

https://dspace.org/

Destructive estimation of canopy chlorophyll content in smallholder crops using UAV and Sentinel-2

Rampyapedi, Napoleon

2025
University of the Witwatersrand, Johannesburg
https://hdl.handle.net/10539/50275
Rampyapedi, Napoleon. (2025). Destructive estimation of canopy chlorophyll content in smallholder crops using UAV and Sentinel-2 [Master's dissertation, University of the Witwatersrand, Johannesburg]. WIReDSpace. https://hdl.handle.net/10539/50275

Downloaded from a DSpace repository. https://dspace.org/



Destructive estimation of canopy chlorophyll content in smallholder crops using UAV and Sentinel-2

Napoleon Rampyapedi, 1769768

Supervisor: Prof. Elhadi Adam

Plagiarism Declaration

I, Napoleon Rampyapedi declare that this dissertation is my own, unaided work. This report is submitted in partial fulfilment of the requirements for the degree Master of Science (GIS and Remote Sensing) in Geography at the University of the Witwatersrand. I declare that it has never been previously submitted for any other degree at any other university.

Signature: 

Date: 09 October 2024

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Prof. Elhadi Adam, for his invaluable guidance, support, and encouragement throughout this research project. His insightful comments and feedback were instrumental in shaping the direction of my research.

I also want to express my gratitude to the South African National Space Agency (SANSA) for their financial assistance in making this research possible.

Finally, I would like to thank my parents and siblings for their unwavering love and encouragement throughout this challenging project. Their belief in me kept me motivated during difficult times.

Abstract

Agriculture plays a crucial role in South Africa's economy and food security, making crop health monitoring essential. Chlorophyll content is a key indicator of crop health, and its assessment can provide valuable insights into crop conditions. This study focused on a smallholder farm in the Free State Province, South Africa, aiming to estimate the chlorophyll content in wheat plants using UAV multispectral data and Sentinel-2 satellite imagery from October 2021. Vegetation indices derived from these remote sensing datasets, along with field-measured chlorophyll data obtained using a SPAD meter, were used in the analysis. Two machine learning algorithms, Support Vector Machine (SVM) and Random Forest (RF) regression, were employed to build predictive models. Four models were developed: SVM-UAV, SVM-Sentinel-2, RF-UAV, and RF-Sentinel-2.

The results indicated that the RF model using Sentinel-2 data outperformed the others, achieving an R^2 of 0.14 and an RMSE of 5.11 The most significant indices in this model were NDVI, GNDVI, and CIgreen. The other models produced weaker predictions: RF-UAV (R^2 = 0.038, RMSE = 6.71), SVM-Sentinel-2 (R^2 = 0.022, RMSE = 5.35), and SVM-UAV (R^2 = 0.020, RMSE = 5.41). This study demonstrates the potential of UAV multispectral and Sentinel-2 imagery for predicting and mapping chlorophyll content in wheat crops, particularly the effectiveness of Sentinel-2 for large-scale assessments.

Contents

Plagiarism Declaration i

Acknowledgements ii

Abstract iii

List of figures vi

List of tables vii

Acronyms viii

Chapter 1 1

General Introduction 1

1.1 Background 1

1.2 Problem statement 2

1.3 Aims and objectives 4

Research report outline 5

Chapter 2 6

Literature review 6

2.1 Crops quality and Chlorophyll 6

2.2 The use of remote sensing for the mapping and monitoring chlorophyll content 6

2.2.1 The use of Hyperspectral datasets and chlorophyll content 6

2.2.2 The use of traditional Multispectral data for chlorophyll content 9

2.2.3 The use of new generation Multispectral data for chlorophyll content 11

2.2.4 The usage of UAV data for chlorophyll content 14

2.3 Conclusion 16

Chapter 3 18

Materials and Methods 18

3.1 Study site 18

3.2 Flowchart of the methodology utilized in this study 19

3.3 Field data and ground data collection 21
3.4 UAV Multispectral Imagery acquisition and preprocessing 22
3.5 Sentinel-2 data acquisition and pre-processing 23
3.6 Data analysis 25
3.6.1 Spectral indices 25
3.6.2 SVM regression 26
3.6.3 RF regression 27
3.6.4 Variable Importance and selection 28
3.6.5 Model validation 29
Chapter 4 30
Results 30
4.1 Description of field data 30
4.2 Correlation between chlorophyll and derived spectral indices 30
4.3 Chlorophyll prediction using SVM and RF machine learning regression algorithms 32
4.4 Variable importance of spectral indices in prediction of chlorophyll content 34
4.5 Spatial distribution of wheat chlorophyll concentration 36
Chapter 5 38
Discussion 38
Chapter 6 42
Conclusions 42
References 43

List of figures

Figure 1: Location of the study region, located in Free State, South Africa.....19
Figure 2: Flowchart of the methodology followed in this study.....20
Figure 3: Correlation plots for spectral indices (UAV and Sentinel – 2 derived) and field measured chlorophyll.....31
Figure 4: Prediction plots for chlorophyll estimation of wheat crops during the maturity stage for the month of October 2021using the test datasets (30%).....33
Figure 5: Permutation based variable importance plots for both the SVM and RF analysis...35
Figure 6: Spatial distribution maps of chlorophyll in the study region generated using all models employed in the study.....37

List of tables

Table 1: MicaSense RedEdge-MX Multispectral sensor characteristics. 22
Table 2: Characteristics of the Sentinel-2 multispectral satellite imagery. 24
Table 3: The study's spectral indices and the formulas they utilize. 26
Table 4: Statistics of the field measured chlorophyll using SPAD-502 meter for the 100 sampling points used in this study. 30

Acronyms

ANN - Artificial Neural Network

ASD - Analytical Spectral Devices

CCCI - Canopy Chlorophyll Content Index

CIgreen - Chlorophyll Index Green

CVI - Chlorophyll Vegetation Index

CIrededge - Chlorophyll Index Red-edge

CRP - Common Remote Sensing Platform

DNN - Deep Neural Network

DLS - Digital Level Sensor

ENVI - Environment for Visualizing Images

EVI - Enhanced Vegetation Index

GPS - Global Positioning System

GNDVI - Green Normalized Difference Vegetation Index

INFORM - Intelligent Non-contact Forestry Measurement

KNN - k-Nearest Neighbors

LAI - Leaf Area Index

MERIS - Medium Resolution Imaging Spectrometer

MODIS - Moderate Resolution Imaging Spectroradiometer

MSAVI - Modified Soil Adjusted Vegetation Index

MSE - Mean Squared Error

NDRE - Normalized Difference Red Edge

NDVI - Normalized Difference Vegetation Index

OLI - Operational Land Imager

OLSR - Ordinary Least Squares Regression

OOB - Out-of-Bag (Random Forest parameter)

PCA – Principal Component Analysis

PLSR - Partial Least Squares Regression

PROSAIL - PROSPECT + SAIL Radiative Transfer Models

PSR - Photochemical Reflectance Index

RMSE - Root Mean Square Error

RF - Random Forest

SAVI - Soil-Adjusted Vegetation Index

SAWS - South African Weather Service

SLR - Simple Linear Regression

SNAP - Sentinel Application Platform

SPOT - Satellite Pour l'Observation de la Terre

SPAD - Soil-Plant Analysis Development

SVR - Support Vector Regression

SVM - Support Vector Machine

SWIR - Shortwave Infrared

TM - Thematic Mapper

UAV - Unmanned Aerial Vehicle

VNIR - Visible and Near-Infrared

Chapter 1

General Introduction

1.1 Background

Agriculture plays a vital role in the economy by maintaining food security and providing employment opportunities (Kamara *et al.,* 2019). Monitoring crop growth and health is crucial for sustaining food security and accurately estimating crop yields (Battude *et al.,* 2016; Zhu *et al.,* 2020). Among the various indicators of crop health, leaf chlorophyll content is particularly important, as it serves as a key measure for assessing crop health, stress, growth, and nitrogen status (Kanning *et al.,* 2018; Singhal *et al.,* 2019; Zhu *et al.,* 2020). Therefore, accurately predicting chlorophyll content is critical for effective crop health assessment and soil fertility management (Shah *et al.,* 2019). Traditional methods for determining chlorophyll content involve in-situ measurements and laboratory analysis. However, these approaches are often labour-intensive, time-consuming, costly, and destructive (Fernández-Marín *et al.,* 2015; Shah *et al.,* 2019). The limitations of these traditional methods highlight the need for more efficient and non-destructive techniques for monitoring chlorophyll content.

To enhance the monitoring of crop growth, remote sensing technologies have been increasingly adopted for estimating chlorophyll content in crops (Croft *et al.,* 2015; Tahir *et al.,* 2018). One such approach involves the use of field-based spectroscopy, a direct measurement technique that utilizes devices like the SPAD chlorophyll meter. This device estimates chlorophyll content by measuring the leaf's absorbance in the red and near-infrared wavelength regions (Pimstein *et al.,* 2011; Süß *et al.,* 2015). Ground-based measurements offer the advantage of proving additional spatial details while simultaneously allowing for spectral and image analysis (Liu *et al.,* 2014).

Hyperspectral remote sensing data is also widely used for estimating canopy chlorophyll content due to its narrow and continuous spectral bands. These bands provide detailed spectra that are highly responsive to crop characteristics, including chlorophyll (Xue and Yang, 2009; An *et al.,* 2020). However, despite its potential, hyperspectral remote sensing has certain limitations, such as the need for specialized skills in the data processing and the high costs associated with data acquisition (Govender *et al.,* 2007; Shoko and Mutanga, 2017).

Broad-band multispectral remote sensing data, known for its high spectral resolution and average spatial resolution, is another approach used for chlorophyll prediction. Its accessibility and capability for larger-scale chlorophyll prediction make it a practical alternative, although it does have limitations in spatial resolution (Oyama *et al.,* 2007; Marshall and Thenkabail, 2015; Shoko and Mutanga, 2017).

High-resolution commercial satellite imagery from sources such as WorldView, QuickBird, SPOT, and GeoEye is also utilized for chlorophyll estimation. These satellites offer high spatial resolution and incorporate spectral bands from the red to near-infrared regions of the electromagnetic spectrum, making them suitable for this purpose. However, they are expensive to acquire, have fewer spectral bands, and require substantial computing capacity (Marshall and Thenkabail, 2015; Tahir *et al.,* 2018). Additionally, the accuracy of airborne and spaceborne remote sensing tools can be compromised by adverse weather conditions, particularly cloud cover.

Unmanned Aerial Vehicles (UAVs) have revolutionized crop monitoring by providing real-time, high temporal, and high spatial resolution remote sensing data (Zhang *et al.,* 2021; Narmilan *et al.,* 2022). Equipped with multispectral sensors, UAVs are effective in predicting chlorophyll content due to their ability to capture data across visible and near-infrared bands (Singhal *et al.,* 2019). For example, Wu *et al.* (2023) used UAV technology to estimate chlorophyll content in wheat utilizing sensors that included red-edge and near-infrared bands. Similarly, Zhu *et al.* (2020) employed hyperspectral UAV data with 125 channels demonstrating high sensitivity to crop characteristics for accurate chlorophyll prediction.

1.2 Problem statement

Smallholder farms across Southern Africa face problems such as lower yields and reduced profits due to changes in climate (Kamara *et al.,* 2019; Salami *et al.,* 2010; Brewer *et al.,* 2022). Smallholder farms are also faced with a challenge of lacking resources to have high yield of crops production (Brewer *et al.,* 2022). Another challenged faced by smallholder farms, particularly in the Free State province, is their reliance on rainfall for crop growth and yield. Variations in rainfall amounts can significantly affect crop health and directly impact yields (Gcayi *et al.,* 2019). It is therefore important to manage the health of crops, this can be done using crop health indicators which include crop yield, nitrogen, and quality of soil (Whitford *et al.,* 1998; Sahu *et al.,* 2019). Chlorophyll concentration is also an indicator of crop health

conditions which has been used by various literature to monitor crop health (Singhal *et al.,* 2019; Brewer *et al.*, 2022). Traditional in-situ and lab measurements of chlorophyll content are deemed time consuming, destructive, costly, require more labour and data collection is complicated due to analytical methods (Fernández-Marín *et al.,* 2015; Shah *et al.,* 2019). Although the open-source satellite imagery such as multispectral data can be utilized in predicting chlorophyll concentration of crops for effective monitoring and management the limitation is their inefficacy to attain the escalating requirement for high temporal and spatial data which is important in observing crop health and growth in smallholder farms (Brewer *et al.*, 2022).

Landsat 7 – 8 and MODIS satellite sensors have pixel resolutions of 30 m × 30m and 500 m × 500m respectively (Xue and Su, 2017). Furthermore, the orbit period of Landsat 7-8 and SPOT satellites are 16 and 26 days respectively (Xue and Su, 2017). Both the pixel resolutions and orbit period of the mentioned multispectral satellites are a challenge for effective monitoring of crops in smallholder farms. Moreover, high spatial resolution such as WorldView 2-3 satellite sensors which can predict chlorophyll content of crops, but they have a limitation of cost, high computation capacity required and they cannot provide information for less time monitoring of crops (Zhang and Kovacs, 2012; Xue and Su, 2017). The other limitation which is posed by satellite imagery for crop chlorophyll estimation is that information regarding crops is mostly needed during growing season, and the growing seasons of crops are generally during rainy seasons as a result weather conditions can make the acquisition of satellite imagery difficult (Zhang and Kovacs, 2012). Multispectral UAV and Sentinel-2 data present a valuable opportunity for assessing chlorophyll content in crops on smallholder farms. This is due to the high temporal and spatial resolution of multispectral UAVs, their ability to capture imagery below cloud cover, and the inclusion of sensors with bands in the visible, NIR, and red-edge zones, which are sensitive to chlorophyll (Narmilan *et al.*, 2022). The amount of chlorophyll in crops could be evaluated using vegetation indices which are derived from multispectral UAV bands and machine learning algorithms (Brewer *et al.,* 2022). Vegetation indices are used to retrieve spectral information of crops such as growth status and vegetation cover (Xue and Su, 2017). The NDVI is widely employed in assessing crop health as it offers insights into changes in crop growth and overall well-being. There are vegetation indices which were specifically developed for estimating chlorophyll content. These indices use spectral bands that are highly responsive to chlorophyll levels in crops. Marshall and Thenkabail (2015) elaborated that crop chlorophyll absorbs in the visible zone (620-700 nm), transmits in the red-edge zone (700-740

nm), and strongly reflects within NIR region (740-1000 nm). Some vegetation metrics that have demonstrated a strong correlation with crop chlorophyll levels are GNDVI, CCCI, modified chlorophyll absorption index, and GCI (Brewer *et al.,* 2022). Furthermore, Sentinel-2 imagery incorporates the visible, NIR, and red-edge bands, which are sensitive to crop chlorophyll. These bands offer good spectral resolution, making Sentinel-2 imagery effective for predicting chlorophyll levels on a large scale. (Shoko and Mutanga, 2017). Using Sentinel-2 and multispectral UAVs addresses the limitation of traditional multispectral satellites, which often struggle to provide the high spatial and temporal resolution needed for monitoring chlorophyll content in crops on smallholder farms. The use of Sentinel-2 and multispectral UAVs can overcome the limitation of multispectral satellites, which is their inability to meet the demand of high spatial and temporal resolution needed to monitor chlorophyll content of crops in smallholder farms. The Sentinel – 2 data will also be used to assess if it can be utilized to map chlorophyll content on a larger scale (Wang *et al.,* 2017).

