severe
	Actual			

Figure 5.5: Confusion Matrix for leaf spectra.

All the trees in Healthy stage are classified correctly. Two trees of Mild stage were classified in Moderate Stage, whereas for one tree each of Moderate was classified in Mild and Severe stage. For Severe stage only one tree is misclassified as moderate. Hence, total number of mis-classifications is 5. Since, the correct classifications are 35 out of 40 trees and the average accuracy is 87.50%. Also, considering the individual accuracy values for each of 10-fold, standard deviation was found to be 5.7. From the confusion matrix, The True Positive (TP), False Positive (FP) and False Negative (FN) for each health condition

are noted in Table 5.1, so as to calculate the Precision, Recall, F1-score. The Precision, Recall and F1-score were calculated from TP, FP and FN as 87.90%, 87.50% and 87.60%, respectively.

Table 5.1: Summary of results of the BPNN model for leaf spectral data.

	TP	FP	FN	Precision	Recall	F1-score	Accuracy
HC1	10	0	0	100.00	100.00	100.00	
HC2	8	1	2	88.89	80.00	84.21	
HC3	8	3	2	72.73	80.00	76.19	
HC4	9	1	1	90.00	90.00	90.00	
Average				87.90	87.50	87.60	87.50

The calculations for Precision, Recall, F1-score for one of the classes (HC2) case are as follows

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{8}{8 + 1} = \frac{8}{9} = 88.89\%$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{8}{8 + 2} = \frac{8}{10} = 80.00\%$$

$$\text{F1 - score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 * 88.89 * 80}{88.89 + 80} = \frac{14222.4}{168.89} = 84.21\%$$

The calculations for Accuracy for the entire classification is

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = \frac{37}{40} = 87.50\%$$

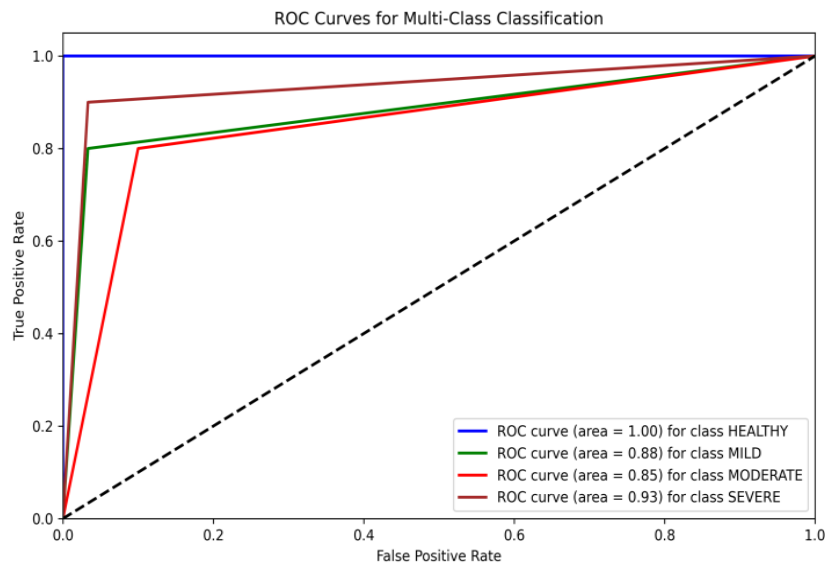


Figure 5.6: Receiver Operating Characteristic Curve for leaf spectra alone. Also, the model performance is validated with another metric parameter i.e Receiver Operating Characteristic Curve as shown in figure5.6. Accordingly, the Area Under curve

(AUC) for Healthy, Mild, Moderate and Severe stages (program code as given in Annexure II) were 1.00, 0.88, 0.85 and 0.93. Since, the values are above 0.7, they are acceptable and the model is very good.

5.2 Bark Spectral analysis

After the preliminary analysis, bark samples from any height of the tree are selected to record the spectral data set. The total number of bark samples used for this study is 80. To create this database, 10 trees were identified for each of the four health categories (40 trees). For each tree, two bark samples were collected (at any heights between the ground level to 1.5m) from points on the circumference which is diametrically apart. The selected wavelength range is 705 to 1400 nm & 1500-2000 nm having 241 wavelength points. To preprocess the spectra, a processing window of 15 and all the polynomial order 2 are selected for Savitzky-Golay smoothing function. Figure 5.7 shows the 40 recorded spectra of bark (Averaged per tree) and the preprocessed spectra of bark, for individual and then averaged for each 10 trees in all the four health categories.

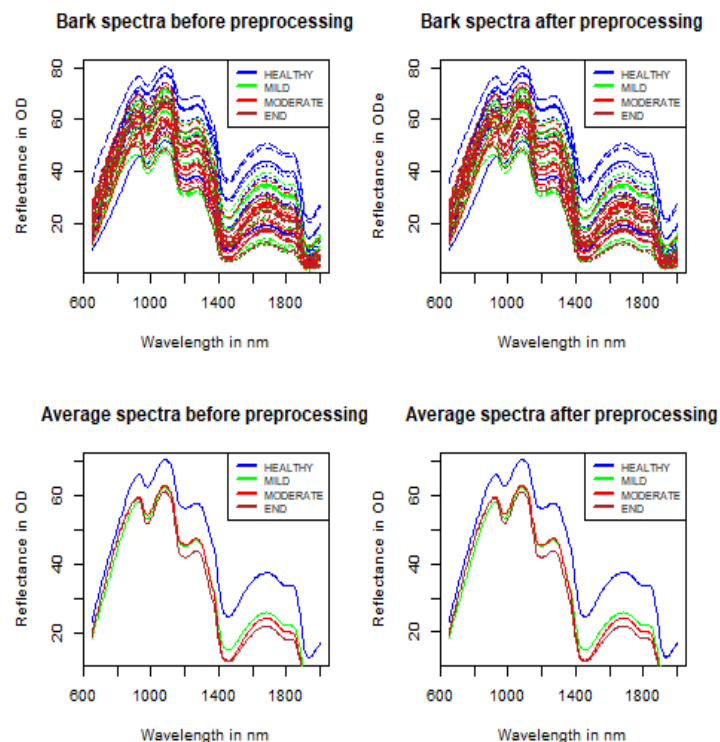


Figure 5.7: the recorded spectra and the preprocessed spectra of bark, for individual and then averaged for each 10 trees in all the four health categories.

The spectra of different health conditions Severe, Moderate, Mild, and Healthy, of bark

data are indicated as points over 3-D space in its score plot (figure 5.8), which has axis of principal components PC1, PC2, and PC3. Individual variances of first ten principal components were 91.601, 5.008, 2.827, 0.386, 0.109, 0.029, 0.014, 0.0042, 0.003, 0.001, summing to summing to 99.99% total variance.

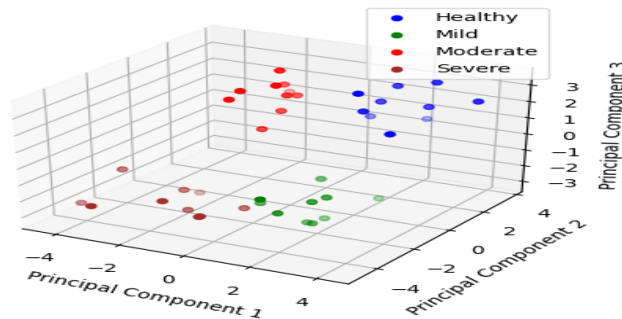


Figure 5.8: Score plot of bark spectral data representing different health conditions.

