DETECTION AND CLASSIFICATION OF WATER, NUTRIENT AND DISEASE STRESS IN WHEAT AND MAIZE USING PROXIMAL AND REMOTE SENSING

Thesis Submitted for the Award of the Degree of

DOCTOR OF PHILOSOPHY

in

Geospatial Information Science and Technology

By

Harpinder Singh

Registration Number:41800522

Supervised By

Dr. Ajay Roy (22652)

Department of Electronics and Electrical Engineering (Professor)

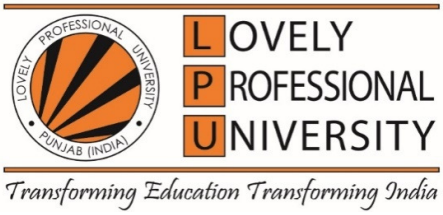
Lovely Professional University

Co-Supervised by

Dr. Brijendra Pateriya, Dr. R.K. Setia

Punjab Remote Sensing Centre (Director, Scientist)

Punjab Remote Sensing Centre



LOVELY PROFESSIONAL UNIVERSITY, PUNJAB

2022

DECLARATION

I hereby declared that the presented work in the thesis entitled "Detection and Classification of Water, Nutrient and Disease Stress in Wheat and Maize using Proximal and Remote Sensing" in fulfillment of the degree of Doctor of Philosophy (Ph.D.) is an outcome of research work carried out by me under the supervision of Dr. Ajay Roy, working as a Professor, in the School of Electronics and Electrical Engineering of Lovely Professional University, Punjab, India. The research work was co-supervised by Dr. Brijendra Pateriya (Director) and Dr. R.K. Setia (Scientist) from Punjab Remote Sensing Centre, Ludhiana, Punjab. In keeping with the general practice of reporting scientific observations, due acknowledgements have been made whenever work described here has been based on findings of other investigators. This work has not been submitted in part or full to any other University or Institute for the award of any degree.

(Signature of Scholar)

Name of the scholar: Harpinder Singh

Registration No.: 41800522

Department/School: School of Electronics and Electrical Engineering

Lovely Professional University,

Punjab, India

CERTIFICATE

This is to certify that the work reported in the Ph.D. thesis entitled "Detection and Classification of Water, Nutrient and Disease Stress in Wheat and Maize using Proximal and Remote Sensing" submitted in fulfilment of the requirement for the reward of degree of Doctor of Philosophy (Ph.D.) in the Geospatial Information Science & Technology, is a research work carried out by Harpinder Singh, 41800522, is bonafide record of his original work carried out under my supervision and that no part of thesis has been submitted for any other degree, diploma or equivalent course.

(Signature of Supervisor)
Dr Ajay Roy (Professor)
School of Electronics and Electrical Engineering
Lovely Professional University

(Signature of Co-Supervisor)
1) Dr. Brijendra Pateriya
(Director)
2) Dr. R.K Setia (Scientist)
Punjab Remote Sensing Centre

ABSTRACT

Losses in agricultural fields are becoming a serious challenge to food security. Both biotic and abiotic stress conditions cause field losses. Since the effect of stress on crop growth is usually only noticed after it becomes visible, monitoring agricultural crop conditions throughout the growing season aids in accurate and timely estimation of yield losses. Regular and continuous field-scale crop condition monitoring (proximal sensing) is time-consuming, labour-intensive, and location-specific. For crop status monitoring, satellite remote sensing could be a viable alternative to field sampling. This study will attempt to identify the biotic and abiotic stress signatures from the vegetation in the Punjab region and further relate it to the multispectral satellite imagery, which is freely available. MODIS Enhanced Vegetation Index (EVI) was analyzed in Google Earth Engine to monitor the crop conditions in Punjab. Crop stress related to biotic and abiotic variables was identified in this study. The separation of abiotic and biotic stress is required for the site-specific management of crops. In order to discriminate the abiotic factors (like nutrients and water), a field experiment with a maize-wheat cropping system under various water and nutrient levels was carried out at Punjab Agricultural University, Ludhiana. The reflectance spectra of maize and wheat were collected at regular intervals using Spectroradiometer (wavelength interval between 400 - 2500 nm) on clear and cloudless days. The leaf samples were collected and analyzed for water content, chlorophyll, nitrogen (N), phosphorus (P) and potassium (K) using standard methods. The measured plant parameters were related to spectra hyperspectral data using machine learning and explainable artificial intelligence techniques to identify the optimum wavelengths for nutrient and water stress in maize and wheat. The selection of optimum wavelengths from field experiments under Indian conditions may not apply to other climatic and soil conditions. So similar studies were conducted on secondary datasets of Israel and the U.S.A. The wavelengths identified from hyperspectral data in abiotic stresses were compared with freely available multispectral satellite imagery (Sentinel-2 MSI, Landsat-8 OLI and Landsat-7 ETM+) for remote monitoring of stresses in crops. These studies identified the optimum spectral wavelengths for abiotic stresses in crops, but the spectral behaviour of biotic and abiotic stresses is quite different. In order to study the biotic stress in crops,

yellow rust of wheat (a major disease) was classified in the parts of Punjab from Sentinel-2 satellite imagery using deep learning artificial neural network.

ACKNOWLEDGEMENT

I would like to first say a huge thank you to my supervisor, Dr Ajay Roy (Lovely Professional University, Jalandhar) and Co-Supervisors Dr R.K. Setia and Dr Brijendra Pateriya (Punjab Remote Sensing Centre, Ludhiana). This research work would not have been possible without their support, supervision, and inspiration. Under their guidance, I successfully overcame many difficulties and learned a lot. Over the last four years, I could see the journey as the most memorable one in terms of learning, knowledge, and taking up a challenge with confidence.

I am grateful to the School of Research Degree Programme (RDP) members for being in my entire research progress review panels and providing valuable suggestions during the period. I am thankful to all anonymous reviewers of my research papers submitted to various International Journals and International conferences.

I am indebted to the Honorable Chancellor, Worthy Pro-Chancellor, the Vice-Chancellor, and the successive Deans, LPU, for facilitating the administrative issues involved and encouraging me.

I was indeed able to improve upon the work contained herein. My deep appreciation goes out to my PRSC colleagues. Their encouragement helped me to come to this level.

I would also like to thank my father, mother, wife, and daughter for always believing in me and encouraging me to follow my dreams.

At the end of my thesis, I would like to thank everyone who made this thesis possible and an unforgettable experience for me.

Date:22/May/22

Harpinder Singh

Table of Contents

Page No

Declaration **i**
Certificate **ii**
Abstract **iii**
Acknowledgement **iv**
Table of Contents **vi**
List of Tables **ix**
List of Figures **xi**
List of Annexures **xii**

Chapter 1 Introduction **1-19**
- 1.1 Background and Introduction
- 1.2 Literature Review
  - 1.2.1 Introduction to Biotic and Abiotic Stress and role of proximal and remote sensing
  - 1.2.2 Thermal, Fluorescence and reflectance proximal/remote sensing techniques
  - 1.2.3 Role of Artificial intelligence (AI) algorithms in crop stress studies
  - 1.2.4 Simulation of multispectral satellite imagery
- 1.3 Objectives
- 1.4 Research Methodology
- 1.5 Thesis Structure

Chapter 2: Spatio-Temporal analysis of crop conditions using MODIS EVI in Google Earth Engine **20-30**
- 2.1 Introduction
- 2.2 Objectives
- 2.3 Study Area and Datasets
  - 2.3.1 Study area
  - 2.3.2 Datasets
- 2.4 Methods
  - 2.4.1 Preparing the MODIS EVI Dataset
  - 2.4.2 Deviation
- 2.5 Results and Discussions

Chapter 3: Detection of nutrient and water stresses in maize and wheat using proximal sensing. **31-42**
- 3.1 Introduction

3.2 Materials and Methods
3.2.1 Field Experiment and collection of hyperspectral data
3.2.2 Hyperspectral Data Processing
3.2.3 Estimation of the status of various parameters of maize and wheat from hyperspectral data using machine learning techniques.
3.2.4 Identification of important features (wavelengths) from hyperspectral data
3.3 Results and Discussion
3.3.1 Selection of optimum machine learning regression models to estimate the crop parameters.
3.3.2 Sensitive bands for estimation of water content, chlorophyll a, chlorophyll b, nitrogen, phosphorus and potassium

Chapter 4: Estimation of nitrogen content in wheat from proximal hyperspectral data using machine learning and explainable artificial intelligence (XAI) approach. 43-54

4.1 Introduction
4.2 Materials and Methods
4.2.1 Study area and Data Acquisition
4.2.2 Hyperspectral Data Processing
4.2.3 Estimation of N status in plants from hyperspectral data using machine learning analysis
4.2.4 Identification of important features (wavelengths) from hyperspectral data
4.3 Results and Discussions
4.3.1 Comparison of spectral pre-processing techniques and machine learning techniques
4.3.2 Machine Learning Model explanation
4.3.2.1 Global Interpretability
4.3.2.2 Local Interpretability

Chapter 5: Estimation of chlorophyll, macronutrients and water content in maize from hyperspectral data using machine learning and explainable artificial intelligence techniques 55-68

5.1 Introduction
5.2 Materials and Methods
5.2.1 Study area and Data Acquisition
5.2.2 Hyperspectral Data Processing
5.2.3 Estimation of the status of various parameters of plants from hyperspectral data using machine learning techniques.

5.2.4 Identification of important features (wavelengths) from hyperspectral data
5.3 Results and Discussions
5.3.1 Comparison of machine learning techniques across three nitrogen regimes
5.3.2 Machine learning model explanations
5.4 Conclusions

Chapter 6: Classification of yellow rust of wheat from Sentinel-2 satellite imagery using deep learning artificial neural network 69-82
6.1 Introduction
6.2 Materials and Methods
6.2.1 Study area
6.2.2 Satellite Data
6.2.3 Pre-processing of Satellite Data
6.2.4 Generation of spectral indices to extract crop data
6.2.5 Preparation of the data to be fed to the ANN
6.2.6 ANN model design and development
6.3 Results and Discussion
6.3.1 Response of spectral indices to yellow rust
6.3.2 Classification of healthy and diseased plants
6.4 Conclusion

Chapter 7: Simulation of multispectral data using hyperspectral data for crop stress studies 83-91
7.1 Objective
7.2 Materials and Methods
7.2.1 Hyperspectral Data
7.2.2 AI Techniques
7.2.3 Suitable bands in multispectral satellite imagery
7.3 Results and Discussion
7.3.1 Suitable bands in multispectral satellite imagery

Chapter 8: Summary and Future Work 92-94

References 95-106

Annexure

List of Tables

Table 1.1	Methodology/ Tools/ Instruments to be used
Table 2.1	Deviation Classes
Table 2.2	District wise area statistics
Table 2.3	District wise wheat yield (Per Hectare in Kilogram) statistics
Table 3.1	Model Parameters
Table 3.2	Coefficient of determination (R^2) and mean square error (MSE) for the Maize models
Table 3.3	Coefficient of determination (R^2) and mean square error (MSE) for the Wheat models
Table 3.4	Important wavelengths associated with the parameters of Maize
Table 3.5	Important wavelengths associated with the parameters of Wheat
Table 4.1	Results of the statistical parameters (Coefficient of determination, R^2 and Mean Square Error, MSE) for the machine learning models
Table 4.2	Hyper parameters selected for the machine learning models
Table 4.3	Important Wavelengths (associated with N) at various growth stages of wheat crop (Local Interpretability of Random Forest Model).
Table 5.1	Results of the statistical parameters for water content, chlorophyll and macronutrients estimated using the six machine learning regression models
Table 5.2	Optimum three important wavelengths estimated for different crop parameters using explainable artificial intelligence for the six machine learning regression models
Table 6.1	Villages in the study area
Table 6.2	Characteristics of the Sentinel-2 satellite data
Table 6.3	Spectral Indices used in the study for discrimination of yellow rust of wheat
Table 6.4	Summary statistics of the ANN model in the areas of Indian Punjab for discrimination of yellow rust of wheat
Table 7.1	Spectral bands and resolutions of Sentinel-2 MSI sensor
Table 7.2	Spectral bands and resolutions of Landsat-8 OLI sensor

Table 7.3 Spectral bands and resolutions of Landsat-7 ETM+ sensor

Table 7.4 Bands of freely available multispectral satellite imagery associated with the maize(Punjab) parameters

Table 7.5 Bands of freely available multispectral satellite imagery associated with the wheat (Punjab) parameters

Table 7.6 Bands of freely available multispectral satellite imagery associated with the wheat (Israel) parameters

Table 7.7 Bands of freely available multispectral satellite imagery associated with the maize(US) parameters

List of Figures

Figure 1.1	Multispectral v/s Hyperspectral data
Figure 1.2	Typical spectral reflectance curves of selected surfaces
Figure 2.1	Research Methodology
Figure 2.2	Generated Datasets (a) Reference Mean of 2014-2019, (b) Current Maximum of 2019-2020
Figure 2.3	Deviation Raster (Red-Green Colour ramp shows areas of negative to positive deviations)
Figure 2.4	Stressed Areas (Red Colour) in Punjab State
Figure 2.5	Stressed Areas (Red Colour) around Dera Baba Nanak in Gurdaspur
Figure 2.6	Stressed Areas (Red Colour) around Bassi Pathanan in Fatehgarh Sahib
Figure 3.1	An overview of the approach
Figure 3.2	Output of XAI
Figure 4.1	Methodology
Figure 4.2	SHAP variable importance plot (Random Forest)
Figure 4.3	SHAP variable importance plot (Gradient Boosting Regression)
Figure 5.1	An overview of the approach
Figure 5.2	Average spectral reflectance of maize leaves under greenhouse, low and high nitrogen datasets
Figure 6.1	Study Area
Figure 6.2	Process Flow of the Research work
Figure 6.3(a)	ANN Layout
Figure 6.3(a)	Model Structure
Figure 6.4	Spectral indices in healthy and diseased wheat plants in (A) Jalandhar/Kapurthala, and (B) Rupnagar area.
Figure 7.1	Comparison of the bands of Sentinel-2, Landsat 7 and 8.

List of Annexures

Annexure 1 Crop Experiment Photographs

Chapter 1

Introduction

1.1: Background and Introduction:

The science and art of obtaining information about an object, area, or phenomenon without coming into physical touch with the object, area, or phenomenon under enquiry is referred to as remote sensing [1]. In remote sensing, information is transferred without direct contact through electromagnetic radiations (EMR). The EMR, which is reflected or emitted from an object, is the usual remote sensing data source. Only a part of the electromagnetic radiation in the wavelength range 0.4–0.7μm is sensed by the human eye. Other reflecting and radiating (e.g. thermal) energy band-length ranges that reach or are emitted by Earth's surface, and even parts of Earth's atmosphere reflect, such as the EM reflective properties of clouds, can be detected using remote sensing technologies.

Remote sensing is the measuring of the object's attributes on the Earth's surface using data collected from electromagnetic energy sensors operated from proximal and spaceborne platforms. Remote sensing sensors operate from proximal, air and spaceborne platforms. These sensors collect data on how various earth surface features emit and reflect electromagnetic radiation (EMR), which is then analyzed to provide information about the resources being studied.

Data collection and processing are the two primary procedures involved in electromagnetic remote sensing of Earth's resources.

The various elements of the data collection process are:

1. Energy Sources (e.g. Sun, Radar etc.)
2. Energy transmission through the atmosphere.
3. Interaction of energy with earth's surface features.
4. Retransmission of the energy through the atmosphere.
5. Proximal/Spaceborne/ Airborne Sensor acquires this energy.
6. Generation of sensor data in the form of tables or images.

The various elements of the data processing module are:

1. The generated data, if in pictorial form, is examined using various interpretation and viewing devices and if in digital form, it is analyzed and interpreted using various specialized software. Reference data (rainfall, soil type, groundwater quality, administrative boundary, crop acreage statistics etc.) about the area under study are used if available to assist in the analysis. The analyst extracts information about the type, extent, condition and location of the resources over which the sensor data was collected.
2. This data is compiled in DBMS/file-based systems as spatial layers, which can be used in a geographic information system (GIS) or printed as hardcopy maps.
3. Users are provided with the information which they use in their decision-making process.

Remote Sensing Technology has proven capability for a quick and efficient survey and monitoring of natural resources like land, water, crops, soil, forests etc. Such information is vital for efficient and optimum utilization of these resources to meet an ever-increasing population's food, fuel, and fibre requirements.

Natural radiations that are either released or reflected from the Earth are detected by passive sensors. Active sensors illuminate the Earth's surface with electromagnetic radiation of a given wavelength or band of wavelengths. The primary parameters of a sensing system that can be utilized as indicators of data quality and have an impact on optimal data utilization for a certain end-use are listed below. [2]:

1. The sensor's capacity to distinguish the tiniest thing on the ground is spatial resolution.
2. The spectral bandwidth with which the data is obtained is spectral resolution.
3. Radiometric resolution refers to a sensor's capacity to distinguish between two objects based on their reflective/emittance differences; it is evaluated in the smallest reflective/emittance that may be detected.
4. The ability to examine the same target at regular intervals under the same conditions is termed temporal resolution.

Satellites, planes, and unmanned aerial vehicles are used in remote sensing to assist map the Earth's surface (UAV). On the other hand, Proximal sensing devices acquire detailed information near the object's surface being studied. These sensing

approaches have a considerable advantage over prior data sources since they allow for low-cost, repeatable wide-area coverage and non-destructive calculations. These sensors can collect a lot of information, which can help farmers and agricultural policymakers make better decisions. Remote and proximal sensing systems can provide a wealth of data for agricultural applications, with the primary goal of mapping, monitoring, and modelling agricultural resources and their environmental implications.

Losses in agricultural fields are becoming a serious challenge to food security. Both biotic and abiotic stress conditions cause field losses. Living species, such as bacteria, viruses, fungi, parasites, beneficial and harmful insects, weeds, and cultivated or native plants, can cause biotic stress in crops. On the other hand, abiotic stress is the negative influence of non-living forces on living organisms in a particular habitat. Abiotic stresses such as water (drought, flooding), fertility (nutrient deficiency), temperature (high/low), salinity, radiation (UV, ionizing radiation), and chemicals have a substantial impact on agricultural performance.

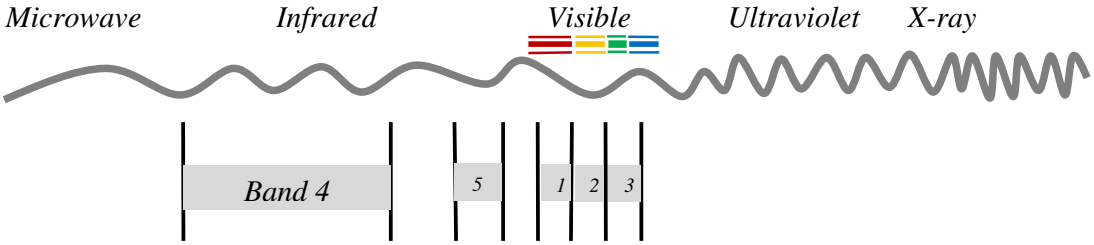
Since the effect of stress on crop growth is usually only noticed after it becomes visible, monitoring agricultural crop conditions throughout the growing season aids in accurate and timely estimation of yield losses. Stress-related losses are critical in nations where agriculture is the primary source of income.

Continuous crop condition evaluation on a field scale is time-consuming, labour-intensive, and location-specific. For crop status monitoring, satellite remote sensing could be a viable alternative to field sampling. It can also provide continuous coverage of a vast area. Hyperspectral proximal field spectroradiometer data has a higher spectral resolution than satellite-based multi and hyperspectral data, allowing for more precise detection of spectral changes related to crop condition changes.

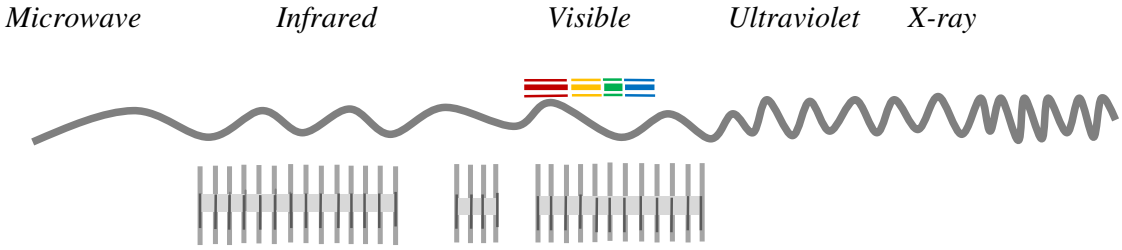
The number of bands and the wavelength width of each band are referred to as the data's spectral resolution. A band is a section of the electromagnetic spectrum that is very narrow. Higher spectral resolution data can distinguish shorter wavelength widths. Multispectral data (many popular satellite imagery datasets) may detect many broad wavelength bands, such as visible green and near-infrared. In comparison to multispectral imagery, hyperspectral data detects energy in narrower and more

numerous bands. Hyperspectral data has narrow bands that are more sensitive to fluctuations in energy wavelengths and hence have a greater ability to identify crop stress than multispectral data.

The difference between a multispectral and hyperspectral dataset can be seen in Figure 1.1.



Multispectral Example: 5 wide bands (Image not drawn to scale)



Hyperspectral Example: Imagine hundreds of narrow bands

Fig 1.1 Multispectral v/s Hyperspectral data

Currently, satellite and airborne hyperspectral data are not readily available compared to multispectral satellite data. Limited multispectral data is also freely available for research studies by a few space agencies, like the Landsat and Sentinel satellite data from NASA (National Aeronautics and Space Administration) and ESA (European Space Agency), respectively. This study will attempt to identify the biotic and abiotic stress signatures from the vegetation in the Punjab region and further relate it to the multispectral satellite imagery, which is freely available. This work will help the researchers identify biotic and abiotic plant stress over large areas with less cost, time, and effort.

1.2 Literature Review

The literature review has been categorized into four parts. The first part introduces the biotic and abiotic stress in plants and how proximal and remote sensing can be helpful in identifying it. The second part provides various studies which have used thermal, fluorescence and reflectance proximal/remote sensing techniques to identify stress in plants. The maximum studies reviewed are in this section. The third part describes the Artificial Intelligence (AI) algorithms and their use in crop stress identification. Finally, the last section reviews studies in which multispectral satellite imagery has been simulated using spectroradiometer data or various other models.

1.2.1 Introduction to Biotic and Abiotic Stress and the role of proximal and remote sensing

Agricultural field losses are becoming a significant threat to food security. Biotic and abiotic stress factors are major causes of crop losses. While the abiotic stress may be due to adverse climatic conditions, salinity and insufficient nutrients or water, the biotic stress is caused by plant diseases and damage due to insects. Various physiological and anatomical changes occur in the plant due to these stress factors. The colour of the leaf changes, leaves may droop or curl, and the plant material may be ingested or detached also. Visual surveys and crop inspections are the most popular methods to detect stress. Although these methods are very accurate, they incur time, costs and human resources. Sensing of the reflected and emitted radiation by the stressed crops can be done by remote and proximal sensing techniques. These approaches are capable of sensing that radiation, allowing for quantitative assessment of plant stress induced by biotic and abiotic causes [3]. These stress factors are accountable for the significant decline in crop yield. Yield decline is critical for countries that are primarily dependent on agriculture.

The effect of stress on the crop is usually identified only when it is noticeable and visible. Hence, continuous monitoring of plant/crop conditions using proximal and remote sensing technologies can assist in the early diagnosis of the problem. Then timely measures can be applied to curb it.

Proximal sensing techniques involve time and human resources to identify stress in crops. The process is also very intricate and location-specific. On the other

hand, remote sensing provides continuous coverage over a larger area[4]

. Proximal sensing systems provide better quality (spatial, spectral and radiometric resolution) data than remote sensing. Proximal sensor data have the upper hand primarily due to less interference of gases, air particles and water vapour in the atmosphere[5], [6]. Examples of various instruments used in the proximal techniques are the field spectroradiometer, camera, thermometer, SPAD etc.

In contrast, the remote sensing instruments are various sensors (optical and non-optical) onboard a satellite, aeroplane or unmanned aerial vehicle (UAV). These sensors sense the electromagnetic waves in the visible, infrared, and microwave regions. Crop cover, crop health, soil moisture, nutrient stress, and crop yield have all been successfully monitored using visible red, green, and blue bands, as well as the red-edge and near-infrared (NIR) regions of the electromagnetic spectrum. According to the literature review, the spectral bands used to identify crop stress are green, red, red edge and NIR.

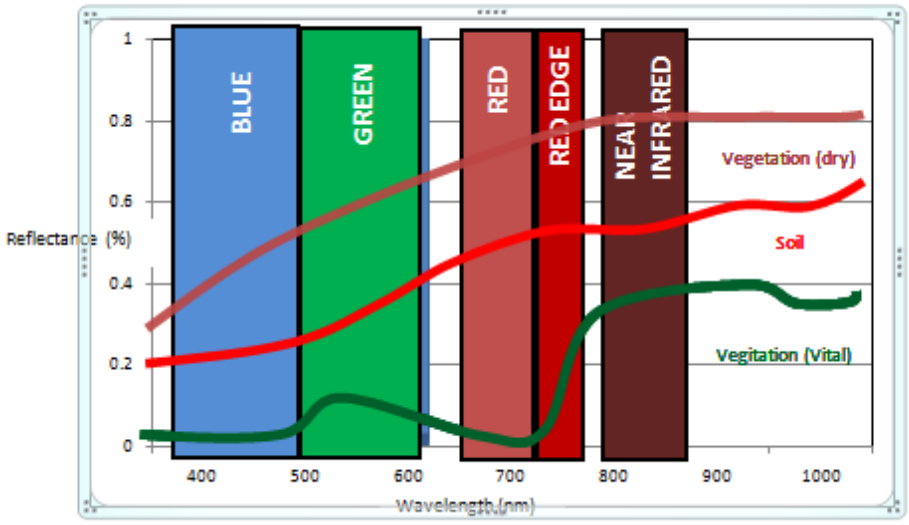


Fig 1.2 Typical spectral reflectance curves of selected surfaces

Figure 1.2 shows the typical spectral reflectance curves of vegetation and soil across various spectral bands. The difference between the dry and healthy vegetation reflectance curves can be seen in the figure above.

Proximal sensors usually output hyperspectral data, while remote sensors output both multi and hyperspectral datasets. A multispectral sensor senses the

electromagnetic energy in a few broad bands; on the other hand, the hyperspectral sensor data comprises numerous narrow bands. Narrow bands of the hyperspectral dataset are extremely sensitive and insightful to even small changes in the energy wavelengths. As a result, hyperspectral has a vast possibility for analyzing crop stress compared to the multispectral datasets. [7] and [8] reviewed the use of hyperspectral remote sensing for agriculture and vegetation studies. They have explained the important modelling and classification algorithms used to estimate and predict crop type using biophysical factors. [9] reviewed a range of hyperspectral crop data analysis techniques. The hyperspectral dataset is huge and intricate; therefore, its analysis techniques are also different from multispectral data.

1.2.2 Thermal, Fluorescence and reflectance proximal/remote sensing techniques

Crop stress is detected using the following proximal and remote sensing techniques: reflectance, thermal and fluorescence. This study will focus more on reflectance techniques.

Thermal sensors measure the temperature of the objects under study. These sensors are generally used to study water-related stress for crop stress research. The probability of water stress is more if the temperature of the crops is higher. According to a study by [10], the crop water stress index (CWSI) is extremely useful to compute and check water-related stress in the corn crop. Proximal sensors have been employed to sense the vapour pressure, canopy and air temperature. These parameters were used to create the index. CSWI has also been successfully used by [11] to schedule the irrigation of soybean crops. They measured the air, plant canopy and vapour pressure using various proximal sensors. They initiated the irrigation of the crop when a particular CSWI threshold value was reached. They found that if the value of the CSWI threshold was higher, the grain yield was less. [12]have used a combination of thermal and optical sensing datasets to monitor the water shortage in the horticultural crop. According to the research, a multi-sensor approach has a great potential to sense water stress and further use it for scheduling irrigation. Thermal data from remotely sensed satellite imagery also identifies water stress. [13] have computed various vegetation and thermal indices from MODIS satellite data of crops in West Africa. They combined the generated indices with crop yield models and found that the

combined data gives better results for yield assessment than using the indices alone.

The second sensing technique is fluorescence. During photosynthesis, plant leaves absorb solar radiation with the help of chlorophyll. Fluorescence is the process of re-emission of absorbed solar radiations at longer wavelengths. It is used to detect nutrient stress by monitoring the photosynthetic efficiency of a crop. [14]detected plant stress in early stages and developed a stress catalogue by combining thermal, fluorescence and video imagery. The research shows that fluorescence helps detect crops' nitrogen, pollutants, heavy metal and water-related stress. According to them, using multiple sensors in a study is more beneficial. [15] collected fluorescence data of crops using a narrow band multispectral airborne camera. The research on the olive and peach plantations established the possibility of detecting water stress using remotely sensed fluorescence data.