1.3 Aims and objectives

The main aim of this study is to predict the canopy chlorophyll content of crops in a smallholder farm using the combination of multispectral UAV and Sentinel 2 imagery.

The specific objectives are to:

- Evaluate the performances of SVM and Random Forest regression algorithms in predicting crop chlorophyll content.
- Identify key spectral indices for predicting chlorophyll content in crops.
- Assess the capability of Sentinel – 2 remote sensing data to predict and map chlorophyll content.

Research report outline

The first chapter of the research report details the background of the research, the importance of chlorophyll content of crops and why the chlorophyll levels should be monitored. The chapter also explains the problem statement which is the reason why this study was conducted, lastly it lists the aims and objectives of the study.

Chapter two focuses on reviewing literature which has applied different remote sensing technologies in the mapping and monitoring of the chlorophyll content of crops. Focusing on the methods the different literature used, the findings of their studies and the limitations or gaps they have identified.

In **Chapter 3,** the research site, data collection methods, and characteristics of the remote sensing data (UAV multispectral and Sentinel-2 data) are detailed. It goes into further detail on the data's statistical analysis and the machine learning methods used, which include Random Forest and Support Vector Machines. It goes on to detail the significance of variables, the validation of the model, and the derivation of the vegetation indexes**.**

Chapter four provides detailed results from the analysis performed in the study including a description of these results.

Chapter five focuses on discussing the results obtained from this study and comparing the results with existing literature.

Chapter six sums up the research drawing conclusion on the results, mentioning the limitations and providing recommendations which can be adopted for future research purposes.

Chapter 2

Literature review

2.1 Crops quality and Chlorophyll

Agricultural systems are key to ensuring food security and driving economic growth, which makes understanding crop health and growth in these contexts important (Zhu *et al.,* 2020). Monitoring of crop health and growth can be done through the measurement of chlorophyll content, this is because chlorophyll is the pigment which is responsible for photosynthesis in crops (Narmilan *et al.,* 2022). Crop nitrogen content is also one of the important components which indicate the process of photosynthesis in crops, and it is closely related to chlorophyll (Tahir *et al.,* 2018). The literature review will examine studies that have employed different types of remote sensing techniques which include hyperspectral and multispectral data to predict the crop chlorophyll concentration while also identifying gaps and opportunities for research in this area.

2.2 The use of remote sensing for the mapping and monitoring chlorophyll content

Remote sensing has become an important tool in agriculture, providing insights into crop health, growth and productivity. The following sections explore different remote sensing technologies used to assess chlorophyll content in crops, a key indicator of plant health. The discussion is structured around different types of remote sensing data, including hyperspectral, multispectral, and UAV data. Each sub-section evaluates the method utilized, their effectiveness, and the challenges associated with these technologies in the context of chlorophyll mapping.

2.2.1 The use of Hyperspectral datasets and chlorophyll content

Hyperspectral data is also being used in chlorophyll monitoring and mapping, this is because this data consists of narrow bands that are continuous and these bands supply continuous spectra which has sensitivity to chlorophyll (An *et al.,* 2020). Haboudane *et al*. (2008) utilized spectral indices such as the Modified Chlorophyll Absorption in Reflectance Index (MCARI), Transformed Chlorophyll Absorption in Reflectance Index (TCARI), MERIS Terrestrial Chlorophyll Index (MTCI), Double Difference Index (DD), Redness Index (RM), Triangular

Chlorophyll Index (TCI), Renormalized Difference Vegetation Index (RDVI), Optimized Soil-Adjusted Vegetation Index (OSAVI), Soil-Adjusted Vegetation Index (SAVI), and Modified Soil-Adjusted Vegetation Index (MSAVI) derived from hyperspectral data to assess the chlorophyll content across crop canopies. Their findings indicated that the indices MCARI, OSAVI, and MSAVI were highly sensitive to changes in leaf pigments and variations in LAI. A study conducted by Nofrizal *et al.* (2022) was aimed at testing the capability of hyperspectral data and dimensionality reduction techniques to predict the chlorophyll content of *Zizania latifolia*. The techniques of dimensionality employed included Independent Component Analysis, PCA and Kernel PCA and the collection of hyperspectral reflectance was undertaken using FieldSpec4 (Nofrizal *et al.,* 2022). Hyperspectral reflectance data collected using ASD FieldSpec4 and four machine learning algorithms (SVM, RF, Kernel based extreme learning machine, and Deep belief nets) were utilized in predicting chlorophyll concentration of tea leaves in the research by Sonobe *et al.* (2020). Due to its low root-mean-squared error (RMSE), the Kernel-based Extreme Learning Machine method outperformed the other three algorithms in terms of chlorophyll content estimation accuracy (Sonobe *et al.,* 2020). This study highlighted the importance of integrating advanced machine learning techniques with hyperspectral data to improve accuracy of predictions. Lu and He (2019) estimated the canopy chlorophyll content of various grassland species using bi-seasonal hyperspectral imaging that used linear regression, PLSR, RF regression, and PROSAIL. The Micro-HyperSpec sensor was used to obtain the hyperspectral data in this study, the sensor has the capacity of measuring spectral signatures for 300 or more bands between the regions 400 to 1000 nm (Lu and He, 2019). The preparation of the hyperspectral dataset was undertaken utilizing SpectralView and ENVI. The results of their study indicated that PLSR and RF regression had high performance accuracies for chlorophyll prediction when compared to linear regression and PROSAIL (Lu and He, 2019). This implied that more complex models may be required to fully utilize the extensive details offered by hyperspectral data.

A study conducted by Peng *et al.* (2019) used hyperspectral data together with ANN, OLSR, SLR, and PLSR to estimate the chlorophyll concentration of naval orange plant leaves. The hyperspectral data in this study was collected using a field hyperspectral system which had a spectral range from the visible to the NIR region (400 to 1000 nm). Regression coefficient curves were generated to identify effective wavelengths, which were then used to model the OLSR and SLR. In contrast, full spectral data was used to modelling PLSR and ANN (Peng *et al.*, 2019). The findings of the research revealed that PLSR model performed best when

compared to the OLSR and SLR models, conversely the ANN prediction model yielded weak results compared to OLSR and SLR. The performance of the PLSR model could be attributed to the fact that the model incorporated the full spectra as compared to OLSR and SLR. Additionally, the poor prediction accuracy of the ANN model could be due to a low residual prediction deviation of less than 3 (Peng *et al.*, 2019). Airborne hyperspectral data, vegetation indices and two regression models which are PLSR and RF were utilized to predict the canopy chlorophyll content in a temperate forest in a study conducted by Hoeppner *et al.* (2020). The sensor which was used in the study was the AisaFenix which has 623 bands with the spectral regions ranging of VNIR to SWIR, and the field chlorophyll was sampled from the top canopy of plants using a CCM-300 handheld chlorophyll content meter which is a non-destructive method that provided reliable measurements of chlorophyll content of leaves. The PLSR regression model produced a high prediction accuracy with low RMSE and high R^2 values as compared to the RF regression model (Hoeppner *et al.*, 2020). The research further identified that the significant spectral regions for chlorophyll prediction are between 390 to 400 nm, 470 to 540, 1050 to 1100, and 2000 to 2270 nm. Hu *et al.* (2022) employed CNN, DNN, and Autoencoder deep learning algorithms to retrieve LAI and leaf chlorophyll content using simulated PROSPECT and SAIL hyperspectral data as well as acquired CASI data. PROSPECT and SAIL were used to obtain reflectance values that matched the CASI hyperspectral data. The performance of the employed deep learning algorithms varied in the prediction of the LAI and leaf chlorophyll content, with CNN yielding high R^2 values as compared to DNN and Autoencoder (Hu *et al.*, 2022). An *et al.* (2020) utilized in-situ hyperspectral data collected using a PSR 3500 spectroradiometer and four machine learning algorithms (RF, SVR, gradient boosting regression tree, and Gaussian process regression for estimating chlorophyll concentration of rice. The field measured chlorophyll in this study was collected using hand-held SPAD. Although all four regression models performed admirably, the analysis found that the RF model yielded the most precise estimations (An *et al.,* 2020).

The studies reviewed in this section show that hyperspectral data, when paired with appropriate machine learning and regression models, may effectively estimate chlorophyll content in various crop kinds and settings (Lu and He, 2019; Sonobe *et al.,* 2020). The efficiency of these techniques, however, is dependent on a number of parameters, including the machine learning algorithms used, the spectral regions investigated, and the dimensionality reduction methods applied. While hyperspectral data provides abundant spectral information, its complexity and high dimensionality requires complex processing approaches to extract valuable information

(Shoko and Mutanga, 2017). The persistent outperformance of advanced methods such as PLSR, RF, and CNN across several studies implies that the future of chlorophyll estimation can benefit from the continued integration of machine learning with hyperspectral data (Peng *et al.,* 2019; An *et al.,* 2020; Hu *et al.,* 2022). However, there are still limitations, notably in terms of standardizing procedures across studies and improving model applicability. Future studies would benefit from focusing on cross-comparisons of different crops and environments, as well as investigating the possibilities of fusing models which combine the characteristics of multiple algorithms.

2.2.2 The use of traditional Multispectral data for chlorophyll content

The use of traditional multispectral data in mapping and monitoring chlorophyll content is also common in the field of remote sensing and was utilized in various studies (Croft *et al.,* 2013; Patra *et al.*, 2017; Tong and He, 2017). Yin *et al.* (2016) employed Landsat 8 (OLI) and two vegetation indices (green chlorophyll index and greenness index) to predict leaf and canopy level chlorophyll content of grasslands in Wutumeiren. It was observed that soil background influences the reflectance captured by satellite data, potentially leading to inaccuracies in estimating chlorophyll content. To mitigate this, leaf chlorophyll content was categorized into three levels: low, moderate, and high (Yin *et al.,* 2016). In the low leaf chlorophyll content level, the ratio between the green chlorophyll index and the greenness index was employed, as this ratio had low sensitivity to the soil background. Findings of the research revealed that the combination of the two indices exhibited high estimation accuracies of leaf and canopy chlorophyll content, demonstrating the potential of carefully selected indices in improving chlorophyll predictions in challenging environments (Yin *et al.,* 2016). Similarly, Patra *et al.* (2017) employed Landsat 8 (OLI) multispectral data to predict chlorophyll-a content in Nalban Lake. Their study's overarching goal was to assess how well the reflectance readings from Landsat 8 for chlorophyll-a prediction worked. Linear regression was employed to build a model between the field collected chlorophyll-a samples and the Landsat 8 reflectance values (Patra *et al*., 2017). Their results indicated that the NIR and R band of Landsat-8 showed strong correlation to chlorophyll-a and its variation across the lake, highlighting the importance of these specific spectral bands in aquatic environments (Patra *et al.,* 2017). However, the dependence on linear regression reveals a significant limitation, such models may simplify the complicated correlations between spectral data and chlorophyll concentration, indicating the need for more robust modelling approaches in future studies (Quang *et al.,* 2022).

A study conducted by Zhou *et al.* (2020) utilized Landsat 8 (OLI) along with four prediction methods (vegetation indices, machine learning, look-up table-based inversion, and hybrid regression approaches) to predict the chlorophyll content of winter wheat in Shunyi District, China. In their results the vegetation index which yielded high prediction accuracy was the modified triangular vegetation index, followed by indices which incorporated the red and NIR bands (Zhou *et al.*, 2020). The machine learning algorithm which performed better was the gaussian process regression, for the hybrid regression approaches implementation of active learning and gaussian regression yielded better prediction results (Zhou *et al.*, 2020). Lastly, with the look up table-based inversion the cost function method produced better estimation accuracies of chlorophyll. The analysis which was conducted in this study highlighted the importance of integrating multiple methods to enhance prediction accuracy, while also highlighting the applicability of multispectral data in agricultural settings.

In contrast to the widespread use of Landsat 8 data, Croft *et al*. (2013) explored the effectiveness of two traditional multispectral sensors, Landsat TM 5 and MERIS, along with a hyperspectral sensor known as CASI. They used these sensors, combined with canopy and leaf modelling, to retrieve chlorophyll content. Findings of the study showed that leaf chlorophyll concentration predicted using the two multispectral sensors (Landsat TM 5 and MERIS) exhibited a better correlation to the measured chlorophyll content even though the sensors had different spatial resolutions, while the hyperspectral data yielded lower accuracy in contrast to multispectral sensors (Croft *et al.,* 2013). This showed that, in some cases, traditional multispectral sensors can perform better than hyperspectral sensors, probably due to simpler processing needs and lower noise in multispectral datasets. This study questions the idea that hyperspectral data always performs better, underscoring the context-dependent aspect of sensor effectiveness (Thenkabail *et al.,* 2004; Simic and Chen, 2008). Wu *et al.* (2012) employed multispectral Landsat TM and hyperspectral Hyperion to generate four vegetation indices CIgreen, EVI, MSAVI, and NDVI to predict chlorophyll content of corn. Using the PROSPECT model, the sensitivity of the vegetation indices to chlorophyll was examined. CIgreen had a linear association with chlorophyll, however MSAVI and NDVI did not exhibit a linear correlation because they reached saturation at low levels of chlorophyll (Wu *et al.,* 2012). The Hyperion and Landsat TM datasets were further used in validation incorporating the field measurements, the results indicated that CI_{green} is the index which obtained high prediction accuracy with a high R^2 and the index which produced the low but better yield was NDVI (Wu *et al.,* 2012).

The evaluated studies reveal that traditional multispectral data, that comprises fewer spectral bands than hyperspectral data, can still produce accurate chlorophyll content predictions when utilized properly. The effective utilization of multispectral data in these studies is frequently dependent on the appropriate selection of vegetation indices and the use of advanced modelling approaches such as machine learning and physical modelling. Multispectral data has the advantage of being more accessible and less expensive, making it a viable option for large-scale agricultural monitoring (Marshall and Thenkabail, 2015; Shoko and Mutanga, 2017). However, the studies highlight multiple limitations, such as the possibility of soil background interference and the saturation of indices like MSAVI and NDVI at low chlorophyll levels. (Wu *et al.,* 2012; Yin *et al.,* 2016). These limitations emphasize the importance of thorough vegetation indices selection, as well as the potential advantages offered by hybrid techniques which combine multispectral data with other data types or models (Wu *et al.,* 2012). Overall, while hyperspectral data may offer finer spectral detail, multispectral data remains a valuable tool for chlorophyll estimation, particularly when processed with advanced algorithms and models (Croft *et al.,* 2013). Future research should continue to explore how these methods can be optimized, potentially by integrating multispectral data with other remote sensing techniques to enhance accuracy and applicability across different environments and crop types.