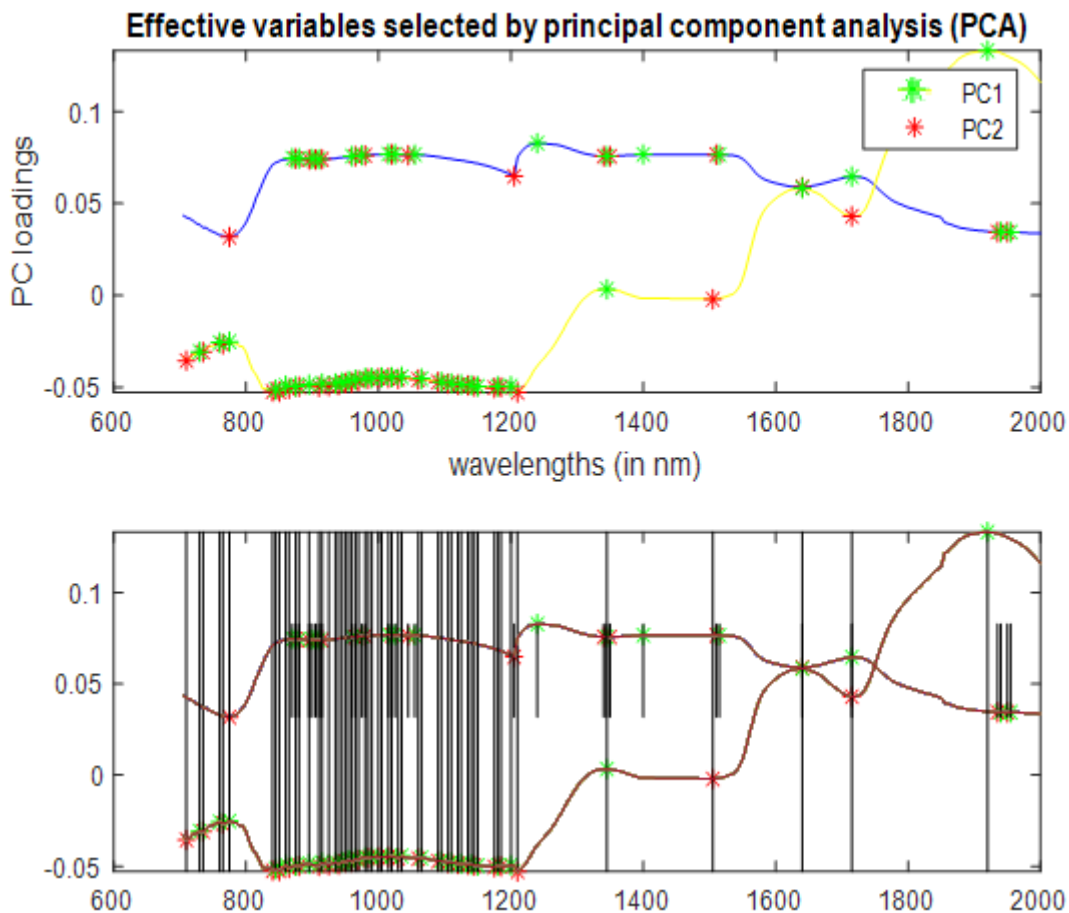


Figure 5.9: PC loadings of first 2 components of the PCA to bark data, with indication of minima and maxima of each.

There are no outliers for the bark data as there are no unusual data points noticed far away from the main clusters of data. These outliers could indicate exceptions or anomalies within the dataset.

Here only the PC1 and PC 2 are selected which accounts for approx. 96.6% variance in order to simplify the model. The loadings plot of bark spectra (as shown in figure 5.8) helped identify the wavelengths with maximum variance (710, 730, 735, 760, 775, 775, 840, 845, 850, 860, 865, 870, 875, 880, 895, 900, 905, 910, 915, 925, 935, 940, 945, 950, 955, 960, 965, 975, 980, 990, 1000, 1005,1015, 1020, 1025, 1030, 1035, 1045, 1055, 1060, 1065, 1090, 1095, 1105, 1110, 1120, 1125, 1135, 1140, 1145, 1150, 1175, 1180, 1185, 1200, 1205, 1210, 1240, 1345, 1350, 1400, 1505, 1510, 1515, 1640, 1715, 1920, 1935, 1940, 1955 nm. These 70 numbers of wavelengths were the selected variables among original 241 wavelength points, and the corresponding spectra of 40 sample trees constituted the input matrix. These 70 selected variables are given as an input to the designed BPNN model for classification of the 4 health conditions of cashew tree. This would help in simplifying the models and improving the rate of operation greatly, due to minimal number of variables.

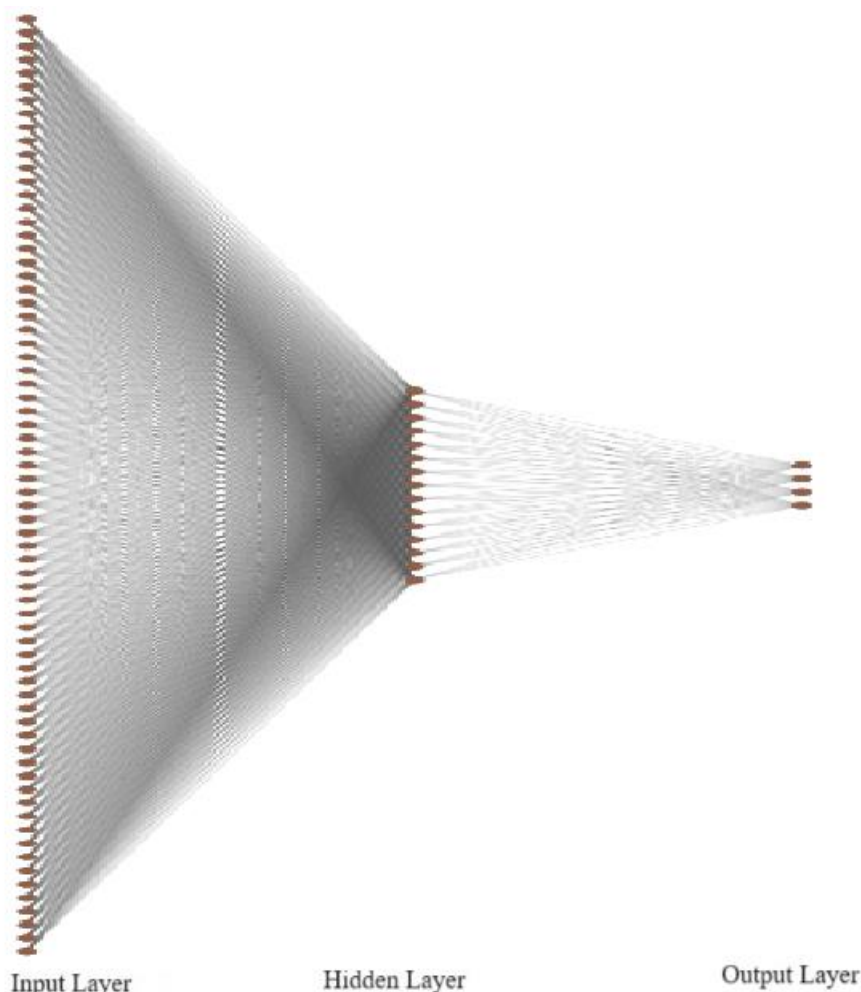


Figure 5.10: Architecture of the BPNN model for bark spectral data alone. The BPNN model of bark alone (architecture as depicted in figure. 5.10) was designed having the inputs layer, hidden layer and output layer with 70,15 and 4 neurons

respectively. After 10-fold cross validations, the classifications results are depicted in form of confusion matrix (program code as given in Annexure III) as shown in figure 5.11.

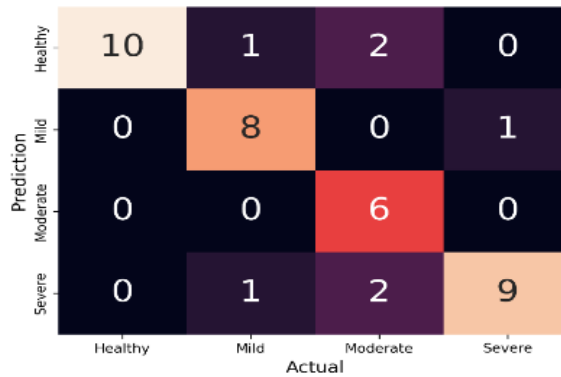


Figure 5.11: Confusion Matrix for bark spectra alone.

All the trees in Healthy stage are classified correctly. The mis-classifications found for Mild stage were one each into the Healthy and Severe Stage, whereas for Moderate Severe stage was two each into Healthy and Severe stage. The only one misclassification found for Severe Stage was into Mild Stage. Since the correct classifications are 33 out of 40 trees of all classes, the average accuracy is 82.5%. Also, considering the individual accuracy values for each of 10-fold, standard deviation was found to be 6.4. From the confusion matrix, The True Positive (TP), False Positive (FP) and False Negative (FN) for each health condition are noted in Table 5.1, so as to calculate the Precision, Recall, F1-score. The Precision, Recall and F1-score were calculated from TP, FP and FN as 85.20%, 82.50% and 82.0%, respectively

Table5.2: Summary of results of the BPNN model for bark spectral data.

	TP	FP	FN	Precision	Recall	F1-score	Accuracy
HC1	10	1	0	76.90	100.00	87.00	
HC2	9	0	1	88.90	80.00	84.20	
HC3	10	0	0	100.00	60.00	75.00	
HC4	10	0	0	75.00	90.00	81.80	
Average				85.20	82.50	82.00	82.50

Also, the classifications results are depicted in form of the Receiver Operating Characteristic Curve as shown in figure 5.12). Accordingly, the Area Under curve (AUC) for Healthy, Mild, Moderate and Severe stages (program code as given in Annexure III) were 0.95, 0.88, 0.80 and 0.90. Since, the values are above 0.7, they are acceptable and the model is very good.

As per the accuracy values of leaf and bark data alone, i.e., 85% and 82.7% respectively, idea of tracking the two distinct thrust zones separately is not promising to detect CSRB attack. Hence, it is decided to combine the data as per designed methodology, and classifying to obtain results using two algorithms - BPNN and SVM.