The third sensing technique is reflectance. [16] reviewed various studies in which the crop reflectance data has been used to identify nutrient and water-related stress. Major bands helpful to sense crop stress are green, red, red-edge and near-infrared. [17] has confirmed the possibility of using indices computed from hyperspectral imagery (airborne) to map water-related stress in maize crops. This research has sensed one hundred twenty-six spectral bands in the spectrum's near-infrared and visible region (400 – 992 nm). [18] has used plant canopy reflectance data (350 - 2500nm) to estimate water stress. Parameters that were measured or computed in the research are 1) hyperspectral reflectance, 2) soil water potential, 3) leaf area index, 4) soil moisture, 5) chlorophyll content, 6) canopy water content, and 7) various environmental factors. The relationship between the water content and reflectance of the crops was studied, and it was found that reflectance information could be beneficial in detecting crop stress. [19] has estimated the nitrogen values of wheat by simulating the reflectance values. For this purpose, the indices computed from the hyperspectral field dataset were compared with the partial least square regression (PLSR) model results. They found that for the estimation of N, PLRS model results are better than the spectral indices. [20] also experimented using PLS and spectroscopy data to calculate wheat's phosphorus(P) and potassium (K). The results concluded that PLS and narrow-band indices give better results than traditional broadband indices. Leaf area index is another parameter that effectively calculates

stress in crops. It is the projected plant leaf area over a unit of land. It is significant as it helps to monitor crop/vegetation properties like photosynthesis, evaporation and transpiration. [21] has studied various methods to estimate the LAI of plants using data from a range of remote sensors.

Many biotic stress factors affect the crops, but only yellow rust is reviewed in this study. Wheat, a vital cereal crop of Punjab state, is vulnerable to the attack of several diseases. Yellow rust is significant biotic stress developed during cool climate, rain, dew, fog, and favourable wind conditions [22]. A variety of techniques, sensors, datasets, and algorithms have been used for the timely identification of this disease. With the advancements in earth observation, new-age satellite data have better spatial, spectral, and radiometric resolutions than their predecessors. There are studies in which multi and hyperspectral remote and proximal sensing datasets have been used (in addition to meteorological data) to identify the yellow rust of wheat, but proximal hyperspectral datasets are more popular.

The field spectroradiometer was used to measure the leaf spectra by [23], [24], and the continuous wavelet analysis was applied on the hyperspectral data to generate the rust signatures. [25] also used the hyperspectral data collected from the field and found that normalized photochemical reflectance index (NPRI) can be used to identify rust. A spectrograph mounted on a spray broom level was used by [26] to capture in-field spectral images (wavelength between 463 and 895 nm. Neural networks were used to develop disease detection algorithms, which classified diseased and non-diseased crops. [27] have also utilized the hyperspectral in-situ data collected in the spectral range of 350 to 2500nm. To estimate the severity of yellow rust in the wheat crop, they created models utilising partial least squares (PLS) and multiple linear regression (MLR). The coefficients of determination (R2) for both models were 0.96 and 0.89, respectively. [28] evaluated ten common narrow-band spectral indices for identifying rusts from individual wheat leaves. These indices were based on the in-situ spectrometer measurements in the electromagnetic spectrum's visible and near-infrared regions. The yellow rust infected crop produced a strong response to all the indices. Using spectroradiometer data, [29] created two spectral disease indices to identify wheat leaf rust. These indices were calculated using reflectance at wavelengths of 605, 695, and 455 nm. The R2 between the estimated and observed

values was as high as 0.94 in both indices. These studies show that the accuracy (classification of yellow rust) from proximal sensing methods was higher than remote sensing methods. However, it requires a lot of resources and time to collect the spectra of yellow rust. Although hyperspectral data is popular and provides accurate results, it requires a lot of resources (time, computing and equipment) to collect, process, and analyze such datasets. There are very few studies in which multispectral remote sensing data have been used for yellow rust identification.

[30] developed the yellow rust forecasting system for the Gurdaspur and Nawanshehar districts of Punjab using weather and land use information extracted from IRS-P6 AWiFS satellite data (spatial resolution = 56 m). In Hebei Province, China, the Red Edge Disease Stress Index (REDSI) [31], was utilized to identify yellow rust infection at various severity levels. Sentinel-2 multispectral bands were simulated using the canopy hyperspectral data. Canopy hyperspectral data were used to simulate the corresponding multispectral bands of Sentinel-2. [32] has also proposed an approach for monitoring yellow rust based on Sentinel-2 multispectral images, two-stage vegetation indices, and meteorological data. Similarly, [33] also asserted that deep learning has a great potential for increased accuracy to identify crop diseases from remote and proximal sensor data. [34] has applied the convolutional deep learning algorithm 'U-Net' on the multispectral images acquired from an unmanned aerial vehicle (UAV) to monitor wheat rust. [35] developed a similar approach to identify the rust, but they have acquired hyperspectral images from a UAV. A new deep convolutional neural network (DCNN)-based technique for automated crop disease detection has been suggested. The overall accuracy of their model was 0.85. [36] conducted field experiments and utilized MODIS satellite data and deep learning algorithms (like ANN, CNN, and recurrent neural networks) to monitor the wheat fungus. Automatically learned features were used for the model development. Most previous studies have used deep learning CNN algorithms for yellow rust identification.

1.2.3 Role of Artificial intelligence (AI) algorithms in crop stress studies:

Artificial intelligence (AI) algorithms have transformed the research on crop stress identification. [37] highlighted the advantages of machine learning over the

conventional statistical techniques for detecting biotic stress in crops. Machine learning algorithms require very few statistical assumptions about the data and can be used to develop linear and non-linear models. [38] emphasized that machine learning is very promising in plant stress analytics. It can be used for Identification (K-Nearest Neighbour (K–NN), SVM, SOM, Bayes Classifier (BC), ANN, K–means, Linear and Quadratic Discriminant Analysis (LDA/QDA)), classification (Random Forest (RF), K-means, SVM, BC, LDA/ QDA, K–NN, SOM and RF), quantification (SVM) and prediction (SVM and ANN) of many parameters. Deep learning algorithms like artificial neural networks (ANN), convolutional neural networks (CNN), and recurrent neural networks (RNN) are gaining popularity due to their accurate results. [39] have reviewed the use of deep learning of images for plant stress phenotyping, and they concluded that deep learning algorithms require fewer data pre-processing before the modelling work.

Once the biophysical parameters have been predicted/classified from the hyperspectral data/images, the challenge is finding out the contribution of each spectral wavelength to the performance of regression or classification algorithm. There are many methods for calculating feature importance based on statistical or machine learning algorithms. These can be classified into model-dependent and model agnostic methods. Model agnostic methods can be applied to any algorithm (e.g. RF, SVM, ANN, kNN, etc.). In contrast, model-dependent methods are specific to a particular algorithm, like the Random Forest feature ranking method. Few studies have used machine learning algorithms to identify essential features. [40] computed a Relief-F value that uses a kNN-based scoring and is considered a reliable metric to calculate the feature score followed by top-scoring features.

The machine learning models are usually considered "black boxes" because they do not explain how they arrive at the predictions. This is due to their non-linear, complicated, and nested structure. According to [41] and [42], the concept of Explainable artificial intelligence (XAI) has been getting very popular recently. XAI helps to visualize, explain and interpret the machine learning models. XAI is being applied in critical fields like biology and finance, where it is necessary to know the details of the models like its working and identification of essential features [43].

These techniques help to build the trust and transparency of the machine learning models[44].

1.2.4 Simulation of multispectral satellite imagery

Below is a review of studies in which multispectral satellite imagery has been simulated using spectroradiometer data or from data generated from various other models. The purpose of the pilot study [45] pilot study, which was carried out in a greenhouse using a hand-held spectroradiometer, was to examine if the same wavebands employed in the Sentinel-2 MSI could be used to assess and model coffee leaf rust (CLR) at the leaf level. [46] also examined Sentinel-2 red-edge bands for estimating green LAI and chlorophyll concentration empirically. They emphasize the significance of red-edge bands for operational biophysical parameter assessment from Sentinel-2. The objective of the [47] study was to see how well Sentinel-2 MSI spectral band reflectances performed in estimating fractional vegetation cover (FVC). Band 4 (Red), band 12 (SWIR2), and band 8a (NIR2) are the three most crucial Sentinel-2 MSI bands for FVC estimation[48][48][48]. The researchers used optical remote sensing data (proximal hyperspectral and Sentinel 2A) with a radiative transfer model (PROSAIL) to estimate leaf Area Index (LAI) and biomass in a dairy farming environment [48]. [49] investigated the Sentinel-2's potential for archaeological research. Known buried archaeological sites and entirely undiscovered locations were uncovered using a simulated Sentinel-2 image. Using high spatial resolution and multispectral WorldView-2 satellite images, the researchers[50]employed a neural network approach to detect water-stressed crops.

Six multispectral satellites' radiometric signals were simulated using spectroradiometry: a) IKONOS, b) Landsat 5 TM, c) Landsat 8, d) Pléiades, e) Sentinel-2, and f) WorldView-2in [51]. The suitability of each sensor was evaluated to calculate the PV cover fraction recorded for five different habitat types during a vegetative cycle from February to October 2013. The most accurate PV is obtained using multivariate regression using Worldview-2 reflectance.

[52] has used high-frequency ground-based hyperspectral canopy observations to simulate satellite reflectance data for in-season grain yield and nitrogen status estimation in winter wheat.

Variations in the spectral response function (SRF), which describes the sensitivity of each spectral band, have been identified as one of the most significant sources of uncertainty when using multisensory data. The SRF differences between 21 Earth observation satellite sensors, as well as their cross-sensor adjustments for red, near-infrared (NIR), and shortwave infrared (SWIR) reflectance and the normalised difference vegetation index (NDVI) for worldwide vegetation monitoring, are shown in [53].

Optical sensors onboard different platforms can record spectral features of vegetation cover. [21] aims to research the various uses and potential challenges linked to mapping leaf area index (LAI) by integrating remote sensing data obtained by several sensors, given the growing popularity of using UAVs to map plant cover.

1.3 Objectives of Research Work

The objectives of the research work are given below:

- Investigate the onset and the degree of water and nutrient stress in crops (maize and wheat) growing in fields using proximal and remote sensing techniques.
- Development of the spectral indices from multi and hyperspectral data for accurately detecting the stress in plants using a spectroradiometer.
- Simulate the multispectral satellite imagery from the data collected from a spectroradiometer so that in future the stress can be identified from the multispectral satellite imagery.

1.4 Research Methodology and Tools

The proposed methodology for the achievement of the objectives is as mentioned in Table-1.1

Table-1.1: Methodology/ Tools/ Instruments to be used

Objective	Analysis to be undertaken	Instruments/ processes/ software to be used	In-house availability (Yes/ No)	Organization/ Institute (where the facility is available)
1	Create an AI model to classify the water and nutrient stress in crops.	Free satellite imagery, Open Proximal Sensors Datasets, SNAP, QGIS, Weka, Python with the machine and deep learning libraries	Yes	NA
2	Calculation of various vegetation spectral indices from multi and hyperspectral data for detecting the stress in plants.	Free satellite imagery, Open Proximal Sensors Datasets, SNAP QGIS, Weka, Python with the machine and deep learning libraries	Yes	NA
3	Identification of suitable wavelengths from the hyperspectral data and find the similar wavelengths in freely available multispectral satellite imagery datasets.	Free satellite imagery, Open Proximal Sensors Datasets, SNAP, QGIS, Weka, Python with the machine and deep learning libraries	Yes	NA

This thesis is a collection of various research activities ranging from field experiments, analysis of multispectral satellite images and also detecting water and nutrient stress in maize and wheat from hyper spectral data using machine learning techniques. There are eight chapters in thesis and the methodology in each chapter is

different from the others. Therefore, details of methodology is given in each chapter, however a brief methodology for chapter is given below.

Mapping the crop stress for a whole state is a complex task involving many resources. The second chapter explores the computing capabilities of Google Earth Engine (GEE) to map the crop stress in the Punjab state using the remotely sensed Enhanced Vegetation Index(EVI) dataset from the MOD13A1.006 Terra Vegetation collection. The deviation of the EVI of the current (2019-2020) period from the reference period (2014-2019) has been analyzed to classify the crop stress as severe, high, no-change and improved.

In the third chapter, field experiments with a maize-wheat cropping system under various water and nutrient levels were carried out at Punjab Agricultural University, Ludhiana during 2019-20 and 2020-21. The reflectance spectra of crops were collected at regular intervals using Spectroradiometer (wavelength interval between 400 - 2500 nm) on clear and cloudless days. The leaf samples were collected and analyzed for water content, chlorophyll, nitrogen (N), phosphorus (P) and potassium (K) using standard methods. The measured plant parameters were related to spectra hyperspectral data using machine learning regression algorithms and explainable artificial intelligence techniques to identify the optimum wavelengths for nutrient and water stress in maize and wheat.

In fourth and fifth chapters, the secondary hyperspectral datasets (wavelength interval between 400 - 2500 nm) of maize and wheat collected from field experiments conducted in Israel and the United States of America were used. The relationship between spectral wavelength and measured plant parameters was studied using machine learning regression algorithms and explainable artificial intelligence techniques. These studies identified the optimum spectral wavelengths for abiotic stresses in crops.

In the seventh chapter, the wavelengths identified from hyperspectral data from the previous three studies were compared with freely available multispectral satellite imagery (Sentinel-2 MSI, Landsat-8 OLI and Landsat-7 ETM+) for remote monitoring of abiotic stresses in crops.

In the sixth chapter, yellow rust of wheat (a major biotic stress) was classified in the parts of Indian Punjab from Sentinel-2 satellite imagery using deep learning artificial neural networks. A classifier was built that differentiated between an infected and a non-infected data point. The input to the neural network was the data generated from popular spectral indices generated from the satellite imagery.

1.5 Thesis Organization

This thesis is organized into eight chapters. **Chapter 1** provides the background and introduction and the previous studies on detecting and classifying water, nutrient and disease stress in crops using proximal and remote sensing. Mapping crop stress using spectral indices is helpful for crop growth monitoring at a coarser scale (1:50,000 scale or higher). Therefore, crop conditions were monitored in the districts of Punjab using MODIS Enhanced Vegetation Index(EVI) in Google Earth Engine **(Chapter 2)**. This study identified crop stress due to biotic and abiotic factors. The separation of abiotic and biotic stress is required for the site-specific management of crops.

In order to discriminate the abiotic factors (like nutrients and water), a field experiment with a maize-wheat cropping system under various water and nutrient levels was carried out at Punjab Agricultural University, Ludhiana. The reflectance spectra of crops were collected at regular intervals using Spectroradiometer (wavelength interval between 400 - 2500 nm) on clear and cloudless days **(Chapter 3)**. The leaf samples were collected and analysed for water content, chlorophyll, nitrogen (N), phosphorus (P) and potassium (K) using standard methods. The measured plant parameters were related to spectra hyperspectral data using machine learning and explainable artificial intelligence techniques to identify the optimum wavelengths for nutrient and water stress in maize and wheat. The selection of optimum wavelengths from field experiments carried out under Indian conditions may not be applicable to other climatic and soil conditions. Therefore, the secondary hyperspectral datasets (wavelength interval between 400 - 2500 nm)collected from field experiments conducted in Israel and the United States of America were used. The relationship between spectral wavelength and measured plant parameters was studied using machine learning and explainable artificial intelligence techniques **(Chapters 4 and 5)**. The results were compared under Indian, Israel and US conditions to identify the optimum wavelength for nutrient and water stress. These studies identified the optimum spectral wavelengths for abiotic stresses in crops, but the spectral behaviour of biotic and abiotic stresses is quite different. In order to study the biotic stress in crops, yellow rust of wheat (a major disease) was classified in the

parts of Punjab from Sentinel-2 satellite imagery using deep learning artificial neural network **(Chapter 6)**. The wavelengths identified from hyperspectral data in abiotic stresses were compared with freely available multispectral satellite imagery (Sentinel-2 MSI, Landsat-8 OLI and Landsat-7 ETM+) for remote monitoring of stresses in crops **(Chapter 7)**.

Chapter 2

Spatio-Temporal analysis of crop conditions using MODIS EVI in Google Earth Engine

2.1 Introduction

The primary threat to food security is the biotic and abiotic stress in crops. Biotic stress occurs due to the harm done to the crops by living organisms like insects, fungus, viruses, weeds etc. Abiotic stress is usually caused by human activity. Examples of abiotic stress are floods, drought, nutrient deficiency, salinity and low and high temperature. These stress factors are the main reasons for crop yield decline [3].

The traditional way to monitor crop stress is by field surveys and proximal sensing techniques, but this process involves time, costs and human resources. Satellite remote sensing is an alternative by providing continuous coverage over large areas [54]. Remote sensing can aid in continuous monitoring of the crops, quickly diagnosing the problem, and applying timely measures.

[55] reviews many remote sensing techniques for detecting various plant diseases. Techniques like multispectral and hyperspectral imaging, multiband and fluorescence spectroscopy, visible and infrared are discussed. [56] also discusses various remote sensing tools for crop health assessment like popular vegetation indices and platforms like UAV. The study by [57] aims to present a comprehensive review of the widely used and popular crop water stress monitoring methods using remote sensing and machine learning.

Remote sensing researcher requires many resources to download, process, analyze and store BIG earth observation datasets [58]. Google earth engine (GEE) is the most popular modern cloud-based remote sensing platform, which provides ready-to-use high computational capabilities for planetary-scale geospatial big data analytics [59]. This platform is free for research work. GEE includes all the popular remote sensing datasets and the algorithms for processing and analyzing them. [60] discusses

numerous application areas in which GEE is being used, like land use, land cover and crop mapping, agriculture, climate change and disaster management.

The remote sensing data is available in various spatial, spectral, temporal and radiometric resolutions. The data required for a crop stress study depends on the nature of the study(parameters to be identified, e.g. nutrients, water, salinity etc.) and its scale, i.e. if the analysis has to be done for a field, village, district, state or a country. For this research, the whole Punjab state was analyzed for crop stress. As the volume of remote sensing datasets was large, the GEE was exploited for this research. The enhanced vegetation index (EVI) MODIS VI satellite imageries are used for this research. Moderate Resolution Imaging Spectroradiometer (MODIS) [61] is based on the Terra and Aqua satellites. The normalised difference vegetation index (NDVI) and enhanced vegetation index (EVI) are two MODIS vegetation indices (VI) created at 1-km and 500-m resolutions with 16-day compositing periods. These indices provide consistent spatial and temporal information about the vegetation canopy greenness, chlorophyll and canopy structure. [62] compared the MODIS data with other remote /proximal datasets. At four intensively studied test locations representing semi-arid grass/shrub, savanna, and tropical forest biomes, their results demonstrated a good correlation between airborne-measured top-of-canopy reflectances and vegetation indices values and those from the MODIS sensor. The MODIS data's scientific significance was confirmed by the derived field biophysical metrics.

The enhanced vegetation index (EVI) reduces canopy-soil variations and enhances sensitivity in dense vegetation. The formula of EVI [61] is given in equation (1),

$$EVI = G((NIR - Red)/(NIR + C1 * Red - C2 * Blue + L)) \quad (1)$$

where Near infra-red (NIR), Red, and Blue bands are the surface reflectances, C1 and C2 are the coefficients of the aerosol resistance, L is the canopy background adjustment, and G is the scaling factor. For the above algorithm, the values of coefficients are L=1, G=2.5, C1=6 and C2=7.5.

MODIS EVI is used for various vegetation and agricultural-related applications like crop mapping, yield estimation, agricultural drought monitoring,

salinity mapping, water and other environmental stress. The time series of EVI was used by [63] to map crop cycles in China. [64] compared the NDVI and EVI products for the estimation of large-scale rice yield. They found that the EVI-based models were slightly more accurate than NDVI-based models. Extracting phenological characteristics from MODIS EVI is a possible approach for detecting soil salinity in rice-growing areas at regional scales, according to a study [65]. [66] used Google Earth Engine and MODIS EVI data to map crop phenology, identify cropping systems, detect land use change and drought risk in Ethiopia. The 16-day Terra-MODIS composite and EVI products are sensitive to stresses associated with drought, according to a study by [67], and these products can aid in drought emergency monitoring.

2.2 Objectives

This research work aims to map the crop stress in the Punjab state using remotely sensed Enhanced Vegetation Index(EVI) datasets from the MOD13A1.006 Terra Vegetation collection. The deviation of the EVI of the current (2019-2020) period from the reference period (2014-2019) has been analyzed to classify the crop stress as severe, high, no-change and improved.

2.3 Study Area and Datasets

In this section, the study area is first introduced and subsequently, the information about the satellite datasets is provided.

2.3.1. Study Area

The whole Punjab (India) state is the study area of this research. It lies between 29^0 33’ & 32^0 31’ N latitude and 73^0 53’ & 76^0 55' E longitude and covers 50,362 sq. km. Three distinct seasons occur in Punjab. The summer season extends from April to June, the monsoon season from July to September and the winter season extends from October to March. The cropping pattern of the state shows a majority of rice, wheat and cotton crops. The vegetation in the rabi season(October - March) has been analyzed for this research. Wheat is the major rabi crop grown in Punjab state.

2.3.2. Datasets

The MODIS VI satellite data MOD13A1.006 Terra Vegetation Indices makes available the global vegetation conditions. This dataset consists of two indices: Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI). Both the datasets are provided every sixteen days at 500 (m) spatial resolution. For this research, the EVI dataset for the rabi season of the period 2014-2020 has been analyzed.

The European Space Agency's World Cover 10m v100 [68] is a 10-meter resolution worldwide land cover map for 2020 based on Sentinel-1 and Sentinel-2 satellite data. This dataset has been used for masking the non-agricultural areas.

2.4 Methods

GEE, a modern cloud-based remote sensing platform, has been used for image analysis for this research. The methods include filtering the MODIS EVI images based on the geographic extent, cropping season and desired time periods, deviation calculation from the reference and current maximum images, reclassifying the deviation raster, masking out the non-cropland areas, and district-wise calculation of the stressed areas. The methodology of the research is shown in Figure 2.1.

2.4.1. Preparing the MODIS EVI Dataset

The MOD13A1.006 Terra Vegetation Indices images were filtered based on three parameters index (EVI), study area (Punjab State) and acquisition dates. The images acquired between November 2014 and March 2019 were selected to create the reference dataset. The current dataset consisted of images between the period November-2019 and March 2020.

The reference dataset was prepared by computing the maximum (pixel value) of the seasonal (Rabi season) images for the last five years (2014-2019) and then calculating the mean of all the five images. The current dataset was created by taking a maximum of all the pixels of the EVI images of 2019-2020. Figure 2.2 represents the generated datasets.

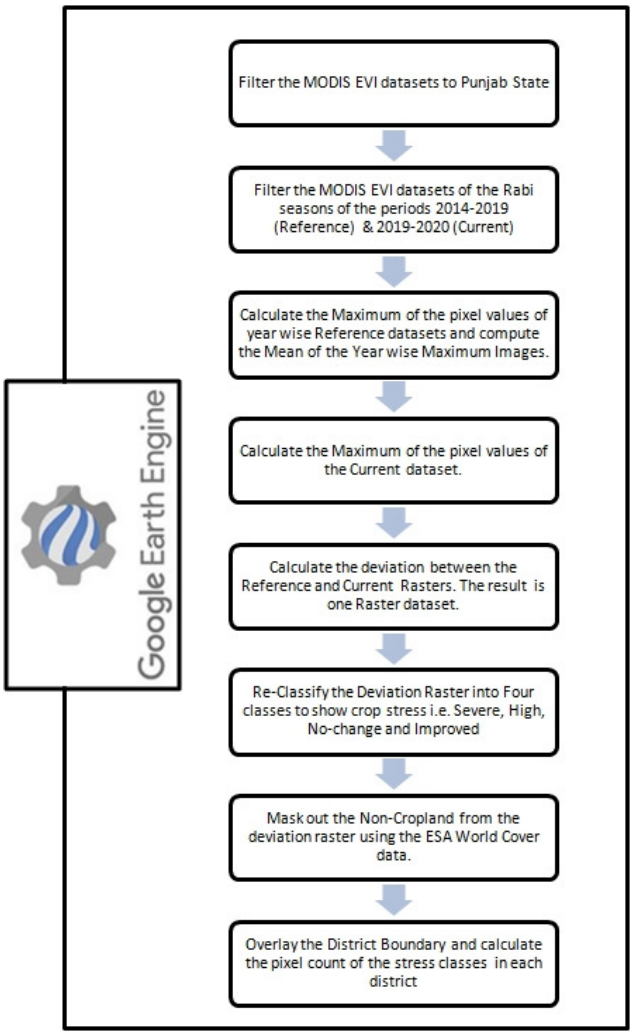


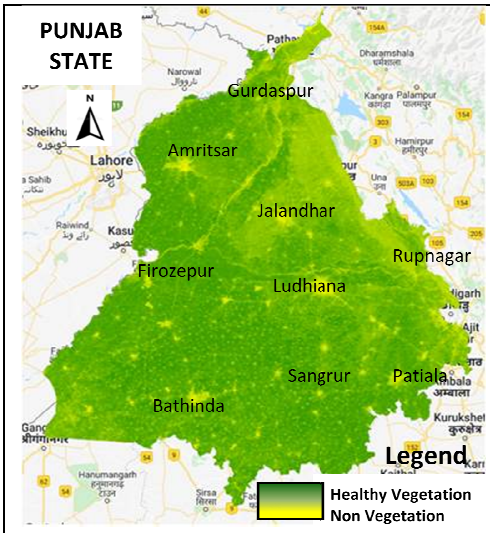
Figure 2.1. Research Methodology

2.4.2. Deviation

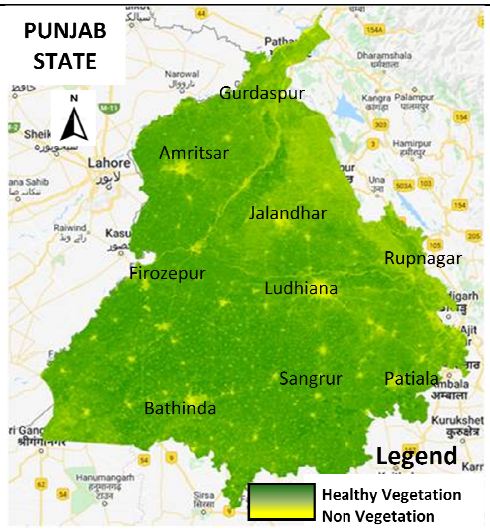
The deviation or anomaly raster was calculated by subtracting the reference from the current dataset. The deviation represents the current crop conditions compared to the conditions from the last five years. This raster was reclassified into the following classes which are summarized in table 2.1.

Table 2.1. Deviation Classes

S.No	Deviation Value	Crop Conditions
1	-0.25 to -1.0	Severe Stress
2	-0.1 to -0.25	High Stress
3	-0.1 to 0.1	Normal (No Change)
4	0.1 to 0.25	Improvement in crop conditions
5	0.25 to 1.0	Very Good Conditions



a) Reference Mean of 2014-2019



b) Current Maximum of 2019-2020

Figure 2.2. Generated Datasets (a) Reference Mean of 2014-2019, (b) Current Maximum of 2019-2020

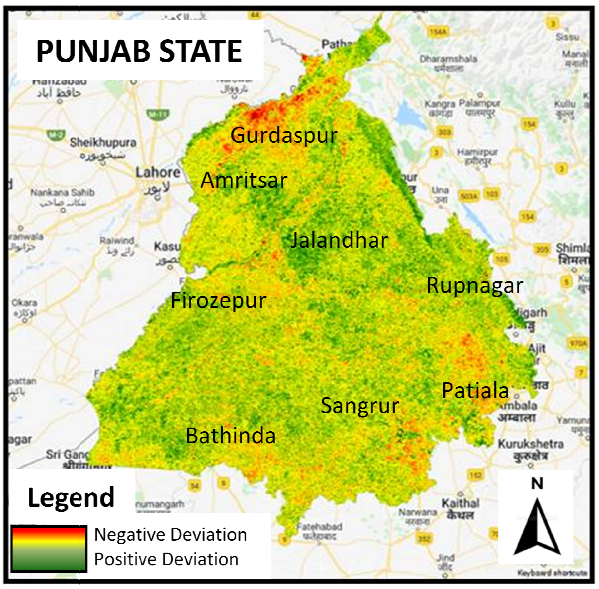


Figure 2.3. Deviation Raster (Red-Green Colour ramp shows areas of negative to positive deviations)

2.4.3. Masking the Non-Cropland area

The non-cropland area, including forests, built-up, water bodies etc., were masked using the ESA World Cover 10m v100 dataset. This step was included so that the final statistics included only agricultural vegetation.

2.4.4. District-Wise Analysis

The Punjab district layer was overlayed on the masked output to generate crop condition class-wise statistics for all the districts.