2.2.3 The use of new generation Multispectral data for chlorophyll content

The new generation multispectral remote sensing has become increasingly utilized in the mapping and monitoring of chlorophyll content as most of the satellite missions offer the red edge region and have relatively high spatial resolutions, while the traditional multispectral data lacks the availability the red-edge zone and have less spatial resolutions (Zillmann *et al.,* 2015; Clevers and Gitelson, 2013). Schlemmer *et al.* (2013) employed Sentinel-2 and 3 to predict canopy nitrogen and chlorophyll in maize crops. In their findings, they discovered that red-edge and green sections of spectrum are suitable for estimating chlorophyll and nitrogen in maize crops. The study's findings are significant because they highlight the benefit of Sentinel-2 and 3 in identifying small changes in crop health that may be overlooked by sensors that lack the red-edge band (Schlemmer *et al.,* 2013). Furthermore, the study demonstrates the potential of these sensors (Sentinel – 2 and 3) to provide useful information for precision agriculture, enabling better management of crops. Similarly, Clevers and Gitelson (2013) also utilized Sentinel-2 and Sentinel-3 to estimate grass and crops chlorophyll and nitrogen concentration incorporating vegetation indices. According to Clevers and Gitelson (2013), the red-edge bands

and Sentinel-2's high spatial resolution (20 m) were highly beneficial for agricultural applications and chlorophyll estimation. The authors noted that the red-edge bands provided a distinct advantage in differentiating between varying levels of chlorophyll, which is often challenging with traditional multispectral data (Clevers and Gitelson, 2013). This study notes the role of vegetation indices in improving the precision of remote sensing-based chlorophyll assessments. A study conducted by Tshabalala (2020) utilized Sentinel-2 MSI and PLSR method to predict chlorophyll content of natural and rehabilitated wetland vegetation species (*Cyperus and Typha*). The results of the study showed that in the natural wetland the *Typha* species showed high accuracies of chlorophyll estimation when compared to the *Cyperus* species. In contrary the *Cyperus* species showed high accuracies of chlorophyll estimation than the *Typha* species in the rehabilitated wetland (Tshabalala, 2020). The study concluded that the red and red-edge bands of the Sentinel-2 MSI are more sensitive to chlorophyll estimation of wetland vegetation species and different wetland conditions than the other spectral bands. This shows that, while Sentinel-2 is an effective tool for estimating chlorophyll in wetland habitats, species-specific characteristics and environment conditions must be carefully evaluated to produce the best results (Tshabalala, 2020). Thus, this study emphasizes the need for adaptive approaches when applying remote sensing technologies to diverse ecological contexts.

Chemura *et al*. (2017) modelled and mapped the chlorophyll content of coffee leaves in coffee plantations using Sentinel-2 data and RF regression. For the purpose of prediction, the Sentinel-2 spectral bands were resampled to spatial resolutions of 20 m and 10 m. Highest accuracies of their results in chlorophyll estimation were exhibited when the spatial resolution was 10 m, even when the band settings were not put into consideration (Chemura *et al.,* 2017). This indicates that Sentinel-2's finer spatial resolution is essential in acquiring significant changes in chlorophyll content, which is particularly important in heterogeneous crop environments like coffee plantations (Chemura *et al.,* 2017). Furthermore, the findings of this study highlighted the importance of spatial resolution in remote sensing applications, particularly when precision is important.

Zillmann *et al.* (2015) utilized commercial multispectral RapidEye satellite data to generate chlorophyll content and green ground cover maps for wheat crops. In this study vegetation indices, which incorporated the red-edge band were derived using the RapidEye data the indices assisted in the generation of chlorophyll content map. The study revealed that the field measured chlorophyll was strongly correlated to vegetation indices which included the red-edge band, highlighting the importance of the red-edge region for accurate chlorophyll

mapping (Zillmann *et al.*, 2015). This study demonstrated how commercial satellites with improved spectral capabilities can be used to enhance agricultural monitoring, particularly in areas where access to open-source satellite data might be limited. Similarly, multispectral RapidEye satellite data was also used in the research by Lee *et al.* (2014) to predict amount of chlorophyll-a. The RapidEye data was used along with spectral mixture analysis. The research was aimed at evaluating whether multitemporal remotely sensed data could be analysed using spectral mixture analysis. They also employed band ratio models which incorporated the red-edge band in estimating chlorophyll-a. Findings highlighted that the spectral mixture analysis performed better than the band ratio model for the prediction of chlorophyll-a, demonstrating that more complex analytical techniques can increase the application of multispectral data in chlorophyll estimation (Lee *et al.,* 2014). This study showed the potential of integrating complex processing techniques, such as spectral mixture analysis, with multispectral data to improve environmental monitoring effectiveness. Darvishzadeh *et al.* (2019) compared RapidEye, Sentinel-2 data employing INvertible FOrest Reflectance Model (INFORM) radiative transfer model to retrieve & map chlorophyll concentrations. INFORM was used to create look up tables from the Sentinel-2 and RapidEye remote sensing data along with field data that was collected. The results of the study showed that the multispectral Sentinel-2 data has high yield for chlorophyll content as compared to RapidEye. It also illustrated the role of the red-edge region in the prediction of chlorophyll content (Darvishzadeh *et al.,* 2019). The findings demonstrate the importance of red-edge bands in enhancing chlorophyll predictions, supporting the advantages of new-generation multispectral sensors for agricultural applications (Darvishzadeh *et al.,* 2019). The study also highlighted the potential for INFORM with multispectral data (RapidEye and Sentinel-2) to enhance the accuracy of chlorophyll mapping.

The studies reviewed show that new-generation multispectral sensors have made substantial improvements in the remote sensing of crop mapping and monitoring. The addition of red-edge bands and improved spatial resolutions in these sensors has enabled more precise and accurate estimation of chlorophyll content, particularly in agricultural contexts (Zillmann *et al.,* 2015; Chemura *et al.,* 2017). The studies demonstrated the importance of the red-edge region in comprehending relatively small changes in chlorophyll levels that are not easily detected by traditional multispectral sensors. Furthermore, integrating cutting-edge analytical approaches such as PLSR, RF regression, and spectral mixture analysis with multispectral data has improved these sensors' accuracy and effectiveness in a range of contexts (Lee *et al.,* 2014; Chemura *et al.,* 2017; Tshabalala, 2020). These approaches are particularly beneficial in

complex ecosystems like wetlands and heterogenous agricultural settings, where traditional methods may fall short. However, these studies show that the effectiveness of multispectral data varies depending on the type of plant or crop, environmental factors, and the analysis methodologies used (Brewer *et al.*, 2022). This implies that, while new-generation multispectral sensors provide significant benefits, their implementation must be carefully adjusted to the specific needs and characteristics of the research region. In conclusion, new-generation multispectral data offer a great tool for chlorophyll mapping and monitoring, outperforming traditional sensors. To optimize their potential, future studies should focus on integrating these sensors with recent modelling and analytical tools, as well as exploring their application in different environmental and agricultural circumstances.

2.2.4 The usage of UAV data for chlorophyll content

UAVs remote sensing data are deemed to be the cutting-edge technology in remote sensing and they have been utilized in most studies which are conducted for crop monitoring (Munghemezulu *et al*., 2023). Several UAV sensors offer high temporal and spectral resolution, which are essential for effectively mapping and monitoring chlorophyll content. (Zhang *et al.*, 2021; Narmilan *et al.*, 2022). Singhal *et al*. (2019) utilised multispectral UAV images in conjunction with various machine learning algorithms to predict the concentration of chlorophyll in maize crops. These algorithms included support vector regression, random forest, relevance vector regression, kernel ridge regression, and gaussian process regression. This comprehensive comparison highlighted that machine learning models can effectively leverage UAV data to generate accurate predictions of chlorophyll content. Notably, the study found that the kernel ridge regression model yielded high prediction accuracy for chlorophyll content, while the red band was the band which showed high correlation to chlorophyll (Singhal *et al.*, 2019). Narmilan *et al.* (2022) expanded on this approach by applying UAV multispectral data in conjunction with machine learning algorithms such as MLR, RF, DT, SVR, XGB, KNN and ANN to predict chlorophyll content in sugarcane crops, also incorporating vegetation indices. The study successfully demonstrated that combining multiple machine learning models can enhance prediction accuracy, with XGB algorithm outperforming the other machine learning algorithms. Furthermore, of the vegetation indices employed in the study RVI and DVI had the highest correlation to chlorophyll (Narmilan *et al.*, 2022). This study is significant as it highlights the potential for UAV data, machine learning algorithms and vegetation indices to support site-specific crop management, allowing for more targeted

interventions based on accurate chlorophyll mapping. UAV multispectral data was used in the research conducted by Yang *et al.* (2022) to predict the chlorophyll amounts in potatoes, to achieve the aim the study examined three factors which were feature selection for chlorophyll prediction, fractional vegetation cover prediction, and evaluating and constructing the prediction model. The study used the method of recursive feature elimination to evaluate the vegetation indices. Among them, the modified SR, NDVI, and ratio vegetation index showed the strongest correlation with chlorophyll content (Yang *et al.,* 2022). For the fractional vegetation cover estimation gaussian mixture model and SVM were used and the index which showed high estimation accuracies for fractional vegetation cover was SAVI. To predict the chlorophyll content four machine learning models were constructed, the model which performed best in predicting chlorophyll content with high accuracies was the stacking model which consisted of the SVM, KNN, and gradient boosting machine (Yang *et al.,* 2022). This study highlighted the importance of selecting appropriate features and combining multiple models to enhance the predictive power of UAV data.

Wu *et al.* (2023) emphasised importance in management of nitrogen fertilization in wheat as this has a significant impact in the yield of wheat crops, as a result they used UAV multispectral data to estimate the chlorophyll content of wheat. They also employed four machine learning techniques which are RF, PLSR, DNN and Adaptive boosting along with 26 vegetation indices derived from the UAV data. The findings of their research revealed that PLSR was the machine learning technique which performed best in the prediction of chlorophyll content as compared to RF, DNN and Adaptive boosting (Wu *et al.,* 2023). This study is particularly important for precision agriculture, where optimizing input use (such as nitrogen) can lead to significant yield improvements and cost savings (Sishodia, *et al.,* 2020). Multisource features were collected along with the UAV multispectral in the study conducted by Zhai *et al.* (2023) to predict chlorophyll content. The study also employed four machine learning algorithms similar to study by Yang *et al.* (2022) although algorithms were different, the algorithms in this study were RF, ridge regression, light gradient boosting and stacked generalization. The findings of their research coincided with the outcomes of Yang *et al.* (2022). with the stacking model being the best performing model as compared to the individual machine learning models. This highlights the view that hybrid models can leverage the strengths of various algorithms to attain excellent prediction accuracy. Ban *et al.* (2022) used both multispectral and hyperspectral UAV remote sensing data to predict chlorophyll content of rice leaves in two different regions, their study also employed PLSR, ANN and SVR. Red, green, and red-edge band correlations

between the chlorophyll content in each region are highly correlated. The findings showed that PLSR performed best for the chlorophyll prediction in both regions compared to ANN and SVR as it yielded high R^2 values (Ban *et al.,* 2022). The analysis of multispectral and hyperspectral UAV data provided valuable insights into the capability of utilized UAV imagery for chlorophyll prediction in different environments.

The evaluated research show that UAV technology has a significant influence on chlorophyll content monitoring, especially when paired with robust machine learning algorithms such as ANN, SVR, RF, DNN and Adaptive boosting (Ban *et al.,* 2020; Wu *et al.,* 2023). UAVs provide high spatial and temporal resolution that traditional remote sensing platforms lack, making them suitable for agricultural applications that require comprehensive, real-time data (Zhang *et al.,* 2021). A constant element in these studies is the effectiveness of machine learning models, particularly when numerous models are integrated or stacked, in improving the predicted accuracy of UAV data (Yang *et al.,* 2022; Zhai *et al.,* 2023). This implies that future research and practical applications should focus on improving these hybrid models to properly leverage the abundant data acquired by UAVs. The selection of spectral indices and characteristics was also identified as an important aspect in the effectiveness of these models. The red band, specifically, was repeatedly identified as a major band with high sensitivity to chlorophyll content in the studies by Singhal *et al.* (2019) and Narmilan *et al.* (2022), reinforcing its significance in agricultural remote sensing. While the studies emphasize the potential of UAV data and machine learning, they also emphasize the necessity for thorough evaluation of context-specific factors, such as crop variety, development stage, and environmental conditions, which might influence the accuracy of chlorophyll predictions (Ban *et al.,* 2022). To summarize, UAV data, when paired with robust machine learning techniques, is a strong tool for precise monitoring and management of chlorophyll concentration in crops. As UAV technology advances and machine learning algorithms improve, the potential for UAVs to transform the mapping and monitoring of corps enhances.

2.3 Conclusion

The gaps which have been identified from the reviewed literature are that some of the satellite imagery such as RapidEye are commercial therefore they are costly to acquire which can be a hinderance for crop monitoring especially in regions with limited resources (Wu *et al.,* 2007). Additionally, image processing takes time, interference of clouds during acquisition and consequently this hinders effective less time-consuming monitoring (Wu *et al.,* 2007).

Chlorophyll prediction is also possible with conventional multispectral remote sensing data, but these datasets have poor spatial resolution and don't include the red-edge region, which is sensitive to chlorophyll concentration (Clevers and Gitelson, 2013). The other gap is that most satellite imagery with high resolutions such as WorldView-2 are readily available, but they have low temporal resolution making it difficult for application on chlorophyll estimation for timely monitoring of agricultural crops (Tahir *et al.,* 2018). To deal with these issues, incorporating multiple remote sensing data sources, combining high-resolution imagery with frequent temporal data from UAVs or higher temporal resolution satellites can help balance the differences between spatial and temporal resolution, allowing for more accurate and effective monitoring (Li *et al.,* 2022). Cloud computing and artificial intelligence (AI) technologies may significantly decrease long image processing times, with AI-driven algorithms automating and speeding up the processing procedures for immediate crop monitoring (Jung *et al*., 2021). Utilizing open-access satellite imagery, such as Sentinel-2, which offers red-edge bands essential to chlorophyll estimation, promoting the use of these freely available datasets, together with improved processing techniques, could increase monitoring efficiency while reducing costs (Ghosh *et al.,* 2018). In summary, while there are still challenges in using remote sensing data for chlorophyll estimation, technological advancements and the integration of various data sources present promising methods to overcome these limitations, allowing for more precise, timely, and less costly monitoring of chlorophyll content as well as contributing to more effective agricultural management practices.

Chapter 3

Materials and Methods

3.1 Study site

The study site is a wheat farm situated in Dihlabeng Local Municipality which falls under the Free State Province, South Africa. The municipality has an area of 4 746 km^2 and it falls under Thabo Mofutsanyane District Municipality. The average annual rainfall in Dihlabeng Local Municipality ranges from 540 to 1 142 mm, during the month of October to March (Musvoto *et al.,* 2009). The rainfall received in the study region from October to March makes the region more suitable for the cultivation of dryland wheat which depends on rainfall for growth (Gcayi *et al.,* 2019). The average annual temperatures rise to 26°C and drop to 13°C in summer and winter respectively (Musvoto *et al.,* 2009). Temperature is also a factor which can have an impact on the growth of wheat (Barnard and Smith, 2009). The lowest and highest temperature which were recorded during 1961 and 1990 where -8°C and 40°C respectively in the municipality (Musvoto *et al.,* 2009). The dominant activity in the study region is agriculture which constitutes an important part towards the economy of Southern Africa. The land which is covered by agricultural fields is approximately 32 000 km^2, while most of the fields are rainfed, 10% of the fields depend on irrigation. The municipality varies from 1 400 to 2 600 meters above sea level in elevation, with central and southern regions having a slope that is greater than 7%. The geology varies across the region with the Molteno formation being dominant. In the southwestern and central parts of the municipality the formations which occur are the Clarens and the Drakensberg (Sekhele, 2018).