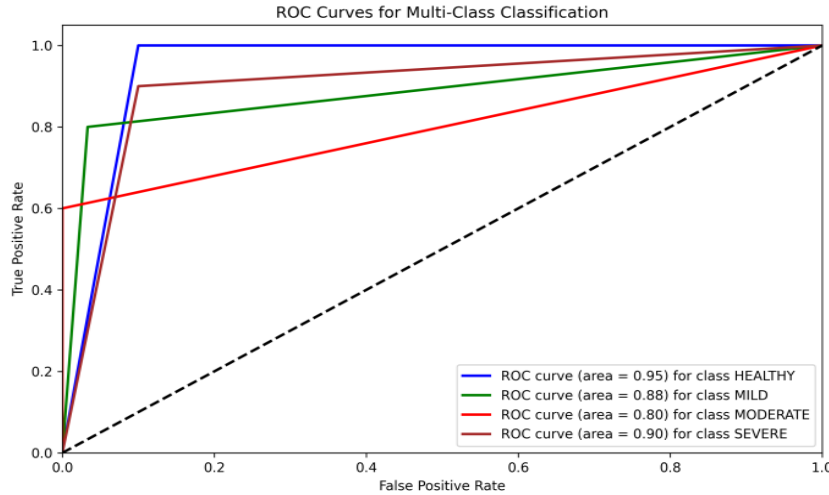


Figure 5.12: Receiver Operating Characteristic Curve for bark spectra alone.

5.3 Data Fusion (Leaf and Bark)

The leaf spectra were recorded in the range 550nm to 670nm & 750nm to 1000nm with resolution of 5nm gives 76 wavelengths points. Similarly bark reflectance spectra was recorded in the range 705-1400nm & 1500-2000 nm having 241 wavelengths points. Both these bark and leaf reflectance spectra are preprocessed with Savitzky-Golay technique. Thereafter, PCA was applied to each of the two datasets of leaves and bark. For the PCA, the number of samples is the number of cashew trees (i.e., 40) and variables are the number of wavelengths. The principal components which meet the total variance of 99.99% i.e., the first seven and ten components of leaf and bark, were retained. Accordingly, the synergizing of both leaf and bark to generate the input to the data fusion models leads to total seventeen variables, due to concatenation of PC1 to PC7 of leaf and PC1 to PC10 of bark, and thus, forms a single dataset of 40*17(program code as given in Annexure II). The other input to the data fusion models is the array of labels (dimension of 40*1), which are the health conditions of the each of the 40 trees known as per noted information during data collection. The output layer of the data fusion models is generated in two formats: - 1) 2 Classes (Class1: desirable (Healthy & Mild) and Class2: undesirable (Moderate and Severe)), 2) 4 Classes - Healthy, Mild, Moderate and Severe.

5.3.1 Classification using BNNN

5.3.1.1 The 2 Class model

The BPNN model for fusion of leaf and bark data (architecture as depicted in figure 5.13) was designed having the inputs layer, hidden layer and output layer with 17,15 and 2 neurons respectively.

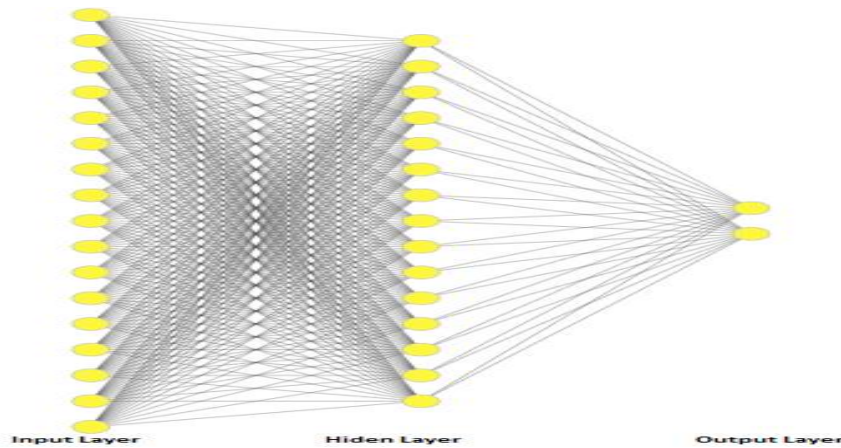


Figure 5.13: Architecture of the BPNN for leaf and bark spectral data together, for 2 class model.

After 10-fold cross validations, the classifications results are depicted in form of confusion matrix as shown in figure 5.14.

Prediction	class 1	17	2
	class 2	3	18
		class 1	class 2
		Actual	

Figure 5.14: Confusion Matrix for leaf and bark spectral data, for 2 class model using BPNN.

The mis-classifications found for Class 1 were three, whereas the mis-classifications for Class 2 were two. Since the correct classifications are 35 out of 40 trees of all classes and the average accuracy is 87.50%. From the confusion matrix, The True Positive (TP), False Positive (FP) and False Negative (FN) for each health condition are noted in Table 5.3. The Precision, Recall and F1-score were calculated from TP, FP and FN as 100%, 100% and 92.45%.

Table 5.3: Summary of results of the BPNN model for leaf and bark spectral data, for 2 Classes.

	TP	FP	FN	Precision	Recall	F1-score	Accuracy
Class 1	17	2	3	89.5	85	87.2	
Class 2	18	3	2	85.7	90	87.8	
	Average			87.6	87.5	87.5	87.5

5.3.1.2 The 4 Class model

The BPNN model for fusion of leaf and bark data (architecture as depicted in figure. 5.15) was designed having the inputs layer, hidden layer and output layer with 17,15 and 4 neurons respectively.

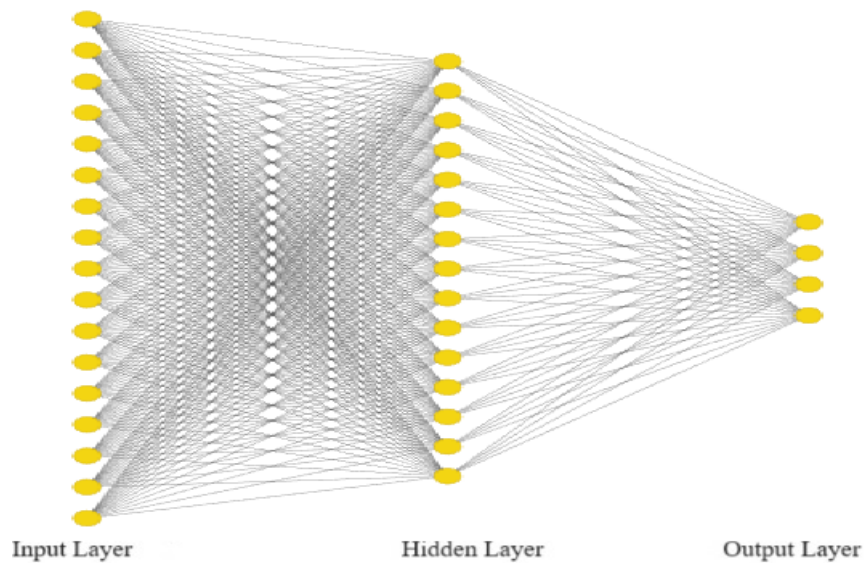


Figure 5.15: Architecture of the BPNN model for leaf and bark spectral data together.

After 10-fold cross validations, the classifications results are depicted in form of confusion matrix (program code as given in Annexure III) as shown in figure 5.16.

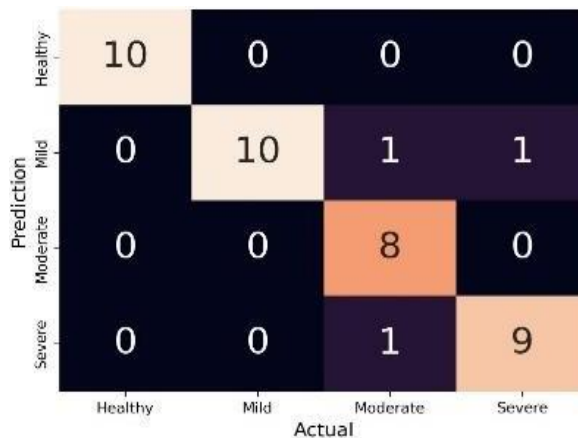


Figure 5.16: Confusion Matrix for leaf and bark together, using BPNN.

All the trees in Healthy and Mild stages are classified correctly. The mis-classifications found for Moderate stage were one each into Mild and Severe, whereas the mis-classification for Severe stage was one into Mild stage. Since the correct classifications are 37 out of 40 trees of all classes and the average accuracy is 92.50%. Also, considering the individual accuracy values for each of 10-fold, standard deviation was found to be 3.1. From the confusion matrix, The True Positive (TP), False Positive (FP) and False Negative (FN) for each health condition are noted in Table 5.4. The Precision, Recall and F1-score were calculated from TP, FP and FN as 93.33%, 92.50% and 92.45%.