2.5 Results & Discussions

The results of the research were in the form of maps and tables. According to the results, the Gurdaspur district experienced maximum crop stress in the rabi cropping season of 2019-2020. Amritsar, Patiala, Pathankot and Fatehgarh Sahib followed the Gurdaspur District. The least stress was identified in the Patiala, Nawan Shehar and SAS Nagar(Mohali) districts. The district-wise area statistics are provided in table 2.2. The ‘very good class‘ was not found in any district during the analysis.

Table 2.2. District-wise area statistics

SNo	District Name	Severe Stress		High Stress		No Change in crop conditions		Improvement in crop conditions	
		Pixel Count	Area (ha)	Pixel Count	Area (ha)	Pixel Count	Area (ha)	Pixel Count	Area (ha)
1	Gurdaspur	2	50	106	2650	2271	56775	47	1175
2	Amritsar			20	500	2142	53550	111	2775
3	Patiala			17	425	2836	70900	150	3750
4	Pathankot			15	375	536	13400	9	225
5	Fatehgarh Sahib			12	300	1138	28450	26	650
6	Bathinda			10	250	2953	73825	128	3200
7	Kapurthala			8	200	1364	34100	148	3700
8	Ludhiana			8	200	2977	74425	164	4100
9	Muktsar			6	150	2323	58075	142	3550
10	Sangrur			6	150	3300	82500	133	3325
11	Hoshiarpur			6	150	2172	54300	86	2150
12	Rupnagar			5	125	865	21625	32	800
13	Mansa			5	125	2062	51550	33	825
14	Jalandhar			5	125	2169	54225	221	5525
15	Moga			4	100	2135	53375	46	1150
16	Faridkot			4	100	1297	32425	53	1325
17	Firozpur			3	75	2278	56950	84	2100
18	Taran Taran			3	75	2288	57200	74	1850
19	Barnala			2	50	1362	34050	44	1100
20	Fazilka			2	50	2423	60575	127	3175
21	SAS Nagar(Mohali)			1	25	742	18550	40	1000
22	Nawan Shehar			1	25	987	24675	70	1750
23	Patiala			0	0	11	275	0	0

Although the stress pixels (Red Colour) were scattered all over the state (figure 2.4), areas around Dera Baba Nanak (figure 2.5) in the Gurdaspur district and Bassi Pathana (figure 2.6) in Fatehgarh Sahib experienced extensive stress. The stress pixels near these places were found in large clusters. It was also witnessed that the northwestern part of Punjab, comprising of districts of Pathankot, Gurdaspur and Amritsar, witnessed more stress than the rest of Punjab. Some parts of the Jalandhar, Ludhiana and Patiala districts also witnessed improvements in crop conditions.

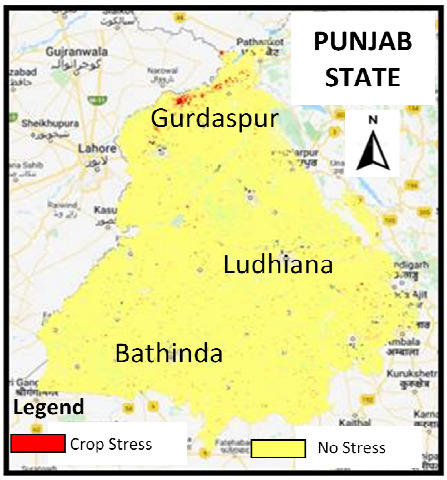


Figure 2.4. Stressed Areas (Red Colour) in Punjab State

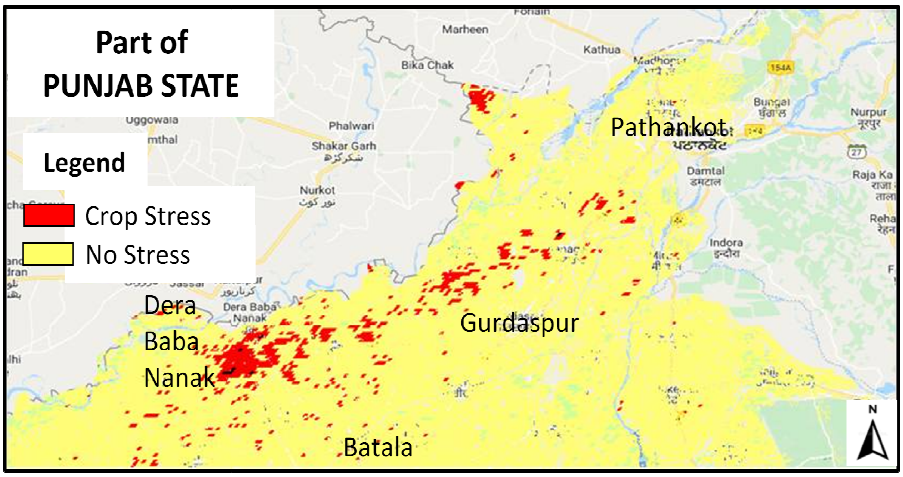


Figure 2.5. Stressed Areas (Red Colour) around Dera Baba Nanak in Gurdaspur

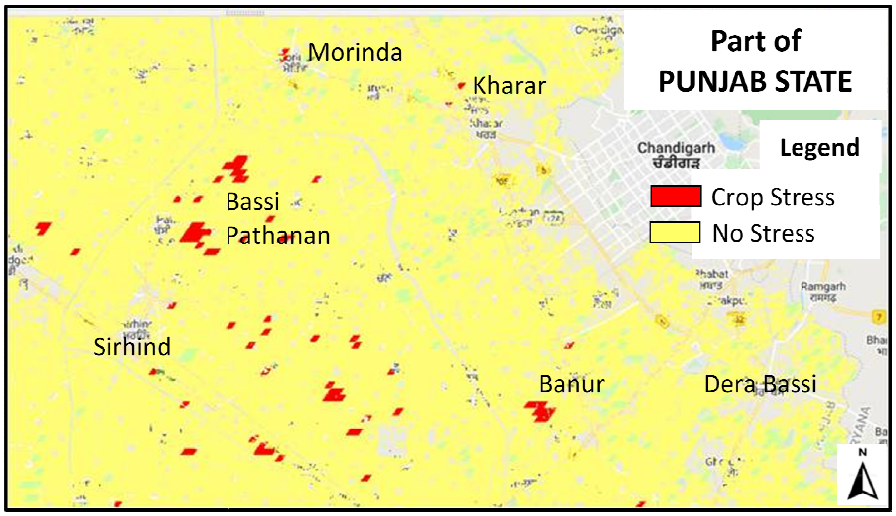


Figure 2.6. Stressed Areas (Red Colour) around Bassi Pathanan in Fatehgarh Sahib

According to the Statistical Abstract of Punjab 2020[69], 2019[70], 2018[71], the wheat crop yield (Per Hectare in Kilogram) for the last few years of Gurdaspur, Amritsar, Patiala, Pathankot and Fatehgarh Sahib Districts is given in table 2.3. According to the table, it can be seen that the crop yield has decreased in the Gurdaspur, Amritsar, Patiala, Pathankot and Fatehgarh Sahib Districts. This proves the relation between the decrease in the EVI and the crop yield.

Table 2.3. District-wise wheat yield (Per Hectare in Kilogram) statistics

Rabi Period	Gurdaspur	Amritsar	Patiala	Pathankot	Fatehgarh Sahib
2017-2018	4733	4866	5272	4018	5341
2018-2019	4717	4813	5472	3984	5180
2019-2020	4175	4533	4771	3944	4394

Although research papers, newspapers and various reports of the study period were reviewed but the exact cause of the decrease in EVI could not be identified. Traditionally, field surveying is used to identify crop stress and estimate its area statistics. Therefore it is essential to develop an efficient method for crop stress mapping using remote sensing datasets. This study demonstrates an effective way of utilizing a cloud-based earth observation tool for producing a crop stress map of

Punjab. Using the freely available MODIS datasets and openly available tools ensures other researchers' work extension. Such a large-scale investigation incorporating multiple satellite images necessitates a lot of resources (computation, storage, specialized software and time). This case study illustrates how to use GEE for massive geospatial data analytics in the cloud with just a normal computer and internet access.

Chapter 3

Detection of nutrient and water stresses in maize and wheat using proximal sensing.

3.1 Introduction

Abiotic stress, particularly water and nutrient stress, is becoming a major challenge to food security [72] and 71% of the yield losses in crops are caused by abiotic factors [73]. Because the impact of abiotic stress on crop growth is usually only noticed after it becomes visible, assessing yield loss due to water and nutrient is critical for long-term agricultural production.

Water content, nitrogen (N), phosphorus (P), potassium (K), chlorophyll a and chlorophyll b, and others are key biochemical components of crop organic matter. Nitrogen is a critical component of the chlorophyll molecule, which allows the plant to capture sunlight energy for carbohydrate generation, resulting in increased plant growth and output [74]. Phosphorus is required to develop new plant cells and transfer the genetic code from one cell to another [75]. Potassium is an activator of essential enzymes, such as protein synthesis, sugar transport, carbon and nitrogen metabolism, and photosynthesis [76].The function of chlorophyll is to absorb light for photosynthesis. Chlorophyll is classified into two types: a and b. The major photosynthetic pigment is chlorophyll a, while chlorophyll b is an accessory pigment that receives energy and transfers it on to chlorophyll a [77]. Chlorophyll a absorbs energy from blue-violet and orange-red light wavelengths, whereas chlorophyll b absorbs energy from green light wavelengths.

Plants require high amounts of these key biochemical components, and estimating these nutrients on a spatiotemporal scale is required to improve macronutrient usage efficiency. Traditionally, these values have been assessed in a laboratory using invasive procedures on plant tissue. The laboratory methods are costly and time-consuming. The non-invasive estimation of these characteristics using multispectral and hyperspectral remote sensing aids in plant health monitoring [7]. The regular monitoring of crop conditions by field surveys is a resource-intensive task [78]. Proximal and remote sensing sensors provide a solution by collecting information about crop conditions non-invasively.

To estimate various plants parameters of plants from spectra, statistical and machine learning algorithms have been extensively applied [8]. Machine learning algorithms are more popular because they produce better results than statistical methods for predicting non-linear correlations [79].Though the machine learning algorithms have a better performance, but they are challenging to interpret due to their non-linear, nested, and complicated structure. Explainable artificial intelligence (XAI) techniques are gaining popularity because they make sophisticated 'black box' machine learning models easier to understand, visualise, and analyse [43]. The explanations and interpretations contribute to the machine learning models' transparency and trustworthiness.

Many investigations have used machine learning approaches to estimate biochemical parameters in plants using hyperspectral images, but only a handful have combined machine learning and XAI techniques to estimate nutrient content in plants. As a result, combining machine learning and XAI approaches, a study was conducted to estimate water content and nutritional content in maize and wheat. The study's particular objectives were to develop machine learning models for estimating water content, N, P, K, chlorophyll a and chlorophyll b in maize and wheat using hyperspectral data, and (ii) apply XAI approaches to determine the best wavelengths for each parameter estimation.

3.2 Materials and Methods

3.2.1 Field Experiment and collection of hyperspectral data

Field experiments with maize and wheat were conducted at Punjab Agricultural University farm, Ludhiana during 2019 and 2020. The experiment was laid out in three replicates in a split plot design with three irrigation treatments (IW/CPE 1.0, 0.75 and 0.5) in main plots with five nutrient treatments (T1 = 100% NPK, T2 = 75% NPK, T3 = 50% NPK, T4 = 100% N, T5 = 100% NP) in subplots (IW: Irrigation Water, CPE: Cumulative Pan Evaporation). The recommended doses of fertilizers (100% NPK) for maize (variety PMH I) were Urea (90 kg / acre), Di-Ammonium Phosphate (55 kg/acre) and Muriate of Potash (20 kg /acre). The recommended fertilizer doses (100% NPK) for wheat (variety PBW725) were Urea (90 kg / acre), Single Superior Phosphate (55 kg/acre) and Muriate of Potash (20 kg /acre). The leaf

samples were collected at monthly interval during August-October 2019 and 2020 for maize, December 2019-March 2020 and December 2020-April 2021 for wheat. The crop field experiment photographs are presented in Annexure-1. The leaf samples were analysed for water content, chlorophyll, nitrogen (N), phosphorus (P) and potassium (K) using standard methods. The reflectance spectra of crops were collected on leaf sampling dates using ASD Field Spec-4 Spectroradiometer (wavelength interval between 350 - 2500 nm) on clear and cloudless days.

3.2.2 Hyperspectral Data Processing

Two data transformation techniques were used to pre-process the raw data: Savitzky-Golay 1st derivative and Deresolve. By correcting the baseline effects in the spectra, Savitzky-1st Golay's derivative transformation removes nonchemical influences and generates strong calibration models. Derivatives can help resolve overlapped bands. Deresolve is commonly used for noise reduction. It smooths the spectra with a resolution function, making it appear as if it was taken with a lower resolution equipment. Both Chlorophyll a and Chlorophyll b are sensitive between 406 and 752 nm, therefore spectral subsetting was performed in this wavelength region to estimate these pigments from hyperspectral data.

3.2.3 Estimation of water, chlorophyll, N, P and K contents in maize and wheat from hyperspectral data using machine learning techniques.

The water, chlorophyll, N, P and K contents in maize and wheat were estimated from hyperspectral data using the four machine learning regression algorithms: Random Forest (RF), Support Vector Regression (SVR), Gradient Boosting Regression (GBR) and Partial Least Square Regression (PLSR). In all these algorithms, reflectance values were used as the input parameters (X variables), and water content, chlorophyll, nutrient concentration and other parameters in maize and wheat were selected as the target variable (Y variable). The machine learning algorithms were implemented in Python [80] programming language using the Scikit-learn [81] library. Scikit-learn is a standard, open-source library that integrates many ML algorithms. Google's Colaboratory [82] cloud-based environment was selected for developing and running the machine learning algorithms. Without installation or

configuration, Colaboratory allows the user to write and execute Python code in a basic browser. The GridSearchCV function was used to tune hyperparameters. This method assisted in determining the appropriate parameters for each regression algorithm. The various parameters used for the development of the models are listed in Table3.1.

The performance of the machine learning algorithms is also influenced by the sampling distribution of the training and the testing datasets. [83], [84] have recommended the 80% and 20% split for the training and validation datasets. The dataset was also randomly split into 80% and 20%. The performances of machine learning algorithms were evaluated using the coefficient of determination (R2) and mean-squared error (MSE). An overview of the methodology is given in Figure 3.1.

Table3.1 Model Parameters

Machine Learning Model	**Model Parameters**
Random Forest	max_depth, random_state, max_features, min_samples_leaf, min_samples_split, n_estimators
Support Vector Regression	kernel, C, gamma, epsilon, coef0
Gradient Boosting Regression	learning_rate, max_features, min_samples_leaf, n_estimators, max_depth
Partial Least Square Regression	copy, max_iter, n_components, scale, tol

3.2.4 Identification of important features (wavelengths) from hyperspectral data

Machine learning models are frequently referred to as "black boxes," making it difficult to decode them and understand how they produced a specific output. Optimal wavelengths to quantify water content, chlorophyll, nutrients, and other characteristics were identified in our work using Explainable artificial intelligence (XAI) tools, which could be useful in interpreting and explaining complex AI models. [85] proposed the SHapley Additive explanations (SHAP) value, a critical XAI tool, for interpreting the machine learning models. SHAP values are based on shapley values that originated from the mathematical game theory. In the context of machine learning, the game is the outcome of a model, and the players are the features

included in the model. SHAP helps to quantify the contribution that each feature (player) brings to the model prediction (game) [86]. It breaks down the prediction to show the impact of each feature. It can explain individual predictions and aggregate them to get powerful insights into the whole model. SHAP values calculations are explained in detail by [87]. We implemented the XAI in Python using the SHAP library. The key benefit of using SHAP values for this study is the model's overall interpretability. The collective SHAP values are interpreted globally to find the critical wavelengths linked with the target variables. The RF and GBR models were explained using the SHAP function "TreeExplainer" while the PLSR and SVR models were explained using the SHAP function "KernelExplainer".

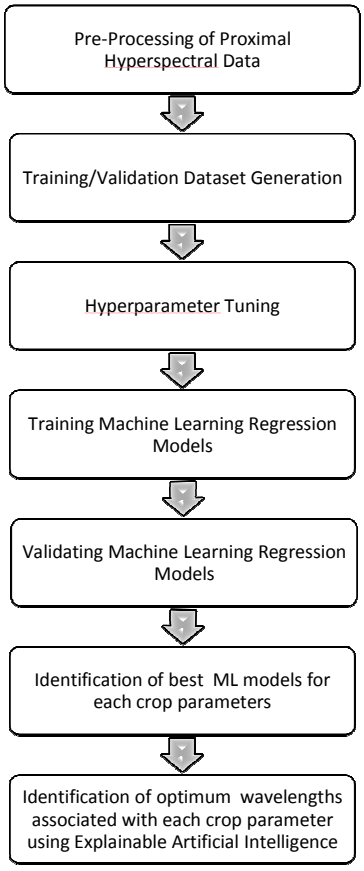


Figure 3.1 An overview of the approach

3.3 Results and Discussions

3.3.1 Selection of optimum machine learning regression models to estimate the crop parameters

(a) Maize

Four machine learning algorithms were used to estimate water, chlorophyll a and chlorophyll b, N, P and K contents in maize and wheat.

The coefficient of determination (R^2) and mean square error (MSE) to estimate the six plant parameters in maize are given in Table 3.2.

The water content in maize leaves was better estimated with GBR followed RF models. The performance of GBR model was better than the other two models for estimation of chlorophyll a in maize leaves. The chlorophyll b was better predicted with the PLSR model. Among the three macronutrients, N in maize was estimated using PLSR, GBR and RF models, total P content using PLSR and RF models, and total K using PLSR and RF models.

Table 3.2 Coefficient of determination (R^2) and mean square error (MSE) for the Maize models

ML Algorithm	Water Content		Nitrogen		Phosphorus		Potassium		Chlorophyll a		Chlorophyll b	
	R^2	MSE	R^2	MSE	R^2	MSE	R^2	MSE	R^2	MSE	R^2	MSE
RF	0.21	17.7	0.55	0.04	0.56	0.001	0.63	0.027	0.55	0.065	0.47	0.005
SVR	0.08	82.1	0.40	0.06	-1.69	0.009	-0.06	0.08	0.5	0.07	-0.4	0.015
GBR	0.36	57.1	0.56	0.04	0.5	0.001	0.61	0.029	0.63	0.05	0.52	0.005
PLSR	0.07	82.5	0.57	0.04	0.59	0.001	0.63	0.028	0.52	0.06	0.57	0.004

(b) Wheat

The water content in wheat leaves was better predicted with RF and GBR models, chlorophyll a and b with PLSR model, and total N, P and K with GBR model. The coefficient of determination (R^2) and mean square error (MSE) to estimate the six plant parameters in wheat are given in Table 3.3.

Table 3.3 Coefficient of determination (R^2) and mean square error (MSE) for the Wheat models

ML Algorithm	Water Content		Nitrogen		Phosphorus		Potassium		Chlorophyll a		Chlorophyll b	
	R^2	MSE	R^2	MSE	R^2	MSE	R^2	MSE	R^2	MSE	R^2	MSE
RF	0.87	27.78	0.72	0.078	0.63	0.004	0.72	0.035	0.49	0.14	0.6	0.005
SVR	0.62	81.52	0.51	0.135	-0.33	0.017	0.46	0.068	-0.004	0.29	-0.78	0.025
GBR	0.86	29.95	0.82	0.048	0.71	0.003	0.78	0.027	0.46	0.15	0.6	0.005
PLSR	0.79	44.24	0.63	0.102	0.63	0.004	0.64	0.044	0.55	0.132	0.75	0.003

3.3.2 Sensitive bands for estimation of water content, chlorophyll a, chlorophyll b, nitrogen, phosphorus and potassium

(a) Maize

The important wavelengths predicted from the best performing machine learning model associated with each crop parameter were identified using the SHAP values. The important wavelengths associated with the parameters are given in Table 3.4 (Maize) and Fig. 3.2.

The most important wavelengths associated with water content for detecting water stress in maize are in the near infrared region (841, 842, 847 and 702 nm) using GBR). In general, weak water absorption bands are centred near 970 and 1200 nm (NIR region) and strong water absorption bands are centred on 1450, 1940, and 2500 nm (SWIR region), but the water stress in our study was detected around 841-847 nm due to major water absorption shoulder around this wavelength because of the combined overtone of vibrational symmetric, asymmetric, and bending stretches of O-H bands [88]. [89] found that leaf water content in maize is predicted from 700 nm, 740 nm and 860 nm. [90] found that the wavelength interval of 689 and 720 nm, and 755–842 nm including the other regions (553–556, 950–970, 1013–1034, and 1055–1075 nm) are useful for predicting water content in plants.

The sensitive bands were 599, 617, 600, 618, and 651 nm for estimation of chlorophyll a, and 486, 691, 599, 690 and 431 nm for chlorophyll b. The reflectance spectra is highly sensitive to chlorophyll content from 680 to 780 nm and the light is strong penetrating to leaves in this region [91]. Our results find support from the previous studies in which chlorophyll concentration within leaves can be better estimated from red and far red region of the spectrum ([92], [93]). [94] also found that red and red-edge bands are helpful for estimation of chlorophyll in maize.

Though nitrogen content in maize leaves was predicted RF, GBR and PLSR models with R^2 >0.50, but the wavelength predicted with the RF models were in the visible, near-infrared and SWIR region (617, 804, 1978, 1979 and 738 nm). Previous studies have also found that visible and near infrared (VNIR) region are mainly used for estimation of N in plants due to close relationship of N and chlorophyll which has deep absorption valleys in this region ([95], [96]). With GBR and PLSR, the most

important wavelength for estimation of N in maize was in the SWIR region. [97] also found that SWIR region of the spectra can also be used for estimation of N in plants. The total P content in maize leaves was better predicted using PLSR and RF models, but the sensitive bands were in the SWIR region using GBR, visible, NIR and SWIR region using RF model. In general, NIR and SWIR regions are helpful for detecting the P deficiency in plants. [98] and [99] also found the sensitive band in these regions for estimation of P in plants. There is a close relationship between P and reflectance in the NIR region because P is an energy supplier in energy-consuming processes such as photosynthesis. In combination with reflectance in visible or NIR region, reflectance in the SWIR range was found to monitor nitrogen, phosphorus, and potassium status of plants [100]**.** The total K in plants were better estimated using the RF amd PLSR models. The sensitive bands were in visible and NIR region with the RF model, but in SWIR region with PLSR model. [101] found that visible and NIR regions are useful for estimation of K stress in plants. The deficiency of K in plants causes reduces the translocation of photosynthates from leaves to fruits which results in lower chlorophyll concentration due to increased hexose, sucrose and starch contents in leaves. This results in changing the reflectance features of maize leaves. The K deficiency in plants is also related with the SWIR region because K ions regulate leaf water content through stomatal conductance. [100] and [102] also found that sensitive bands for K deficiency in plants are around the SWIR region.

. Table 3.4 Important wavelengths associated with the parameters of Maize

ML Algorithm	**Water Content**	**Nitrogen**	**Phosphorus**	**Potassium**	**Chlorophyll a**	**Chlorophyll b**
Random Forest				617, 1180, 653, 651, 560		
GBR	841, 842, 847, 702				599, 617, 600, 618, 651	
PLSR		1726, 1725, 1547, 1724, 751	1724, 1725, 1726, 1723, 1547	1725, 1726, 1724, 751, 1344		486, 691, 599, 690, 431

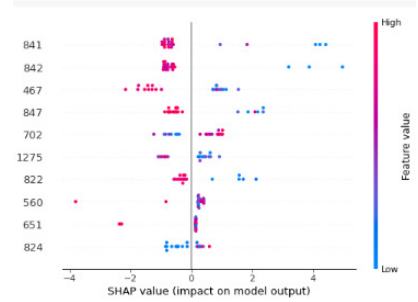
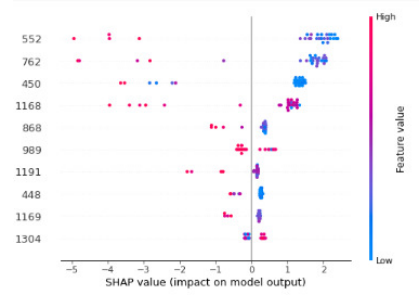
(b) Wheat

The important wavelengths associated with the parameters are given in Table 3.5 (Maize) and Fig. 3.2. The optimum bands for estimation of water content in wheat were visible, NIR and SWIR regions. The top most wavelength for estimation of water in wheat was in visible region due to fifth and sixth overtone of O–H bands at 401 and 449 nm which cause water absorption bands in this region [103]. The optimum wavelength of 552 nm for estimation of water content in wheat leaves is due to third overtone of O–H bands at this wavelength (small absorption peak) [88]. The sensitive bands to estimate water content in SWIR region were 1168 and 1305 nm. [104] also found that the region between 1150 and 1260 nm is useful for estimation of leaf water content. The most important bands for estimation of chlorophyll a and b were in the visible region (554, 553, 656, 555 and 405 nm for chlorophyll a, and 486, 691, 599, 690 and 431 nm for chlorophyll b). [105] found that wavelength from 520 to 550 nm and 695 to 705 nm was closely related with chlorophyll content in plant leaves. The sensitive bands to estimate N in wheat leave were 443 nm> 449 nm >597 nm> 885 nm > 450 nm. Since nitrogen is a major component of the chlorophyll pigments which are sensitive in 406 and 752 nm, therefore VIS-NIR region of the spectra was useful for estimation of N content in leaves [106]. [107] developed two spectral indices(canopy chlorophyll content index from 670, 720, 790 nm and spectral ratio planer index from 445, 705 and 750 nm) to detect the N stress in crops. [108]used the hyperspectral data to identify the N status of wheat and they selected few features in the NIR region (> 750 nm) for estimating plant N content. The reflectance values 650–680 nm, adjacent to the Chl absorption peak at 675 nm coupled with PLSR are suitable for estimation of chlorophyll and N content in plants [109], which concurs with previous studies on spectroscopic estimation of plant N content**.** According to [110], shortwave infrared is associated with the leaf water content and biochemicals, near-infrared with cell structure and visible spectrum is with leaf pigments. According to [111], chlorophyll and N contents can be accurately remotely estimated from green and red-edge chlorophyll indices using near-infrared (780–800 nm) and either green (540–560 nm) or red-edge (730–750 nm) spectral bands.

The optimum bands for estimation of total P in wheat leaves were in 1304, 1305, 810, 1259 and 1505 nm. Among these five wavelengths, four are in the SWIR region and one in the NIR region. [20] also found that SWIR region is useful for estimation of P deficiency in wheat. [99] also found the sensitive band in NIR region for estimation of P in plant leaves. The sensitive bands for detecting K deficiency in wheat leaves were in the SWIR region because K ions regulate leaf water content through stomatal conductance. [100] and [102] also found that sensitive bands for K deficiency in plants are around the SWIR region. [20] developed a vegetation index using wavelengths 1645 and 1715 nm (SWIR) to determine the potassium concentration in wheat.