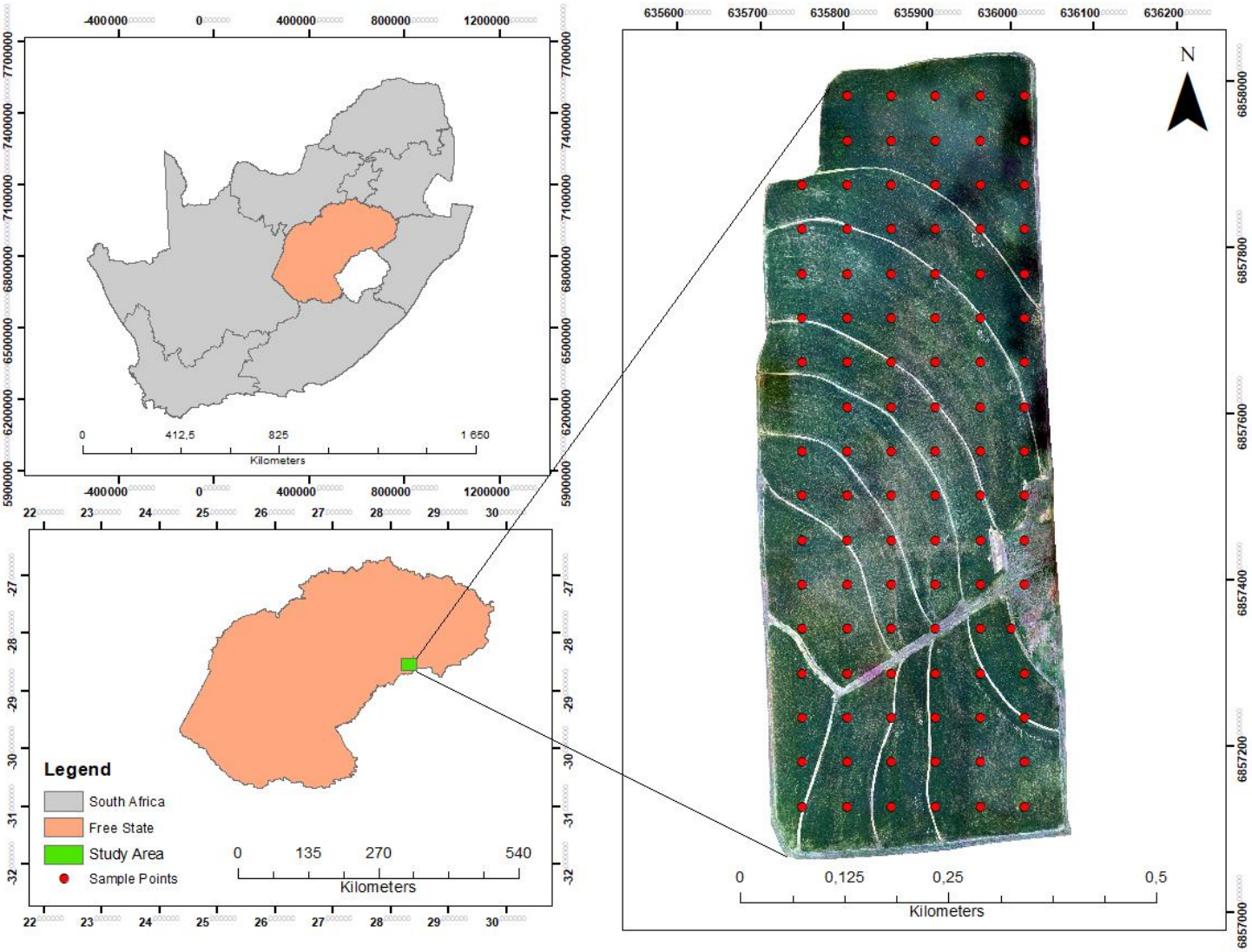


Figure 1: Shows the study region. Top left showing the extent of the South African Country, Bottom left shows the Free State province where the study region is location and lastly to the right is the study area which depicts a smallholder dryland wheat farm and the sampling points.

3.2 Flowchart of the methodology utilized in this study

The study utilized UAV and Sentinel-2 data to predict the chlorophyll content of wheat crops. The flowchart of the methodology is presented in Figure 2. The field chlorophyll measurement was collect using SPAD-502, during the maturity stage of the wheat crops (Süß *et al.,* 2015). during the same period for the collection of the UAV multispectral data and Sentinel-2. Pre-processing of the UAV data was undertaken using the Pix4D mapper software, and SNAP software for the Sentinel-2 imagery (Chemura *et al.,* 2018; Qiao *et al.,* 2022). The field measurement was then used to extract reflectance values from the remote sensing imagery

using the ArcMap software. Within the R studio software vegetation indices (NDVI, EVI, NDRE, GNDVI, Simple Ratio, CVI, CCCI, CIgreen, and CIred-edge) for both datasets were computed for analysis purposes. The datasets combined with the field measured chlorophyll were used to run Pearson correlation to establish the correlation between the vegetation indices used in this study and chlorophyll content. The vegetation indices were then utilized for data analysis employing SVM and Random Forest to develop models. The data was split into training (70%) and testing (30%) sets, the training data was using to train the models with all the vegetation indices. Variable Importance was assessed to identify the most influential spectral indices for chlorophyll prediction using permutation based variable importance method, after that the testing data was used to build models using most important variables (Joshi *et al.,* 2023).

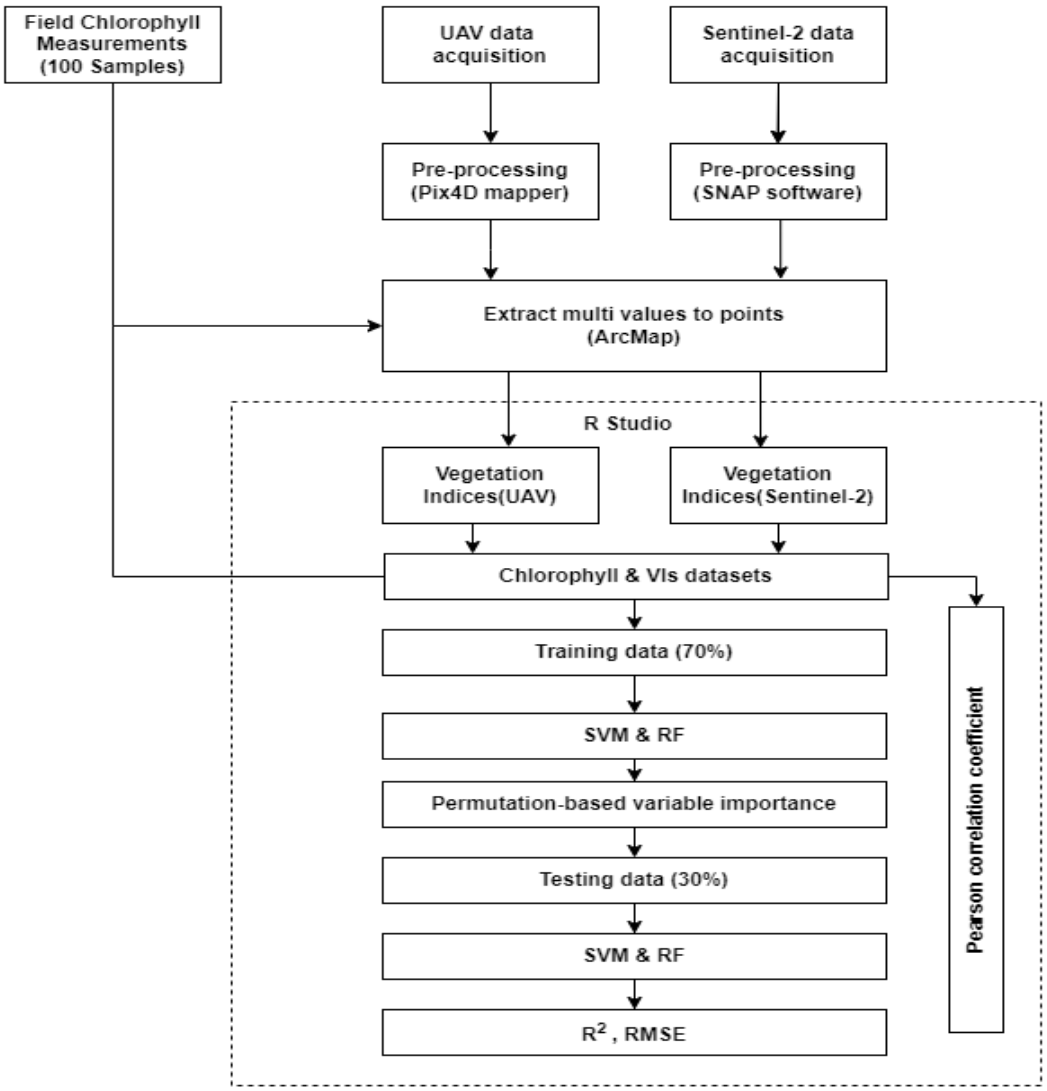


Figure 2: Flowchart of the methodology followed in this study

3.3 Field data and ground data collection

The field data collection for the chlorophyll measurements was carried out during the maturity stage of the wheat growth which was on the 19th of October 2021 which is the same date the UAV multispectral data was acquired. The chlorophyll content of the crops was measured using SPAD-502 plus chlorophyll meter (Narmilan *et al.,* 2022). Prior to the measurement of chlorophyll of crops, the SPAD-502 was calibrated by clamping the measuring heads together with the meter empty, ensuring there was no sample inside during the calibration (Süß *et al.,* 2015). The SPAD-502 estimates the amount of chlorophyll present by measuring the leaf's absorbance in two wavelength regions which are red and near-infrared. The accuracy of the SPAD-502 is ± 1.0 SPAD units (Süß *et al.,* 2015). In the study area 100 sampling plots of 1 m by 1m were systematically established. Within each sampling plot, 5 wheat crops were selected using the random sampling method and location of the samples were geo-located using a hand-held Garmin Global Positioning System (Kasim *et al.,* 2018). In each plot measurements of the chlorophyll content were undertaken for each of the 5 wheat crops. the upper most fully expanded leaves of the wheat crops where selected, from each plot 5 SPAD values were measured, and the average of the measurements was recorded as the chlorophyll of the specific plot. A total of 100 samples were collected during the field survey (Fig. 1) these collected samples were used in the development of prediction models for chlorophyll prediction.

3.4 UAV Multispectral Imagery acquisition and preprocessing

The UAV multispectral imagery was acquired using a multi-rotor DJI-Matrice 600 pro mounted with the MicaSense RedEdge-MX sensor which is a 5-band multispectral sensor (Munghemezulu *et al.,* 2023). The multispectral sensor consists of 5 bands from visible to red edge regions of the electromagnetic spectrum; these bands are blue, green, red, near infrared, and red edge (Munghemezulu *et al.,* 2023). The center wavelengths of these bands are 475, 560, 668, 840 and 717 nm, respectively (MicaSense 2020). The characteristics of the MicaSense sensor can be seen on Table 1. below. The flight mission was undertaken on the 19th of October 2021 which was on a clear sunny day between 12h30 to 14h00 during the maturity stage of the wheat crops. The altitude of the flight missions was 100 m and ground sample distance was 8.1 cm/pixel (Narmilan *et al.,* 2022; Qiao *et al.,* 2022). The UAV multispectral imagery was selected for this study due to its high spatial resolution of 8.1 cm and its capability to capture cloud-free images, which are crucial for effective crop monitoring. (Qiao *et al.*, 2022). The forward image and side line overlap will be 80% and 70% respectively.

Table 1: MicaSense RedEdge-MX Multispectral sensor characteristics.

Name of Band	**Center Band (nm)**	**Bandwidth (nm)**
Blue	475	20
Green	560	20
Red	668	10
Near-Infrared	840	40
Red edge	717	10

The correction of radiometric values in the UAV multispectral imagery involved utilizing measurements from the Downwelling Light Sensor (DLS) and a calibration reflectance panel (CRP). The CRP which is unique and well-defined reflectance values calibrated in a laboratory setting, was utilized only for calibrating the multispectral images. The DLS gave additional information on the flight's lighting conditions, allowing the calibration procedure to be refined even further. The combined use of CRP and DLS guaranteed that the surface reflectance values produced from UAV data were precise and reliable (Munghemezulu *et al.,* 2023). The

atmospheric correction of the UAV imagery was undertaken through the consideration of the sensor gain, exposure settings, sensor sensitivity, gradient dimming, and sensor black level within the Pix4D mapper software (Qiao *et al.,* 2022). Fixing the atmospheric distortion in the UAV footage required adjusting several parameters in the Pix4D mapper programme, including the gain, exposure, sensitivity, gradient dimming, and black level of the sensors.

3.5 Sentinel-2 data acquisition and pre-processing

The Sentinel-2 imagery for the study area was obtained from the https://scihub.copernicus.eu/ which is an open-source site for Sentinel data. The data was acquired on 23 October 2021 which was 4 days after the UAV multispectral data (19 October 2021) was collected. Sentinel-2 is made up of two satellites, Sentinel-2A and 2B, each of which has a 10-day revisit period and a 5-day revisit duration when there are no clouds in the sky (Zhou *et al.,* 2021). The satellite consists of 13 spectral bands of different spectral bands with 4 bands (2, 3, 4 and 8) with 10m resolution, 6 bands (5, 6, 7, 8A, 11 and 12) having 20 m resolution and 3 bands (1, 9, and 10) 60 m resolution (Frampton *et al.,* 2013; Zhou *et al.,* 2021). The summary of the specifications of the Sentinel-2 are presented on Table 2. The sensor increases the possibility to monitor crops using satellites because of the band(s) incorporated in the red-edge zone since they increase the accuracy of chlorophyll prediction (Zhou *et al.,* 2021). In regions with medium latitudes, the satellite revisits every 2-3 days, offering more detailed updates on crop conditions over vast areas in a shorter period.

The atmospheric correction of the Sentinel-2 data for the removal of the impacts which might be caused by the atmosphere was performed using the Sen2Cor tool which is a plugin in the SNAP software (Chemura *et al.,* 2018). The SNAP Reproject program was utilized to resample the bands to a resolution of 10 m. The tool uses the method of bilinear interpolation to compute the process of resampling (Zhou *et al.,* 2021).

Table 2: Characteristics of the Sentinel-2 multispectral satellite imagery.