Table 5.4: Summary of results of the BPNN model for leaf and bark spectral data, for 4 Classes

	TP	FP	FN	Precision	Recall	F1-score	Accuracy
HC1	10	0	0	100.00	100	100.00	
HC2	10	2	0	83.33	100	90.91	
HC3	8	0	2	100.00	80	88.89	
HC4	9	1	1	90.00	90	90.00	
	Average			93.33	92.50	92.45	92.50

The calculations for Precision, Recall, F1-score for one of the classes (HC4) case are as follows

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{9}{9 + 1} = \frac{9}{10} = 90.00\%$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{9}{9 + 1} = \frac{9}{10} = 90.00\%$$

$$\text{F1 - score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 * 90 * 90}{90 + 90} = \frac{16200}{180} = 90.00\%$$

The calculations for Accuracy for the entire classification is

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = \frac{37}{40} = 92.5\%$$

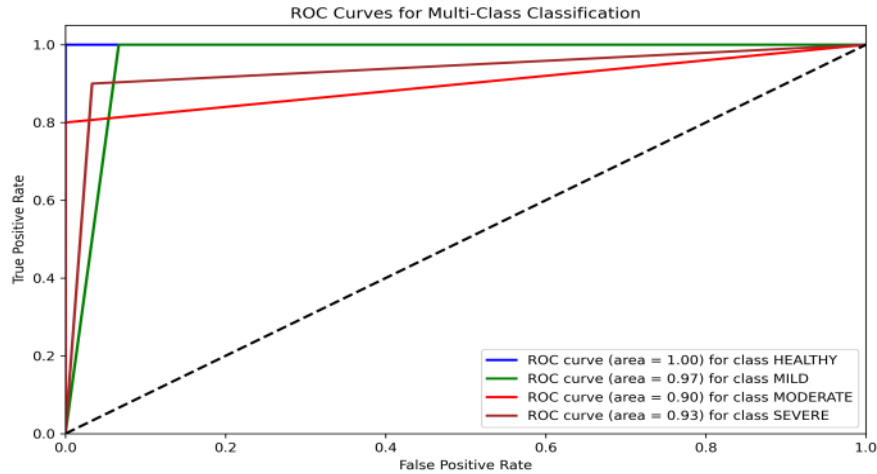


Figure 5.17: Receiver Operating Characteristic Curve for leaf and bark together, using BPNN.

Also, the classifications results are depicted in form of the Receiver Operating Characteristic Curve (figure 5.17). Accordingly, the Area Under curve (AUC) for Healthy, Mild, Moderate and Severe stages (program code as given in Annexure II) were 1.00, 0.97, 0.90 and 0.93. Since, the values are above 0.90, which is well above 0.7 level of expectation, they are acceptable and the model is very good.

5.3.2 Classification Using SVM

For combining the leaf and bark data, principal components which meet the total variance of 99.99% were utilized. The data fusion had input with first seven and ten components of leaf and bark, respectively. Hence the input to the SVM model is concatenated matrix of total seventeen variables (PC1 to PC7 of leaf and PC1 to PC10 of bark) along with the 40 number of samples.

5.3.2.1 The 2-Class model

The outputs of the SVM are two classes. After the 10-fold cross validations, the classifications results are depicted in form of confusion matrix as shown in figure 5.18.

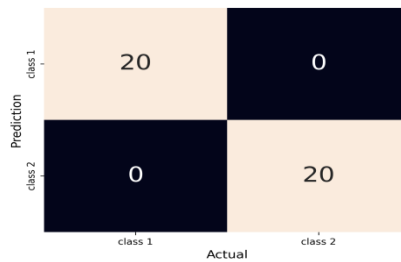


Figure 5.18: Confusion Matrix for leaf and bark spectral data, for 2 class model using SVM.

Since both the classes are classified correctly, i.e. correct classifications as 40 out of 40 trees, the accuracy is 100%. From the confusion matrix, The True Positive (TP), False Positive (FP) and False Negative (FN) for each health condition are noted in Table 5.5. The Precision, Recall and F1-score were calculated from TP, FP and FN as FN as 100%, 100% and 100%.

Table 5.5: Summary of results of SVM for leaf and bark spectral data, for 2 Class model.

	TP	FP	FN	Precision	Recall	F1-score	Accuracy
Class 1	20	0	0	100	100	100	
Class 2	20	0	0	100	100	100	
	Average			100	100	100	100

5.3.2.2 The 4-Class model

The output of the SVM was four classes. After the 10-fold cross validations, the classifications results are depicted in form of confusion matrix (program code as given in Annexure III) as shown in figure 5.19.

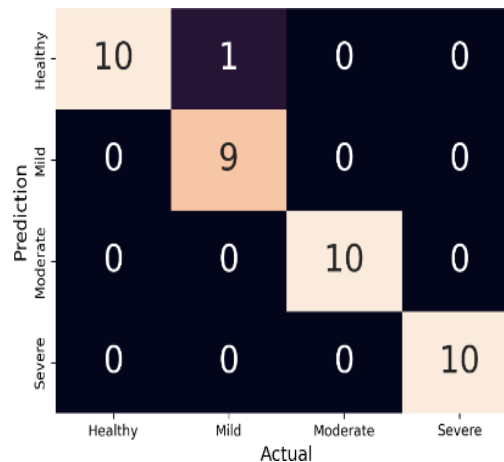


Figure 5.19: Confusion Matrix for leaf and bark together, using SVM.

All the trees in Healthy, Moderate and Severe stages are classified correctly. The only one mis-classification found for Mild stage was into Healthy. Since the correct classifications are 39 out of 40 trees of all classes and the average accuracy is 97.50%. Also, considering the individual accuracy values for each of 10-fold, standard deviation was found to be 2.5. From the confusion matrix, The True Positive (TP), False Positive (FP) and False Negative (FN) for each health condition are noted in Table 5.6. The Precision, Recall and F1-score were calculated from TP, FP and FN as 97.73%, 97.50% and 97.49%, respectively.

Table 5.6: Summary of results of SVM for leaf and bark spectral data, for 4 Class model.

	TP	FP	FN	Precision	Recall	F1-score	Accuracy
HC1	10	1	0	90.91	100	95.24	
HC2	9	0	1	100.00	90	94.74	
HC3	10	0	0	100.00	100	100.00	
HC4	10	0	0	100.00	100	100.00	
	Average			97.73	97.5	97.49	97.50

The calculations for Precision, Recall, F1-score for one of the classes (HC1) case are as follows

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{10}{10 + 1} = \frac{10}{11} = 90.91\%$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{10}{10 + 0} = \frac{10}{10} = 100.00\%$$

$$\text{F1 - score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 * 90.91 * 100}{90.91 + 100} = \frac{18182}{190.91} = 95.24\%$$

The calculations for Accuracy for the entire classification is

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = \frac{39}{40} = 97.5\%$$

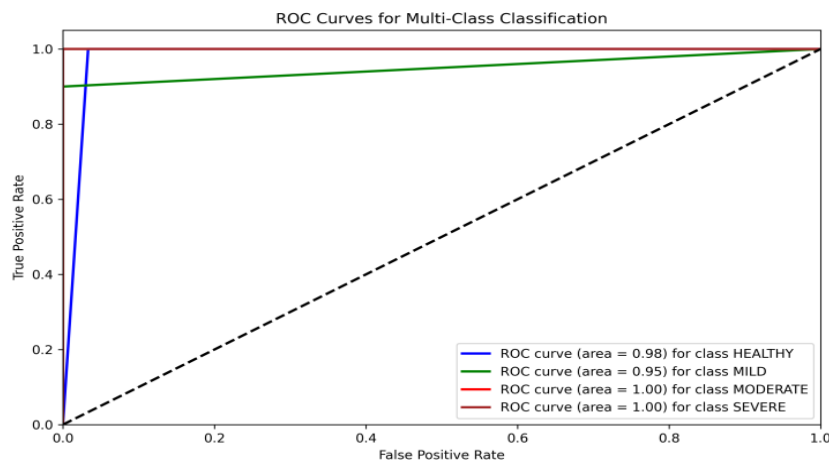


Figure 5.20: Receiver Operating Characteristic Curve for leaf and bark together, using SVM.

Also, the classifications results are depicted in form of the Receiver Operating Characteristic Curve as shown in figure 5.20. Accordingly, the Area Under curve (AUC) for Healthy, Mild, Moderate and Severe stages (program code as given in Annexure III) were 0.98, 0.95, 1.00 and 1.00.

Table 5.7: Summary of results of all three modalities, for 4 class models.

	Leaf alone	Bark alone	Leaf and bark - BPNN	Leaf and bark - SVM
Accuracy	87.5 ± 5.7	82.5 ± 6.4	92.5 ± 3.1	97.5 ± 2.5
Precision	87.9	85.2	93.33	97.7
Recall	87.5	82.5	92.5	97.5
F1-score	87.6	82	92.45	97.5

Since, the accuracy, precision, recall and f1-score are the highest and standard deviation is lowest amongst all for models, the SVM is best suitable to detect the CSRB attack in cashew trees.

Chapter 6

Conclusion

As per the overall conduct of methodology following are the breakthrough achievements of the research over all the previous plant disease detection systems: -

1. Spectral recordings of the trunk zone of the cashew trees.
2. Spectral recordings of the 4 Health Condition stages of the cashew trees.
3. The computational model for fusion of Spectroscopic data of two thrust zones (i.e., leaves and bark) of cashew tree.

A model for early detection for CSRB detection has been designed and analyzed using three modalities. The data collection was done using two spectrophotometers- USB 2000+ and Jasco- V770. The entire design consisting of preprocessing the spectra, effective wavelength selection, PCA and the Machine learning algorithms-BPNN and SVM, has been conducted using MATLAB R2021a and Python 3.10.12.

In case I (Leaf Spectral Analysis): The input dataset to BPNN model was 40×35 , which was after reducing the recorded dataset of 40×76 , as per effective wavelengths selections conducted due to PCA. The accuracy, precision, recall and f1-score were 87.9%, 87.5% and 87.6%, for the leaf spectral analysis, whereas AUC for the Health, Mild, Moderate and Severe health conditions were 1.00, 0.88, 0.85 and 0.93.