Table 3.5 Important wavelengths associated with the parameters of Wheat

ML Algorithm	Water Content	Nitrogen	Phosphorus	Potassium	Chlorophyll a	Chlorophyll b
Random Forest	552, 762, 450, 1168, 868					
GBR		443, 449, 597, 885, 450	1304, 1305, 810, 1259, 1505	1305, 1504, 1091, 1259, 1505		
PLSR					554, 553, 656, 555, 405	404, 656, 405, 554, 553

	Maize	Wheat
Water Content		

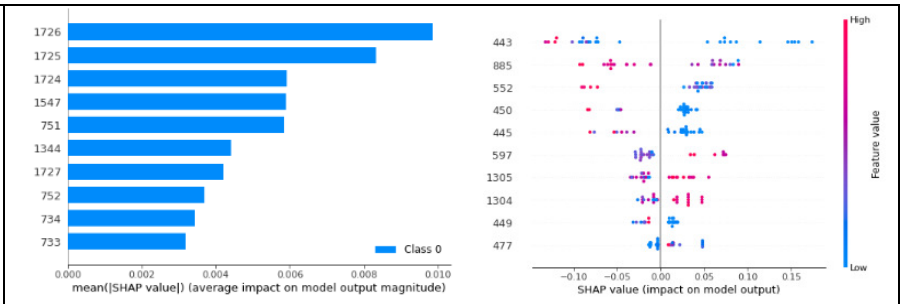
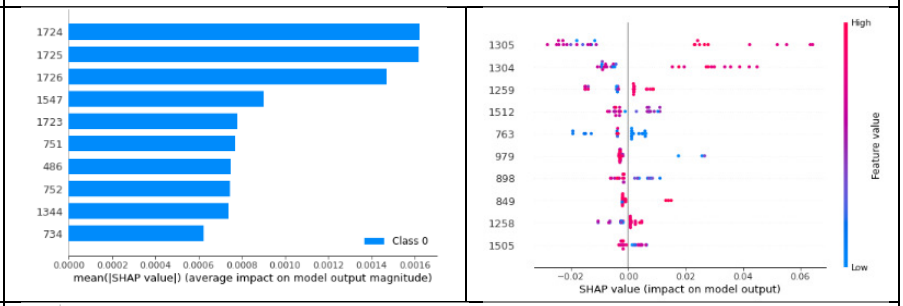
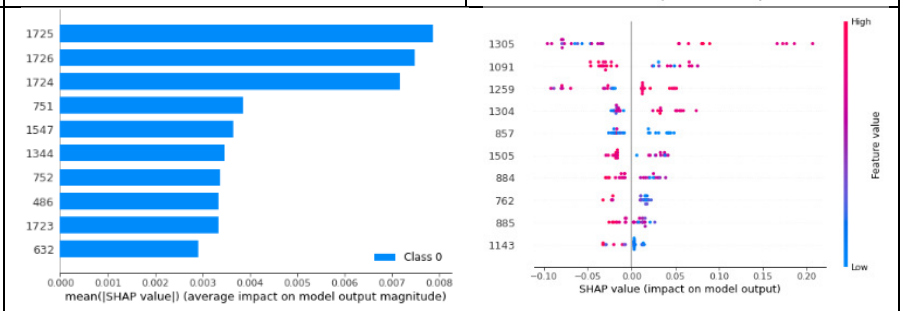
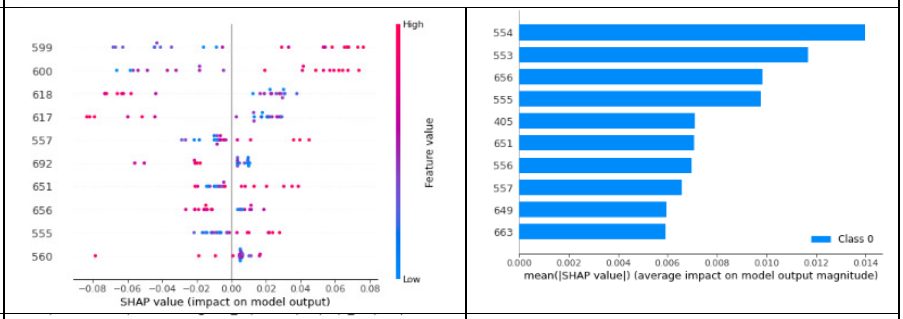
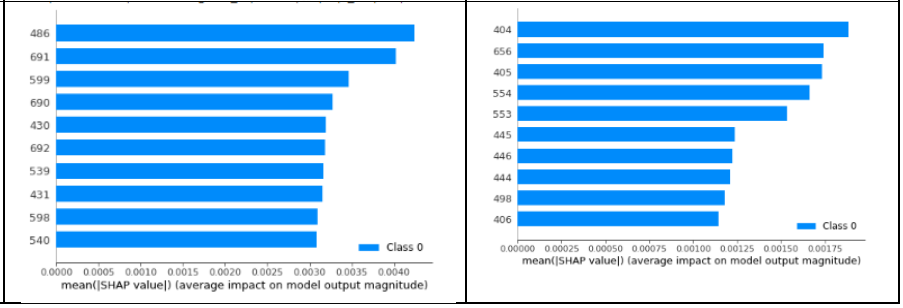
Nitrogen	
Phosphorus	
Potassium	
Chlorophyll a	
Chlorophyll b	

Figure 3.2 Output of XAI

The results of this study suggest that VIS-NIR–SWIR leaf reflectance can be a valuable tool for low-cost, non-destructive, high-throughput investigation of leaf physiological and biochemical characteristics (including N, P and K)**.** The wavelengths identified for various parameters are useful for designing future multispectral satellites/super-resolution for multispectral and hyperspectral images which may help in remote monitoring of water and nutrient stress.

Chapter -4

Estimation of nitrogen content in wheat from proximal hyperspectral data using machine learning and explainable artificial intelligence (XAI) approach. 40-52

4.1 Introduction

Nitrogen (N), a primary macronutrient, is a principal component of many plant structures (e.g., chlorophyll and amino acids etc.) and also required for metabolic processes (e.g., photosynthesis) of plants. It promotes rapid vegetation growth, gives plants a healthy green colour, and produces the greatest yield response in crop plants[112][113][114]. The insufficient supply of N leads to critical plant disorders. One of the main adverse effects is the reduction of chlorophyll, which damages the starch and protein contents. Therefore, it is essential to manage the N in the agriculture produce effectively. However, the over-application of N can also be a problem as it may pollute the underground water[115].

The precision management of N is vital for increasing the N use efficiency. Its accurate measurement is required to optimize the application of N fertilizers in soils[116]. Traditionally, N concentration in plant tissues is determined in the laboratory using invasive procedures such as Kjeldahl-digestion and Dumas-combustion, but these procedures are time-consuming and costly[117]. Remote sensing has a great potential to non-invasively assess the nutrient stress from the leaf to the landscape scale. Several studies have found significant relationships between crop physiological parameters and spectral reflectance. Hyperspectral (narrowband) remote sensing has been found more beneficial in assessing crop stress than multispectral (broadband) sensors[118]. Hyperspectral data, mostly from proximal sensors, provide an ideal opportunity to estimate the N content from the plants[119]. There are several tools and techniques to process and analyze hyperspectral data.[8]discussed various approaches to analyze the hyperspectral data for vegetation. These techniques include artificial intelligence, statistical regressions, physical modelling, and spectral positioning. Artificial intelligence techniques listed various studies that have exploited neural networks and machine learning based on regression trees. Among various statistical techniques, partial least square regression (PLSR) is the most popular technique for estimating plant biophysical parameters from

hyperspectral data[120]. [121] used PLSR to model the leaf N concentration and density, leaf area index, and dry leaf weight from hyperspectral reflectance of the plants.[122] computed the normalized difference vegetation indices of all possible combinations of wavelengths in the region from 438 to 884 nm. They found that leaf chlorophyll and nitrogen concentration can be estimated using linear regression and PLSR. [19]also estimated the canopy nitrogen content in winter wheat using optimised hyperspectral spectral indices coupled with partial least squares regression.[123] also used PLSR to estimate the plant biomass from the spectra (350-2500nm) collected using a field spectroradiometer. According to the authors, the prediction quality of the estimation model could be increased by selecting suitable wavelengths and applying the derivatives to the raw spectra. The plant biomass was predicted with R^2 of 0.70 using the PLSR model.

Besides PLSR, other machine learning algorithms are becoming more popular in analyzing many plant parameters from hyperspectral data.[37] highlighted the advantages of machine learning over the conventional statistical techniques for detecting biotic stress in crops. Machine learning algorithms require very few statistical assumptions about the data, and they can be used to develop both linear and non-linear models. [38] emphasised that machine learning is very promising in plant stress analytics. It can be used for Identification (K-Nearest Neighbour (K–NN), SVM, SOM, Bayes Classifier (BC), ANN, K–means, Linear and Quadratic Discriminant Analysis (LDA/QDA)), classification (Random Forest (RF), K-means, SVM, BC, LDA/ QDA, K–NN, SOM and RF), quantification (SVM) and prediction (SVM and ANN) of many parameters.

Once the biophysical parameters have been predicted/classified from the hyperspectral data/images, the challenge is finding out the contribution of each spectral wavelength to the performance of regression or classification algorithm. There are many methods for calculating feature importance based on statistical or machine learning algorithms. These can be classified into model-dependent and model agnostic methods. Model agnostic methods can be applied to any algorithm (e.g. RF, SVM, ANN and kNN etc.), whereas model-dependent methods are specific to a particular algorithm, like the feature ranking method of Random Forest. Few studies have used machine learning algorithms, especially to identify the essential

features. [40]computed a Relief-F value that uses a kNN based scoring and is considered a reliable metric to calculate the feature score followed by top-scoring features.

The machine learning models are usually considered “black boxes” because they do not provide information about how they arrive at the predictions. This is due to their non-linear, complicated, and nested structure. According to[41] and [42]the concept of Explainable artificial intelligence (XAI) is getting very popular recently. XAI helps to visualize, explain and interpret the machine learning models. XAI is being applied in critical fields like biology and finance, where it is necessary to know the details of the models like its working and identification of essential features[43]. These techniques help to build the trust and transparency of the machine learning models [44].

There are many studies in which hyperspectral data have been used to predict the N status of plants, but few studies have used machine learning techniques to estimate the N status of plants. Therefore, a study was carried out to estimate the N status of wheat using machine learning techniques. Specific objectives of the study were (i) Identification of spectra pre-processing techniques and their effect on prediction accuracy of the machine learning regression algorithms, (ii) development of machine learning models and (iii) the use of Explainable artificial intelligence (XAI) methods to explain the best performing machine learning model and identification of important hyperspectral wavelengths associated with N.

4.2. Material and Methods

4.2.1 Study area and Data Acquisition

The secondary hyperspectral data of wheat was downloaded for the Gilat Agricultural Research Center, Israel. The dataset[124] was downloaded from the Ecological Spectral Information System (EcoSIS) website https://ecosis.org/. The proximal canopy measurements were taken in the year 2004 and 2005 using a spectroradiometer (ASD FieldSpec bare fiber). The data was collected throughout the growing season at 6, 45, 63-72 and 91-97 days of wheat growth. The plant tissues were collected and analysed for water content, dry weight, leaf area index (LAI) and

nitrogen content. The spectral range of the data is 400 to 2400 nm with a spectral sampling resolution of 2 nm (330 measurements).

4.2.2 Hyperspectral Data Processing

Based on the data transformation techniques, four input datasets were generated in this study:

i. Raw spectra

ii. 1st Derivative dataset: Savitzky-Golay 1st derivative transformation was applied on the raw spectra. Derivatives are applied to correct the baseline effects in spectra for removing nonchemical effects and creating robust calibration models. Derivatives may also aid in resolving overlapped bands which can provide a better understanding of the data, emphasizing small spectral variations not evident in the raw data.

iii. Deresolved Dataset: Deresolve uses a triangle kernel filter for smoothing to convolve spectra with a resolution function to make it appear as if it had been taken on a lower resolution instrument. The inputs are the high-resolution spectra to be deresolved and the number of channels to convolve them over. The output is the estimate of the lower resolution spectra with the original number of variables maintained. Deresolve is also be used for noise reduction.

iv. Deresolve and 1st Derivative dataset: In this dataset, both the data transformation techniques were applied.

4.2.3 Estimation of N status in plants from hyperspectral data using machine learning analysis

In the computational environment, reflectance values were used as the input parameters (X variables), and N concentration in wheat was selected as the target variable (Y variable). The spectra was pre-processed using first derivative, de-resolve, and both first derivative and de-resolve. Six regression machine learning algorithms (Support Vector Regression (SVR), Random Forest (RF), k-nearest neighbours (kNN), Multilayer Perceptron (MLP) and Gradient Boosting Regression (GBR) and partial least square regression (PLSR)) were used for retrieval of N from hyperspectral data. Cloud-based Google Colaboratory (Colab)[82] environment was

selected for developing and running the algorithms. It allows writing and executing Python[80] code in the browser without installation and configuration. The machine learning algorithms were implemented using the Python language and the Scikit-learn[81] package. Scikit-learn is a popular open-source library that integrates many ML algorithms[125]. The hyperparameter tuning was performed (using the GridSearchCV function) for each regression algorithm to select the parameters which gave the most accurate model in terms of root-mean-square error (RMSE).

The sampling distribution of the training and the testing datasets impacts the machine learning algorithms. Many previous studies by[83] and [84]have advised splitting training and validation datasets into 80% and 20%, respectively. In our study, the dataset was randomly split into 80% and 20%. The performances of machine learning algorithms were evaluated using the coefficient of determination (R2) and mean-squared error (MSE). The methodology is given in Figure-4.1.

Table-4.1: Results of the statistical parameters (Coefficient of determination, R^2 and Mean Square Error, MSE) for the machine learning models

Machine Learning Model	**Raw**		**De-Resolve**		**Ist Derivative**		**De-Resolve + 1st Derivative**	
	R^2	**MSE**	**R^2**	**MSE**	**R^2**	**MSE**	**R^2**	**MSE**
Support Vector Regression	0.72	2.42	0.79	2.36	0.74	3	0.74	3
Random Forest	0.60	4.53	0.59	4.71	0.88	1.38	0.89	1.26
k-nearest neighbours	0.42	6.65	0.42	6.65	0.73	3.04	0.73	3.01
Multilayer Perceptron	0.69	3.52	0.32	7.81	0.07	0.0001	0.2	9.2
Gradient Boosting Regression	0.64	4.13	0.62	4.36	0.87	1.49	0.87	1.39
Partial Least Square Regression	0.73	2.89	0.74	2.9	0.78	2.51	0.79	2.4

Table-4.2: Hyper parameters selected for the machine learning models

Machine Learning Model	Raw	De-Resolve	Ist Derevative	De-Resolve + 1st Derevative
Support Vector Regression	kernel='linear', C=5, gamma=0.1, epsilon=.0001	kernel='linear', C=10, coef0=0, gamma=0.1, epsilon=.0001	kernel='linear', C=100, gamma=0.1, epsilon=.0001	kernel='poly', C=5, gamma=0.1, epsilon=.0001, coef0=10
Random Forest	max_depth=40, max_features='auto', min_samples_leaf=1,min_samples_split=2, n_estimators=200	max_depth=40, max_features='auto', min_samples_leaf=1, min_samples_split=2, n_estimators=400	max_depth=20, max_features='sqrt', min_samples_leaf=1, min_samples_split=5, n_estimators=200	max_depth=20, random_state=0, max_features='sqrt', min_samples_leaf=2, min_samples_split=2, n_estimators=200
k-nearest neighbours	algorithm= 'brute', leaf_size=30, metric='euclidean', n_neighbors=5, weights='distance'	algorithm= 'brute', leaf_size=30, metric='euclidean', n_neighbors=5, weights='distance'	algorithm= 'auto', leaf_size=30, metric='manhattan', n_neighbors=5, weights='distance'	algorithm= 'brute', leaf_size=30, metric='manhattan', n_neighbors=3, weights='distance'
Multilayer Perceptron	activation: 'relu', solver: 'lbfgs', alpha: 0.0001, hidden_layer_sizes: (1000,), 'learning_rate': 'constant',	activation = 'relu', solver='lbfgs', alpha=0.0001, learning_rate='constant', hidden_layer_sizes=(1000,)	activation = 'tanh', solver='adam', alpha=0.05, learning_rate='constant', hidden_layer_sizes=(1000,)	activation = 'relu', solver='adam', alpha=0.0001, random_state=1
Gradient Boosting Regression	learning_rate= 0.1, max_features=0.3, min_samples_leaf=5, n_estimators=1000, max_depth=4	learning_rate= 0.1, max_features=0.5, min_samples_leaf=3, n_estimators=500, max_depth=4	learning_rate= 0.1, max_features=0.3, min_samples_leaf=3,n_estimators=500, max_depth=6	learning_rate= 0.1, max_features=0.3, min_samples_leaf=3, n_estimators=500, max_depth=4
Partial Least Square Regression	copy=True, max_iter=500, n_components=8, scale=True, tol=1e-06	copy=True, max_iter=500, n_components=8, scale=True, tol=1e-06	copy=True, max_iter=500, n_components=8, scale=True, tol=1e-06	copy=True, max_iter=500, n_components=5, scale=True, tol=1e-06

4.2.4 Identification of important features (wavelengths) from hyperspectral data

SHapley Additive explanations (SHAP) value is an essential tool of the Explainable artificial intelligence (XAI) which was proposed by[85]. It is used to interpret and explain any sophisticated machine learning model SHAP values are based on shapley values, a concept originating from the game theory. In the game theory there is a game and players. In the context of machine learning, the game is the outcome of a model, and the players are the features included in the model. SHAP helps to quantify the contribution that each feature (player) brings to the model prediction (game)[86]. It can explain individual predictions and aggregate these predictions to get powerful insights into the whole model. The details of the calculation of SHAP values are explained by [87].

The XAI was implemented in Python using the SHAP library[126]. There are two main advantages of using SHAP values in this study:

a) Global Interpretability: SHAP provides a global interpretation of the whole model. The collective SHAP values can show how much each wavelength (predictor/variable) contributes positively or negatively to the target variable (N status in plants).
b) Local Interpretability: Each observation (or spectra) gets its own set of SHAP values. Using these SHAP values, the contributions of the predictors is explained. It also helps to pinpoint and contract the impact of factors at various growth stages of wheat.

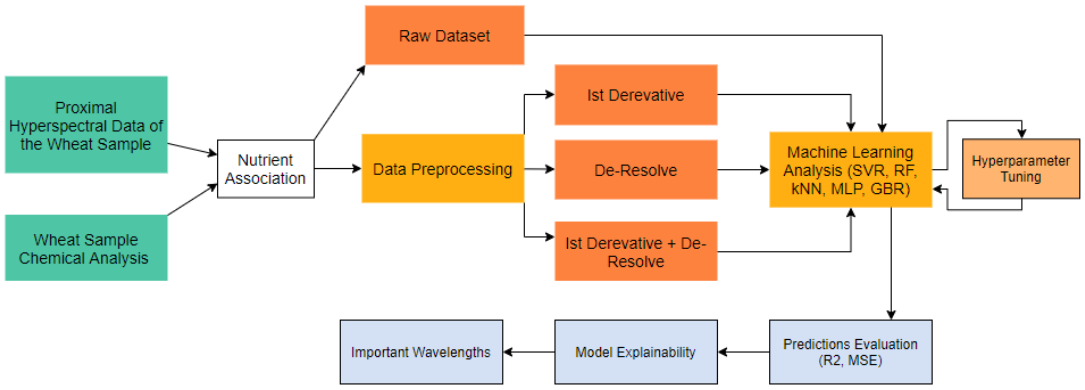


Figure-4.1: Methodology

Table-4.3: Important Wavelengths (associated with N) at various growth stages of wheat crop (Local Interpretability of Random Forest Model).

	Wheat sample				
	1	2	3	4	5
3.	**6 Days**				
	806	806	806	806	806
	794	794	794	804	820
	804	804	804	794	804
	672	672	820	820	794
	820	820	438	796	672
	45 Days				
	806	806	806	806	806
	804	804	794	860	794
	794	862	1422	794	804
	790	796	716	804	790
	796	794	804	820	796
	63- 72 Days				
	1554	1556	806	806	724
	1662	1582	820	724	1182
	1732	2080	804	820	720
	1556	1344	822	862	1272
	1582	1734	816	816	1300
	91 -97 Days				
	724	806	1272	724	1556
	1300	884	724	718	1272
	720	740	1300	716	1736
	1182	942	718	1278	1288
	816	1728	1276	1272	1586

4.3 Results & Discussions

4.3.1 Comparison of spectral pre-processing techniques and machine learning techniques

Using the four pre-processing techniques, six machine learning techniques produced 24 models with different performances for estimating N status in plants. It was found that coefficient of determination (R^2) was the lowest for raw spectra, irrespective of any machine learning algorithm. Among all the 24 models, random forest (RF) model with both De-Resolve plus 1st derivative pre-processing techniques was better than the other models (R^2 = 0.89). The RF model was followed by gradient boosting regression (GBR) model with both De-resolve plus 1st derivative pre-processed for predicting N status of plants (R2 of 0.87) on data. The performance of PLSR with 1st derivative and both De-resolve and 1st derivative pre-processed was almost similar. The R^2 and MSE of all the models is given in Table 4.1. The ML models hyper parameters which gave the best predictions, are given in Table 4.2.

4.3.2 Machine Learning Model explanation

4.3.2.1 Global Interpretability:

The top four machine learning models (RF, GBR, PLSR and SVR) with higher R^2 values were explained using SHAP values. The important wavelengths related to N, lies between 790 and 862 nm for both RF and GBR models, between 658 - 672nm for PLSR model, and between 600-2000 nm for SVR. The SHAP function "TreeExplainer" has been used to explain the RF and GBR models, whereas "KernelExplainer" has been used to explain the PLSR and SVR models.

The global interpretability for RF model showed that the wavelengths predicting N status of the plant were 790,794, 796, 804, 806, 816, 820, 862, 848, 860 nm, irrespective of the growth stage. These wavelengths lie in between reference wavelength (typically between 750 and 900 nm). The reference wavelengths are independent of the stage of leaf development due to negligible chlorophyll absorption, and reflectance at these wavelengths is mostly regulated by light scattering properties of the leaves[104]. For GBR model, the important wavelengths predicting N status of

the plant were 406, 408, 790, 792, 796, 802, 806, 804, 816 and 820 nm. For PLSR model, the bands predicting N status of the plant were 658, 662, 664, 666, 668, 670, 672, 766, 1662, 2258 nm. The selected bands in the PLSR model were primarily within the range of 650–680 nm, adjacent to the Chl absorption peak at 675 nm [109]which concurs with previous studies on spectroscopic estimation of plant N content. In addition to these bands, 1662 and 2258 nm were also selected as the important bands to predict N status of plants. These bands correspond to leaf dry matter content, which is often correlated with the N concentrations[127].

In the SHAP variable importance plot of the random forest model (Figure-4.2), every feature's SHAP values for every sample are plotted to get an overview of the important features. The important features are ranked in descending order (vertically). A wider spread of SHAP values signifies more differentiation in model output, and therefore the feature has higher importance. This plot also shows how each feature tends to influence the model predictions. The red and blue colour of the scale shows the high and low values of a feature, respectively. In the RF model plot, the lower values of the wavelengths (blue) results in increasing the N status of plants. The SHAP variable importance plot for GBR model is shown in Figure 4.3.

4.3.2.2 Local Interpretability:

The random forest model was selected for the local interpretability for identifying the important wavelengths at various growth stages of wheat. It was found that wavelengths 672, 794, 804, 806, 816 and 820 nm were important predicting N status of the plant during the first six days of wheat growth (Table 4.3), 716, 794, 804 and 806 nm after 45 days of wheat growth, 724, 806, 820, 1556 and 1582 after 63-72 days of wheat growth and 718, 720, 724 and 1272 nm at the end of the wheat growth (91 - 97 days). Most of these wavelengths estimated at different growth stages of wheat have been used for predicting the N status of wheat crop.

[107]developed two spectral indices (canopy chlorophyll content index from 670, 720, 790 nm and spectral ratio planer index from 445, 705 and 750 nm) to detect the N stress in crops.[108]used hyperspectral data to identify the nitrogen status of wheat. The genetic algorithm selected relatively few features in the NIR region (>750

nm) for estimating plant N content. Reflectance features were commonly selected between 583 and 722 nm in the two growing seasons, encompassing the primary wavelengths for visible red light absorption by chlorophyll.

The study results suggest that ML can be effectively used to estimate the N status of wheat from the hyperspectral data. The eXplainable Artificial Intelligence (XAI) tools can be efficiently used to explain the complex black box ML models, which will help remote monitoring of N nutrition of wheat crop.

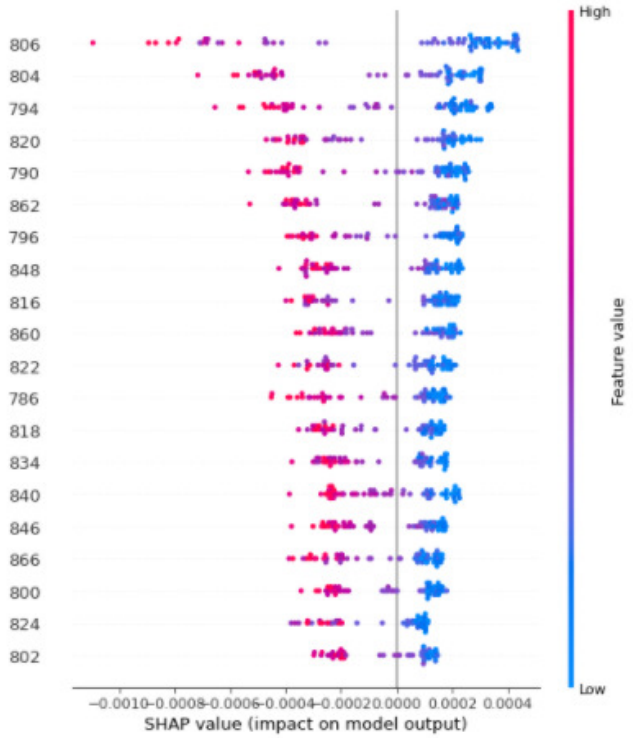


Fig.4.2. SHAP variable importance plot (Random Forest)

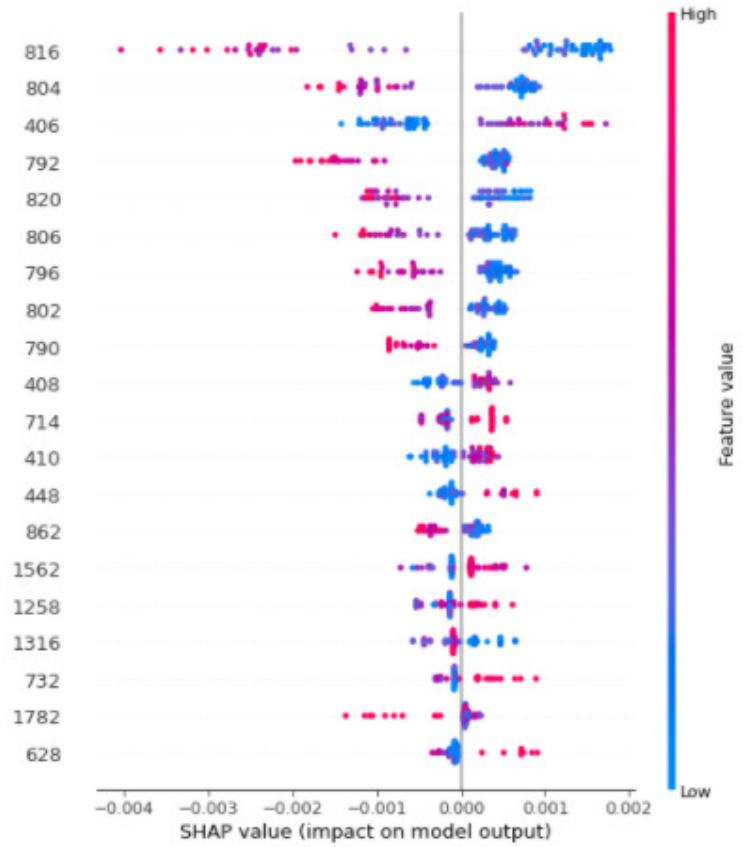


Fig.4.3. SHAP variable importance plot (Gradient Boosting Regression)

Chapter 5

Estimation of chlorophyll, macronutrients and water content in maize from hyperspectral data using machine learning and explainable artificial intelligence techniques

5.1 Introduction

The abiotic stress, mainly water and nutrient stress, is becoming a major threat to food security[72], and 71% of the yield losses in crops are caused by abiotic factors [73]. The effect of abiotic stress on crop growth is usually detected only after it becomes visually apparent; therefore, assessment of yield loss due to water and nutrient stresses is essential for sustainable crop production.

The essential biochemical components of crop organic matter are water content, chlorophyll, nitrogen (N), phosphorus (P), and potassium (K). Chlorophyll helps plants to make their food through photosynthesis. Nitrogen is an essential component of the chlorophyll molecule, which enables the plant to capture sunlight energy by photosynthesis for carbohydrate formation, driving plant growth and grain yield [74]. Phosphorus is required to develop new plant cells and transfer the genetic code from one cell to another [75]. Potassium is an activator of essential enzymes, such as protein synthesis, sugar transport, carbon and nitrogen metabolism, and photosynthesis [76]. These essential macronutrients are required by plants in large amounts, and estimation of these nutrients on a spatio-temporal scale is required for increasing macronutrient use efficiency. Traditionally these parameters are determined from the plant tissue using invasive methods in a laboratory. The laboratory methods are costly and time-consuming. Estimation of these parameters non-invasively using multispectral and hyperspectral remote sensing helps monitor the plants' health [7]. The regular monitoring of crop conditions by field surveys is a resource-intensive task [78]. Proximal and remote sensing sensors provide a solution by collecting information about crop conditions non-invasively.

Visible (VIS, 400–700 nm), near-infrared (NIR, 700–1100 nm), and short-wave infrared (SWIR, 1100–2500 nm) spectroscopy is quite helpful in measuring the leaf biochemical and nutrient properties quickly and non-destructively [9].[118]

discussed the advantages of analyzing hyperspectral data to detect various types of crop stress for leaf, canopy and field levels. [128] studied the effects of N on corn canopies' growth, yield, and reflectance characteristics. They found decreased reflectance in the near-infrared band and increased reflectance in the red band in the N-stressed canopies. [129] studied the estimation of N and chlorophyll content of maize leaves using hyperspectral data from 400 to 1100 nm. The 520-600nm wavebands gave a better estimate of N and chlorophyll concentrations. [130] used the regression analysis to identify the sensitive wavelengths for detecting N deficiencies in a corn canopy using a spectroradiometer in the range of 350 -1100 nm. They found that the radiations near 550 and 710 nm were superior for detecting N deficiencies. [99] found that the red and green regions were best suited for predicting N stress, NIR and blue regions for early season P stress in maize.