		Sentinel-2A		**Sentinel-2B**	
Spatial resolution	**Bands**	**Central wavelength (nm)**	**Bandwidth (nm)**	**Central Wavelength (nm)**	**Bandwidth (nm)**
10	Band 2—Blue	492.4	66	492.1	66
	Band 3—Green	559.8	36	559	36
	Band 4—Red	664.6	31	664.9	31
	Band 8—NIR	832.8	106	832.9	106
20	Band 5—Red edge	704.1	15	703.8	16
	Band 6—Red edge	740.5	15	739.1	15
	Band 7—Red edge	782.8	20	779.7	20
	Band 8A—Narrow NIR	864.7	21	864	22
	Band 11—SWIR	1613.7	91	1610.4	94
	Band 12—SWIR	2202.4	175	2185.7	185
60	Band 1—Coastal aerosol	442.7	21	442.2	21
	Band 9—Water vapour	945.1	20	943.2	21
	Band 10—SWIR—Cirrus	1373.5	31	1376.9	30

3.6 Data analysis

3.6.1 Spectral indices

The R environment was utilized to calculate the spectral indices for Sentinel-2 and UAV multispectral imagery. The reflectance values that were taken from the UAV and Sentinel-2 datasets were utilized to calculate the indices that were employed in this investigation and are shown in Table 3 (Yin *et al.,* 2023). In this study, nine spectral indices were utilised: NDVI, EVI, NDRE, GNDVI, Simple Ratio, CVI, CCCI, CIgreen, and CIred-edge. The specific indices were adopted in this study because they play an important part in assessing vegetation health and their sensitivity to chlorophyll concentration (Qi *et al.,* 2021; Priyanka *et al.,* 2023). The other reason for choosing these indices in this study was because of the spectral bands which were available (red-edge, NIR, red, green, and blue). The previously published indices which were selected based on their sensitivity to crop health attributes are NDVI, EVI, GNDVI, CVI and CIgreen (Kasim *et al.,* 2018; Wu *et al.,* 2023). The red-edge band which is a band that is sensitive to chlorophyll also had an influence in selecting some of the indices used in the study, these indices are NDRE, Simple Ratio, CCCI, and CIrededge (Singhal *et al.,* 2019). The NDVI is the widely used index and high values of NDVI indicate healthy vegetation and more greenness, GNDVI has a similar formula as NDVI although it encompasses green band and not red band and has sensitivity to green vegetation (Gitelson and Merzlyak, 1994; Ramos *et al.,* 2020). CIgreen estimates chlorophyll using the green band and it is sensitive to chlorophyll in leaves. EVI enhances sensitivity to chlorophyll and canopy structure and adjusts the atmospheric influences (Qiao *et al.,* 2022). CVI's focus is on the absorption features of chlorophyll in the red and NIR regions (Zheng *et al.,* 2018). To determine the degree of linear relationship between chlorophyll and the derived spectral indices from UAV and Sentinel -2 data, the Pearson correlation coefficient was utilized. This statistical method was specifically chosen because it effectively measures the strength and direction of the linear relationship between two variables, making it well suited for assessing the sensitivity of spectral indices to chlorophyll content (Narmilan *et al.,* (2022). The derived spectral indices and the field chlorophyll were used to build the models using the two algorithms (SVR and Random Forest) employed in the study.

Table 3: The study's spectral indices and the formulas they utilize.

Spectral Index	Formula	Reference
NDVI	$\frac{NIR - Red}{NIR + Red}$	Rouse *et al.*, 1974.
EVI	$2.5 \times \frac{(NIR-Red)}{(NIR+ 6 \times Red - 7.5 \times Blue + 1)}$	Qiao *et al.*, 2022.
NDRE	$\frac{NIR - Rededge}{NIR + Rededge}$	Li *et al.*, 2014.
GNDVI	$\frac{NIR - Green}{NIR + Green}$	Gitelson and Merzlyak, 1994.
Simple Ratio	$\frac{Rededge}{NIR}$	Song, 2013.
CVI	$NIR \times (\frac{Red}{Green^2})$	Zheng *et al.*, 2018.
CCCI	$\frac{(NIR - Rededge)(NIR + Rededge)}{(NIR - Red)(NIR + Red)}$	Li *et al.*, 2014.
CIgreen	$\frac{NIR}{Green} - 1$	Gitelson *et al.*, 2003; Clevers *et al.*, 2017.
CIred-ege	$\frac{NIR}{Rededge} - 1$	Gitelson *et al.*, 2003; Clevers *et al.*, 2017.

3.6.2 SVM regression

SVM algorithm was initially developed to solve problems relating to classification by Cortes and Vaanpik (1995). They were later expanded towards the solving of problems relating to regression (Shawe-Taylor and Cristianini, 2004). SVM utilises kernels and if there are no minimum values, the thin distribution of the solution and capacity control is attained through manipulating the boundary or a set of support vectors (Omer *et al.*, 2016). In the support vector regression procedure kernel functions are used to transform nonlinear regression problems into linear relationships (Cortes and Vaanpik, 1995). SVM algorithm utilises structural minimization principle that comprises precise generalisation performances for unique data capacities (Akande *et al.*, 2014). For optimising the algorithm there are two parameters which need to be set correctly, these parameters are (C) regularization parameter and (ε) insensitive zone (Cherkassky and Ma, 2004). The estimation accuracy of SVM is dependent on adequate setting of the parameters, regularization parameter, kernel parameters and insensitive zone. The chlorophyll content of wheat crops was predicted using the Support Vector Machine (SVM)

algorithm in this study, which used a linear kernel. In order to improve the predictive power of the model, two crucial SVM regression parameters the ε insensitive zone and the C regularization were incorporated. The ε – insensitive zone was implemented to establish a range of tolerance (ε), within which errors are not penalized (Mutanga *et al.,* 2015). To identify the optimal value for this parameter, a robust methodology involving 10 – fold cross-validation was employed. This technique ensured a thorough assessment of the model's execution across various subsets utilized data, leading to a more reliable determination of the ε parameter (Liang *et al.,* 2016; Ban *et al.,* 2022). Simultaneously, the C – regularization parameter was introduced to accurately classify training points while striking a balance to create a smooth decision boundary. The trade-off between a well-fitted decision boundary and precise classification was carefully managed (Akande *et al.,* 2014). The selection of the best value for the C parameter was also carried out through the method of 10-fold cross-validation. This iterative process was repeated until the best-performing value of C, which demonstrated optimal performance on the training set, was identified (Liang *et al.,* 2016). The adoption of 10-fold cross-validation in both parameter tuning processes enhances the robustness and generalizability of the SVM model in predicting chlorophyll content for wheat crops. The implementation of the SVM regression was undertaken within the R Studio environment (Chambers, 2008)

3.6.3 RF regression

The RF regression was used for predicting the chlorophyll content of crops, this is because the algorithm has been adopted by most remote sensing studies for regression purposes (Singhal *et al.,* 2019; Priyanka *et al.,* 2023). As a component of an ensemble learning strategy, RF is a machine learning technique that combines several decision trees. (Breiman, 2001). The algorithm is robust and can deal with noisy data, and it also works well with high dimensional data (Yang *et al.,* 2021). Each tree in the ensemble is trained using a randomly chosen portion of the training data. Also, during the splitting process of each tree, a random subset of features is considered (Singhal *et al.,* 2019; Yang *et al.,* 2021). Decision trees in RF are constructed by repeatedly splitting the data into subsets based on a feature that provides the best split at each node (Mtry). This process is repeated until the specified criteria is reached. To obtain a prediction in random forest all the trees are averaged (Joshi *et al.,* 2023). To estimate the error associated with each variable in a random forest the MSE difference between OOB dataset and dataset used for the training the decision trees is calculated (Singhal *et al.,* 2019). In this study, the Random Forest (RF) regression algorithm was employed to predict the chlorophyll content

of wheat crops using the 9 spectral indices adopted. To optimize the RF regression model for chlorophyll prediction, two important parameters were considered, the Ntree and Mtry. The performance and complexity of the random forest model are largely controlled by these factors. To identify the optimal Ntree value the method of 10 – fold cross validation was incorporated, model validation was conducted using 1-fold, while model training was conducted using the remaining 9 folds (Prado Osco *et al.*, 2019). This process was iterated until each of the ten folds was assessed separately. The optimal Ntree value was determined to be 500. The Mtry parameter, which indicates the number of variables randomly sampled at each split, also significantly affects the performance of the random forest model (Li *et al.*, 2018). In this study, the selected Mtry value was set to 3, denoting the number of variables considered at each split during the model's training process. The RF analysis was performed within R using the package called RandomForest.

3.6.4 Variable Importance and selection

In this study to assess the variable importance of the variables in both the SVM models and the Random Forest models the permutation-based variable feature importance method was employed within the R software (Joshi *et al.*, 2023). The two datasets employed in this study were used to generate the variable feature importance in the prediction of chlorophyll using the two regression techniques. In total there were four variable rankings, for the UAV multispectral dataset two variable rankings were performed using the SVM and the RF regression models. Furthermore, for Sentinel-2 dataset two variable rankings were also performed using the SVM and RF regression models. The permutation based variable feature importance method is commonly used in machine learning for assessment of the performance of variables in regression models. This method requires a trained model and corresponding test data to compute the variable rankings. The variables are shuffled randomly in this technique for the calculation of change in performance, as a result the independent variables which have a great influence on the performance of the model is deemed the most important variable (Joshi *et al.*, 2023). The spectral indices which are the independent variables in our study were positioned according to their feature importance to indicate their influence in chlorophyll prediction. The variable importance graphs which were generated from this process were used to select the most important variables which were then used to construct the final models for the prediction of chlorophyll.

3.6.5 Model validation

The measured chlorophyll ground dataset was split at random to 70% and 30% for training and testing to calibrate and validate the performance of the RF and SVM regression models (Adelabu *et al.,* 2015; Zha *et al.,* 2020). The two regression algorithms were optimised utilizing the train data and the test data was utilized for assessing performance of the prediction models (Adelabu *et al.,* 2015; Zha *et al.,* 2020). The models created with the two machine learning methods were assessed by plotting the field-measured chlorophyll against the projected values and calculating the coefficient of estimation (R^2) and root mean square error (RMSE). (Zha *et al.,* 2020).

Chapter 4

Results

4.1 Description of field data

The summary of 100 samples of wheat field measured chlorophyll statistics is presented in Table 4. The minimum chlorophyll value from the sampled points is 33.42 and the maximum chlorophyll value is 62.40. The average of all the filed chlorophyll values is 52.67 and the median of the values is 54.24. The standard deviation for the dataset is 6.15 which tells us that the chlorophyll values are clustered around the mean.

Table 4: Statistics of the field measured chlorophyll using SPAD-502 meter for the 100 sampling points used in this study.

Field Chlorophyll (SPAD) Statistics	
No. of Samples	100
Minimum	33.32
Median	54.24
Mean	52.67
Maximum	62.40
Std	6.15

4.2 Correlation between chlorophyll and derived spectral indices

In order to determine the degree of linear link between the field chlorophyll and the spectral indices for the UAV and Sentinel-2 data, the Pearson correlation coefficient was utilised. Figure 3a) and 3b). depict the correlation between the nine spectral indices employed in the study and the wheat chlorophyll content for the UAV multispectral and Sentinel-2 datasets respectively. All spectral indices except EVI, NDRE, CVI and CIrededge in Figure 3a). have positive correlation to chlorophyll with the highest correlation being that of CIgreen having a correlation coefficient of 0.19 and the lowest being that of Simple Ratio (0.05). In Figure 3b). 8 spectral indices which are NDVI, EVI, NDRE, GNDVI, CVI, CCCI, CIgreen and CIrededge have positive correlation to chlorophyll with their correlation coefficients being above 0.10 with the highest correlation being 0.22 for the CVI index. The Simple Ratio index is the index with a negative correlation to chlorophyll in the correlation plot for the Sentinel – 2 data with a correlation coefficient of – 0.12.

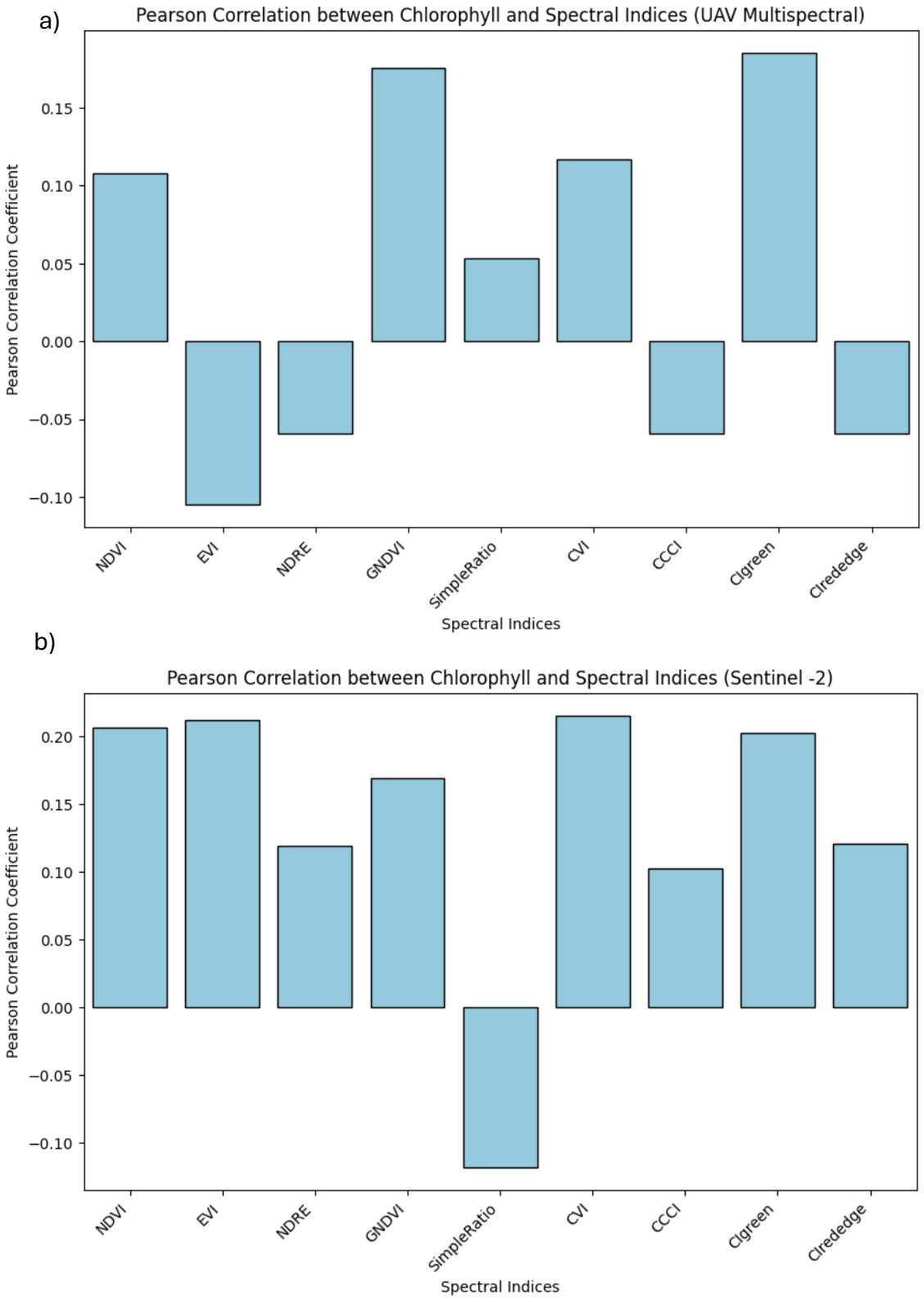


Figure 3: Correlation plots for a) The extracted spectral indices from the UAV multispectral image and the field measured chlorophyll b) The extracted spectral indices from the Sentinel – 2 imagery and the field chlorophyll content

4.3 Chlorophyll prediction using SVM and RF machine learning regression algorithms

The wheat chlorophyll concentration in this study was predicted using spectral indices that were generated using two remote sensing datasets UAV multispectral and Sentinel - 2 data as well as by combining SVM and Random Forest regression machine learning methods. The models were constructed using the indices of high importance to the prediction of chlorophyll, obtained using permutation based variable importance (Figure 5). The model with the best performance in this research was the RF model for the Sentinel – 2 multispectral image, this model (Figure 4c.) yielded a R^2 of 0.14 (RMSE = 5.11). The remaining models employed exhibited low accuracies compared to the random forest for the Sentinel – 2 dataset. The random forest for the UAV data yielded a low R^2 of 0.038 & RMSE of 6.71, subsequently the SVM model for the Sentinel – 2 data which produced an R^2 of 0.022 (RMSE = 5.35). These models are represented by Figure(s) 4d). and 4a). respectively. As shown in Figure 4b, the SVM model for UAV data was the last model to perform poorly compared to the others. The model's R^2 value was 0.020 and its RMSE was 5.41. The RF model for the Sentinel – 2 data used 3 spectral indices randomly selected from the permutation based variable importance which were NDVI, GNDVI, and CIgreen. The RF model of the UAV also used 3 indices which were randomly selected from the permutation based variable importance and the indices were CIgreen, CIrededge, and GNDVI. The SVM model for the Sentinel-2 utilized four variables selected based on their importance ranking. These indices, listed in order of decreasing importance, were CIgreen, NDVI, GNDVI, and CIrededge. Lastly the SVM model for the UAV was implemented using CVI, CIgreen, NDVI, and CIrededge.