In case II (Bark Spectral Analysis): The input dataset to BPNN model was 40×70 , which was after reducing the recorded dataset of 40×241 , as per effective wavelengths selections conducted due to PCA. The accuracy, precision, recall and f1-score were 82.5%, 85.2%, 82.2% and 82.0%, whereas AUC for the Health, Mild, Moderate and Severe health conditions were 0.95, 0.88, 0.80 and 0.90.

In case III (Data fusion): The model of BPNN had 17, 15 and 4 neurons in its three layer of input, hidden and output. The input dataset to BPNN and SVM was 40×17 (consisting PC1 to PC7 of leaf and PC1 to PC10 of bark). The accuracy, precision, recall and f1-score for the model of leaf and bark data fusion using BPNN were 92.5%, 93.33%, 92.50% and 92.45% respectively, whereas AUC for the Health, Mild, Moderate and Severe health conditions were 1.00, 0.97, 0.90 and 0.93. The accuracy, precision, recall and f1-score for the model of leaf and bark data fusion using SVM were 97.5%, 97.73%, 97.50% and 97.49% respectively, whereas AUC for the Health, Mild, Moderate and Severe health

conditions were 0.98, 0.95, 1.00 and 1.00.

Table 6.1: Summary of results of all three modalities.

	Leaf alone	Bark alone	Leaf and bark - BPNN	Leaf and bark - SVM
Accuracy	87.5 ± 5.7	82.5 ± 6.4	92.5 ± 3.1	97.5 ± 2.5
Precision	87.9	85.2	93.33	97.7
Recall	87.5	82.5	92.5	97.5
F1-score	87.6	82	92.45	97.5

As per the individual analysis of leaf and bark, the correctly classified trees were 33 and 35. With the data fusion and using BPNN model, the number improved to 37. Ultimately, the fused data given to the SVM model classified correct trees as 39. These results were obtained after the 10-fold cross validation so that test and train sets do not produce a single result constantly, as the train-test sets are different over 10 iterations in the cross-validation. Thus, the standard deviations are noted as 5.7, 6.4, 3.1 and 2.5.

Considering the lowest accuracy (lower bound) of data fusion model of leaf and bark using SVM i.e., 95.0%, is not exceeded either by the highest accuracies (upper bound) of leaf alone or bark alone i.e., 93.2% or 88.9%, and also that the metrics of Precision, Recall and F1-score are the highest amongst all for models, the SVM is best suitable to detect the CSR attack in cashew trees.

Future scope

One future scope of the research lies in designing detection at other locations of CSR attacks than the Mollem-Goa.

This would have the following three major appendages: -

1. At certain times, the characteristics of the symptoms of the infestation may vary with the geographic position. So, the basis of detection design being concretely dependant on symptoms details, tracking them with supposedly effect of variation in nutritional scenario is of prime importance. The utmost focus is attributed to the variety of soil composition type of different geographic locations; the various states in India viz. Andhra Pradesh, Goa, Maharashtra, Karnataka have soils ranging from loomy to rocky, laterite to sandy and, this

variety moreover persists within a given state too. There is a whole zone of study called ‘**domain adaption**’, devoted to relieving this problem. This incorporates procedures such as support vector machines, Bayesian divergence prior, PAC-Bayesian analysis, transfer component analysis etc.

2. In this study, the data carried its corresponding labels comfortably, as the different trees could be unanimously spotted out for their health status. This was possible greatly for the reason that, the location of data collection was well known for a large span of time (during and before data collection). Hence, the knowledge of case of each tree was exclusively at the researchers’ disposal. And thus, the different stages of infestation were spotted amongst them. This privilege will not be possible for any other place of location of CSRB. It can be tedious to do guess-work upon the labels of the collected samples, and eventually end up inaccurate results because an ML model is only good as its data. Therefore, in those cases, we would resort to neural networks based on **unsupervised classification**, known as Clustering or segmentation. In clustering problems, a neural network groups data by similarity. Since the self-organizing feature map (SOFM) is the most popular network for clustering, it can be applied to track CSRB attacks on locations than the Mollem.

3. To implement the methodology at other locations of CSRB attacks than the Mollem, we can even resort to a different step of data collection, which could be the procuring or obtaining of images instead of spectral data. Even the issues stated above could be incorporating this step, so they will revise or escalate on same grounds. The images collection includes different levels of complexities for obtaining the data – ranging from a simple camera mounted at heights (using some setup to hold via scaffolding), to UAV (Unmanned Aerial Vehicles) or drones, to satellite images. The images can be of type hyper-multi spectral or thermal. They are collectively called **Remote Sensing**.

Concisely, the pondering over detection design must be expanded to consider information from different other locales of the nation as well, conceivably utilizing a viable one amongst any of the above-mentioned methods.

There could be a study to investigate and confirm that, even in cashew trees, the VOCs caused by the mechanical damage are different from those caused by CSRB, possibly using **MOS-based E-nose**.

Another future study could be on temporal frontiers, i.e., to estimate the rate of deterioration due to infestation in tress, and the rate of the spread over a plantation. In that context, a **forecasting** kind of approach will have to be resorted. The following are its reverberations:

1. At any point of time, when the health condition of a particular tree is traced, the next case of infestation is obviously known i.e., the upcoming stage of infestation. But it is vital to track how fast will be the occurrence of upcoming stage. Specifically for the case of mild stage of infestation – how much sooner the given tree will be encountering the impending moderate stage is critical; since the preventive cure measures have to be initiated, and moreover, also completed.
2. Apart from time constraint, the estimation of optimum quantity of pesticides or fertilizers (whichever applicable) without disturbing the soil chemistry profile, other CSRB non-infested parts of same tree and the other CSRB non-infested trees in the same plantation, can also be forecasted.

The main aim is undoubtedly to prevent the tree to reach the severe stage (dead end), or even a moderate stage (at and after which all measures are futile).

All the powerful preventive procedures will be fruitful only if the mild case detection is effective. Consequently, only by marking and tagging the mild cases as an ‘alarm situation’, the preventive measures would be successfully accomplished within appropriate duration of time and optimum quantity of chemicals.

ANNEXURE I

MATLAB Program for PCA : Principal components to total variance 99.99 and variances of each Principal components

```
load('arcane_DATASETS_exact');

%% PCA part: to get the features of leaves and trunk data
[~,~,~,l_variances]=pca(pca_inputs_BD);
l_sum_explained = 0; l_idx = 0;
while l_sum_explained < 99.99
    l_idx = l_idx + 1; l_sum_explained = l_sum_explained + l_variances(l_idx);
end
[l_loadings,l_score,~,~,~]=pca(pca_inputs_BD,'NumComponents',l_idx); % leaf features

[~,~,~,b_variances]=pca(pca_inputs_BD);
b_sum_explained = 0; b_idx = 0;
while b_sum_explained < 99.99
    b_idx = b_idx + 1; b_sum_explained = b_sum_explained + b_variances(b_idx);
end
[~,b_score,~,~,~]=pca(pca_inputs_D,'NumComponents',b_idx % bark f); % Bark features

PC1=l_loadings(:,1);
PC2=l_loadings(:,2);
pc1=l_score(:,1);
pc2=l_score(:,2);
pc3=l_score(:,3);
```

MATLAB Program for Effective Wavelengths selection and plotting them on loading plot of PCA

```
% Find local min and max
min_PC1=islocalmin (PC1); max_PC1=islocalmax (PC1);
min_PC2=islocalmin (PC2); max_PC2=islocalmax (PC2);
Wavelengths=wavelengthT'
% plots of the positions of local min and max, and plot with the corresponding line
indicators

figure;
subplot(2,1,1);
```

```

plot(Wavelengths,PC1,'b',Wavelengths,PC2,'y',Wavelengths(max_PC1),PC1(max_PC1),'g
*');
                                hold
                                on;
plot(Wavelengths(min_PC1),PC1(min_PC1),'r*',Wavelengths(min_PC2),PC2(min_PC2),'r
*',Wavelengths(max_PC2),PC2(max_PC2),'g*');
legend('PC1','PC2','Orientation','horizontal', 'location','northwest'); title('Effective variables
selected by principal component analysis (PCA)'); xlabel('wavelengths (in nm)');
ylabel('PC loadings');
subplot(2,1,2);
% mark positions of local max and min
plot(Wavelengths,PC1,'b',Wavelengths(min_PC1),PC1(min_PC1),'r*',Wavelengths(max_P
C1),PC1(max_PC1),'g*'); hold on;
plot(Wavelengths,PC2,'y',Wavelengths(min_PC2),PC2(min_PC2),'r*',Wavelengths(max_P
C2),PC2(max_PC2),'g*');

% Part 1: insert the minima lines ----
% PC1 --
    Array_LocMinWavelengths_PC1=Wavelengths(min_PC1); % locations of minima on
x-axis
    Array_LocMinPC1=PC1(min_PC1); % values of the minima of PC1
    for i = 1:241
        Matrix_LocMinWavelengths_PC1(:,i)=Array_LocMinWavelengths_PC1; end
    for i=1:length(Array_LocMinWavelengths_PC1);
plot(Matrix_LocMinWavelengths_PC1(i,:),PC1,'k- '); hold on; plot(Wavelengths,PC1,'b');
end
% PC2 --
    Array_LocMinWavelengths_PC2=Wavelengths(min_PC2); % locations of minima on
x-axis
    Array_LocMinPC2=PC2(min_PC2); % values of the minima of PC1
    for i = 1:241
        Matrix_LocMinWavelengths_PC2(:,i)=Array_LocMinWavelengths_PC2; end
    for i=1:length(Array_LocMinWavelengths_PC2);
plot(Matrix_LocMinWavelengths_PC2(i,:),PC2,'k- '); hold on; plot(Wavelengths,PC2,'y');
end

% Part 2: create maxima lines ----
% % PC1 --
    Array_LocMaxWavelengths_PC1=Wavelengths(max_PC1); % locations of maxima on
x-axis
    Array_LocMaxPC1=PC1(max_PC1); % values of the maxima of PC1
    for i = 1:241
        Matrix_LocMaxWavelengths_PC1(:,i)=Array_LocMaxWavelengths_PC1;
end
    for i=1:length(Array_LocMaxWavelengths_PC1);