[131] found the 11 wavelengths (609,647,651,654,669,675,676,680,721,727 and 760 nm) for estimating N, two wavelengths (675 and 680 nm) for P and 15 wavelength (410, 411, 417, 422, 460, 463, 468, 646, 651, 658, 669, 670, 674, 676 and 682 nm) for K in corn leaves. [89] found that 700 nm, 740 nm and 860 nm bands were related to leaf water content and 540 nm, 780 nm and 860 nm with leaf N in maize.

The statistical and machine learning algorithms have been extensively used to analyze the crop spectra and estimate various elements [8]. The machine learning algorithms are more popular as they give better results than statistical methods as non-linear relationships are better predicted with machine learning approaches [79]. Though the machine learning algorithms have a better performance, they are challenging to interpret due to their non-linear, nested, and complicated structure. Explainable artificial intelligence (XAI) techniques are becoming very attractive as they facilitate explaining, visualizing, and interpreting complex ‘black box’ machine learning models [43]. The explanations and interpretations help build transparency and trust in the machine learning models.

Many studies have used machine learning techniques to estimate the biochemical, N, P and K content in plants from hyperspectral images [132], [133][134][135], but there are very few studies that have applied machine learning along with XAI techniques on hyperspectral data to estimate the nutrient content in plants. Therefore, a study was conducted to estimate water content, chlorophyll, N, P

and K content in maize using machine learning and XAI techniques. Specific objectives of the study were (i) to develop the machine learning models for estimation of water content, chlorophyll, N, P and K content in maize using hyperspectral data, (ii) the use of XAI methods to identify the optimum wavelengths for estimation of each parameter.

5.2. Material and Methods

5.2.1 Study area and Data Acquisition

The secondary hyperspectral dataset of maize crop in this study was the output of a field experiment conducted on Havelock Research Farm and Greenhouse Innovation Centre of the University of Nebraska-Lincoln (U.S.). This dataset [136] was downloaded from the Ecological Spectral Information System (EcoSIS) website https://ecosis.org/. Maize association panel was grown in the Havelock Research Farm in 2018 and 2019 under low nitrogen conditions (–N) and high N conditions (+N). In addition, the panel was also grown in the Greenhouse Innovation Center with optimum conditions in 2018. Hyperspectral reflectance was measured at the leaf level. Ground truth data were collected for fresh and dry leaf weight, chlorophyll, nitrogen (N), phosphorus (P), potassium (K) and leaf area. The spectral range of the data is 350 to 2500 nm with a spectral sampling resolution of 1 nm (1210 measurements). The details of this experiment are given in [137].

5.2.2 Hyperspectral Data Processing

The raw data was pre-processed with two data transformation techniques: Savitzky-Golay 1st derivative and Deresolve. Savitzky-Golay's 1st derivative transformation removes the nonchemical effects and creates robust calibration models by correcting the baseline effects in the spectra. Derivatives may also help resolve the overlapped bands, and Deresolve is generally used for noise reduction. It uses a kernel filter to smooth the spectra with a resolution function, appearing as if it had been taken on a lower resolution instrument.

5.2.3 Estimation of the status of various parameters of plants from hyperspectral data using machine learning techniques.

[137] reported the results of the dataset used in this study by employing partial least squares regression and support vector regression to estimate the leaf properties from hyperspectral data. In our study, the chemical properties of maize leaves were derived from hyperspectral data using six machine learning regression algorithms: Random Forest (RF), Support Vector Regression (SVR), k-nearest neighbours (kNN), Multilayer Perceptron (MLP), Gradient Boosting Regression (GBR) and Partial Least Square Regression (PLSR). In all these algorithms, reflectance values were used as the input parameters (X variables), and water content, chlorophyll, N, P and K concentration in maize were selected as the target variable (Y variable). The machine learning models were built for all the chemical parameters in low nitrogen, high nitrogen and glasshouse regimes. The machine learning algorithms were implemented in Python [80] programming language using the Scikit-learn (Developers n.d.) library. Scikit-learn is a standard, open-source library that integrates many ML algorithms (Pedregosa et al., 2011). Google's Colaboratory (Google Research n.d.) cloud-based environment was selected for developing and running the machine learning algorithms. Colaboratory allows writing and executing Python code in a simple browser without installation and configuration. GridSearchCV function was employed for hyperparameter tuning. This process helped in the selection of the best parameters for each regression algorithm.

The performance of the machine learning algorithms is also influenced by the sampling distribution of the training and the testing datasets. Verrelst et al. (2012) and Verrelst et al. (2011) have recommended the 80% and 20% split for the training and validation datasets. The dataset was also randomly split into 80% and 20%. The performances of machine learning algorithms were evaluated using the coefficient of determination (R2) and mean-squared error (MSE). An overview of the methodology is given in Figure-5.1.

5.2.4 Identification of important features (wavelengths) from hyperspectral data

Machine learning models are usually considered "black boxes"; therefore, it is challenging to interpret them and get an inside view of how the model resulted in a particular output. [137] predicted the leaf chemical parameters from hyperspectral used in our study by employing partial least squares regression and support vector

regression, but they didn't identify the optimum wavelengths to estimate the chemical parameters of leaves. However, they used the Green Normalized Difference Vegetation Index (GNDVI), Red-edge Normalized Difference Vegetation Index (RENDVI, and Normalized Difference Water Index (NDWI) to quantify the chemical traits of leaves. In our study, optimum wavelengths to quantify the N, P, K, chlorophyll and water were identifying using the Explainable artificial intelligence (XAI) tools which may be helpful to interpret and explain the complex AI models. Lundberg and Lee (2017) proposed the SHapley Additive explanations (SHAP) value, a critical XAI tool, for interpreting the machine learning models. SHAP values are based on shapley values that originated from the mathematical game theory. In the context of machine learning, the game is the outcome of a model, and the players are the features included in the model. SHAP helps to quantify the contribution that each feature (player) brings to the model prediction (game) (Mazzanti 2020). It breaks down the prediction to show the impact of each feature. It can explain individual predictions and aggregate them to get powerful insights into the whole model. SHAP values calculations are explained in detail by [87]. We implemented the XAI in Python using the SHAP library (Lundberg n.d.). The main advantage of using SHAP values for this study is the Global Interpretability of the whole model. Global interpretation of the collective SHAP values helps identify the important wavelengths associated with the target variables.

5.3. Results & Discussions

5.3.1 Comparison of machine learning techniques across three nitrogen regimes

[137]estimated the water content, chlorophyll, N, P and K in maize leaves under the three regimes of N (Green House, Low and High N) using SVM and PLSR. Since we used six machine learning algorithms, therefore these parameters were estimated using 90 models (30 for each N regime, 5 chemical parameters x 6 machine learning algorithms) with different performances. Among the 90 models, RF and PLSR performed better than the other models for estimating N, P, K and chlorophyll content across the three N regimes. The coefficient of determination (R^2) and mean square error (MSE) of all the models are given in Table 5.1.

The water content in leaves was better predicted using SVR under high N status. However, the performance of the algorithms under high N status were in the order: SVR>RF>PLSR>GBR. The RF algorithm performed better than the other models under the greenhouse and low N conditions. The MLP algorithm was unsuitable for estimating water content under greenhouse, low and high N conditions. For chlorophyll content, the three algorithms (RF, SVR and PLSR) performed better than the other three algorithms. However, chlorophyll content was estimated using PLSR and RF algorithms with good accuracy under high N and greenhouse conditions (Table 5.1).

Among greenhouse, low N and high N status, nitrogen was best predicted for low N status and the performance of the algorithms were in the order: PLSR>RF>GBR. These three models also predicted N content in leaves under high N and greenhouse samples with similar performance order, except GBR, which performed better than RF under greenhouse conditions. The coefficient of determination was less than 0.30 for all the models to estimate P concentration in leaves. Among the three conditions, R^2 was 0.29 to estimate P content in leaves under greenhouse conditions (Table 5.1). The algorithms' performance was in the order PLSR>GBR>RF. For low N status, the PLSR model was better than the remaining algorithms, but RF and PLSR performed better than the other models under high N status. Potassium content in the leaves under high N was better predicted using the PLSR and RF, followed by GBR. However, under low N status, the algorithms' performance was in the order RF>PLSR>GBR. The GBR was better than RF and PLSR under greenhouse conditions, and MLP and SVR were unsuitable for this dataset.

The partial least-square(PLSR) and support vector machine (SVM) regression methods were compared by [138]to predict N, P and K in crops, and they found that SVM achieved better accuracies than PLSR in estimating the nutrient concentration. A gradient-boosting machine model was developed by [139] to estimate N from the hyperspectral images. The red-edge bands (700–725 nm) helped estimate N in maize. [137] found that PLSR and SVR were the best algorithms for estimating chlorophyll, N, P and K in maize leaves. In their study, N and chlorophyll were estimated

accurately, K and leaf water content with moderate accuracy, and P with low accuracy.

5.3.2 Machine learning model explanations

Out of the 90 machine learning models developed in this study, the two models with higher R^2 values for all the five parameters under greenhouse, low and high N status were explained globally using the SHAP values to identify the important wavelengths associated with the crop parameters. The three most important wavelengths associated with the crop parameters are given in Table 5.2. The SHAP function "TreeExplainer" was used to explain the RF and GBR models, whereas "KernelExplainer" was used to explain the PLSR, kNN, MLP and SVR models.

Under greenhouse conditions, important wavelengths for estimating water content in maize leaves are 1688, 2200 and 2202 nm using the RF model. However, the important wavelengths were 1979, 1982 and 1978 (SWIR region) under low N status using RF and GBR algorithms, and 1894, 1967 and 1894 nm under high N status using RF algorithm. The SVR and kNN algorithms showed similar wavelengths in the violet region under high N and greenhouse conditions. Since the average wavelength was 0.07% in the violet region (Fig. 5.2), therefore this band is not useful for detecting water stress in plants. It is well known that water absorption by leaf occurs in the SWIR range; thus, reflectance in the SWIR band is negatively correlated with leaf water content. According to [110] also, short-wave infrared is associated with the leaf water content and biochemicals. The study by [140] suggests that the IR and SWIR bands help estimate leaf water content.

For the low N treatments, the important wavelengths for estimating chlorophyll in maize leaves were 741, 750 and 751 nm. This red edge region indicates the sharp increase in reflectance from the VIS to NIR regions associated with strong chlorophyll absorptions and internal leaf structure. Using the SVR model, the important wavelengths were 367, 363 and 378 nm, which correspond to the violet region of the visible spectrum. The important wavelengths identified using the SVR model were not important as the reflectance was 0.06% in the violet region. These bands are not important for detecting chlorophyll in plants. However, with the RF

model, the average reflectance was 0.43% in this red edge/NIR region, which is helpful in determining chlorophyll in plants. Under greenhouse and high N status, the important wavelengths were in the red edge region using the RF algorithm, but wavelengths identified using PLSR were 539, 540, 741, 745 and 747 nm. The green peak (central wavelength of 550 nm), as a representation of chlorophyll content, has demonstrated importance in classifying different crop species. Additionally, absorption in wavebands within the green region adjacent to the reflectance peak is associated with xanthophylls and anthocyanins. Anthocyanins can be estimated using anthocyanin's absorption maximum near 550 nm and a band from the red edge, usually 700 nm. According to [111], chlorophyll and N contents can be accurately remotely estimated using green and red-edge chlorophyll indices using near-infrared (780–800 nm) and either green (540–560 nm) or red-edge (730–750 nm) spectral bands. [94] also asserted that the red and red-edge bands are useful for the estimation of chlorophyll in maize.

For the low N treatments, the optimum wavelengths for estimating N in maize leaves were 528, 527 and 584 nm with the RF model, and all these wavelengths correspond to the green region of the visible spectrum. The reflectance was 0.16% in these regions. The plants with N stress are subjected to degradation of active pigments with an increase in reflectance in the visible region. However, with the PLSR model, the wavelengths were mainly in the SWIR region (2145, 2139 and 2136 nm), and the corresponding reflectance was 0.14% in this region. Under greenhouse conditions, the wavelengths were in the SWIR region with GBR and PLSR algorithms. These bands correspond to leaf dry matter content, which is often correlated with N concentration in plants. The important wavelengths under high N status were 772, 773 and 775 nm using the RF model and 2218, 2151, and 2158 nm using the PLSR model. The important wavelength in the NIR region are related to the cellular structure of maize leaves. Higher reflectance due to high N in the NIR region is related to the spreading of radiation from this spectrum region due to spongy mesophiles and a higher volume of vacant spaces. Canopy nitrogen and chlorophyll content of maize were closely related [111]. Green and red-edge bands are useful for the estimation of N in maize.

Results from a study by [141] reveal a firm advantage for the SWIR-based indices in their ability to predict and in their sensitivity to N content in potatoes.

The important wavelengths for estimation of P in maize leaves were 1088, 1262 and 1263 nm under low N status, 2217, 2158, 2159 nm under high N status and 2153, 2154 and 2166 nm under greenhouse conditions using the PLSR model. These results show that NIR and SWIR region is useful for detecting P deficiency in plants.There is a close relationship between P and reflectance in the NIR region because P plays a critical role as an energy supplier in energy-consuming processes such as photosynthesis. In combination with reflectance in the visible or NIR regions, reflectance in the SWIR range was found to monitor nitrogen, phosphorus, sulphur and potassium status in plants [100].

The important wavelengths of detecting K deficiency in maize leaves were 2239, 2238, 2239 and 2260 nm (SWIR region) under low N status with the RF algorithm. Similar wavelengths in the SWIR regions were also found for high N and greenhouse conditions. Potassium in plants is responsible for activating several enzymes in plants that affect starch and protein content. The SWIR band identified for K deficiency in plants corresponds to starch and protein absorption features. [20] developed a vegetation index using wavelengths of 1645 and 1715 nm (SWIR) to determine wheat's potassium concentration. VIS–NIR–SWIR leaf reflectance can be a valuable tool for low-cost, non-destructive, high-throughput investigation of leaf physiological and biochemical characteristics (including N, P and K), according to [137].

5.4. Conclusions

This suggests that machine learning algorithms coupled with eXplainable Artificial Intelligence (XAI) tools can be effectively used to estimate the status of various crop parameters from the hyperspectral data and explain the complex black-box machine learning models. Among various machine learning techniques applied on the low nitrogen dataset (N stress), the water content may be better predicted with RF, chlorophyll with SVR, N with PLSR, P and K with RF algorithm. It was found that the SWIR band was helpful in the estimation of water content; red & IR for chlorophyll; red, IR & SWIR for nitrogen; NIR & SWIR for P and SWIR for K. The

hyperspectral data can be used to differentiate nutrient deficiency in plants, and the proposed framework of machine learning and XAI tools have the potential to apply in different agricultural management regions on various spatial scales (within a field/region).

Table-5.1: Results of the statistical parameters for water content, chlorophyll and macronutrients estimated using the six machine learning regression models

Green House Dataset										
Machine Learning Algorithm	**Water Content**		**Chlorophyll**		**Nitrogen**		**Phosphorus**		**Potassium**	
	R2	**MSE**	**R2**	**MSE**	**R2**	**MSE**	**R2**	**MSE**	**R2**	**MSE**
RF	0.12	117	0.86	1172	0.44	0.175	0.17	0.01	0.18	0.26
SVR	-0.39	186	0.85	1216	0.2	0.25	0.003	0.015	-0.02	0.33
kNN	0.12	116	0.65	2890	0.17	0.26	0.12	0.013	0.005	0.328
MLP	-0.06	141	0.84	1355	-0.009	0.32	-0.04	0.01	-0.02	0.33
GBR	0.03	128	0.84	1339	0.5	0.158	0.24	0.011	0.23	0.25
PLSR	0.09	120	0.86	1134	0.51	0.15	0.29	0.01	0.17	0.27

Low Nitrogen Dataset										
Machine Learning Algorithm	**Water Content**		**Chlorophyll**		**Nitrogen**		**Phosphorus**		**Potassium**	
	R2	**MSE**	**R2**	**MSE**	**R2**	**MSE**	**R2**	**MSE**	**R2**	**MSE**
RF	0.21	22.6	0.95	287	0.74	0.067	0.12	0.003	0.33	0.09
SVR	0.05	27.4	0.96	245	0.68	0.08	-0.25	0.004	0.06	0.13
kNN	-0.03	29.9	0.87	807	0.64	0.09	-0.05	0.004	0.2	0.11
MLP	-0.09	31.6	0.94	388	0.66	0.27	-0.03	0.004	0.21	0.11
GBR	0.19	23.2	0.95	320	0.73	0.07	-0.07	0.004	0.29	0.1
PLSR	0.19	23.3	0.95	301	0.78	0.05	0.21	0.003	0.32	0.09
High Nitrogen Dataset										

Machine Learning Algorithm	Water Content		Chlorophyll		Nitrogen		Phosphorus		Potassium	
	R2	**MSE**	**R2**	**MSE**	**R2**	**MSE**	**R2**	**MSE**	**R2**	**MSE**
RF	0.38	41.4	0.74	1291	0.64	0.07	0.21	0.003	0.34	0.13
SVR	0.41	39.2	0.74	1301	0.35	0.13	-0.07	0.004	0.05	0.19
kNN	0.24	50.7	0.55	2261	0.4	0.12	0.1	0.003	0.2	0.15
MLP	0.01	66.2	0.67	1644	0.36	0.13	-0.0001	0.004	0.03	0.19
GBR	0.32	45.2	0.73	1324	0.62	0.07	0.15	0.003	0.3	0.14
PLSR	0.36	42.6	0.79	1020	0.67	0.06	0.2	0.003	0.34	0.13

RF : Random Forest, SVR: Support Vector Regression, kNN: k-nearest neighbours, MLP: Multilayer Perceptron, GBR: Gradient Boosting Regression, PLSR: Partial Least Square Regression

Table-5.2: Optimum three important wavelengths estimated for different crop parameters using explainable artificial intelligence for the six machine learning regression models

Green House Dataset					
Machine Learning Algorithm	**Water Content**	**Chlorophyll**	**Nitrogen**	**Phosphorus**	**Potassium**
RF	1688, 2202, 2200	743, 745, 742			1844, 2251, 2247
kNN	355, 359, 351				
GBR			2139, 2137, 539	1833, 1846, 623	2251, 2241, 1844
PLSR		540, 539, 745	2163, 2164, 2165	2154, 2153, 2166	

Low Nitrogen Dataset					
Machine Learning Algorithm	**Water Content**	**Chlorophyll**	**Nitrogen**	**Phosphorus**	**Potassium**
RF	1979, 1982, 1978	750, 751, 741	528, 527, 584	1183, 1443, 1444	2239, 2238, 2260
SVR		367, 363, 378			
GBR	1983, 2261, 1657				
PLSR			2145, 2139, 2136	1262, 1263, 1088	2249, 2255, 2253

High Nitrogen Dataset					
Machine Learning Algorithm	**Water Content**	**Chlorophyll**	**Nitrogen**	**Phosphorus**	**Potassium**
RF	1967, 1894, 1978	755, 741, 750	773, 772, 775	1440, 1458, 1443	2247, 2249, 2248
SVR	364, 369, 361				
PLSR		540, 751, 747	2218, 2151, 2158	2217, 2158, 2159	2251, 2249, 2254

(RF : Random Forest, SVR: Support Vector Regression, kNN: k-nearest neighbours, MLP: Multilayer Perceptron, GBR: Gradient Boosting Regression, PLSR: Partial Least Square Regression)

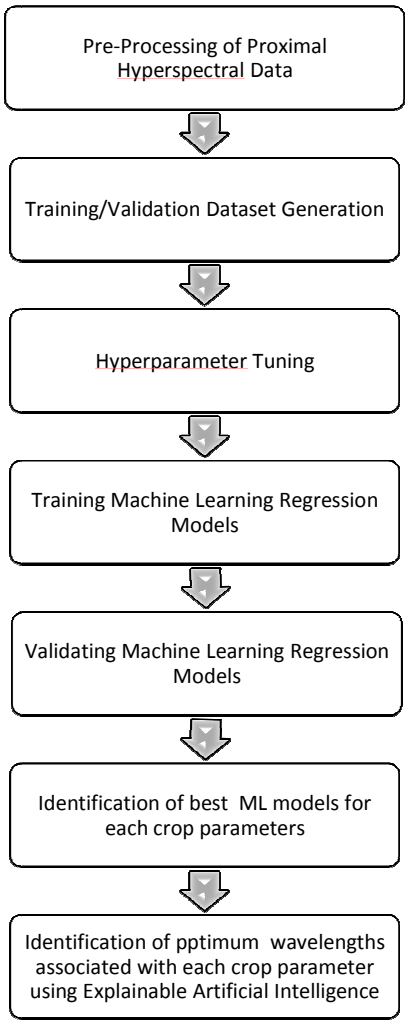


Figure5.1: An overview of the approach

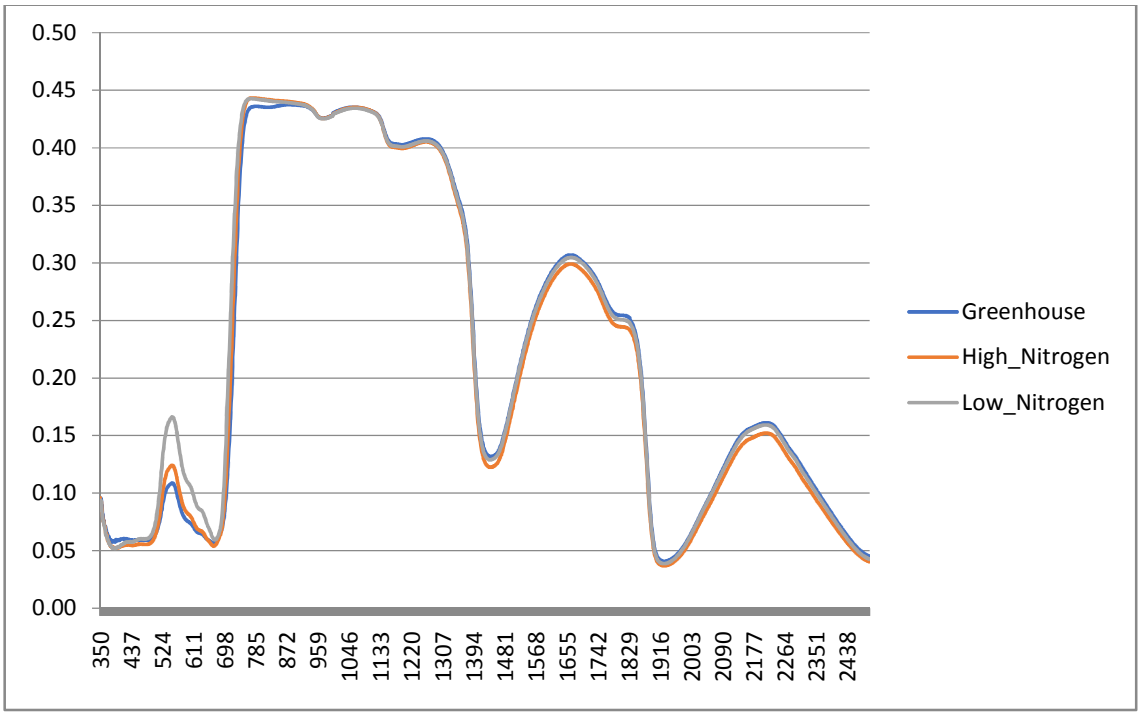


Figure-5.2: Average spectral reflectance of maize leaves under greenhouse, low and high nitrogen datasets

Chapter 6

Classification of yellow rust of wheat from Sentinel-2 satellite imagery using deep learning artificial neural network

6.1. Introduction

Wheat, an important cereal crop of Punjab state, is vulnerable to the attack of several diseases. Among these, yellow rust is major biotic stress which is developed during cool climate, rain, dew, fog, and favourable wind conditions [22]. The main symptom of the disease is the appearance of stripes of yellow dust on the leaves. It reduces the green leaf area, followed by a decline in the crop yield [142]. There are reports that crop yield losses due to fungal infection can be significant as it is one of the most destructive of the various rust diseases in wheat crop [143]. The northern Indian region (including Punjab and Himachal Pradesh) are in severe danger of this disease during the winter season when the climate is constructive for it [144].

A variety of techniques, sensors, datasets, and algorithms have been used for the timely identification of this disease. With the advancements in the field of earth observation, new-age satellite data have better spatial, spectral, and radiometric resolutions than their predecessors. There are studies in which multi and hyperspectral remote and proximal sensing datasets have been used (in addition to meteorological data) for identification of the yellow rust of wheat, but proximal hyperspectral datasets are more popular.

The field spectroradiometer was used to measure the leaf spectra by [23], [24], and the continuous wavelet analysis was applied on the hyperspectral data to generate the rust signatures. [25] also used the hyperspectral data collected from the field and found that normalized photochemical reflectance index (NPRI) can be used to identify rust. A spectrograph mounted on a spray broom level was used by [26] to capture in-field spectral images (wavelength between 463 and 895 nm). Neural networks were used to develop disease detection algorithms, which classified diseased and non-diseased crops. [27] have also utilized the hyperspectral in-situ data collected in the spectral range of 350 to 2500nm. They developed models using partial least squares (PLS) and multiple linear regression (MLR) to assess the severity of yellow rust in the wheat crop. Both the models yielded a coefficient of determination (R2) of 0.96 and

0.89, respectively. [28] evaluated ten common narrow-band spectral indices for the identification of rusts from individual wheat leaves. These indices were based on the in-situ spectrometer measurements in the visible and the near-infrared regions of the electromagnetic spectrum. The yellow rust infected crop produced a strong response to all the indices. Using spectroradiometer data [29] created two spectral disease indices to identify wheat leaf rust. These indices were calculated using reflectance at wavelengths of 605, 695, and 455 nm. The R2 between the estimated and observed values was as high as 0.94 in both indices. These studies show that the accuracy (classification of yellow rust) from proximal sensing methods was higher than remote sensing methods but it requires a lot of resources and time to collect the spectra of yellow rust. Although hyperspectral data is popular and provides accurate results, it requires a lot of resources (time, computing and equipment) to collect, process, and analyze such datasets. There are very few studies in which multispectral remote sensing data have been used for yellow rust identification.

Due to its high spatial, spectral, and temporal resolution, freely available Sentinel-2 satellite data has made revolutions in the field of agriculture. [30] developed the yellow rust forecasting system for the Gurdaspur and Nawanshehar districts of Punjab using weather and land use information extracted from IRS-P6 AWiFS satellite data (spatial resolution = 56 m) satellite data. In Hebei Province, China, the Red Edge Disease Stress Index (REDSI) [31], was utilised to detect yellow rust infection at various severity levels. Sentinel-2 multispectral bands were simulated using the canopy hyperspectral data. [32] have also proposed a model for yellow rust monitoring based on Sentinel-2 multispectral images and a series of two-stage vegetation indices and meteorological data. Though few studies have used multispectral data to identify disease, there are very few studies in which Sentinel-2 data has been used to identify yellow rust using machine learning techniques.

Artificial intelligence (AI) algorithms have transformed the research on crop stress identification. Deep learning algorithms like artificial neural networks (ANN), convolutional neural networks (CNN), and recurrent neural networks (RNN) are gaining popularity due to their accurate results. [39] have reviewed the use of deep learning of images for plant stress phenotyping, and they concluded that deep learning

algorithms require less data pre-processing before the modelling work. These algorithms are faster, give better performance, and are more reliable. Similarly, [33] also asserted that deep learning has a great potential for increased accuracy to identify crop disease from remote and proximal sensor data. [34] has applied the convolutional deep learning algorithm 'U-Net' on the multispectral images acquired from an unmanned aerial vehicle (UAV) to monitor wheat rust. [35] developed a similar approach to identify the rust, but they have acquired hyperspectral images from a UAV. A new deep convolutional neural network (DCNN)-based technique for automated crop disease diagnosis has been suggested. The overall accuracy of their model was 0.85. [36] conducted field experiments and utilized MODIS satellite data along with deep learning algorithms (like ANN, CNN, and recurrent neural networks) to monitor the wheat fungus. Automatically learned features were used for the model development. Most of the previous studies have used deep learning CNN algorithms for yellow rust identification, but there is no study in which Sentinel-2 images have been analyzed for yellow rust of wheat using deep learning ANN's. Therefore, a study was planned to identify the yellow rust of wheat in the parts of Indian Punjab using Sentinel-2 and deep learning ANN's. Various spectral indices were used to build a model which is a classifier differentiating between the yellow rust infected and the non-infected crop data points using free and open-source data, software, and programming tools.

6.2. Materials and Methods

6.2.1. Study Area

The villages in the northeast zone (Rupnagar district) and Central zone (Jalandhar and Kapurthala districts) of Indian Punjab were selected in this study (Table 6.1 and Figure 6.1). There are reports that the wheat crop in the villages of these districts is prone to yellow rust. In these districts, south-western monsoon showers constitute about 73 per cent of the annual rainfall. The monsoon begins in the first week of July and extends up to mid-September. The hottest and coldest months in the study area are June and January, respectively.