Figure 4: Prediction plots for chlorophyll estimation of wheat crops during the maturity stage for the month of October 2021 using the test datasets (30%) a) Support vector machine plot between predicted and actual chlorophyll content from the Sentinel-2 multispectral image for CIgreen, GNDVI, NDVI, and CIrededge b) Support vector machine plot between the field and estimated chlorophyll content from the UAV data for CVI, CIgreen, NDVI, and CIrededge optimal vegetation indices. c) Random forest regression plot for observed versus predicted chlorophyll content from the Sentinel - 2 multispectral image for 3 spectral indices (NDVI, GNDVI, and CIgreen), and d) The random forest regression plot for the one-to-one relationship between the actual and estimated chlorophyll against the predicted chlorophyll for UAV image indices: CVI, CIgreen, NDVI, and CIrededge spectral indices.

4.4 Variable importance of spectral indices in prediction of chlorophyll content

Figure 5. shows the variable importance plots generated using the permutation-based variable feature importance for the four machine learning models used in the study (2 support vector machine models and 2 random forest regression models). In the support vector machine model for predicting chlorophyll using the UAV image the variable importance shown by Figure 5a). indicates that the spectral indices which had high importance in predicting chlorophyll in the model were CVI, CIgreen, NDVI, and CIrededge. The spectral index which had high importance was CVI with an importance score of 1.21. The other indices in the model were least important as their importance score was low. Figure 5b). shows the importance of indices in the SVM regression model for Sentinel -2 image, in this model the indices with better variable importance scores for chlorophyll prediction were CIgreen, NDVI, GNDVI and CIrededge similar to those of the SVM model for the UAV multispectral image except for the GNDVI index. The index with a better variable importance score was the CIgreen which had a score of 1.1. The variable importance for the random forest model for the UAV image depicted by Figure 5c). shows that all the spectral indices had an importance greater than 0.5 in chlorophyll prediction with the highest importance score being for the EVI index having a score of 0.81. Lastly the variable importance for the random forest model for predicting chlorophyll content using the Sentinel – 2 image (10 m resolution) represented by Figure 5d). showed that all the spectral indices in the model had an importance greater than 0.5 in contribution to the chlorophyll prediction except for the Simple Ratio, NDRE and CIrededge indices which had importance scores which are less than 0.5. The index with a higher variable importance score was CVI with an importance score of 1.7.

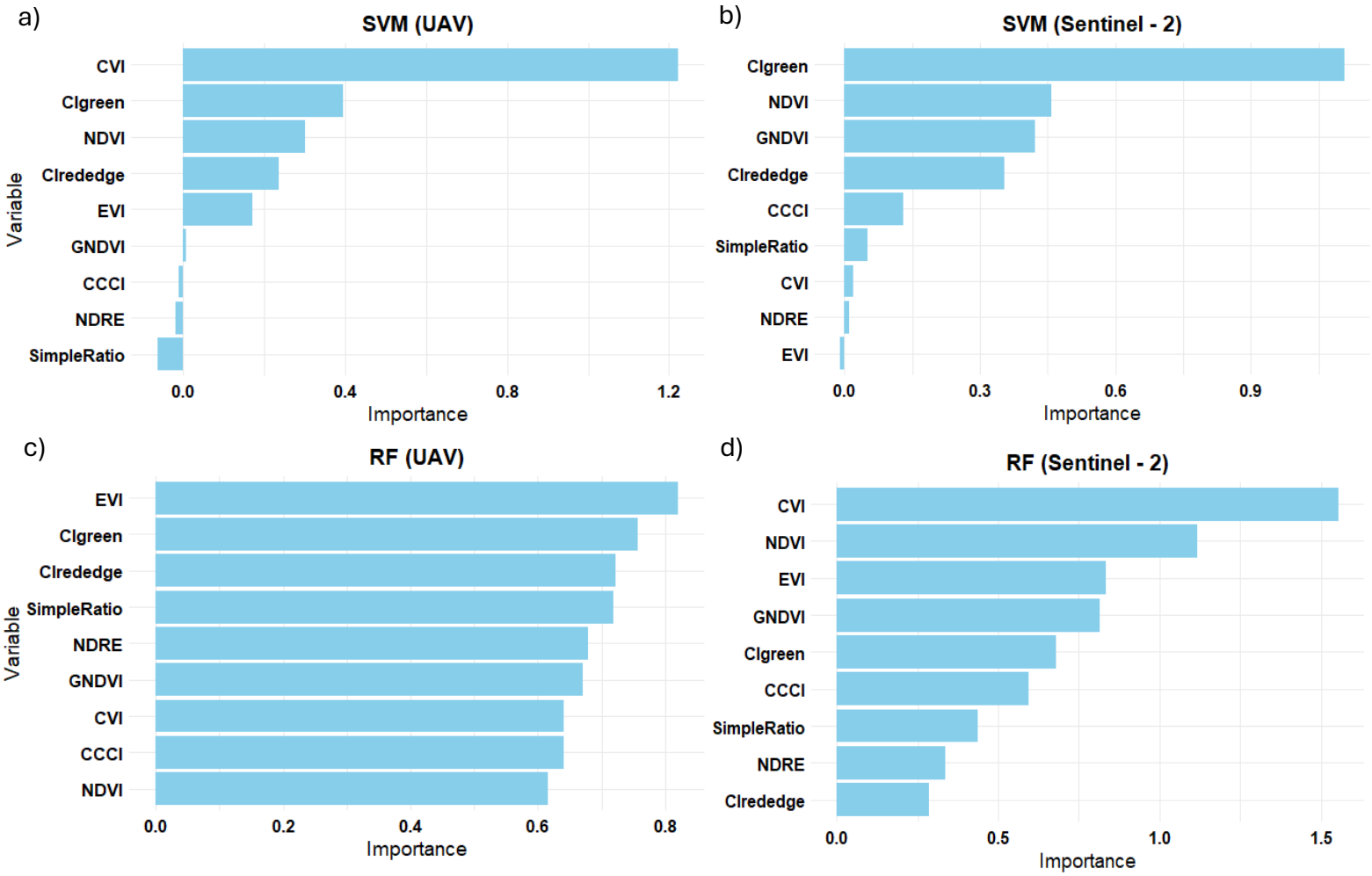


Figure 5: Permutation based variable importance plots for both the SVM and RF analysis a) The support vector machine variable importance for UAV image, b) SVM variable importance for the Sentinel – 2 image, c) Random Forest variable importance for the UAV image and d) Random Forest variable importance for the Sentinel- 2 image.

4.5 Spatial distribution of wheat chlorophyll concentration

Adopting the two SVM models and the two RF models including the two datasets (UAV and Sentinel -2) utilized in this research, the spatial distribution maps of the chlorophyll content throughout the study region were produced. The spatial distribution maps are presented in Figure 6. The distribution of the chlorophyll content across the study area follows a similar pattern this can be seen in all the distribution maps generated using the different machine learning algorithms and different datasets. In Figure 6a). it is evident that the chlorophyll values across the entire region are higher, lower chlorophyll values depicted by the red colour are as a result of less to no vegetation at all. The random forest derived spatial distribution map (Figure 6b). for the UAV multispectral shows that from the predictions, there is low chlorophyll content as we move towards the center of the study area and there is high chlorophyll in northern and southern sections of study area this can also be exhibited in Figure 6a). Figure 6c) and 6d) exhibit a similar pattern although Figure 6c) the SVM derived spatial distribution shows to have more low chlorophyll values as compared to the Random Forest derived spatial resolution for the upscaled UAV image. The boundary located in the southeastern part of the study area is characterised by low wheat chlorophyll content values. The distribution maps can help in understanding the patterns of wheat health and how the chlorophyll is distributed across the field.

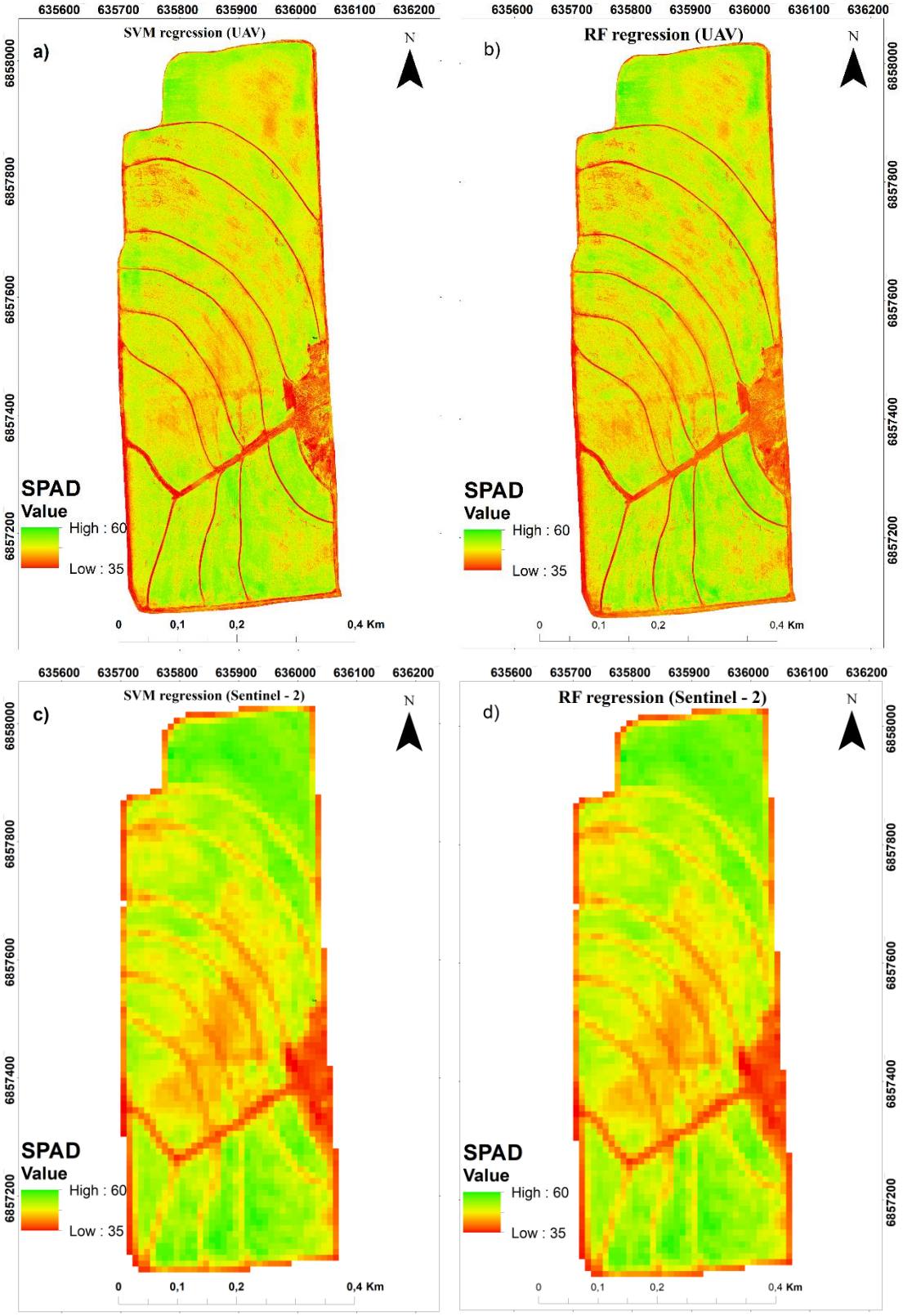


Figure 6: Spatial distribution maps of chlorophyll in the study region generated using a) The SVM model and the UAV multispectral image, b) Random Forest model and the UAV multispectral image c) Support vector machine model and Sentinel – 2 and d) Random Forest incorporating the Sentinel -2 imagery.

Chapter 5

Discussion

The goal of this research was to employ UAV multispectral with 5 spectral bands and Sentinel-2 remote sensing data and field measured chlorophyll to predict chlorophyll content of smallholder wheat crops. Two machine learning methods, which are SVM and Random Forest were also adopted in this study. The model which performed better in chlorophyll content prediction in this study was the RF model for the Sentinel – 2 multispectral data, followed by RF model for UAV, then SVM for Sentinel – 2 and SVM for the UAV data respectively.

The Pearson correlation analysis of the Sentinel-2 and UAV spectral indices and chlorophyll content yielded a correlation between chlorophyll and the two datasets ranging from 0.20 to -0.10. The correlation between chlorophyll and UAV derived spectral indices reported NDVI, GNDVI, SR, CVI and CIgreen as variables with weak positive correlation to chlorophyll whilst EVI, NDRE, CCCI and CIrededge exhibited weak negative correlation to chlorophyll. The negative correlation of EVI in this study may have been caused by variations in the wheat growth stage, the surrounding environment, or spatial resolutions (Figueiredo *et al.*, 2016). The positive correlation of NDVI, GNDVI, SR, CVI and CIgreen to chlorophyll content can be attributed to the study conducted by Tong and He (2017), where these indices were utilized to predict the chlorophyll content when performing the correlation analysis, the spectral indices exhibited positive correlations to chlorophyll content although the correlation in their study were high as compared to our study. The high correlation could be attributed to the fact that their study was focused on grassland species, and the chlorophyll was measured in a laboratory and not collected using SPAD as in our study (Tong and He, 2017). Madonsela *et al.* (2023) also reported a positive relationship with spectral indices used and chlorophyll concentration in their study, the spectral indices are CVI, NDVI, GNDVI, CIrededge and CIgreen, although in our study the CIrededge index had negative correlation to chlorophyll. The negative correlation of CIrededge could be as a result that it is less effective to chlorophyll prediction of wheat or different growth stages (Lin *et al.,* 2019). These similarities tell us that the correlation of vegetation indices used in this study and chlorophyll play a role in the chlorophyll predictions. The results of this study's Pearson correlation analysis for Sentinel - 2 showed that, in contrast to all other indices, which were CIgreen, NDVI, GNDVI, NDRE, and EVI, the Simple Ratio had a negative correlation with chlorophyll content. This is consistent with findings by Narmilan *et al.* (2022), who observed that these indices exhibited positive

correlations with chlorophyll content in their study. The similarity in results across various datasets highlights the reliability of these indices in chlorophyll prediction, but the negative correlation of the Simple Ratio could be due to specific factors in this study, such as crop growth stage or the spectral sensitivity of Sentinel-2 (Vincini and Frazzi, 2011). Although the contradiction was that the Simple Ratio index had negative correlation in our study, in this research it showed substantial correlation to chlorophyll. The positive correlation of the Sentinel -2 derived indices highlights the potential of using these indices for chlorophyll prediction in wheat (Zhou *et al.,* 2020). Also, the weak to moderate correlation results for both UAV and Sentinel – 2 imagery indicate to us that using single indices for predictions will not yield accurate results.