```

```

plot(Matrix_LocMaxWavelengths_PC1(i,:),PC1,'k- '); hold on; plot(Wavelengths,PC1);
end
% % PC2 --
Array_LocMaxWavelengths_PC2=Wavelengths(max_PC2); % locations of minima on
x-axis
Array_LocMaxPC2=PC2(max_PC2); % values of the minima of PC1
for i = 1:241
    Matrix_LocMaxWavelengths_PC2(:,i)=Array_LocMaxWavelengths_PC2;
end
for i=1:length(Array_LocMaxWavelengths_PC2);
plot(Matrix_LocMaxWavelengths_PC2(i,:),PC2,'k- '); hold on; plot(Wavelengths,PC2);
end

selWavelengths_PC1=cat(1,Array_LocMaxWavelengths_PC1,Array_LocMinWavelengths
_PC1);selWavelengths_PC2=cat(1,Array_LocMaxWavelengths_PC2,Array_LocMinWave
lengths_PC2);selWavelengths_allPCs=cat(1,selWavelengths_PC1,selWavelengths_PC2);

```

Python program for plotting 3-D score plot of PCA components

```

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler
from scipy.signal import savgol_filter

raw = pd.read_excel("/content/drive/MyDrive/Colab Notebooks/PHD/leaf.xlsx")
raw1 = pd.read_excel("/content/drive/MyDrive/Colab Notebooks/PHD/bark-proper.xlsx")
X_L = raw.values[:,1:].astype('float32')

Xs_L = StandardScaler().fit_transform(X_L)
L_SG = savgol_filter(X_L, 21, polyorder = 2, deriv=2)
pca = PCA(n_components=4)
Xpca = pca.fit_transform(L_SG)
df = pd.DataFrame(Xpca, columns=['PC1', 'PC2', 'PC3', 'PC4'])

X_B = raw1.values[:,1:].astype('float32')
B_SG = savgol_filter(X_B, 21,polyorder=2)
pca = PCA(n_components=6)
Xs_B = StandardScaler().fit_transform(X_B)
Xpca1 = pca.fit_transform(B_SG)
df1 = pd.DataFrame(Xpca1, columns=['PC1', 'PC2', 'PC3', 'PC4', 'PC5','PC6'])

import matplotlib.pyplot as plt
import numpy as np

```

```

# ENER pc1, pc2, and pc3 into x, y, and z
#x,y,z = df1.PC1,df1.PC2,df1.PC3

x,y,z = df.PC1,df.PC2,df.PC3

# Create a 3D scatter plot
fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')
# Assign specific colors to each group of 10 points
colors = ['brown', 'red', 'green', 'blue']

# Determine the total number of data points
num_points = len(x)
# Create an array of colors for each data point
data_point_colors = np.repeat(colors, num_points // len(colors))

# Scatter plot the data points with the specified colors
scatter = ax.scatter(x, y, z, c=data_point_colors, marker='o')

# Create a legend with labels for each group
legend_labels = ['Severe', 'Moderate', 'Mild', 'Healthy']
legend_handles = [plt.Line2D([0], [0], marker='o', color='w', label=label, markersize=8,
markerfacecolor=color) for label, color in zip(legend_labels, colors)]
ax.legend(handles=legend_handles, labels=legend_labels)

# Set labels and title
ax.set_xlabel('PC 1')
ax.set_ylabel('PC 2')
ax.set_zlabel('PC 3')
ax.set_title('Effect of SG-filter for leaf')

# Save the plot as a PNG image
plt.savefig('cluster_leaf.png', dpi=300)
plt.savefig('cluster_leaf.pdf')
# Show the plot
plt.show()

```

MATLAB code for Individual LeafSpectral Analysis

```

%%% selWavelengths_allPCs' : run code for EW on the bark data (40*76)
input= selWavelengths_allPCs'; %double

% hidden layers : = 10,15,10 for leaf, bark and data fusion .
hiddenLayerSize = 10; net = patternnet(hiddenLayerSize);

net.trainFcn = 'trainscg';

```

```

% Choose the Performance Function

% Division of Data for Training, Validation : K-Fold cross validation (stratified)
k = 10;
cvFolds1 = crossvalind('Kfold',10, k);cvFolds2 = crossvalind('Kfold',10, k);cvFolds3 =
crossvalind('Kfold',10, k);cvFolds4 = crossvalind('Kfold',10, k);
cvFolds=[cvFolds1;cvFolds2;cvFolds3;cvFolds4];

%% check for all categories inclusions - CAN BE EASILY SEEN VISUALLY
%E1= countcats(categorical(cvFolds1));E2= countcats(categorical(cvFolds2));E3=
countcats(categorical(cvFolds3));E4= countcats(categorical(cvFolds4));

for i = 1:k
net = configure(net,input,target);
testIdx = (cvFolds == i);
trainIdx = ~testIdx;
trInd = find(trainIdx)
tstInd = find(testIdx)
net.trainParam.epochs = 100;
net.divideParam.trainInd = trInd
net.divideParam.testInd = tstInd

% Performance Function
net.performFcn = 'mse'; % Mean squared error

% Train the network
[net, tr] = train(net,input,target);
net.trainParam.showCommandLine=true; % training information will be displayed in the
command line

% test using test instances
output = net(input);
errors = gsubtract(target, output);
performance = perform(net,target,output)

% Recalculate Training, Validation and Test Performance
trainTargets = target .* tr.trainMask{1};
testTargets = target .* tr.testMask{1};
trainPerformance = perform(net,trainTargets,output)
testPerformance = perform(net,testTargets,output)
test(k)=testPerformance;

% Plot confusion
save net

```

```
figure, plotconfusion(target,output)
end
accuracy=mean(test);
accuracy_percentage=accuracy*100;

% View the Network
view(net)
genFunction(net,'myNeuralNetworkFunction','MatrixOnly','yes');
    y = myNeuralNetworkFunction(input);
genFunction(net,'myNeuralNetworkFunction');
    y1 = myNeuralNetworkFunction(input);

gensim(net);
```


ANNEXURE II

MATLAB code for Individual Bark Spectral Analysis

```
%%% selWavelengths_allPCs' : run code for EW on the bark data (40*241)
% input to be set to selected wavelengths matrix out of leaves , and bark resctly
input= selWavelengths_allPCs';

% hidden layers
hiddenLayerSize = 15; net = patternnet(hiddenLayerSize);

net.trainFcn = 'trainscg';
% Choose the Performance Function

% Division of Data for Training, Validation
% K-Fold cross validation (stratified)
k = 10;
cvFolds1 = crossvalind('Kfold',10, k);cvFolds2 = crossvalind('Kfold',10, k);cvFolds3 =
crossvalind('Kfold',10, k);cvFolds4 = crossvalind('Kfold',10, k);
cvFolds=[cvFolds1;cvFolds2;cvFolds3;cvFolds4];

%%% check for all categories inclusions - CAN BE EASILY SEEN VISUALLY
%E1= countcats(categorical(cvFolds1));E2= countcats(categorical(cvFolds2));E3=
countcats(categorical(cvFolds3));E4= countcats(categorical(cvFolds4));

%%% crossval = Loss estimate using cross-validation
% % 1
% classf = @(XTRAIN, ytrain,XTEST)(classify(XTEST,XTRAIN,ytrain));
% cvMCR = crossval('mcr',input,target,'predfun',classf,'partition',cp)
% % 2
% f = @(xtr,ytr,xte,yte)confusionmat(yte, classify(xte,xtr,ytr),'order',order);
% cfMat = crossval(f,input,target,'partition',cp);

for i = 1:k

net = configure(net,input,target);
testIdx = (cvFolds == i);
trainIdx = ~testIdx;
trInd = find(trainIdx)
tstInd = find(testIdx)
net.trainParam.epochs = 100;
net.divideParam.trainInd = trInd
net.divideParam.testInd = tstInd

% Performance Function
net.performFcn = 'mse'; % Mean squared error
```

```

% Train the network
[net, tr] = train(net,input,target);

net.trainParam.showCommandLine=true;
% training information will be displayed in the command line

% test using test instances
output = net(input);
errors = gsubtract(target, output);
performance = perform(net,target,output)

% Recalculate Training, Validation and Test Performance
trainTargets = target .* tr.trainMask{1};
testTargets = target .* tr.testMask{1};
trainPerformance = perform(net,trainTargets,output)
testPerformance = perform(net,testTargets,output)
test(k)=testPerformance;

% Plot confusion
save net
figure, plotconfusion(target,output)
end
accuracy=mean(test);
accuracy_percentage=accuracy*100;

% View the Network
view(net)

genFunction(net,'myNeuralNetworkFunction','MatrixOnly','yes');
y = myNeuralNetworkFunction(input);

genFunction(net,'myNeuralNetworkFunction');
y1 = myNeuralNetworkFunction(input);

gensim(net);