Rupnagar district is part of the Kandi belt of Punjab and lies between latitudes of 30° 44' & 31° 26' N and longitudes of 76° 17' & 76° 44' E covering an area of

1376 sq. km. It is situated on the eastern side of Punjab and shares its borders with Himachal Pradesh and Haryana in the north and northeast, respectively. Jalandhar district is situated in the central part of Punjab. It lies between latitudes of 30° 58' & 31° 37' N and longitudes of 75° 04' & 75° 58' E. Kapurthala district lies between the latitudes of 31° 07' & 31 39' N and the longitudes of 74° 58' & 75 55' E. It is bounded by the Beas River in the North West and Jalandhar and Firozepur districts in the South and southwest, respectively.

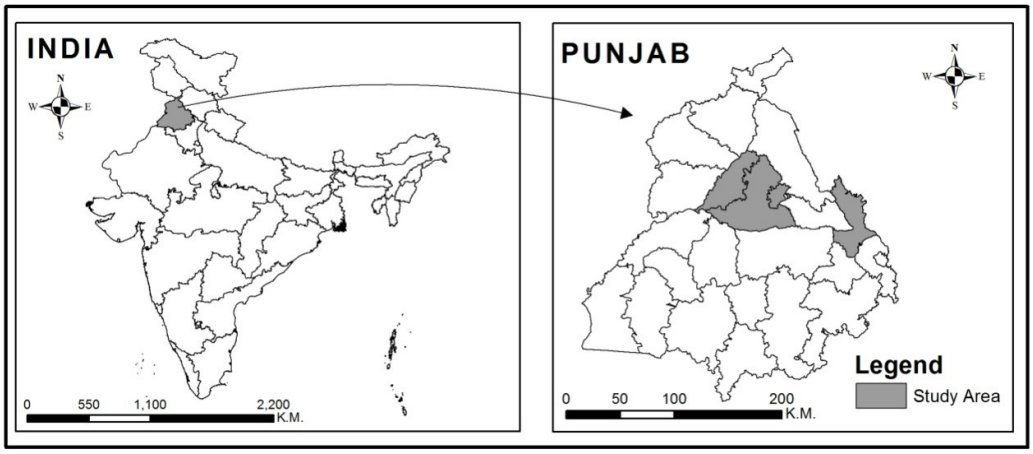


Figure 6.1: Study Area

Table 6.1: Villages in the study area

SNo	District	Villages	Yellow Rust Infestation Status	Date of Satellite Pass
1	Rupnagar	Nikku Nangal, Dokli and Patti	Yes	20-Jan-2020
2	Rupnagar	Takhtgarh, Bajroor, Bahman Majra, Asalatpur, Lahrian, Lalpur, Chanauli, Bhatauli, Kheri, Gopalpur, Dumewal	No	20-Jan-2020
3	Jalandhar	Mansurpur, Badala	Yes	07-Feb-2020
4	Kapurthala	Maqsoodpur, Raipur Pirbax Wala, Ibrahimpur, Nangal Labana, Fatehpur	No	07-Feb-2020

6.2.2 Satellite Data

Severe yellow rust of wheat was reported in the Rupnagar region [145] on 29th January 2020, but it was on 9th February 2020 in the Jalandhar district [146]. The normalized difference vegetation index (NDVI) time series profile of wheat crop in the villages (Table 1) was assessed from December 2019 to March 2020 using the Sentinel-2 data in the Sentinel Hub EO Browser portal (https://apps.sentinel-hub.com/eo-browser). It was observed that there was a dip in the NDVI values around the same period. A ground truth survey was also carried out to confirm the disease infestation. The Sentinel-2 satellite data (Level-2A product) of January and February 2020 was downloaded from the "Copernicus Open Access Hub: https://scihub.copernicus.eu/" portal (Table 6.2). The datum and projection of the data are WGS 84 and UTM, respectively. The methodology or the process flow of the research work is given in Figure 6.2.

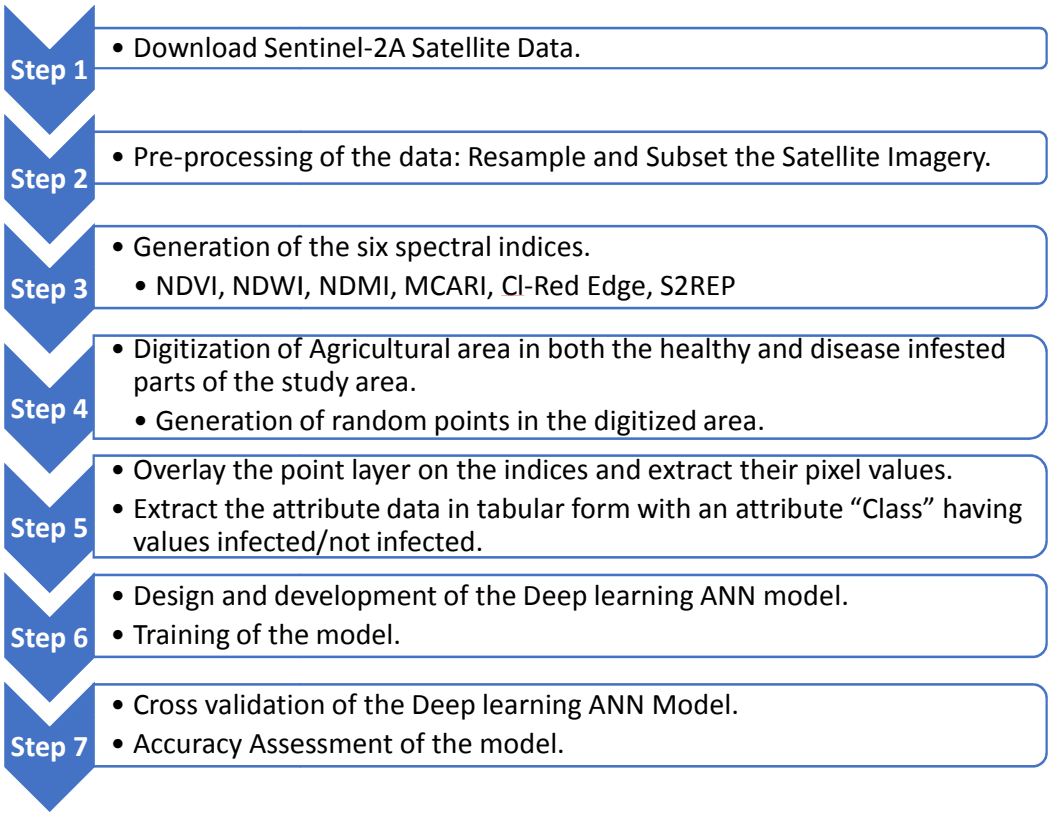


Figure 6.2: Process Flow of the Research work

Table 6.2: Characteristics of the Sentinel-2 satellite data

Band Number	Sentinel-2 MSI Bands	Spatial Resolution (m)	Central Wavelength (nm)	Band Width (nm)
B1	Band 1: Coastal Aerosol	60	443	20
B2	Band 2: Blue	10	490	65
B3	Band 3: Green	10	560	35
B4	Band 4: Red	10	665	30
B5	Band 5: Red-edge 1	20	705	15
B6	Band 6: Red-edge 2	20	740	15
B7	Band 7: Red-edge 3	20	783	20
B8	Band 8: NIR	10	842	115
B8a	Band 8a: NIR narrow	20	865	20
B9	Band 9: Water Vapor	60	945	20
B10	Band 10: SWIR Cirrus	60	1375	30
B11	Band 11: SWIR	20	1610	90
B12	Band 12: SWIR	20	2190	180

6.2.3. Pre-processing of Satellite Data

All thirteen Sentinel-2 bands were resampled to 10 m using the SNAP 7.0 (Sentinel application programme) tool before the data was analysed. Resampling was done to combine all the bands of Sentinel in a single raster stack with uniform pixel size. The data was subset to the geographic extent of the study area. Quantum GIS 2.14.3 (QGIS) was used to analyze the composite dataset, which was converted to GeoTIFF format.

6.2.4. Generation of spectral indices to extract crop data

The six spectral indices were generated using the "Raster Calculator" tool in QGIS: NDVI, MCARI, Cl-Red Edge, S2REP, NDWI and NDMI, and the details of these indices are given in Table 6.3.

Normalized difference vegetation index (NDVI) is used for the assessment of crop or plant health. Its value ranged between -1 and 1. Higher values of NDVI indicate healthy and dense plants [147]. Modified chlorophyll absorption in reflectance Index (MCARI) helps in enhancing the relative abundance of chlorophyll in plants. It is sensitive to the variations in the chlorophyll concentrations and leaf area index. It is also not affected by the background reflectance of the soil and non-photosynthetic materials [148]. Chlorophyll red edge (Cl-Red Edge) is also an indicator of vegetation growth [149]. Sentinel-2 red-edge position (S2REP) is useful to quantify chlorophyll contents [150]. Normalized difference water index (NDWI) reflects the moisture status in crops and soil. Higher NDWI values correspond to high plant water content and coating of high plant fraction [151]. Normalized difference moisture index (NDMI) is used to determine the vegetation water content or the crop's water stress levels [152].

Table 6.3: Spectral Indices used in the study for discrimination of yellow rust of wheat

SNo	Index	Details	General Formula	Formula using Sentinel-2 bands
1	NDVI	Normalized difference vegetation index	(NIR - RED) / (NIR + RED)	(B08 - B04) / (B08 + B04)
2	MCARI	Modified Chlorophyll Absorption in Reflectance Index	((700nm - 670nm) - 0.2 * (700nm - 550nm)) * (700nm /670nm)	((B05 - B04) - 0.2 * (B05 - B03)) * (B05 / B04)
3	Cl-Red Edge	Chlorophyll red-edge	([760:800][690:720])pow (−1)	$(B7/B5)^{(-1)}$
4	S2REP	Sentinel-2 red-edge position	705 + 35 * ((((NIR + R)/2) − RE1)/(RE2 − RE1))	705 + 35 * ((((B08 + B4)/2) − B5)/(B6 − B5))
5	NDMI	Normalized Difference Moisture Index	(820nm - 1600nm) / (820nm + 1600nm)	(B08 - B11) / (B08 + B11)
6	NDWI	Normalized Difference Water Index	(NIR - SWIR) / (NIR + SWIR)	(B08 - B011)/(B08 + B011)

6.2.5. Preparation of the data to be fed to the ANN

The steps by which the data was prepared for input to the ANN are given below:

1) The cropped area was classified using Sentinel-2 data for the four study locations using QGIS, and 4000 points were randomly generated in each of the four digitized cropped areas (infected and non-infected regions of both the study areas).
2) The points were overlaid on the spectral indices. The pixel values for all the indices were extracted using the vector points. Finally, four vector datasets (shapefiles) were generated, which contained seven attributes each for every index pixel value. The Point Sampling Tool, a QGIS plug-in, was used for this work.
3) The attributes of the point layer generated in the previous step were extracted in a tabular form. A new attribute, "Class", was added, and the class had the values 1 (infected) or 0 (non-infected). The data of infected and non-infected

areas were appended in a single file for 14 villages in Rupnagar and seven in Jalandhar and Kapurthala districts, respectively.

4) The final experimental data were categorized into training and test datasets in the ratio of 80:20. The training dataset was used to train the neural network, learn from the input data and identify the biases and weights. The test dataset was used to measure the performance of the network by classifying unseen data. It can also help in the calibrations to prevent the overtraining of the network.
5) Finally, data scaling was performed to remove the effects of the attributes having large values.

6.2.6. ANN model design and development

A Backpropagation neural network (Figure 6.3) was designed for this study. In this network, the input layer has six input neurons. Each neuron corresponds to the six index attributes generated from satellite imagery. The output layer has one neuron which represents whether the point is yellow rust infected (1) or not (0). Two hidden layers were added to the network to get higher accuracy. There is no rule of thumb to decide about the number of neurons in the hidden layers, but we used a value of four (average of the sum of the total number of input attributes plus one). Rectified linear unit (ReLu) activation function was used in both the hidden layers to avoid negative values. The Sigmoid activation function was used in the output layer because of its soft-switching ability and simplicity in derivatives. Adam optimizer, along with the binary cross-entropy function, was used during the compilation of the network. This optimizer can efficiently adjust the weights. Being a binary classification problem, a binary cross-entropy function was used along with the optimizer. During the fitting of the model, epoch's value of two hundred and batch size of five were used. The generated model can be visualized as given in Figures 6.3(a) and 6.3(b). Implementation of these models was performed in Python language using Jupyter notebooks. Standard python libraries like NumPy, Pandas and Keras were also used. Training of the model was performed using 80% of the experimental dataset and validation using the remaining 20% datasets.

Accuracy, precision, recall and F1 were calculated to find out the efficiency of multispectral data for accurately detecting the disease using deep learning models.

Accuracy is a metric that is used to evaluate classification models. It is the fraction of predictions the model got right. Precision answers the question about the proportion of positive identifications which were actually correct. Recall calculates the proportion of actual positives which were identified correctly. F1-score is interpreted as the weighted average of precision and recall. A sampling procedure, k-fold cross-validation, was also used to evaluate the above models. The 10-fold cross-validation was used to train and validate the model. The whole dataset is randomly partitioned into 10 folds (depending on the data size), and the model is fitted with nine folds which are used as a training set and validate the model using the remaining set. This process is repeated until every 10-fold has been served as the test set.

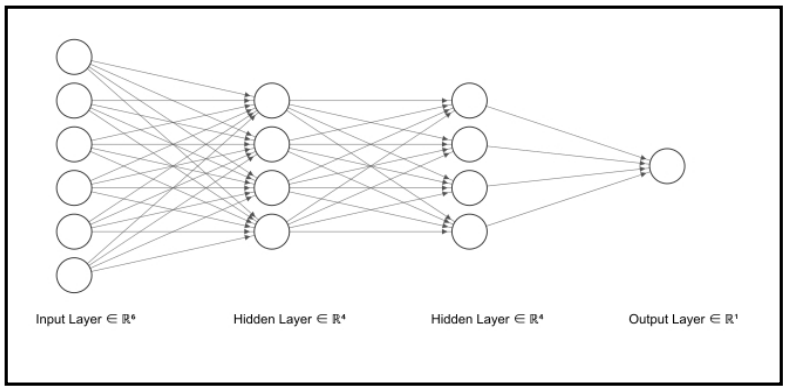


Figure 6.3 (a): ANN Layout

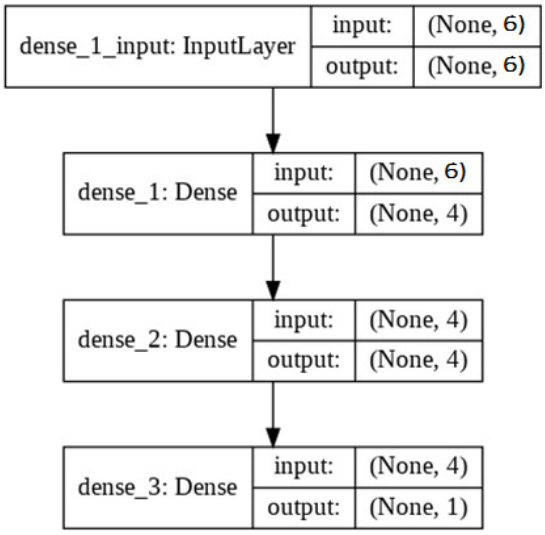


Figure 6.3(b): Model Structure

6.3. Results and Discussion

6.3.1 Response of spectral indices to yellow rust

Spectral indices are used for the identification and monitoring of crop diseases due to changes in the biochemical content of diseased and healthy crops. Upon infestation, the yellow rust rapidly consumes the chlorophyll, nutrients and water content of the plants, causing changes in the chlorophyll content of leaves leading to decrease in crop biomass [143]. Sentinel-2 has the red edge band which makes it unique for monitoring crop growth and diseases [31]. In this study, it was found that the six spectral indices (NDVI, MCARI, Cl-Red, S2REP, NDMI and NDWI) are useful for detecting yellow rust crop disease and these indices use the three Red-edge, NIR, Red and Green bands which capture the destruction in chlorophyll and tissues of crops followed by shifting of the spectrum from visible to NIR band. NDMI and NDWI were used to study the effect of a decrease in moisture of the crop and soils due to disease.

The responses of spectral indices to yellow rust are given in Fig 6.4 (A) and (B), where the average of each spectral index has been compared in diseased and healthy plants. The values of the NDVI, MCARI, Cl-Red and S2REP vegetation indices were decreased in the disease-infested plants of both areas. The NDVI was decreased by 27 % in disease-infested plants than healthy plants of Jalandhar/Kapurthala and 15% in the Rupnagar area. The MCARI was 1.5 times higher in healthy plants also in the Jalandhar/Kapurthala and 1.25 times% in the Rupnagar area. The Cl-Red was decreased by 30% in diseased plants than healthy plants in Jalandhar/Kapurthala and 5% in the Rupnagar area. The S2REP was 14% higher in healthy plants than diseased plants in the Jalandhar/Kapurthala and 26% in the Rupnagar area. Similarly, there was a decrease in the NDMI and NDWI in disease-infested plants for both Rupnagar and Jalandhar/Kapurthala areas.

In order to maintain the same magnitude of spectral indices between -1 and +1, the scaling of spectral indices for disease-infested and healthy plants showed that the magnitude of the difference was higher with MCARI followed by NDVI, Cl-Red, NDWI, S2REP and NDMI for the Jalandhar/Kapurthala, but it was higher with MCARI, S2REP, NDVI, NDMI, Cl-Red and NDWI for the Rupnagar area.

6.3.2. Classification of healthy and diseased plants

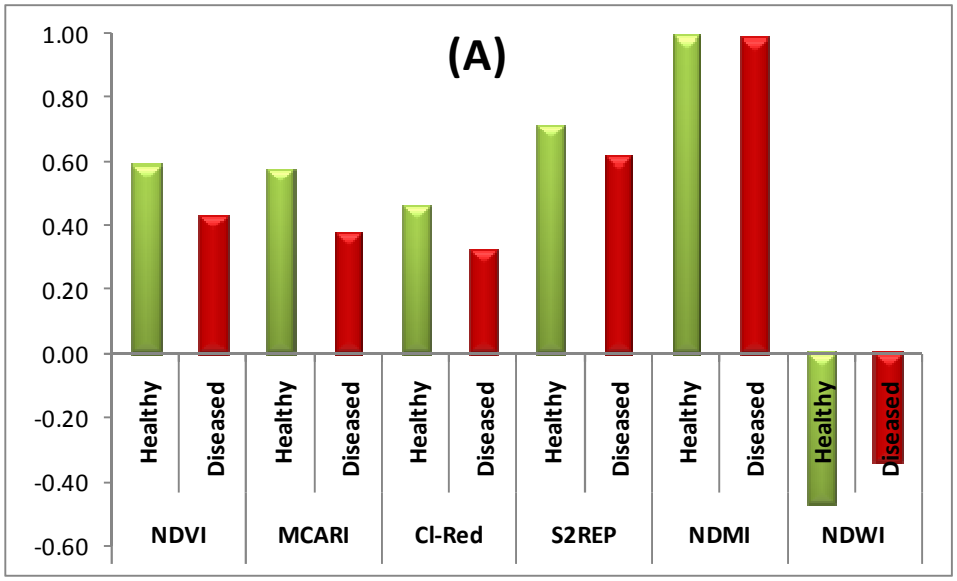
The classified data of healthy and diseased plants using deep learning ANN showed that the accuracy and F1-score for the Rupnagar area were 0.91, whereas there were 0.95 for the Jalandhar/Kapurthala. The classification reports of the Rupnagar and Jalandhar/Kapurthala studies are given in Table 6.4. The data of Jalandhar/Kapurthala could not be classified using the Rupnagar trained model and vice-versa. The 10-fold cross-validation results were 0.97 with a variance of 0.12 for the Roopnagar area and 0.91 with a variance of 0.13 for the Jalandhar/Kapurthala.

The ANN created in this study can be trained to classify the data points from healthy and diseased plants based on the spectral indices values. Proximal sensing, high-resolution images, classical machine learning algorithms, and CNN's have been of primary interest to researchers for the identification of wheat rust, but our findings advance the understanding of AI for the identification of biotic stress in plants using freely available Sentinel 2 data. It offers a novel perspective by demonstrating that multispectral satellite imagery along with ANN's could be used effectively to identify the yellow rust of wheat. The various advantages of the current approach compared with the conventional ones are: (i) high spectral resolution and proximal data are costly, (ii) The multispectral imagery used in this study is open and freely available. (iii) Training a CNN (Image data) is computationally very expensive, whereas the ANN (Tabular Data) used in this study has been trained on a simple desktop computer without a GPU. Classical statistical and machine learning algorithms require an extensive workflow for data pre-processing, feature engineering, and feature selection. In contrast, there are very few requirements in deep learning, with a little bit of pre-processing, the data can be fed into the network. However, there are a few limitations in this study. The yellow rust usually attacks during winters when there is fog and cloud cover in Indian Punjab, and it is difficult to get a cloud-free satellite image during this period. There is a small period between the onset of the yellow rust and the application of fungicide to control it. Therefore, finding a satellite image on a suitable satellite pass date is a big challenge. Finally, the model varies with the season. In our study, the data of Jalandhar/Kapurthala could not be classified using the Rupnagar trained model as there was a time difference between the satellite

images of both study locations. The yellow rust infected the crop during January in the Rupnagar area, whereas wheat was infected during February in the Jalandhar/Kapurthala area crops. The spectral indices change with time [147] which is one of the major reasons for not developing a global model for yellow rust of wheat.

Class	Precision	Recall	F1-score	Accuracy
Jalandhar/Kapurthala				
Healthy	0.99	0.92	0.95	
Diseased	0.92	0.99	0.95	0.95
Rupnagar				
Healthy	0.92	0.88	0.90	
Diseased	0.89	0.93	0.91	0.91

Table 6.4: Summary statistics of the ANN model in the areas of Indian Punjab for discrimination of yellow rust of wheat



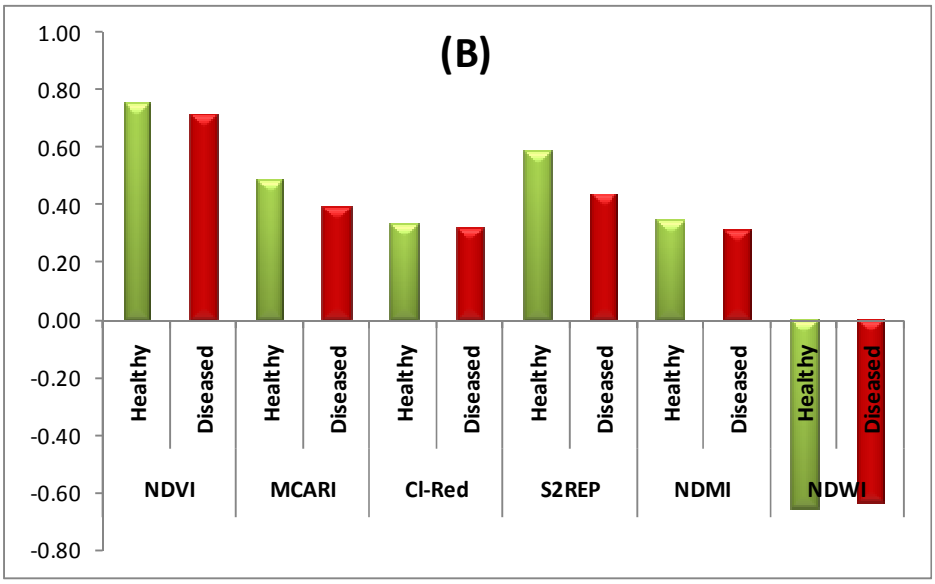


Figure 6.4: Spectral indices in healthy and diseased wheat plants in (A) Jalandhar/Kapurthala, and (B) Rupnagar area.

6.4. Conclusions

The results of this study showed that deep learning ANN can be used to classify the wheat yellow rust extracted from the spectral indices generated from Sentinel-2 satellite imagery. The proposed model is suitable for monitoring of yellow rust of wheat in crops on a local scale, but a global model couldn't be developed in this data. Future work is required to develop a regional/global model for monitoring of disease. However, this disease is also affected by meteorological variables (like sunshine hours, relative humidity and temperature etc.); therefore, coupling meteorological variables with spectral indices may help to develop reliable crop disease, forecasting models.

Chapter 7

Simulation of multispectral data using hyperspectral data for crop stress studies

7.1 Objective

This research aims to find the suitable bands of freely available multispectral satellite imagery associated with the crop parameters, which will help in crop stress identification. Important wavelengths identified in Chapters 3, 4 and 5 for various crop parameters using AI techniques have been used for the analysis. The satellite imagery selected for this analysis is Sentinel-2 MSI[153], Landsat-8 OLI [154]and Landsat-7 ETM+[155]. These multispectral imageries are chosen because they have the best spatial and spectral resolutions among today's freely available earth observation datasets. The results under Indian, Israel and US conditions were compared to identify the optimum wavelength for nutrient and water stress.

7.2 Material and Methods

7.2.1 Hyperspectral Data

Wheat and Maize Data (Punjab) from Chapter 3: Field experiments with maize and wheat were conducted at Punjab Agricultural University farm, Ludhiana, in 2019, 2020 and 2021. The experiment was laid out in three replicates in a split-plot design with three irrigation treatments (IW/CPE 1.0, 0.75 and 0.5) in main plots with five nutrient treatments (T1 = 100% NPK, T2 = 75% NPK, T3 = 50% NPK, T4 = 100% N, T5 = 100% NP) in subplots. The recommended doses of fertilizers (100% NPK) for maize (variety PMH I) were Urea (90 kg / acre), Di-Ammonium Phosphate (55 kg/acre) and Muriate of Potash (20 kg /acre). The recommended fertilizer doses (100% NPK) for wheat (variety PBW725) were Urea (90 kg / acre), Single Superior Phosphate (55 kg/acre) and Muriate of Potash (20 kg /acre). The leaf samples were collected at monthly intervals from August-October 2019 and 2020 for maize, December 2019-March 2020 and December 2020-April 2021 for wheat. The leaf samples were analysed for water content, chlorophyll, nitrogen (N), phosphorus (P) and potassium (K) using standard methods. The reflectance spectra of crops were collected on leaf sampling dates using ASD Field Spec-4 Spectroradiometer

(wavelength interval between 350 - 2500 nm) on clear and cloudless days. There were 103 measurements for maize and 135 for the wheat experiment.

Wheat (Israel) from Chapter 4: The secondary hyperspectral data of wheat was downloaded for the Gilat Agricultural Research Center, Israel. The dataset [97] was downloaded from the Ecological Spectral Information System (EcoSIS) website https://ecosis.org/. The proximal canopy measurements were taken in the year 2004 and 2005 using a spectroradiometer (ASD FieldSpec bare fiber). The data was collected throughout the growing season at 6, 45, 63-72 and 91-97 days of wheat growth. The plant tissues were collected and analysed for water content, dry weight, leaf area index (LAI) and nitrogen content. The spectral range of the data is 400 to 2400 nm with a spectral sampling resolution of 2 nm (330 measurements).

Maize (USA) from Chapter 5: The secondary hyperspectral dataset of maize crop used for this work was the output of the field experiment conducted on Havelock Research Farm and Greenhouse Innovation Centre of the University of Nebraska-Lincoln (U.S.). This dataset [136] was downloaded from the EcoSIS website https://ecosis.org/. Maize association panel was grown in the Havelock Research Farm in 2018 and 2019 under control and nitrogen deficiency conditions. In addition, the panel was also grown in the Greenhouse Innovation Center in 2018. Hyperspectral reflectance was measured at the leaf level. Ground truth data were collected for chlorophyll, nitrogen, phosphorus, potassium, fresh and dry leaf weight, and leaf area. The data consisting of 1210 measurements has a spectral range of 350 to 2500 nm and a spectral sampling resolution of 1 nm.

7.2.2AI Techniques

The raw hyperspectral data was pre-processed with two data transformation techniques, Savitzky-Golay 1st derivative and Deresolve.

Reflectance data were utilized as input parameters (X variables) in the machine learning analysis environment, and water content, N, P, K, and Chlorophyll concentrations in maize and wheat were chosen as the target variable (Y variable). Six machine learning regression algorithms (k-nearest neighbours (kNN), Random Forest

(RF), Support Vector Regression (SVR), Multilayer Perceptron (MLP), Gradient Boosting Regression (GBR), and Partial Least Square Regression (PLSR)) were used to retrieve various maize parameters (from hyperspectral data). The machine learning algorithms were developed using the Python [80] programming language and the Scikit-learn[81] library. Scikit-learn is an open-source library that integrates various machine learning methods[125]. The ML algorithms were developed and executed in a cloud-based environment, Google's Colaboratory[82]. Without the need for installation or configuration, Colaboratory allows creating and executing Python code in a simple browser. GridSearchCV function was employed for hyperparameter tuning. This process helped select the best parameters for each regression algorithm. Finally, the models with the best coefficient of determination (R2) and mean-squared error (MSE) were chosen.