The study used nine spectral indices which were calculated using both the UAV and Sentinel-2 data separately, the indices were then used in the development of 4 machine learning model (2 SVM and 2 RF) for predicting the chlorophyll content of wheat. The final RF model constructed using the Sentinel – 2 multispectral data consisted only of 3 optimal spectral indices (NDVI, GNDVI, and CIgreen). This model yielded high R^2 of 0.14 and low RMSE value of 5.11 in our study. The NDVI was the traditional NDVI which consisted of the NIR and the Red bands, the GNDVI index incorporated the NIR and Green bands while the CIgreen index consisted of the NIR and the Green bands. This result of RF being the best performing model for the Sentinel – 2 dataset agrees with Priyanka *et al.* (2023). Their study also predicted chlorophyll content using UAV and Sentinel-2 data fusion, also employed SVM and Random Forest, their study concluded that the best performing regression algorithm for chlorophyll prediction was the Random Forest which yielded high R^2 and low RMSE values (Priyanka *et al.,* 2023). The overall performance of the regression models employed in this study research was suboptimal, due to various factors. With the UAV imagery, errors were associated with the high variability in pixel resolution, where the finer spatial resolution (8.1 cm) of UAV images captured greater detail but might have also introduced noise and variability. The variability was influenced by temporal variability and GPS accuracy (2 m) which affected the consistency and alignment of the data (Forkuor *et al.,* 2017). However, for Sentinel-2 data, the errors were minimal due to its 10 m coarse resolution, which, although less detailed, provided more stable and consistent results across larger areas (Shoko and Mutanga, 2017). Wang *et al.* (2022) reported that the crop spectrum can be affected by a variety of factors such as the angle of the Sun, speed of wind, direction of reflection, and aerosols. The low R^2 values obtained in this study can be compared to results obtained by Nguyen *et al.* (2021), in the study very low R^2

values of RF and SVM were obtained in prediction of Chl-a. The importance of the NIR region for chlorophyll prediction was mentioned in various remote sensing published literature (Zhang *et al.,* 2016; Delloye *et al.,* 2018). All spectral indices utilized in this research consisted of the NIR band, it was mentioned by Vincini and Frazzi (2011) and Delloye *et al.* (2018) that NIR band is sensitive to Leaf Area Index (LAI) which is closely related to chlorophyll content of crops. This supports the incorporation of the NIR band in the spectral indices which were adopted in this study. The sample size utilized in this study was 100 and the prediction accuracies were low. This is consistent with studies conducted by Hu *et al.* (2018) and Nguyen *et al.* (2021) were they utilized sample sizes of 2871 and 229 respectively and yielded low prediction accuracies for RF and SVM model.

The permutation-based variable importance analysis showed valuable insights in the roles of the different spectral indices employed in this study for chlorophyll prediction. In the SVM model for UAV data, CVI, CIgreen, NDVI, and CIrededge were identified as the most important spectral indices for chlorophyll prediction. CVI exhibited a high importance rating, suggesting its important contribution for estimating chlorophyll content from UAV data. The importance of CVI is consistent with the findings of Vincini and Frazzi (2011) who also identified CVI as a reliable measure of crop vigour and chlorophyll levels. Furthermore, the SVM model for the Sentinel – 2 data the spectral indices with high importance were CIgreen, NDVI, GNDVI, and CIrededge. While these indices are similar to the ones in the UAV model, CIgreen was the index with high importance. A study by Brewer *et al.* (2022) highlighted that as the chlorophyll content of crops increases the sensitivity on the green band also increases. CIgreen and NDVI indices were among the variables with high importance to chlorophyll predictions across all the models. These indices were widely recognized for their importance in vegetation health monitoring, as revealed in study conducted by Li *et al.* (2018) who reported that they were strong predictors of crop properties such as chlorophyll content and yielded high prediction accuracies of chlorophyll. The low variable importance of indices such as Simple Ratio and CIrededge in the RF model for Sentinel-2 data is consistent with discoveries from other studies, such as those by Lin *et al.* (2019) which imply that these indices are less effective for chlorophyll estimation in particular crop types or phases of growth. However, the significant importance of CIgreen across models suggests it's reliability as a predictor of chlorophyll content in wheat, which is consistent with findings from previous studies, such as Wu *et al.* (2012), who discovered CIgreen to be a strongly sensitive to crop chlorophyll. The variable importance scores for all the spectral indices in the RF models were higher when

compared to those of the SVM models. This observation is consistent with the widely reported flexibility of RF models in capturing complex patterns in data (Cavallo *et al.,* 2017; Shah *et al.,* 2019). The four spectral indices which incorporated the red-edge band were CIrededge, CCCI, NDRE, Simple Ratio, although the red-edge region has high sensitivity to chlorophyll its contribution to the prediction of chlorophyll was not highly significant in this study (Jiang *et al.,* 2017). The overall prediction results of this study were relatively low for both SVM and RF regression model, but the results indicate that it is possible to predict the chlorophyll content of wheat crops using SVM and RF machine learning algorithms.

The capability of mapping chlorophyll content was also assessed using UAV multispectral and Sentinel – 2 imageries. To map the chlorophyll content within the study area the SVM and RF regression models were used. The distribution maps created using the RF models were the once showing a better distribution of the chlorophyll as compared to the SVM created distribution maps. The results of the Sentinel – 2 imagery distribution maps are comparable to those of the UAV imagery. This suggests that Sentinel – 2 can be used to predict and map chlorophyll content of crops in a larger scale. The use of Sentinel – 2 imageries to map chlorophyll content in our study coincides Ali *et al.* (2020). The study also used Sentinel – 2 data to map the chlorophyll content in a mixed mountain forest. Significance of NIR and red-edge band towards increasing the predictive performance was also noted in this study. It was also noted that of the SVM and RF models the best performing machine learning algorithm was the RF. RF's best performance could be because of the incorporation of multiple different decision trees, and resistance to overfitting (Breiman, 2001; Hu *et al.,* 2018). The UAV multispectral image spatial distribution maps for chlorophyll showed much detail, due to high spatial resolution of the UAV sensors, in contrast, the Sentinel – 2, with its 10 m resolution, did not display as much detail. However, despite these differences, the distribution of chlorophyll content across all remote sensing sensors exhibited similar patterns. The models which were used to create the distribution maps had low accuracy, although the results of the distribution maps indicate that Sentinel -2 indeed can map chlorophyll content of crops, meaning it can also be used to map chlorophyll content in large scale areas.

Chapter 6

Conclusions

The study aimed to predict the chlorophyll concentration of smallholder wheat crops using UAV and Sentinel-2 data, along with machine learning algorithms including SVM and Random Forest. Although the models succeeded in predicting chlorophyll concentration, their accuracy was notably low. The spectral indices incorporating red-edge and NIR were generally sensitive to chlorophyll content, exhibiting positive correlations to chlorophyll content, as observed for indices such as NDVI, EVI, NDRE, GNDVI, CVI, CCCI, CIgreen, and CIrededge except for Simple Ratio. In contrast, for UAV data, indices such as NDVI, GNDVI, Simple Ratio, CVI, and CIgreen showed positive correlation, while CIrededge, EVI, NDRE and CCCI showed negative correlations.

Sentinel – 2 data outperformed UAV data in chlorophyll prediction. Although the models mapped chlorophyll concentration across the study region, the spatial accuracy was compromised due to the poor predictive performance. This indicates that while Sentinel – 2 data hold promise for large-scale chlorophyll mapping, improvements in model accuracy are needed to enhance the reliability of predictions.

The study's limitations, including a small field dataset and limited UAV data from other time periods, may have introduces biases and affected model performance. The lack of temporal UAV data constrained the potential for comparison and validation, contributing to the poor correlations and prediction results. Future research should explore other multispectral and hyperspectral datasets to improve the precision of chlorophyll monitoring. This will enhance the field of crop monitoring and employ advanced technologies for crop chlorophyll prediction both in small- and large-scale geographic areas. Increasing the sample size of field data is also essential to reduce bias and improve model robustness in crop monitoring.

References

Adelabu, S., Mutanga, O. and Adam, E., 2015. Testing the reliability and stability of the internal accuracy assessment of random forest for classifying tree defoliation levels using different validation methods. *Geocarto International*, *30*(7), pp.810-821.

Akande, K.O., Owolabi, T.O., Twaha, S. and Olatunji, S.O., 2014. Performance comparison of SVM and ANN in predicting compressive strength of concrete. *IOSR Journal of Computer Engineering*, *16*(5), pp.88-94.

Ali, A.M., Darvishzadeh, R., Skidmore, A., Gara, T.W., O'Connor, B., Roeoesli, C., Heurich, M. and Paganini, M., 2020. Comparing methods for mapping canopy chlorophyll content in a mixed mountain forest using Sentinel-2 data. *International Journal of Applied Earth Observation and Geoinformation*, *87*, p.102037.

An, G., Xing, M., He, B., Liao, C., Huang, X., Shang, J. and Kang, H., 2020. Using machine learning for estimating rice chlorophyll content from in situ hyperspectral data. *Remote Sensing*, *12*(18), p.3104.

Ban, S., Liu, W., Tian, M., Wang, Q., Yuan, T., Chang, Q. and Li, L., 2022. Rice leaf chlorophyll content estimation using UAV-based spectral images in different regions. *Agronomy*, *12*(11), p.2832.

Barnard, A. and Smith, M.F., 2009. The effect of rainfall and temperature on the preharvest sprouting tolerance of winter wheat in the dryland production areas of the Free State Province. *Field Crops Research*, *112*(2-3), pp.158-164.

Battude, M., Al Bitar, A., Morin, D., Cros, J., Huc, M., Sicre, C.M., Le Dantec, V. and Demarez, V., 2016. Estimating maize biomass and yield over large areas using high spatial and temporal resolution Sentinel-2 like remote sensing data. *Remote Sensing of Environment*, *184*, pp.668-681.

Breiman, L., 2001. Random forests. *Machine learning*, *45*, pp.5-32.

Brewer, K., Clulow, A., Sibanda, M., Gokool, S., Naiken, V. and Mabhaudhi, T., 2022. Predicting the chlorophyll content of maize over phenotyping as a proxy for crop health in smallholder farming systems. *Remote Sensing*, *14*(3), p.518.

Cavallo, D.P., Cefola, M., Pace, B., Logrieco, A.F. and Attolico, G., 2017. Contactless and non-destructive chlorophyll content prediction by random forest regression: A case study on fresh-cut rocket leaves. *Computers and electronics in agriculture*, *140*, pp.303-310.

Chambers, J.M., 2008. *Software for data analysis: programming with R* (Vol. 2). New York: Springer.

Chemura, A., Mutanga, O. and Odindi, J., 2017. Empirical modeling of leaf chlorophyll content in Coffee (Coffea Arabica) plantations with Sentinel-2 MSI data: Effects of spectral settings, spatial resolution, and crop canopy cover. *IEEE Journal of selected topics in applied earth observations and remote sensing*, *10*(12), pp.5541-5550.

Chemura, A., Mutanga, O., Odindi, J. and Kutywayo, D., 2018. Mapping spatial variability of foliar nitrogen in coffee (Coffea arabica L.) plantations with multispectral Sentinel-2 MSI data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *138*, pp.1-11.

Cherkassky, V. and Ma, Y., 2004. Practical selection of SVM parameters and noise estimation for SVM regression. *Neural networks*, *17*(1), pp.113-126.

Clevers, J.G. and Gitelson, A.A., 2013. Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and-3. *International Journal of Applied Earth Observation and Geoinformation*, *23*, pp.344-351.

Clevers, J.G., Kooistra, L. and Van den Brande, M.M., 2017. Using Sentinel-2 data for retrieving LAI and leaf and canopy chlorophyll content of a potato crop. *Remote Sensing*, *9*(5), p.405.

Cortes, C. and Vapnik, V., 1995. Support-vector networks. *Machine learning*, *20*(3), pp.273-297.

Croft, H., Chen, J.M., Zhang, Y. and Simic, A., 2013. Modelling leaf chlorophyll content in broadleaf and needle leaf canopies from ground, CASI, Landsat TM 5 and MERIS reflectance data. *Remote Sensing of Environment*, *133*, pp.128-140.

Croft, H., Chen, J.M., Zhang, Y., Simic, A., Noland, T.L., Nesbitt, N. and Arabian, J., 2015. Evaluating leaf chlorophyll content prediction from multispectral remote sensing data within a physically based modelling framework. *ISPRS Journal of Photogrammetry and Remote Sensing*, *102*, pp.85-95.

Darvishzadeh, R., Skidmore, A., Abdullah, H., Cherenet, E., Ali, A., Wang, T., Nieuwenhuis, W., Heurich, M., Vrieling, A., O'Connor, B. and Paganini, M., 2019. Mapping leaf chlorophyll content from Sentinel-2 and RapidEye data in spruce stands using the invertible forest reflectance model. *International Journal of Applied Earth Observation and Geoinformation*, *79*, pp.58-70.

Delloye, C., Weiss, M. and Defourny, P., 2018. Retrieval of the canopy chlorophyll content from Sentinel-2 spectral bands to estimate nitrogen uptake in intensive winter wheat cropping systems. *Remote Sensing of Environment*, *216*, pp.245-261.

Fernández-Marín, B., Artetxe, U., Barrutia, O., Esteban, R., Hernández, A. and García-Plazaola, J.I., 2015. Opening Pandora's box: cause and impact of errors on plant pigment studies. *Frontiers in plant science*, *6*, p.148.

Figueiredo, G.K.D.A., Brunsell, N.A., Higa, B.H., Rocha, J.V. and Lamparelli, R.A.C., 2016. Correlation maps to assess soybean yield from EVI data in Paraná State, Brazil. *Scientia Agricola*, *73*(5), pp.462-470.

Forkuor, G., Hounkpatin, O.K., Welp, G. and Thiel, M., 2017. High resolution mapping of soil properties using remote sensing variables in south-western Burkina Faso: a comparison of machine learning and multiple linear regression models. *PloS one*, *12*(1), p.e0170478.

Frampton, W.J., Dash, J., Watmough, G. and Milton, E.J., 2013. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS journal of photogrammetry and remote sensing*, *82*, pp.83-92.

Gcayi, S.R., Chirima, G.J., Adelabu, S.A., Adam, E. and Abutaleb, K., 2019. Evaluating the Potential of Narrow-Band Indices to Predict Soybean (Glycine Max L. Merr) Grain Yield in The Free State and Mpumalanga of South Africa. *Open Access Journal Of Environmental & Soil Science*, *3*(1), pp.265-278.

Ghosh, P., Mandal, D., Bhattacharya, A., Nanda, M.K. and Bera, S., 2018. Assessing crop monitoring potential of Sentinel-2 in a spatio-temporal scale. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *42*, pp.227-231.

Gitelson, A. and Merzlyak, M.N., 1994. Spectral reflectance changes associated with autumn senescence of Aesculus hippocastanum L. and Acer platanoides L. leaves. Spectral features and relation to chlorophyll estimation. *Journal of plant physiology*, *143*(3), pp.286-292.

Gitelson, A.A., Gritz, Y. and Merzlyak, M.N., 2003. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of plant physiology*, *160*(3), pp.271-282.

Govender, M., Chetty, K. and Bulcock, H., 2007. A review of hyperspectral remote sensing and its application in vegetation and water resource studies. *Water Sa*, *33*(2), pp.145-151.