```

ANNEXURE III

MATLAB Program for data fusion and BP-NN

```
% % l_score b_score to be taken from
input=[l_score b_score]'; %double

% hidden layers
hiddenLayerSize = 10; net = patternnet(hiddenLayerSize);

net.trainFcn = 'trainscg';
% Choose the Performance Function

% Division of Data for Training, Validation
% K-Fold cross validation (stratified)
k = 10;
cvFolds1 = crossvalind('Kfold',10, k);cvFolds2 = crossvalind('Kfold',10, k);cvFolds3 =
crossvalind('Kfold',10, k);cvFolds4 = crossvalind('Kfold',10, k);
cvFolds=[cvFolds1;cvFolds2;cvFolds3;cvFolds4];

%% check for all categories inclusions - CAN BE EASILY SEEN VISUALLY
%E1= countcats(categorical(cvFolds1));E2= countcats(categorical(cvFolds2));E3=
countcats(categorical(cvFolds3));E4= countcats(categorical(cvFolds4));

for i = 1:k
net = configure(net,input,target);
testIdx = (cvFolds == i);
trainIdx = ~testIdx;
trInd = find(trainIdx)
tstInd = find(testIdx)
net.trainParam.epochs = 100;
net.divideParam.trainInd = trInd
net.divideParam.testInd = tstInd

% Performance Function
net.performFcn = 'mse'; % Mean squared error

% Train the network
[net, tr] = train(net,input,target);

net.trainParam.showCommandLine=true; % training information will be displayed in the
command line
```

```

% test using test instances
output = net(input);
errors = gsubtract(target, output);
performance = perform(net,target,output)

% Recalculate Training, Validation and Test Performance
trainTargets = target .* tr.trainMask{1};
testTargets = target .* tr.testMask{1};
trainPerformance = perform(net,trainTargets,output)
testPerformance = perform(net,testTargets,output)
test(k)=testPerformance;

% Plot confusion
save net
figure, plotconfusion(target,output)
end
accuracy=mean(test);
accuracy_percentage=accuracy*100;

% View the Network
view(net)

genFunction(net,'myNeuralNetworkFunction','MatrixOnly','yes');
y = myNeuralNetworkFunction(input);

genFunction(net,'myNeuralNetworkFunction');
y1 = myNeuralNetworkFunction(input);

gensim(net);

```

Python Program for drawing confusion matrices, using outputs of MATLAB program

```

#Import the necessary libraries
import numpy as np
from sklearn.metrics import confusion_matrix
import seaborn as sns
import matplotlib.pyplot as plt

# Constants
H1="HEALTHY"
H2="MILD"
H3="MODERATE"

```

```

H4="SEVERE"
# True values
y_true_LB =
np.array([H1,H1,H1,H1,H1,H1,H1,H1,H1,H1,H1,H2,H2,H2,H2,H2,H2,H2,H2,H2,H2,H2,H3,H3,
H3,H3,H3,H3,H3,H3,H3,H3,H4,H4,H4,H4,H4,H4,H4,H4,H4,H4,H4,H4])
# Predicted values
y_pred_LB =
np.array([H1,H1,H1,H1,H1,H1,H1,H1,H1,H1,H1,H2,H2,H2,H2,H2,H2,H2,H2,H2,H2,H2,H2,H4,
H3,H3,H3,H3,H3,H3,H3,H3,H4,H4,H4,H4,H4,H4,H4,H4,H4,H4,H4,H2])

#compute the confusion matrix.
cm_LB = confusion_matrix(y_pred_LB,y_true_LB)

#Plot the confusion matrix.
sns.heatmap(cm_LB,
            annot=True,
            fmt='g', annot_kws={'size': 25},
            xticklabels=['Healthy','Mild','Moderate','Severe'],
            yticklabels=['Healthy','Mild','Moderate','Severe'],
            cbar=False)
plt.ylabel('Prediction',fontsize=13)
plt.xlabel('Actual',fontsize=13)
#plt.show()
plt.savefig('leaf-bark.pdf')
plt.savefig('leaf-bark1.jpg', dpi=300, transparent=True)
# True values
y_true_L =
np.array([H1,H1,H1,H1,H1,H1,H1,H1,H1,H1,H1,H2,H2,H2,H2,H2,H2,H2,H2,H2,H2,H2,H3,H3,
H3,H3,H3,H3,H3,H3,H3,H3,H4,H4,H4,H4,H4,H4,H4,H4,H4,H4,H4,H4])
# Predicted values
y_pred_L =
np.array([H1,H1,H1,H1,H1,H1,H1,H1,H1,H1,H1,H2,H2,H2,H2,H2,H2,H2,H2,H2,H3,H3,H2,H4,
H3,H3,H3,H3,H3,H3,H3,H3,H4,H4,H4,H4,H4,H4,H4,H4,H4,H4,H4,H3])

#compute the confusion matrix.
cm_L = confusion_matrix(y_pred_L,y_true_L)

#Plot the confusion matrix.
sns.heatmap(cm_L,
            annot=True,
            fmt='g', annot_kws={'size': 25},
            xticklabels=['Healthy','Mild','Moderate','Severe'],
            yticklabels=['Healthy','Mild','Moderate','Severe'],
            cbar=False)

```

```

plt.ylabel('Prediction',fontsize=13)
plt.xlabel('Actual',fontsize=13)
#plt.show()
plt.savefig('leaf.pdf')
plt.savefig('leaf.jpg', dpi=300, transparent=True)

# True values
y_true_B =
np.array([H1,H1,H1,H1,H1,H1,H1,H1,H1,H1,H1,H2,H2,H2,H2,H2,H2,H2,H2,H2,H3,H3,
H3,H3,H3,H3,H3,H3,H3,H3,H4,H4,H4,H4,H4,H4,H4,H4,H4,H4])
# Predicted values
y_pred_B =
np.array([H1,H1,H1,H1,H1,H1,H1,H1,H1,H1,H1,H2,H2,H2,H2,H2,H2,H2,H2,H1,H4,H3,H3,
H3,H3,H3,H3,H1,H1,H4,H4,H4,H4,H4,H4,H4,H4,H4,H4,H4,H2])

#compute the confusion matrix.
cm_B = confusion_matrix(y_pred_B,y_true_B)

#Plot the confusion matrix.
sns.heatmap(cm_B,
            annot=True,
            fmt='g', annot_kws={'size': 25},
            xticklabels=['Healthy','Mild','Moderate','Severe'],
            yticklabels=['Healthy','Mild','Moderate','Severe'],
            cbar=False)
plt.ylabel('Prediction',fontsize=13)
plt.xlabel('Actual',fontsize=13)
#plt.show()
plt.savefig('bark.pdf')
plt.savefig('bark.png', dpi=300)

# True values
y_true_S =
np.array([H1,H1,H1,H1,H1,H1,H1,H1,H1,H1,H1,H2,H2,H2,H2,H2,H2,H2,H2,H2,H3,H3,
H3,H3,H3,H3,H3,H3,H3,H3,H4,H4,H4,H4,H4,H4,H4,H4,H4,H4])
# Predicted values
y_pred_S =
np.array([H1,H1,H1,H1,H1,H1,H1,H1,H1,H1,H1,H2,H2,H2,H2,H2,H2,H2,H2,H2,H3,H3,
H3,H3,H3,H3,H3,H3,H3,H3,H4,H4,H4,H4,H4,H4,H4,H4,H4,H4])

#compute the confusion matrix.
cm_LB = confusion_matrix(y_pred_S,y_true_S)

#Plot the confusion matrix.