Machine learning models are frequently referred to as "black boxes," meaning that deciphering them and understanding how they produced a specific output is difficult. Explainable AI (XAI) technologies can aid in the interpretation and explanation of sophisticated AI models.[85] The SHapley Additive explanations (SHAP) value, an important XAI tool, was proposed by[86]. SHAP values are derived from shapley values found in mathematical game theory. The game is the result of a model in machine learning, and the players represent the features contained in the model. SHAP aids in quantifying each feature's (player's) contribution to the model prediction (game)[86]. It dissects the prediction to show how each element affects the outcome. It can explain individual forecasts and aggregate them to provide valuable insights into the entire model. SHAP values calculations are explained in detail by [87].

The SHAP library[126]was used to implement XAI in Python [126]for this study. The key benefit of adopting SHAP values for this study is the model's overall interpretability. The collective SHAP values are interpreted globally to find the important wavelengths related to the target variables.

7.2.3 Suitable bands in multispectral satellite imagery

Suitable bands of freely available multispectral satellite imagery associated with the crop parameters are discussed in this section. The satellite imagery selected

for this analysis is Sentinel-2 MSI, Landsat-8 OLI and Landsat-7 ETM+. These multispectral imageries are selected because they have the best spatial and spectral resolutions among today's freely available earth observation datasets. Table-7.1, Table-7.2, and Table-7.3 outline the spectral bands and resolutions of the Sentinel-2 MSI, Landsat-8 OLI, and Landsat-7 ETM+ datasets, respectively. Figure-7.1 compares the spectral bands of Sentinel-2, Landsat 7 ETM+, and Landsat 8 OLI.

Those bands are identified in these freely available datasets, similar to those identified by the explanation of the ML models in the previous steps.

Table-7.1: Spectral bands of Sentinel-2 MSI sensor

Band	**Spectral Region/Band Description**	**Wavelength range (nm)**	**Central Wavelength (nm)**	**Resolution (m)**
B1	Coastal Aerosol	433 - 453	443	60
B2	Blue	458 - 523	490	10
B3	Green Peak	543 - 578	560	10
B4	Red	650 - 680	665	10
B5	Red-Edge 1	698 - 713	705	20
B6	Red-Edge 2	733 - 748	740	20
B7	Red-Edge	773 - 793	783	20
B8	Near-infrared (NIR)	785 - 900	842	10
B8A	Near-infrared narrow (NIRn)	855 - 875	865	20
B9	Water vapour	935 - 955	945	60
B10	Shortwave infrared /Cirrus	1360 - 1390	1375	60
B11	Shortwave infrared 1 (SWIR1)	1565 - 1655	1610	20
B12	Shortwave infrared 2 (SWIR2)	2100 - 2280	2190	20

Table-7.2: Spectral bands of Landsat-8 OLI sensor

Band	**Spectral Region/Band Description**	**Wavelength range (nm)**	**Central Wavelength (nm)**	**Resolution (m)**
1	Coastal Aerosol	433 - 453	443	30
2	Blue	450 - 515	483	30
3	Green	525 - 600	560	30
4	Red	630 - 680	660	30
5	Near-infrared (NIR)	845 - 885	865	30
6	Shortwave infrared 1 (SWIR1)	1560 - 1660	1650	30
7	Shortwave infrared 2 (SWIR2)	2100 - 2300	2220	30
8	Panchromatic	500 - 680	640	15
9	Shortwave infrared /Cirrus	1360 - 1390	1375	30

Table-7.3: Spectral bands of Landsat-7 ETM+ sensor

Band	Spectral Region/Band Description	Wavelength range (nm)	Central Wavelength (nm)	Resolution (m)
1	Blue	450 - 520	483	30
2	Green	520 - 600	560	30
3	Red	630 - 690	662	30
4	Near-infrared (NIR)	770 - 900	835	30
5	Shortwave infrared 1 (SWIR1)	1550 - 1750	1648	30
6	Thermal	10400 - 12500	11335	30
7	Shortwave infrared 2 (SWIR2)	2090 - 2350	2206	30
8	Panchromatic	520 - 900	706	15

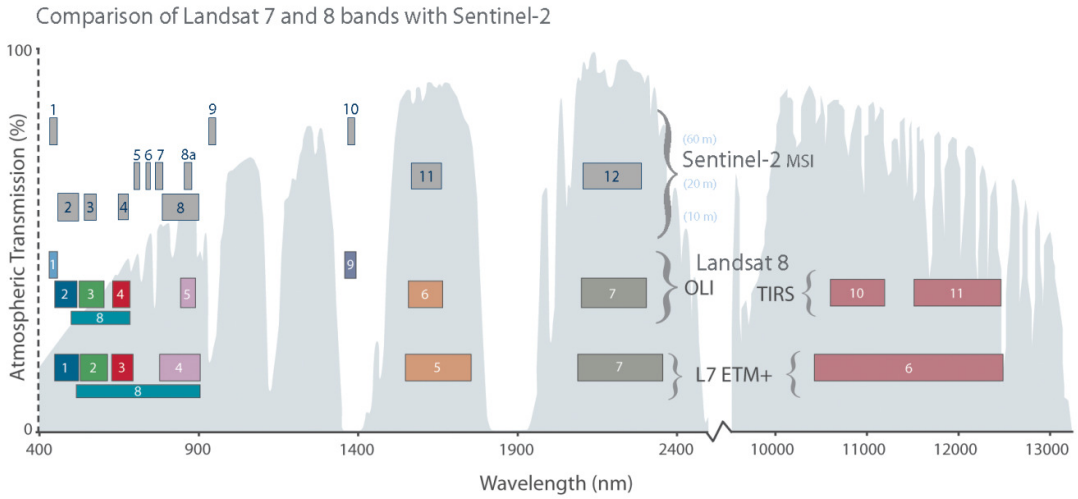


Figure-7.1: Comparison of the bands of Sentinel-2, Landsat 7 and 8.

7.3 Results & Discussions

7.3.1Suitable bands in multispectral satellite imagery

Maize and Wheat (Punjab): The bands of freely available multispectral satellite imagery associated with the six crop (maize and wheat of Punjab) parameters are given in Table-7.4 (Maize) and Table-7.5 (Wheat). According to the analysis, for maize, the SWIR1 band in all the three earth observation datasets is associated with the water nitrogen, phosphorus and potassium parameters.NIR and blue bands are associated with the water content, and red and green bands are with chlorophyll.

For the wheat dataset, SWIR1, NIR and Shortwave infrared /Cirrus are associated with the phosphorus and potassium parameters. Green, NIR and coastal aerosols are

associated with nitrogen. Green peak, green and red are associated with chlorophyll. Green, NIR, blue and green peak are associated with the water content.

Table-7.4: Bands of freely available multispectral satellite imagery associated with the maize(Punjab) parameters

Crop Parameter	Sentinel-2	LANDSAT-8 OLI	LANDSAT-7 ETM+
Water Content	Near-infrared (NIR), Blue	NIR, SWIR2, Blue	NIR, Blue
Nitrogen	SWIR1 (B11), Red-Edge2 (B6)	SWIR1	SWIR1
Phosphorus	SWIR1 (B11)	SWIR1	SWIR1
Potassium	SWIR1 (B11), Red-Edge2 (B6), Red (B4)	SWIR1	SWIR1
Chlorophyll a	Red (B4)	Red	Green, Red
Chlorophyll b	Red (B4)	Green, Red	Green, Red

Table-7.5: Bands of freely available multispectral satellite imagery associated with the wheat (Punjab) parameters

Crop Parameter	Sentinel-2	LANDSAT-8 OLI	LANDSAT-7 ETM+
Water Content	Green peak (B3), Red Edge2 (B6), Coastal Aerosol (B1), Water Vapour (B9)	Green, NIR	Green, NIR, Blue
Nitrogen	Coastal Aerosol (B1), Green Peak (B3), NIR	Coastal Aerosol, Green, NIR,	Green, NIR, Blue
Phosphorus	Shortwave infrared /Cirrus, SWIR1, NIR	Shortwave infrared /Cirrus, SWIR1, NIR	SWIR1, NIR
Potassium	Shortwave infrared /Cirrus, SWIR1	Shortwave infrared /Cirrus, SWIR1	SWIR1, NIR
Chlorophyll a	Green Peak (B3), Red(B4)	Green, Red	Green, Red
Chlorophyll b	Green Peak (B3), Red(B4)	Green, Red	Green, Red

Wheat (Israel): The bands of freely available multispectral satellite imagery associated with Nitrogen are given in Table-7.6. The random forest(RF) algorithm was the best performing model, and according to it the important wavelengths were: 790,794, 796, 804, 806, 816, 820, 862, 848 and 860. According to the analysis, the SWIR2 band in all the three earth observation datasets is associated with the water content, nitrogen, phosphorus and potassium parameters.

Table-7.6: Bands of freely available multispectral satellite imagery associated with the wheat (Israel) parameters

Crop Parameter	Sentinel-2	LANDSAT-8 OLI	LANDSAT-7 ETM+
Nitrogen	NIR (B8), Red-Edge (B7), Near-infrared narrow (NIRn) (B8A)	Near-infrared (NIR)	Near-infrared (NIR)

Maize(US): The bands of freely available multispectral satellite imagery associated with the five crop (maize US) parameters are given in Table-7.7. According to the analysis, the SWIR2 band in all the three earth observation datasets is associated with the water content, nitrogen, phosphorus and potassium parameters. The red-edge2 band of the Sentinel-2 is associated with chlorophyll and nitrogen. In comparison, the Green and NIR bands of both Landsat datasets are associated with chlorophyll. For the water content parameter, the most related bands of Landsat are SWIR1 and SWIR2.

Table-7.7: Bands of freely available multispectral satellite imagery associated with the maize(US) parameters

Crop Parameter	Sentinel-2	LANDSAT-8 OLI	LANDSAT-7 ETM+
Water Content	SWIR2 (B12)	SWIR2, SWIR1	SWIR1, SWIR2,
Nitrogen	SWIR2 (B12), Red-Edge (B7),	SWIR2, Green	SWIR2, Green, NIR
Phosphorus	SWIR2 (B12)	SWIR2	SWIR2
Potassium	SWIR2 (B12)	SWIR2	SWIR2
Chlorophyll	Red-Edge 2 (B6), Green Peak (B3)	Green, NIR	PAN, NIR, Green

According to [110] also, shortwave infrared is associated to the leaf water content and biochemicals, near-infrared with cell structure and visible spectrum is with leaf pigments. [107]developed two spectral indices (canopy chlorophyll content index from 670, 720, 790 nm and spectral ratio planer index from 445, 705 and 750 nm) to detect the N stress in crops. [108] used hyperspectral data to identify the nitrogen status of wheat. The genetic algorithm selected relatively few features in the NIR region (>750 nm) for estimating plant N content. Reflectance features were commonly selected between 583 and 722 nm in the two growing seasons, encompassing the primary wavelengths for visible red light absorption by chlorophyll.

The results show that NIR and SWIR region is useful for detecting P deficiency in plants. There is a close relationship between P and reflectance in the NIR region because P plays a critical role as an energy supplier in energy-consuming processes such as photosynthesis. In combination with reflectance in the visible or NIR regions, reflectance in the SWIR range was found to monitor nitrogen, phosphorus, sulphur and potassium status in plants [100].

The SWIR band identified for K deficiency in plants corresponds to starch and protein absorption features. [20] developed a vegetation index using wavelengths of 1645 and 1715 nm (SWIR) to determine wheat's potassium concentration. VIS–NIR–SWIR leaf reflectance can be a valuable tool for low-cost, non-destructive, high-

throughput investigation of leaf physiological and biochemical characteristics (including N, P and K), according to [137].

According to [111], chlorophyll and N contents can be accurately remotely estimated using green, red and red-edge chlorophyll indices using near-infrared (780–800 nm) and either green (540–560 nm) or red-edge (730–750 nm) spectral bands. [94] also asserted that the red and red-edge bands are useful for the estimation of chlorophyll in maize.

Although various bands are associated with the crop parameters, there are certain limitations of the available datasets. The spatial and spectral resolution of the freely available datasets is not suitable for crop or farm-level studies. Almost all spectral bands' bandwidth ranges in reference datasets are quite wide for precision or higher-resolution studies. Also, according to Table 4, the SWIR2 band overlaps in the four crop parameters, so it is difficult to distinguish between various parameters individually. Thus, due to the broader bandwidth range, the generalization of information takes place. The minimum spatial resolution dataset is Sentinel-2 with 10m; such data may be used for large farms or a complete administrative unit like a block or a district. It will not be suitable for precision farming applications.

The findings imply that machine learning may be utilized to accurately evaluate the status of various crop parameters using hyperspectral data. The eXplainable Artificial Intelligence (XAI) tools can be used to effectively explain sophisticated black box ML models, which will aid remote monitoring of various agricultural parameters and maize crop nutrition. Furthermore, currently available earth observation datasets are appropriate for medium/small scale studies but not for large-scale studies or precision farming to diagnose crop stress.

Chapter 8

Summary and Future Work

The effect of abiotic and biotic stresses on the crop is usually identified only when these are noticeable and visible. Therefore, continuous monitoring of crop conditions using proximal and remote sensing technologies can assist in the early diagnosis of the problem. Therefore, a study was carried out to detect the water, nutrient and disease stress in wheat and maize using proximal and remote sensing. Mapping and monitoring crop stress using spectral indices is helpful for understanding the crop growth trends. In order to understand the crop conditions on 1;50,000 scale, MODIS Enhanced Vegetation Index(EVI) was used in Google Earth Engine. The results of the study showed that maximum crop stress was observed in the Gurdaspur district of Punjab during the Rabi (Winter) cropping season of 2019-2020, and the least stress was identified in Patiala, Nawan Shehar and SAS Nagar(Mohali) districts of Punjab.

In order to identify the stress due to abiotic factors (like nutrients and water), a field experiment with a maize-wheat cropping system under various water and nutrient levels was carried out at Punjab Agricultural University, Ludhiana. The reflectance spectra of crops were collected at regular intervals using Spectroradiometer (wavelength interval between 400 - 2500 nm) on clear and cloudless days. The leaf samples were collected and analyzed for water content, chlorophyll, nitrogen (N), phosphorus (P) and potassium (K) using standard methods. The measured plant parameters were related to spectra hyperspectral data using machine learning (RF, SVR, GBR, PLSR) and explainable artificial intelligence (XAI) techniques to identify the optimum wavelengths for nutrient and water stress in maize and wheat. It was found that one machine learning technique was not suitable for every parameter. For example, the water content in maize leaves was better estimated with GBR followed RF models. N in maize was estimated using PLSR, GBR and RF models, total P content using PLSR and RF models, and total K using PLSR and RF models. The most important wavelengths associated with water content for detecting water stress in maize are in the near-infrared region (841, 842, 847 and 702 nm) using GBR. **T**he optimum bands for estimation of water content in wheat were visible, NIR and SWIR

regions. The sensitive bands were 599, 617, 600, 618, and 651 nm for estimation of chlorophyll a, and 486, 691, 599, 690 and 431 nm for chlorophyll b in maize leaves. The most important bands for estimation of chlorophyll a and b in wheat leaves were in the visible region (554, 553, 656, 555 and 405 nm for chlorophyll a, and 486, 691, 599, 690 and 431 nm for chlorophyll b). The wavelength predicted with the RF models for estimation of N in maize leaves were in the visible, near-infrared and SWIR region (617, 804, 1978, 1979 and 738 nm). The sensitive bands to estimate N in wheat leaves were 443 nm> 449 nm >597 nm> 885 nm > 450 nm. The important bands for detecting P stress in maize were in the SWIR region using GBR, visible, NIR and SWIR region using the RF model. The optimum bands for estimation of total P in wheat leaves were in 1304, 1305, 810, 1259 and 1505 nm. The total K in maize were better using visible and NIR region with the RF model, but in the SWIR region with the PLSR model. The sensitive bands for detecting K deficiency in wheat leaves were in the SWIR region

These studies identified the optimum spectral wavelengths for abiotic stresses in crops, but the spectral behaviour of biotic and abiotic stresses is quite different. In order to study the biotic stress in crops, yellow rust of wheat (a major disease) was classified in the parts of Punjab from Sentinel-2 satellite imagery using deep learning artificial neural network. The values of the normalized difference vegetation index (NDVI), modified chlorophyll absorption in reflectance Index (MCARI), chlorophyll red edge (Cl-Red) and sentinel-2 red-edge position (S2REP) vegetation indices were decreased in the disease-infested plants of both areas. Similarly, there was a decrease in the normalized difference moisture index (NDMI) and normalized difference water index (NDWI) in disease-infested plants for both Rupnagar and Jalandhar/Kapurthala areas. The classified data of healthy and diseased plants using deep learning artificial neural network (ANN) showed that the accuracy and F1-score for the Rupnagar area were 0.91, whereas there were 0.95 for the Jalandhar/Kapurthala.

The wavelengths identified from hyperspectral data in abiotic stresses were compared with freely available multispectral satellite imagery (Sentinel-2 MSI, Landsat-8 OLI and Landsat-7 ETM+) for remote monitoring of stresses in crops. The SWIR2 band in all three earth observation datasets is associated with the water

content, nitrogen, phosphorus and potassium parameters. The red-edge2 band of Sentinel-2 is associated with chlorophyll and nitrogen. In comparison, the green and NIR bands of both Landsat datasets are associated with chlorophyll. The most related bands of Landsat were SWIR1 and SWIR2 for the estimation of water content.

Although various bands are associated with the crop parameters, the available datasets have certain limitations. The spatial and spectral resolution of the freely available datasets is unsuitable for crop or farm-level studies. All spectral bands' bandwidth ranges in reference datasets are wide for precision or higher-resolution studies. Also, the SWIR2 band overlaps in the four crop parameters, so it isn't easy to distinguish between various parameters individually. Thus, due to the broader bandwidth range, information generalization occurs. The minimum spatial resolution dataset is Sentinel-2 with 10m; such data may be used for large farms or a complete administrative unit like a block or a district. It will not be suitable for precision farming applications.

The major findings of this study are that machine learning techniques are useful for estimating the various crop parameters using hyperspectral data. The eXplainable Artificial Intelligence (XAI) tools can be used to explain the ML models, which will aid remote monitoring of various crop growth parameters and crop nutrition.

In order to better understand the implications of these results, future studies could address the issues of the generalization of the results due to the broader bandwidth of the freely available satellite imagery. New AI techniques and future higher resolution (spatial and spectral) earth observation datasets could be used to lessen the gap between the proximal and remote sensing datasets.

Bibliography

[1] T. Lillesand, R. W. Kiefer, and J. Chipman, *Remote sensing and image interpretation*. John Wiley & Sons, 2015.

[2] G. Joseph, *Fundamentals of remote sensing*. Universities Press, 2005.

[3] R. D. Jackson, "Remote Sensing of Biotic and Abiotic Plant Stress," *Annu. Rev. Phytopathol.*, vol. 24, no. 1, pp. 265–287, 1986.

[4] S. Ullah, Y. Si, M. Schlerf, A. K. Skidmore, M. Shafique, and I. A. Iqbal, "Estimation of grassland biomass and nitrogen using MERIS data," *Int. J. Appl. earth Obs. Geoinf.*, vol. 19, pp. 196–204, 2012.

[5] T. Rumpf, A.-K. Mahlein, U. Steiner, E.-C. Oerke, H.-W. Dehne, and L. Plümer, "Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance," *Comput. Electron. Agric.*, vol. 74, no. 1, pp. 91–99, 2010.

[6] J. Van Beek *et al.*, "Temporal dependency of yield and quality estimation through spectral vegetation indices in pear orchards," *Remote Sens.*, vol. 7, no. 8, pp. 9886–9903, 2015.

[7] A. H. Prasad S. Thenkabail, John G. Lyon, *Hyperspectral Remote Sensing of Vegetation*. 2016.

[8] J. Im and J. R. Jensen, "Hyperspectral Remote Sensing of Vegetation," *Hyperspectral Remote Sens. Veg.*, vol. 6, pp. 1943–1961, 2016.

[9] R. N. Sahoo, S. S. Ray, and K. R. Manjunath, "Hyperspectral remote sensing of agriculture," *Curr. Sci.*, vol. 108, no. 5, pp. 848–859, 2015.

[10] S. Irmak, D. Z. Haman, and R. Bastug, "Determination of crop water stress index for irrigation timing and yield estimation of corn," *Agron. J.*, vol. 92, no. 6, pp. 1221–1227, 2000.

[11] D. C. Nielsen, "Scheduling irrigations for soybeans with the Crop Water Stress Index (CWSI)," *F. Crop. Res.*, vol. 23, no. 2, pp. 103–116, 1990.

[12] H. G. Jones, "Remote detection of crop water 'stress' and distinguishing it from other stresses," *Acta Hortic.*, vol. 922, pp. 23–34, 2011.

[13] L. Leroux, C. Baron, B. Zoungrana, S. B. Traore, D. Lo Seen, and A. Begue, "Crop Monitoring Using Vegetation and Thermal Indices for Yield Estimates: Case Study of a Rainfed Cereal in Semi-Arid West Africa," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 9, no. 1, pp. 347–362, 2016.

[14] L. Chaerle, S. Lenk, I. Leinonen, H. G. Jones, D. Van Der Straeten, and C. Buschmann, "Multi-sensor plant imaging: Towards the development of a stress-catalogue," *Biotechnol. J.*, vol. 4, no. 8, pp. 1152–1167, 2009.

[15] P. J. Zarco-Tejada, J. A. J. Berni, L. Suárez, G. Sepulcre-Cantó, F. Morales,

and J. R. Miller, “Imaging chlorophyll fluorescence with an airborne narrow-band multispectral camera for vegetation stress detection,” *Remote Sens. Environ.*, vol. 113, no. 6, pp. 1262–1275, 2009.

[16] B. Das, G. R. Mahajan, and R. Singh, “Hyperspectral Remote Sensing: Use in Detecting Abiotic Stresses in Agriculture,” *Adv. Crop Environ. Interact.*, pp. 317–335, 2018.

[17] M. Rossini *et al.*, “Assessing canopy PRI from airborne imagery to map water stress in maize,” *ISPRS J. Photogramm. Remote Sens.*, vol. 86, pp. 168–177, 2013.

[18] X. Wang, C. Zhao, N. Guo, Y. Li, S. Jian, and K. Yu, “Determining the canopy water stress for spring wheat using canopy hyperspectral reflectance data in loess plateau semiarid regions,” *Spectrosc. Lett.*, vol. 48, no. 7, pp. 492–498, 2015.

[19] F. Li, B. Mistele, Y. Hu, X. Chen, and U. Schmidhalter, “Reflectance estimation of canopy nitrogen content in winter wheat using optimised hyperspectral spectral indices and partial least squares regression,” *Eur. J. Agron.*, vol. 52, pp. 198–209, 2014.

[20] A. Pimstein, A. Karnieli, S. K. Bansal, and D. J. Bonfil, “Exploring remotely sensed technologies for monitoring wheat potassium and phosphorus using field spectroscopy,” *F. Crop. Res.*, vol. 121, no. 1, pp. 125–135, 2011.

[21] X. M. Zongyao Sha, Yuwei Wang, Yongfei Bai, Yujin Zhao, Hua Jin, Ya Na, “Comparison of Leaf Area Index Inversion for Grassland Vegetation through Remotely Sensed Spectra by Unmanned Aerial Vehicle (UAV) and Field-based Spectroradiometer,” *J. Plant Ecol.*, vol. 12, no. 3, pp. 395–408, 2018.

[22] X. M. Chen, “Epidemiology and control of stripe rust [Puccinia striiformis f. sp. tritici] on wheat,” *Can. J. Plant Pathol.*, vol. 27, no. 3, pp. 314–337, 2005.

[23] Z. Jingcheng, L. Juhua, H. Wenjiang, and W. Jihua, “Continuous wavelet analysis based spectral feature selection for winter wheat yellow rust detection,” *Intell. Autom. Soft Comput.*, vol. 17, no. 5, pp. 531–540, 2011.

[24] J. Zhang, R. Pu, R. W. Loraamm, G. Yang, J. Wang, and others, “Comparison between wavelet spectral features and conventional spectral features in detecting yellow rust for winter wheat,” *Comput. Electron. Agric.*, vol. 100, pp. 79–87, 2014.

[25] H. Wenjiang, H. Muyi, L. Liangyun, W. Jihua, Z. Chunjiang, and W. Jindi, “Inversion of the severity of winter wheat yellow rust using proper hyper spectral index,” *Trans. Chinese Soc. Agric. Eng.*, vol. 21, no. 4, pp. 97–103, 2005.

[26] 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,” *Comput. Electron. Agric.*, vol. 44, no. 3, pp. 173–188, 2004.

[27] G. Krishna *et al.*, “Assessing wheat yellow rust disease through hyperspectral remote sensing,” *Int. Arch. Photogramm. Remote Sens. Spat. Inform. Sci*, pp. 1413–1416, 2014.

[28] R. Devadas, D. W. Lamb, S. Simpfendorfer, and D. Backhouse, “Evaluating ten spectral vegetation indices for identifying rust infection in individual wheat leaves,” *Precis. Agric.*, vol. 10, no. 6, pp. 459–470, 2009.

[29] D. Ashourloo, M. R. Mobasheri, and A. Huete, “Developing two spectral disease indices for detection of wheat leaf rust (Pucciniatriticina),” *Remote Sens.*, vol. 6, no. 6, pp. 4723–4740, 2014.

[30] S. Dutta, S. K. Singh, and M. Khullar, “A case study on forewarning of yellow rust affected areas on wheat crop using satellite data,” *J. Indian Soc. Remote Sens.*, vol. 42, no. 2, pp. 335–342, 2014.

[31] Q. Zheng, W. Huang, X. Cui, Y. Shi, and L. Liu, “New spectral index for detecting wheat yellow rust using sentinel-2 multispectral imagery,” *Sensors*, vol. 18, no. 3, p. 868, 2018.

[32] Q. Zheng *et al.*, “Integrating Spectral Information and Meteorological Data to Monitor Wheat Yellow Rust at a Regional Scale: A Case Study,” *Remote Sens.*, vol. 13, no. 2, p. 278, 2021.

[33] M. H. Saleem, J. Potgieter, and K. M. Arif, “Plant Disease Detection and Classification by Deep Learning,” *Plants*, vol. 8, no. 11, p. 468, 2019.

[34] J. Su *et al.*, “Aerial Visual Perception in Smart Farming: Field Study of Wheat Yellow Rust Monitoring,” *IEEE Trans. Ind. Informatics*, 2020.

[35] X. Zhang *et al.*, “A deep learning-based approach for automated yellow rust disease detection from high-resolution hyperspectral UAV images,” *Remote Sens.*, vol. 11, no. 13, p. 1554, 2019.

[36] R. Pryzant, S. Ermon, and D. Lobell, “Monitoring ethiopian wheat fungus with satellite imagery and deep feature learning,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 39–47.

[37] J. Behmann, A. K. Mahlein, T. Rumpf, C. Römer, and L. Plümer, “A review of advanced machine learning methods for the detection of biotic stress in precision crop protection,” *Precis. Agric.*, vol. 16, no. 3, pp. 239–260, 2015.

[38] A. Singh, B. Ganapathysubramanian, A. K. Singh, and S. Sarkar, “Machine Learning for High-Throughput Stress Phenotyping in Plants,” *Trends Plant Sci.*, vol. 21, no. 2, pp. 110–124, 2016.

[39] A. K. Singh, B. Ganapathysubramanian, S. Sarkar, and A. Singh, “Deep Learning for Plant Stress Phenotyping: Trends and Future Perspectives,” *Trends Plant Sci.*, vol. 23, no. 10, pp. 883–898, 2018.

[40] L. P. Osco *et al.*, “A machine learning framework to predict nutrient content in valencia-orange leaf hyperspectral measurements,” *Remote Sens.*, vol. 12, no.

6, p. 906, 2020.

[41] A. Holzinger, “From machine learning to explainable AI,” in *2018 world symposium on digital intelligence for systems and machines (DISA)*, 2018, pp. 55–66.

[42] W. Samek and K.-R. Müller, “Towards explainable artificial intelligence,” in *Explainable AI: interpreting, explaining and visualizing deep learning*, Springer, 2019, pp. 5–22.

[43] V. Belle and I. Papantonis, “Principles and practice of explainable machine learning,” *arXiv Prepr. arXiv2009.11698*, 2020.

[44] A. Adadi and M. Berrada, “Peeking inside the black-box: a survey on explainable artificial intelligence (XAI),” *IEEE access*, vol. 6, pp. 52138–52160, 2018.