```

```

sns.heatmap(cm_LB,
            annot=True,
            fmt='g', annot_kws={'size': 25},
            xticklabels=['Healthy','Mild','Moderate','Severe'],
            yticklabels=['Healthy','Mild','Moderate','Severe'],
            cbar=False)
plt.ylabel('Prediction',fontsize=13)
plt.xlabel('Actual',fontsize=13)
plt.savefig('SVM.pdf')
plt.savefig('SVM.png', dpi=300)

```

Python Program for confirming metrics- Accuracy, Precision, Recall, F1-score

```

from sklearn import metrics
print(metrics.classification_report(y_true_LB, y_pred_LB, digits=3))
print(metrics.classification_report(y_true_L, y_pred_L, digits=3))
print(metrics.classification_report(y_true_B, y_pred_B, digits=3))
print(metrics.classification_report(y_true_S, y_pred_S, digits=3))

```

Python Program for drawing ROC curve and AUC

```

import numpy as np
from sklearn.preprocessing import label_binarize
from sklearn.metrics import roc_curve, auc
import matplotlib.pyplot as plt

# Constants
H1 = "HEALTHY"
H2 = "MILD"
H3 = "MODERATE"
H4 = "SEVERE"

# True values
y_true_LB = np.array([H1, H1, H1, H1, H1, H1, H1, H1, H1, H1, H1, H2, H2, H2, H2, H2,
H2, H2, H2, H2, H2, H2, H3, H3, H3, H3, H3, H3, H3, H3, H3, H3, H3, H3, H3, H3, H4, H4, H4, H4, H4, H4,
H4, H4, H4, H4])

# Predicted values
y_pred_LB = np.array([H1, H1, H1, H1, H1, H1, H1, H1, H1, H1, H1, H2, H2, H2, H2, H2,
H2, H2, H2, H2, H2, H2, H2, H4, H3, H3, H3, H3, H3, H3, H3, H3, H3, H3, H4, H4, H4, H4, H4, H4,
H4, H4, H4, H2])

# One-hot encode the true labels
classes = [H1, H2, H3, H4]

```

```

y_true_one_hot = label_binarize(y_true_LB, classes=classes)

# Compute ROC curve and ROC area for each class
fpr = dict()
tpr = dict()
roc_auc = dict()
n_classes = len(classes)

for i in range(n_classes):
    fpr[i], tpr[i], _ = roc_curve(y_true_one_hot[:, i], (y_pred_LB == classes[i]).astype(int))
    roc_auc[i] = auc(fpr[i], tpr[i])

# Plot ROC curves for each class
plt.figure(figsize=(10, 6))
colors = ['blue', 'green', 'red', 'brown']

for i, color in zip(range(n_classes), colors):
    plt.plot(fpr[i], tpr[i], color=color, lw=2, label=f'ROC curve (area = {roc_auc[i]:0.2f}) for
class {classes[i]}')

plt.plot([0, 1], [0, 1], 'k--', lw=2)
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('ROC Curves for Multi-Class Classification')
plt.legend(loc="lower right")
plt.savefig('ROC-LB.pdf')
plt.savefig('ROC-LB.png', dpi=300)
plt.show()

# True values
y_true_L =
np.array([H1,H1,H1,H1,H1,H1,H1,H1,H1,H1,H1,H2,H2,H2,H2,H2,H2,H2,H2,H2,H2,H3,H3,
H3,H3,H3,H3,H3,H3,H3,H3,H4,H4,H4,H4,H4,H4,H4,H4,H4,H4])
# Predicted values
y_pred_L =
np.array([H1,H1,H1,H1,H1,H1,H1,H1,H1,H1,H1,H2,H2,H2,H2,H2,H2,H2,H2,H3,H3,H2,H4,
H3,H3,H3,H3,H3,H3,H3,H3,H4,H4,H4,H4,H4,H4,H4,H4,H4,H3])

# One-hot encode the true labels
classes = [H1, H2, H3, H4]
y_true_one_hot = label_binarize(y_true_L, classes=classes)

```



```

# Compute ROC curve and ROC area for each class
fpr = dict()
tpr = dict()
roc_auc = dict()
n_classes = len(classes)

for i in range(n_classes):
    fpr[i], tpr[i], _ = roc_curve(y_true_one_hot[:, i], (y_pred_L == classes[i]).astype(int))
    roc_auc[i] = auc(fpr[i], tpr[i])

# Plot ROC curves for each class
plt.figure(figsize=(10, 6))
colors = ['blue', 'green', 'red', 'brown']

for i, color in zip(range(n_classes), colors):
    plt.plot(fpr[i], tpr[i], color=color, lw=2, label=f'ROC curve (area = {roc_auc[i]:0.2f}) for class {classes[i]}')

plt.plot([0, 1], [0, 1], 'k--', lw=2)
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('ROC Curves for Multi-Class Classification')
plt.legend(loc="lower right")
plt.savefig('ROC-L.pdf')
plt.savefig('ROC-L.png', dpi=300)
plt.show()

# True values
y_true_B =
np.array([H1,H1,H1,H1,H1,H1,H1,H1,H1,H1,H1,H2,H2,H2,H2,H2,H2,H2,H2,H2,H3,H3,
H3,H3,H3,H3,H3,H3,H3,H4,H4,H4,H4,H4,H4,H4,H4,H4,H4])
# Predicted values
y_pred_B =
np.array([H1,H1,H1,H1,H1,H1,H1,H1,H1,H1,H1,H2,H2,H2,H2,H2,H2,H2,H2,H1,H4,H3,H3,
H3,H3,H3,H3,H1,H1,H4,H4,H4,H4,H4,H4,H4,H4,H4,H4,H4,H2])

# One-hot encode the true labels
classes = [H1, H2, H3, H4]
y_true_one_hot = label_binarize(y_true_B, classes=classes)

# Compute ROC curve and ROC area for each class
fpr = dict()

```

```

tpr = dict()
roc_auc = dict()
n_classes = len(classes)

for i in range(n_classes):
    fpr[i], tpr[i], _ = roc_curve(y_true_one_hot[:, i], (y_pred_B == classes[i]).astype(int))
    roc_auc[i] = auc(fpr[i], tpr[i])

# Plot ROC curves for each class
plt.figure(figsize=(10, 6))
colors = ['blue', 'green', 'red', 'brown']

for i, color in zip(range(n_classes), colors):
    plt.plot(fpr[i], tpr[i], color=color, lw=2, label=f'ROC curve (area = {roc_auc[i]:0.2f}) for
class {classes[i]}')

plt.plot([0, 1], [0, 1], 'k--', lw=2)
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('ROC Curves for Multi-Class Classification')
plt.legend(loc="lower right")
plt.savefig('ROC-B.pdf')
plt.savefig('ROC-B.png', dpi=300)
plt.show()

# True values
y_true_S =
np.array([H1,H1,H1,H1,H1,H1,H1,H1,H1,H1,H1,H2,H2,H2,H2,H2,H2,H2,H2,H2,H3,H3,
H3,H3,H3,H3,H3,H3,H3,H4,H4,H4,H4,H4,H4,H4,H4,H4,H4])

# Predicted values
y_pred_S =
np.array([H1,H1,H1,H1,H1,H1,H1,H1,H1,H1,H1,H2,H2,H2,H2,H2,H2,H2,H2,H2,H3,H3,
H3,H3,H3,H3,H3,H3,H3,H4,H4,H4,H4,H4,H4,H4,H4,H4,H4])

# One-hot encode the true labels
classes = [H1, H2, H3, H4]
y_true_one_hot = label_binarize(y_true_S, classes=classes)

# Compute ROC curve and ROC area for each class
fpr = dict()
tpr = dict()
roc_auc = dict()

```

```

n_classes = len(classes)

for i in range(n_classes):
    fpr[i], tpr[i], _ = roc_curve(y_true_one_hot[:, i], (y_pred_S == classes[i]).astype(int))
    roc_auc[i] = auc(fpr[i], tpr[i])

# Plot ROC curves for each class
plt.figure(figsize=(10, 6))
colors = ['blue', 'green', 'red', 'brown']

for i, color in zip(range(n_classes), colors):
    plt.plot(fpr[i], tpr[i], color=color, lw=2, label=f'ROC curve (area = {roc_auc[i]:0.2f}) for
class {classes[i]}')

plt.plot([0, 1], [0, 1], 'k--', lw=2)
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('ROC Curves for Multi-Class Classification')
plt.legend(loc="lower right")
plt.savefig('ROC-S.pdf')
plt.savefig('ROC-S.png', dpi=300)
plt.show()

```

ANNEXURE IV

MATLAB Program for data fusion and SVM

```
load('SVM_CVdata_labels');

%% Division of Data for Training, Validation : K-Fold cross validation (stratified)
k = 10;
cvFolds1 = crossvalind('Kfold',10, k);cvFolds2 = crossvalind('Kfold',10, k);cvFolds3 =
crossvalind('Kfold',10, k);cvFolds4 = crossvalind('Kfold',10, k);
cvFolds22=[cvFolds2;cvFolds3;cvFolds1;cvFolds4];

k = 10;
for i = 1:k
    testIdx = (cvFolds == i);
trainIdx = ~testIdx; trInd = find(trainIdx); tstInd = find(testIdx);
end
% SPECIFYING DATA AND LABELS-----
train_Data=input_SVM(trInd,:); train_labels=classes_labels(trInd,:);
test_Data=input_SVM(tstInd,:); test_labels=classes_labels(tstInd,:);

SVM_Model = fitcecoc(train_Data,train_labels,'Coding','onevsone','Learners','linear');
%for 60s epoch for 32 channel
labels=predict(SVM_Model,input_SVM);

trueLabels=classes_labels;
Accuracy=sum(strcmp(trueLabels,labels))/numel(trueLabels)*100;

%% REVERSED
% [m,order] = confusionmat(trueLabels,predictedLabels);
% figure
% cm = confusionchart(m,order);

%% Plot confusion matrix- wanted
[m1,order1] = confusionmat(trueLabels,predictedLabels);
figure
cm2 = confusionchart(m1,order1,'XLabel','True Class', 'YLabel','Predicted Class');
sortClasses(cm2,['Healthy', 'Mild', 'Moderate', 'Severe']);

figure
plotconfusion(trueLabels,predictedLabels)
```