[45] A. Chemura, O. Mutanga, M. Sibanda, and P. Chidoko, “Machine learning prediction of coffee rust severity on leaves using spectroradiometer data,” *Trop. Plant Pathol.*, vol. 43, no. 2, pp. 117–127, 2018.

[46] J. Delegido, J. Verrelst, L. Alonso, and J. Moreno, “Evaluation of sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content,” *Sensors*, vol. 11, no. 7, pp. 7063–7081, 2011.

[47] B. Wang *et al.*, “Assessment of Sentinel-2 MSI spectral band reflectances for estimating fractional vegetation cover,” *Remote Sens.*, vol. 10, no. 12, pp. 1–20, 2018.

[48] S. M. Punalekar, A. Verhoef, T. L. Quaife, D. Humphries, L. Bermingham, and C. K. Reynolds, “Application of Sentinel-2A data for pasture biomass monitoring using a physically based radiative transfer model,” *Remote Sens. Environ.*, vol. 218, no. October, pp. 207–220, 2018.

[49] A. Agapiou, D. D. Alexakis, A. Sarris, and D. G. Hadjimitsis, “Evaluating the potentials of sentinel-2 for archaeological perspective,” *Remote Sens.*, vol. 6, no. 3, pp. 2176–2194, 2014.

[50] and V. C. Dubravko ´Culibrk, Predrag Lugonja, Vladan Mini´c, “Neural Network Approach to Water-Stressed Crops Detection Using Multispectral WorldView-2 Satellite Imagery Dubravko,” in *12th Engi- neering Applications of Neural Networks (EANN 2011) and 7th Artificial Intelligence Applications and Innovations (AIAI)*, 2011, pp. 323–331.

[51] J. B. Feret, C. Corbane, and S. Alleaume, “Detecting the Phenology and Discriminating Mediterranean Natural Habitats with Multispectral Sensors-An Analysis Based on Multiseasonal Field Spectra,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 8, no. 5, pp. 2294–2305, 2015.

[52] L. Prey and U. Schmidhalter, “Simulation of satellite reflectance data using high-frequency ground based hyperspectral canopy measurements for in-season estimation of grain yield and grain nitrogen status in winter wheat,” *ISPRS J. Photogramm. Remote Sens.*, vol. 149, no. July 2018, pp. 176–187, 2019.

[53] A. Gonsamo and J. M. Chen, “Spectral response function comparability among 21 satellite sensors for vegetation monitoring,” *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 3, pp. 1319–1335, 2013.

[54] M. A. Wulder, N. C. Coops, D. P. Roy, J. C. White, and T. Hermosilla, “Land cover 2.0,” *Int. J. Remote Sens.*, vol. 39, no. 12, pp. 4254–4284, 2018.

[55] N. K. Gogoi, B. Deka, and L. C. Bora, “Remote sensing and its use in detection and monitoring plant diseases: A review.,” *Agric. Rev.*, vol. 39, no. 4, 2018.

[56] R. Ferguson, D. Rundquist, D. K. Shannon, D. E. Clay, and N. R. Kitchen, “Remote sensing for site-specific crop management,” *Precis. Agric. basics*, 2018.

[57] S. S. Virnodkar, V. K. Pachghare, V. C. Patil, and S. K. Jha, “Remote sensing and machine learning for crop water stress determination in various crops: a critical review,” *Precis. Agric.*, vol. 21, no. 5, pp. 1121–1155, 2020.

[58] M. Chi, A. Plaza, J. A. Benediktsson, Z. Sun, J. Shen, and Y. Zhu, “Big data for remote sensing: Challenges and opportunities,” *Proc. IEEE*, vol. 104, no. 11, pp. 2207–2219, 2016.

[59] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, “Google Earth Engine: Planetary-scale geospatial analysis for everyone,” *Remote Sens. Environ.*, vol. 202, pp. 18–27, 2017.

[60] O. Mutanga and L. Kumar, “Google earth engine applications.” Multidisciplinary Digital Publishing Institute, 2019.

[61] K. Didan, A. B. Munoz, R. Solano, and A. Huete, “MODIS vegetation index user’s guide (MOD13 series),” *Univ. Arizona Veg. Index Phenol. Lab*, 2015.

[62] A. Huete, K. Didan, T. Miura, E. P. Rodriguez, X. Gao, and L. G. Ferreira, “Overview of the radiometric and biophysical performance of the MODIS vegetation indices,” *Remote Sens. Environ.*, vol. 83, no. 1–2, pp. 195–213, 2002.

[63] L. Li, M. A. Friedl, Q. Xin, J. Gray, Y. Pan, and S. Frolking, “Mapping crop cycles in China using MODIS-EVI time series,” *Remote Sens.*, vol. 6, no. 3, pp. 2473–2493, 2014.

[64] N. T. Son, C. F. Chen, C. R. Chen, V. Q. Minh, and N. H. Trung, “A comparative analysis of multitemporal MODIS EVI and NDVI data for large-scale rice yield estimation,” *Agric. For. Meteorol.*, vol. 197, pp. 52–64, 2014.

[65] A. Paliwal, A. Laborte, A. Nelson, and R. K. Singh, “Salinity stress detection in rice crops using time series MODIS VI data,” *Int. J. Remote Sens.*, vol. 40, no. 21, pp. 8186–8202, 2019.

[66] K. S. Kibret, C. Marohn, and G. Cadisch, “Use of MODIS EVI to map crop phenology, identify cropping systems, detect land use change and drought risk in Ethiopia–an application of Google Earth Engine,” *Eur. J. Remote Sens.*, vol. 53, no. 1, pp. 176–191, 2020.

[67] I. R. Orimoloye, J. A. Belle, and O. O. Ololade, “Drought disaster monitoring using MODIS derived index for drought years: A space-based information for ecosystems and environmental conservation,” *J. Environ. Manage.*, vol. 284, p. 112028, 2021.

[68] D. Zanaga *et al.*, “ESA WorldCover 10 m 2020 v100,” 2021.

[69] G. of P. Economic and Statistical Organization, “Statistical Abstract of Punjab - 2020,” 2020.

[70] G. of P. Economic and Statistical Organization, “Statistical Abstract of Punjab - 2019,” 2019.

[71] G. of P. Economic and Statistical Organization, “Statistical Abstract of Punjab - 2018,” 2018.

[72] J. R. McWilliam, “The dimensions of drought,” 1989.

[73] J. S. Boyer, “Plant productivity and environment,” *Science (80-.).*, vol. 218, no. 4571, pp. 443–448, 1982.

[74] D. Bassi, M. Menossi, and L. Mattiello, “Nitrogen supply influences photosynthesis establishment along the sugarcane leaf,” *Sci. Rep.*, vol. 8, no. 1, pp. 1–13, 2018.

[75] J. A. Raven, “RNA function and phosphorus use by photosynthetic organisms,” *Front. Plant Sci.*, vol. 4, p. 536, 2013.

[76] X. Xu *et al.*, “Effects of potassium levels on plant growth, accumulation and distribution of carbon, and nitrate metabolism in apple dwarf rootstock seedlings,” *Front. Plant Sci.*, vol. 11, p. 904, 2020.

[77] D. Haboudane, J. R. Miller, N. Tremblay, P. J. Zarco-Tejada, and L. Dextraze, “Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture,” *Remote Sens. Environ.*, vol. 81, no. 2–3, pp. 416–426, 2002.

[78] P. L. Hatfield and P. J. Pinter Jr, “Remote sensing for crop protection,” *Crop Prot.*, vol. 12, no. 6, pp. 403–413, 1993.

[79] K. Dineva and T. Atanasova, “SYSTEMATIC LOOK AT MACHINE LEARNING ALGORITHMS–ADVANTAGES, DISADVANTAGES AND PRACTICAL APPLICATIONS,” *Int. Multidiscip. Sci. GeoConference SGEM*, vol. 20, no. 2.1, pp. 317–324, 2020.

[80] P. S. Foundation, “Python.” [Online]. Available: https://www.python.org/. [Accessed: 16-Mar-2021].

[81] S.-L. Developers, “Scikit-learn.” [Online]. Available: https://scikit-learn.org/stable/index.html. [Accessed: 16-Mar-2021].

[82] Google Research, “Colaboratory (Colab).” [Online]. Available: https://colab.research.google.com/. [Accessed: 16-Mar-2021].

[83] J. Verrelst *et al.*, “Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for Sentinel-2 and-3,” *Remote Sens. Environ.*, vol. 118, pp. 127–139, 2012.

[84] J. Verrelst, L. Alonso, G. Camps-Valls, J. Delegido, and J. Moreno, “Retrieval of vegetation biophysical parameters using Gaussian process techniques,” *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1832–1843, 2011.

[85] S. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” *arXiv Prepr. arXiv1705.07874*, 2017.

[86] S. Mazzanti, “SHAP Values Explained Exactly How You Wished Someone Explained to You,” *https://towardsdatascience.com/*, 2020. [Online]. Available: https://towardsdatascience.com/shap-explained-the-way-i-wish-someone-explained-it-to-me-ab81cc69ef30. [Accessed: 05-Jul-2021].

[87] C. Molnar, *Interpretable machine learning*. Lulu. com, 2020.

[88] C. L. Braun and S. N. Smirnov, “Why is water blue?,” *J. Chem. Educ.*, vol. 70, no. 8, p. 612, 1993.

[89] B. B. Naik, H. R. Naveen, G. Sreenivas, K. K. Choudary, D. Devkumar, and J. Adinarayana, “Identification of Water and Nitrogen Stress Indicative Spectral Bands Using Hyperspectral Remote Sensing in Maize During Post-Monsoon Season,” *J. Indian Soc. Remote Sens.*, vol. 48, no. 12, pp. 1787–1795, 2020.

[90] Q. Zhang, Q. Li, and G. Zhang, “Rapid determination of leaf water content using VIS/NIR spectroscopy analysis with wavelength selection,” *Spectrosc. An Int. J.*, vol. 27, no. 2, pp. 93–105, 2012.

[91] S. L. Ustin *et al.*, “Retrieval of foliar information about plant pigment systems from high resolution spectroscopy,” *Remote Sens. Environ.*, vol. 113, pp. S67–S77, 2009.

[92] P. H. Sampson, P. J. Zarco-Tejada, G. H. Mohammed, J. R. Miller, and T. L. Noland, “Hyperspectral remote sensing of forest condition: Estimating chlorophyll content in tolerant hardwoods,” *For. Sci.*, vol. 49, no. 3, pp. 381–391, 2003.

[93] S. Zhang *et al.*, “Structural basis for enzymatic photocatalysis in chlorophyll biosynthesis,” *Nature*, vol. 574, no. 7780, pp. 722–725, 2019.

[94] V. Ciganda, A. Gitelson, and J. Schepers, “Non-destructive determination of maize leaf and canopy chlorophyll content,” *J. Plant Physiol.*, vol. 166, no. 2, pp. 157–167, 2009.

[95] Y. Wang, D. Wang, P. Shi, and K. Omasa, “Estimating rice chlorophyll content and leaf nitrogen concentration with a digital still color camera under natural light,” *Plant Methods*, vol. 10, no. 1, pp. 1–11, 2014.

[96] Q. Cao *et al.*, “Active canopy sensing of winter wheat nitrogen status: An evaluation of two sensor systems,” *Comput. Electron. Agric.*, vol. 112, pp. 54–67, 2015.

[97] D. Li *et al.*, "Estimation of area-and mass-based leaf nitrogen contents of wheat and rice crops from water-removed spectra using continuous wavelet analysis," *Plant Methods*, vol. 14, no. 1, pp. 1–20, 2018.

[98] Z. Peng *et al.*, "MiR399d and epigenetic modification comodulate anthocyanin accumulation in Malus leaves suffering from phosphorus deficiency," *Plant. Cell Environ.*, vol. 43, no. 5, pp. 1148–1159, 2020.

[99] S. L. Osbourne, J. S. Schepers, D. Francis, and M. R. Schlemmer, "Detection of phosphorus and nitrogen deficiencies in corn using spectral radiance measurements," 2002.

[100] G. R. Mahajan, R. N. Sahoo, R. N. Pandey, V. K. Gupta, and D. Kumar, "Using hyperspectral remote sensing techniques to monitor nitrogen, phosphorus, sulphur and potassium in wheat (Triticum aestivum L.)," *Precis. Agric.*, vol. 15, no. 5, pp. 499–522, 2014.

[101] Y. Özyiğit and M. Bilgen, "Use of spectral reflectance values for determining nitrogen, phosphorus, and potassium contents of rangeland plants," *J. Agric. Sci. Technol.*, vol. 15, no. 7, pp. 1537–1545, 2013.

[102] M. Sibanda, O. Mutanga, M. Rouget, and J. Odindi, "Exploring the potential of in situ hyperspectral data and multivariate techniques in discriminating different fertilizer treatments in grasslands," *J. Appl. Remote Sens.*, vol. 9, no. 1, p. 96033, 2015.

[103] R. M. Pope and E. S. Fry, "Absorption spectrum (380–700 nm) of pure water. II. Integrating cavity measurements," *Appl. Opt.*, vol. 36, no. 33, pp. 8710–8723, 1997.

[104] D. A. Sims and J. A. Gamon, "Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages," *Remote Sens. Environ.*, vol. 81, no. 2–3, pp. 337–354, 2002.

[105] A. A. Gitelson, Y. Gritz, and M. N. Merzlyak, "Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves," *J. Plant Physiol.*, vol. 160, no. 3, pp. 271–282, 2003.

[106] J. G. Tallada and M. A. Ramos, "Visible-near-infrared absorbance spectroscopy for rapid estimation of leaf nitrogen contents of Philippine rice cultivars," *Cogent Food Agric.*, vol. 4, no. 1, p. 1487254, 2018.

[107] D. Rodriguez, G. J. Fitzgerald, R. Belford, and L. K. Christensen, "Detection of nitrogen deficiency in wheat from spectral reflectance indices and basic crop eco-physiological concepts," *Aust. J. Agric. Res.*, vol. 57, no. 7, pp. 781–789, 2006.

[108] K. R. Thorp, G. Wang, K. F. Bronson, M. Badaruddin, and J. Mon, "Hyperspectral data mining to identify relevant canopy spectral features for estimating durum wheat growth, nitrogen status, and grain yield," *Comput.*

Electron. Agric., vol. 136, pp. 1–12, 2017.

[109] M. N. Merzlyak, A. E. Solovchenko, and A. A. Gitelson, "Reflectance spectral features and non-destructive estimation of chlorophyll, carotenoid and anthocyanin content in apple fruit," *Postharvest Biol. Technol.*, vol. 27, no. 2, pp. 197–211, 2003.

[110] J. Segarra, M. L. Buchaillot, J. L. Araus, and S. C. Kefauver, "Remote sensing for precision agriculture: Sentinel-2 improved features and applications," *Agronomy*, vol. 10, no. 5, p. 641, 2020.

[111] M. Schlemmer *et al.*, "Remote estimation of nitrogen and chlorophyll contents in maize at leaf and canopy levels," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 25, pp. 47–54, 2013.

[112] D. S. Green, J. E. Erickson, and E. L. Kruger, "Foliar morphology and canopy nitrogen as predictors of light-use efficiency in terrestrial vegetation," *Agric. For. Meteorol.*, vol. 115, no. 3–4, pp. 163–171, 2003.

[113] S. V Ollinger *et al.*, "Canopy nitrogen, carbon assimilation, and albedo in temperate and boreal forests: Functional relations and potential climate feedbacks," *Proc. Natl. Acad. Sci.*, vol. 105, no. 49, pp. 19336–19341, 2008.

[114] D. Stroppiana, M. Boschetti, P. A. Brivio, and S. Bocchi, "Plant nitrogen concentration in paddy rice from field canopy hyperspectral radiometry," *F. Crop. Res.*, vol. 111, no. 1–2, pp. 119–129, 2009.

[115] M. N. Almasri and J. J. Kaluarachchi, "Assessment and management of long-term nitrate pollution of ground water in agriculture-dominated watersheds," *J. Hydrol.*, vol. 295, no. 1–4, pp. 225–245, 2004.

[116] M. Diacono, P. Rubino, and F. Montemurro, "Precision nitrogen management of wheat. A review," *Agron. Sustain. Dev.*, vol. 33, no. 1, pp. 219–241, 2013.

[117] R. F. Muñoz-Huerta, R. G. Guevara-Gonzalez, L. M. Contreras-Medina, I. Torres-Pacheco, J. Prado-Olivarez, and R. V Ocampo-Velazquez, "A review of methods for sensing the nitrogen status in plants: advantages, disadvantages and recent advances," *sensors*, vol. 13, no. 8, pp. 10823–10843, 2013.

[118] J. C. Craig, "Multi-scale remote sensing techniques for vegetation stress detection.," University of Florida, Gaineville, Florida, 2002.

[119] I. B. Strachan, E. Pattey, and J. B. Boisvert, "Impact of nitrogen and environmental conditions on corn as detected by hyperspectral reflectance," *Remote Sens. Environ.*, vol. 80, no. 2, pp. 213–224, 2002.

[120] T. Shi, L. Cui, J. Wang, T. Fei, Y. Chen, and G. Wu, "Comparison of multivariate methods for estimating soil total nitrogen with visible/near-infrared spectroscopy," *Plant Soil*, vol. 366, no. 1, pp. 363–375, 2013.

[121] H. T. Nguyen and B. W. Lee, "Assessment of rice leaf growth and nitrogen status by hyperspectral canopy reflectance and partial least square regression," *Eur. J. Agron.*, vol. 24, no. 4, pp. 349–356, 2006.

[122] P. M. Hansen and J. K. Schjoerring, “Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression,” *Remote Sens. Environ.*, vol. 86, no. 4, pp. 542–553, 2003.

[123] A. Karnieli, “Wheat and maize monitoring based on ground spectral measurements and multivariate data analysis,” *J. Appl. Remote Sens.*, vol. 1, no. 1, p. 013530, 2007.

[124] Agustin Pimstein David J Bonfil Arnon Karnieli, “Wheat Canopy Spectra Collected Throughout Growing Season. Data set. Available on-line [http://ecosis.org] from the Ecological Spectral Information System (EcoSIS).” [Online]. Available: https://ecosis.org/package/wheat-canopy-spectra-collected-throughout-growing-season. [Accessed: 20-Jan-2021].

[125] F. Pedregosa *et al.*, “Scikit-learn: Machine learning in Python,” *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[126] S. Lundberg, “SHAP documentation.” [Online]. Available: https://shap.readthedocs.io/en/latest/. [Accessed: 05-Jul-2021].

[127] G. P. Asner *et al.*, “Taxonomy and remote sensing of leaf mass per area (LMA) in humid tropical forests,” *Ecol. Appl.*, vol. 21, no. 1, pp. 85–98, 2011.

[128] G. Walburg, M. E. Bauer, C. S. T. Daughtry, and T. L. Housley, “Effects of nitrogen nutrition on the growth, yield, and reflectance characteristics of corn canopies 1,” *Agron. J.*, vol. 74, no. 4, pp. 677–683, 1982.

[129] L. Ercoli, M. Mariotti, A. Masoni, and F. Massantini, “Relationship between nitrogen and chlorophyll content and spectral properties in maize leaves,” *Eur. J. Agron.*, vol. 2, no. 2, pp. 113–117, 1993.

[130] T. M. Blackmer, J. S. Schepers, G. E. Varvel, and E. A. Walter-Shea, “Nitrogen deficiency detection using reflected shortwave radiation from irrigated corn canopies,” *Agron. J.*, vol. 88, no. 1, pp. 1–5, 1996.

[131] S. Graeff, W. Claupein, and S. Schubert, “Identification of nutrient deficiencies in corn using reflectance measurements,” *Digit. imaging Spectr. Tech. Appl. to Precis. Agric. Crop Physiol.*, vol. 66, pp. 165–175, 2004.

[132] P. D. Dao, Y. He, and C. Proctor, “Plant drought impact detection using ultra-high spatial resolution hyperspectral images and machine learning,” *Int. J. Appl. Earth Obs. Geoinf.*, vol. 102, p. 102364, 2021.

[133] L. Wang *et al.*, “Estimation of paddy rice nitrogen content and accumulation both at leaf and plant levels from UAV hyperspectral imagery,” *Remote Sens.*, vol. 13, no. 15, p. 2956, 2021.

[134] M. Grieco *et al.*, “Dynamics and genetic regulation of leaf nutrient concentration in barley based on hyperspectral imaging and machine learning,” *Plant Sci.*, vol. 315, p. 111123, 2022.

[135] S. Weksler, O. Rozenstein, N. Haish, M. Moshelion, R. Wallach, and E. Ben-

Dor, “Detection of Potassium Deficiency and Momentary Transpiration Rate Estimation at Early Growth Stages Using Proximal Hyperspectral Imaging and Extreme Gradient Boosting,” *Sensors*, vol. 21, no. 3, p. 958, 2021.

[136] M. G. N. K. W. A. A. Y. G. and J. C. Schnable, “Leaf spectra and physiological and chemical traits from maize grown under nitrogen stress. Data set. Available on-line [http://ecosis.org] from the Ecological Spectral Information System (EcoSIS),” 2020. [Online]. Available: https://ecosis.org/package/leaf-spectra-and-physiological-and-chemical-traits-from-maize-grown-under-nitrogen-stress. [Accessed: 10-Oct-2021].

[137] Y. Ge *et al.*, “High-throughput analysis of leaf physiological and chemical traits with VIS–NIR–SWIR spectroscopy: a case study with a maize diversity panel,” *Plant Methods*, vol. 15, no. 1, pp. 1–12, 2019.

[138] Y. Zhai, L. Cui, X. Zhou, Y. Gao, T. Fei, and W. Gao, “Estimation of nitrogen, phosphorus, and potassium contents in the leaves of different plants using laboratory-based visible and near-infrared reflectance spectroscopy: comparison of partial least-square regression and support vector machine regression met,” *Int. J. Remote Sens.*, vol. 34, no. 7, pp. 2502–2518, 2013.

[139] R. Raj, J. P. Walker, R. Pingale, B. N. Banoth, and A. Jagarlapudi, “Leaf nitrogen content estimation using top-of-canopy airborne hyperspectral data,” *Int. J. Appl. Earth Obs. Geoinf.*, vol. 104, p. 102584, 2021.

[140] E. R. Hunt Jr, C. S. T. Daughtry, and L. Li, “Feasibility of estimating leaf water content using spectral indices from WorldView-3’s near-infrared and shortwave infrared bands,” *Int. J. Remote Sens.*, vol. 37, no. 2, pp. 388–402, 2016.

[141] I. Herrmann, A. Karnieli, D. J. Bonfil, Y. Cohen, and V. Alchanatis, “SWIR-based spectral indices for assessing nitrogen content in potato fields,” *Int. J. Remote Sens.*, vol. 31, no. 19, pp. 5127–5143, 2010.

[142] J. G. Ellis, E. S. Lagudah, W. Spielmeyer, and P. N. Dodds, “The past, present and future of breeding rust resistant wheat,” *Front. Plant Sci.*, vol. 5, p. 641, 2014.

[143] M. S. Hovmøller, C. K. Sørensen, S. Walter, and A. F. Justesen, “Diversity of Puccinia striiformis on cereals and grasses,” *Annu. Rev. Phytopathol.*, vol. 49, pp. 197–217, 2011.

[144] S. K. Sandhu, L. K. Dhaliwal, and P. P. S. Pannu, “Role of microclimate in management of yellow rust (Puccinia striformis f sp tritici) of wheat (Triticum aestivum) under Ludhiana conditions,” *Indian J. Agric. Sci.*, vol. 86, no. 7, 2016.

[145] Punjab Agricultural University, “Regularly monitor wheat crop to manage yellow rust,” 2020. [Online]. Available: https://www.pau.edu/index.php?_act=manageEvent&DO=viewEventDetail&intEventID=4971. [Accessed: 23-Jun-2020].

[146] T. T. N. Service, “Yellow rust hits wheat in some areas of dist,” *The Tribune*, Jalandhar, Feb-2020.

[147] J. W. Rouse, R. H. Haas, J. A. Schell, and D. W. Deering, “Monitoring vegetation systems in the Great Plains with ERTS,” *NASA Spec. Publ.*, vol. 351, p. 309, 1974.

[148] C. S. T. Daughtry, C. L. Walthall, M. S. Kim, E. B. De Colstoun, and J. E. McMurtrey Iii, “Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance,” *Remote Sens. Environ.*, vol. 74, no. 2, pp. 229–239, 2000.

[149] E. Nestola *et al.*, “Are optical indices good proxies of seasonal changes in carbon fluxes and stress-related physiological status in a beech forest?,” *Sci. Total Environ.*, vol. 612, pp. 1030–1041, 2018.

[150] G. Guyot and F. Baret, “4th International Colloquium" Spectral signatures of objects in remote sensing,” *Paris ESA Publ. Aussois*, pp. 279–286, 1988.

[151] B.-C. Gao, “NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space,” *Remote Sens. Environ.*, vol. 58, no. 3, pp. 257–266, 1996.

[152] E. H. Wilson and S. A. Sader, “Detection of forest harvest type using multiple dates of Landsat TM imagery,” *Remote Sens. Environ.*, vol. 80, no. 3, pp. 385–396, 2002.

[153] European Space Agency Signature, “Sentinel-2 MSI Technical Guide.” [Online]. Available: https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-2-msi. [Accessed: 22-Apr-2022].

[154] NASA, “OPERATIONAL LAND IMAGER.” [Online]. Available: https://landsat.gsfc.nasa.gov/satellites/landsat-8/spacecraft-instruments/operational-land-imager/. [Accessed: 22-Apr-2022].

[155] USGS, “USGS EROS Archive - Landsat Archives - Landsat 7 Enhanced Thematic Mapper Plus (ETM+).” [Online]. Available: https://www.usgs.gov/centers/eros/science/usgs-eros-archive-landsat-archives-landsat-7-enhanced-thematic-mapper-plus-etm. [Accessed: 22-Apr-2022].

Annexure-1

Crop Experiment Photographs

Wheat at P.A.U. Field- 2020-21

100%NPK, IW 1.0

December **January** **February** **March**

Wheat at P.A.U. Field- January 2021

100% NPK **75% NPK** **50% NPK**

100% NP **100% N**

IW/CPE = 1.0

N: Nitrogen, P: Phosphorus, K: Potassium, IW: Irrigation Water, CPE: Cumulative Pan Evaporation, PAU: Punjab Agricultural University

Maize at P.A.U. Field- August 2020

100%NPK, IW 1.0

August | September | October

Maize at P.A.U. Field- 2020

100% NPK | 75% NPK | 50% NPK | 100% NP | 100% N

IW/CPE = 1.0

N: Nitrogen, P: Phosphorus, K: Potassium, IW: Irrigation Water, CPE: Cumulative Pan Evaporation, PAU: Punjab Agricultural University

List of Publications

1) **Singh, H**., Roy, A., Setia, R.K. and Pateriya B**.** (2021) Estimation of nitrogen content in wheat from proximal hyperspectral data using machine learning and explainable artificial intelligence (XAI) approach. ***Model. Earth Syst. Environ.*** *https://doi.org/10.1007/s40808-021-01243-z*(SCOPUS Indexed)
2) **Singh H**., Roy A., Setia R.K., Pateriya B. (2022) Estimation of chlorophyll, macronutrients and water content in maize from hyperspectral data using machine learning and explainable artificial intelligence techniques. ***Remote Sensing Letters*****,** 13:10, 969-979, *https://doi.org/10.1080/2150704X.2022.2114108*
3) **Singh, H.**, Roy, A., Setia, R. K., & Pateriya, B. (2021). 9 Proximal and remote sensing of abiotic stress in crops. *Intelligent Circuits and Systems*, 54, CRC Press. *https://doi.org/10.1201/9781003129103* (SCOPUS Indexed)
4) **Singh H.,** Roy A., Patel S., Pateriya B. (2021) Object-Based Classification of Sentinel-2 Data Using Free and Open-Source Machine Learning and GIS Tools. In: Tiwari A., Ahuja K., Yadav A., Bansal J.C., Deep K., Nagar A.K. (eds) Soft Computing for Problem Solving. Advances in Intelligent Systems and Computing, vol 1392. Springer, Singapore. *https://doi.org/10.1007/978-981-16-2709-5_58* (SCOPUS Indexed)
5) **Singh, H**., Roy, A., Setia, R.K. and Pateriya B**.** Classification of yellow rust of wheat from Sentinel-2 satellite imagery using deep learning artificial neural network. Submitted to a SCOPUS Journal, **Under-Review**
6) Spatio-Temporal analysis of crop conditions using MODIS EVI in Google Earth Engine ***Singh H****., Roy A., Patel S., Pateriya B. 4th International Conference on Intelligent Circuits and Systems (ICICS 2022).* 8-9April 2022.SCOPUS Indexed Conference.
7) Simulation of multispectral data using hyperspectral data for crop stress studies ***Singh H****., Roy A., Setia R., Pateriya B. International Conference on Small Satellites (ICSS- 2022)* 29-30April 2022.SCOPUS Indexed Conference.