Haboudane, D., Tremblay, N., Miller, J.R. and Vigneault, P., 2008. Remote estimation of crop chlorophyll content using spectral indices derived from hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, *46*(2), pp.423-437.

Hoeppner, J.M., Skidmore, A.K., Darvishzadeh, R., Heurich, M., Chang, H.C. and Gara, T.W., 2020. Mapping canopy chlorophyll content in a temperate forest using airborne hyperspectral data. *Remote sensing*, *12*(21), p.3573.

Hu, B., Jung, W.M., Liu, J. and Shang, J., 2022. Retrieval of Leaf Area Index and Leaf Chlorophyll Content from Hyperspectral Data Using Deep Learning Networks. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *43*, pp.397-404.

Hu, S., Liu, H., Zhao, W., Shi, T., Hu, Z., Li, Q. and Wu, G., 2018. Comparison of machine learning techniques in inferring phytoplankton size classes. *Remote Sensing*, *10*(3), p.191.

Jiang, C., Johkan, M., Hohjo, M., Tsukagoshi, S. and Maruo, T., 2017. A correlation analysis on chlorophyll content and SPAD value in tomato leaves. *HortResearch*, *71*, pp.37-42.

Joshi, A., Pradhan, B., Chakraborty, S. and Behera, M.D., 2023. Winter wheat yield prediction in the conterminous United States using solar-induced chlorophyll fluorescence data and XGBoost and random forest algorithm. *Ecological Informatics*, *77*, p.102194.

Jung, J., Maeda, M., Chang, A., Bhandari, M., Ashapure, A. and Landivar-Bowles, J., 2021. The potential of remote sensing and artificial intelligence as tools to improve the resilience of agriculture production systems. *Current Opinion in Biotechnology*, *70*, pp.15-22.

Kamara, A., Conteh, A., Rhodes, E.R. and Cooke, R.A., 2019. The relevance of smallholder farming to African agricultural growth and development. *African Journal of Food, Agriculture, Nutrition and Development*, *19*(1), pp.14043-14065.

Kanning, M., Kühling, I., Trautz, D. and Jarmer, T., 2018. High-resolution UAV-based hyperspectral imagery for LAI and chlorophyll estimations from wheat for yield prediction. *Remote sensing*, *10*(12), p.2000.

Kasim, N., Sawut, R., Abliz, A., Qingdong, S., Maihmuti, B., Yalkun, A. and Kahaer, Y., 2018. Estimation of the relative chlorophyll content in spring wheat Based on an optimized spectral index. *Photogrammetric Engineering & Remote Sensing*, *84*(12), pp.801-811.

Lee, H., Nam, G., Kang, T. and Yoon, S., 2014. Estimating Chlorophyll-a Concentration using Spectral Mixture Analysis from RapidEye Imagery in Nak-dong River Basin. *Journal of Korean Society on Water Environment*, *30*(3), pp.329-339.

Li, C., Zhu, X., Wei, Y., Cao, S., Guo, X., Yu, X. and Chang, C., 2018. Estimating apple tree canopy chlorophyll content based on Sentinel-2A remote sensing imaging. *Scientific reports*, *8*(1), p.3756.

Li, F., Miao, Y., Feng, G., Yuan, F., Yue, S., Gao, X., Liu, Y., Liu, B., Ustin, S.L. and Chen, X., 2014. Improving estimation of summer maize nitrogen status with red edge-based spectral vegetation indices. *Field Crops Research*, *157*, pp.111-123.

Li, M., Shamshiri, R.R., Weltzien, C. and Schirrmann, M., 2022. Crop monitoring using Sentinel-2 and UAV multispectral imagery: A comparison case study in Northeastern Germany. *Remote Sensing*, *14*(17), p.4426.

Liang, Y., Xu, Q.S., Li, H.D. and Cao, D.S., 2016. *Support vector machines and their application in chemistry and biotechnology*. CRC Press.

Lin, S., Li, J., Liu, Q., Li, L., Zhao, J. and Yu, W., 2019. Evaluating the effectiveness of using vegetation indices based on red-edge reflectance from Sentinel-2 to estimate gross primary productivity. *Remote Sensing*, *11*(11), p.1303.

Liu, B., Yue, Y.M., Li, R., Shen, W.J. and Wang, K.L., 2014. Plant leaf chlorophyll content retrieval based on a field imaging spectroscopy system. *Sensors*, *14*(10), pp.19910-19925.

Lu, B. and He, Y., 2019. Evaluating empirical regression, machine learning, and radiative transfer modelling for estimating vegetation chlorophyll content using bi-seasonal hyperspectral images. *Remote Sensing*, *11*(17), p.1979.

Madonsela, S., Cho, M.A., Naidoo, L., Main, R. and Majozi, N., 2023. Exploring the utility of Sentinel-2 for estimating maize chlorophyll content and leaf area index across different growth stages. *Journal of Spatial Science*, *68*(2), pp.339-351.

Marshall, M. and Thenkabail, P., 2015. Advantage of hyperspectral EO-1 Hyperion over multispectral IKONOS, GeoEye-1, WorldView-2, Landsat ETM+, and MODIS vegetation indices in crop biomass estimation. *ISPRS Journal of Photogrammetry and Remote Sensing*, *108*, pp.205-218.

MicaSense. (2020). Available from : https://support.micasense.com/hc/en-us/articles/360011389334-RedEdge-MX-Integration-Guide [Accessed 06 December 2023]

Munghemezulu, C., Mashaba-Munghemezulu, Z., Ratshiedana, P.E., Economon, E., Chirima, G. and Sibanda, S., 2023. Unmanned Aerial Vehicle (UAV) and Spectral Datasets in South Africa for Precision Agriculture. *Data*, *8*(6), p.98.

Musvoto, C., Maherry, A. and Adeyemo, O., 2009. report: Environmental Management Framework for Dihlabeng Local Municipality. *Forestry*, *701*, pp.0-15.

Mutanga, O., Adam, E., Adjorlolo, C. and Abdel-Rahman, E.M., 2015. Evaluating the robustness of models developed from field spectral data in predicting African grass foliar nitrogen concentration using WorldView-2 image as an independent test dataset. *International Journal of Applied Earth Observation and Geoinformation*, *34*, pp.178-187.

Narmilan, A., Gonzalez, F., Salgadoe, A.S.A., Kumarasiri, U.W.L.M., Weerasinghe, H.A.S. and Kulasekara, B.R., 2022. Predicting canopy chlorophyll content in sugarcane crops using machine learning algorithms and spectral vegetation indices derived from UAV multispectral imagery. *Remote Sensing*, *14*(5), p.1140.

Nguyen, H.Q., Ha, N.T., Nguyen-Ngoc, L. and Pham, T.L., 2021. Comparing the performance of machine learning algorithms for remote and in situ estimations of chlorophyll-a content: A case study in the Tri An Reservoir, Vietnam. *Water Environment Research*, *93*(12), pp.2941-2957.

Nofrizal, A.Y., Sonobe, R., HIROTO, Y., Morita, A., and Takashi, I.K.K.A., 2022. Estimating chlorophyll content of Zizania latifolia with hyperspectral data and random forest. *International Journal of Engineering and Geosciences*, *7*(3), pp.221-228.

Omer, G., Mutanga, O., Abdel-Rahman, E.M. and Adam, E., 2016. Empirical prediction of leaf area index (LAI) of endangered tree species in intact and fragmented indigenous forests ecosystems using WorldView-2 data and two robust machine learning algorithms. *Remote Sensing*, *8*(4), p.324.

Oyama, Y., Matsushita, B., Fukushima, T., Nagai, T. and Imai, A., 2007. A new algorithm for estimating chlorophyll-a concentration from multi-spectral satellite data in case II waters: a simulation based on a controlled laboratory experiment. International Journal of Remote Sensing, 28(7), pp.1437-1453.

Patra, P.P., Dubey, S.K., Trivedi, R.K., Sahu, S.K. and Rout, S.K., 2017. Estimation of chlorophyll-a concentration and trophic states in Nalban Lake of East Kolkata Wetland, India from Landsat 8 OLI data. *Spatial Information Research*, *25*, pp.75-87.

Peng, Z., Guan, L., Liao, Y. and Lian, S., 2019. Estimating total leaf chlorophyll content of gannan navel orange leaves using hyperspectral data based on partial least squares regression. *IEEE Access*, *7*, pp.155540-155551.

Pimstein, A., Karnieli, A., Bansal, S.K. and Bonfil, D.J., 2011. Exploring remotely sensed technologies for monitoring wheat potassium and phosphorus using field spectroscopy. *Field Crops Research*, *121*(1), pp.125-135.

Prado Osco, L., Marques Ramos, A.P., Roberto Pereira, D., Akemi Saito Moriya, É., Nobuhiro Imai, N., Takashi Matsubara, E., Estrabis, N., de Souza, M., Marcato Junior, J., Gonçalves, W.N. and Li, J., 2019. Predicting canopy nitrogen content in citrus-trees using random forest algorithm associated to spectral vegetation indices from UAV-imagery. *Remote Sensing*, *11*(24), p.2925.

Priyanka D., Srivastava, P.K. and Rawat, R., 2023. Retrieval of leaf chlorophyll content using drone imagery and fusion with Sentinel-2 data. *Smart Agricultural Technology*, *6*, p.100353.

Qi, H., Wu, Z., Zhang, L., Li, J., Zhou, J., Jun, Z. and Zhu, B., 2021. Monitoring of peanut leaves chlorophyll content based on drone-based multispectral image feature extraction. *Computers and Electronics in Agriculture*, *187*, p.106292.

Qiao, L., Tang, W., Gao, D., Zhao, R., An, L., Li, M., Sun, H. and Song, D., 2022. UAV-based chlorophyll content estimation by evaluating vegetation index responses under different crop coverages. *Computers and Electronics in Agriculture*, *196*, p.106775.

Quang, N.H., Quinn, C.H., Carrie, R., Stringer, L.C., Hackney, C.R. and Van Tan, D., 2022. Comparisons of regression and machine learning methods for estimating mangrove above-ground biomass using multiple remote sensing data in the red River Estuaries of Vietnam. *Remote Sensing Applications: Society and Environment*, *26*, p.100725.

Ramos, T.B., Castanheira, N., Oliveira, A.R., Paz, A.M., Darouich, H., Simionesei, L., Farzamian, M. and Gonçalves, M.C., 2020. Soil salinity assessment using vegetation indices derived from Sentinel-2 multispectral data. application to Lezíria Grande, Portugal. *Agricultural Water Management*, *241*, p.106387.

Sahu, P.K., Singh, D.P., Prabha, R., Meena, K.K. and Abhilash, P.C., 2019. Connecting microbial capabilities with the soil and plant health: Options for agricultural sustainability. *Ecological Indicators*, *105*, pp.601-612.

Salami, A., Kamara, A.B. and Brixiova, Z. (2010) *Smallholder agriculture in East Africa: Trends, constraints and opportunities*. African Development Bank Tunis, Tunisia.

Schlemmer, M., Gitelson, A., Schepers, J., Ferguson, R., Peng, Y., Shanahan, J. and Rundquist, D., 2013. Remote estimation of nitrogen and chlorophyll contents in maize at leaf and canopy levels. *International journal of applied earth observation and geoinformation*, *25*, pp.47-54.

Sekhele, N.M., 2018. Assessing the effects of grazing on vegetation cover and associated socio-economic livelihoods in the Clarens Nature Reserve in the Free State, South Africa.

Shah, S.H., Angel, Y., Houborg, R., Ali, S. and McCabe, M.F., 2019. A random forest machine learning approach for the retrieval of leaf chlorophyll content in wheat. *Remote Sensing*, *11*(8), p.920.

Shawe-Taylor, J. and Cristianini, N., 2004. *Kernel methods for pattern analysis*. Cambridge university press.

Shoko, C. and Mutanga, O., 2017. Examining the strength of the newly-launched Sentinel 2 MSI sensor in detecting and discriminating subtle differences between C3 and C4 grass species. *ISPRS Journal of Photogrammetry and Remote Sensing*, 129, pp.32-40.

Simic, A. and Chen, J.M., 2008. Refining a hyperspectral and multiangle measurement concept for vegetation structure assessment. *Canadian Journal of Remote Sensing*, *34*(3), pp.174-191.

Singhal, G., Bansod, B., Mathew, L., Goswami, J., Choudhury, B.U. and Raju, P.L.N., 2019. Chlorophyll estimation using multi-spectral unmanned aerial system based on machine learning techniques. *Remote Sensing Applications: Society and Environment*, *15*, p.100235.

Sishodia, R.P., Ray, R.L. and Singh, S.K., 2020. Applications of remote sensing in precision agriculture: A review. *Remote sensing*, *12*(19), p.3136.

Sonobe, R., Hirono, Y. and Oi, A., 2020. Non-destructive detection of tea leaf chlorophyll content using hyperspectral reflectance and machine learning algorithms. *Plants*, *9*(3), p.368.

Süß, A., Danner, M., Obster, C., Locherer, M., Hank, T., Richter, K. and EnMAP Consortium, 2015. Measuring leaf chlorophyll content with the Konica Minolta SPAD-502Plus.

Tahir, M.N., Naqvi, S.Z.A., Lan, Y., Zhang, Y., Wang, Y., Afzal, M., Cheema, M.J.M. and Amir, S., 2018. Real time estimation of chlorophyll content based on vegetation indices derived from multispectral UAV in the kinnow orchard. *International Journal of Precision Agricultural Aviation*, *1*(1).pp 24-32

Thenkabail, P.S., Enclona, E.A., Ashton, M.S. and Van Der Meer, B., 2004. Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. *Remote sensing of environment*, *91*(3-4), pp.354-376.

Tong, A. and He, Y., 2017. Estimating and mapping chlorophyll content for a heterogeneous grassland: Comparing prediction power of a suite of vegetation indices across scales between years. *ISPRS Journal of Photogrammetry and Remote Sensing*, *126*, pp.146-167.

Tshabalala, N.N., 2020. *The utility of Sentinel-2 MSI to assess wetland vegetation chlorophyll content and leaf area index in wetland areas in Pietermaritzburg, KwaZulu-Natal* (Doctoral dissertation).

Vincini, M. and Frazzi, E., 2011. Comparing narrow and broad-band vegetation indices to estimate leaf chlorophyll content in planophile crop canopies. *Precision Agriculture*, *12*(3), pp.334-344.

Wang, T., Gao, M., Cao, C., You, J., Zhang, X. and Shen, L., 2022. Winter wheat chlorophyll content retrieval based on machine learning using in situ hyperspectral data. *Computers and Electronics in Agriculture*, *193*, p.106728.

Wang, Y., Ryu, D., Park, S., Fuentes, S. and O'Connell, M., 2017, December. Upscaling UAV-borne high-resolution vegetation index to satellite resolutions over a vineyard. In *2nd International Congress on Modelling and Simulation, Hobart, Tasmania, Australia* (Vol. 3).

Whitford, W.G., De Soyza, A.G., Van Zee, J.W., Herrick, J.E. and Havstad, K.M., 1998. Vegetation, soil, and animal indicators of rangeland health. *Environmental Monitoring and Assessment*, *51*(1), pp.179-200.

Wu, C., Niu, Z. and Gao, S., 2012. The potential of the satellite derived green chlorophyll index for estimating midday light use efficiency in maize, coniferous forest and grassland. *Ecological Indicators*, *14*(1), pp.66-73.

Wu, J., Wang, D., Rosen, C.J. and Bauer, M.E., 2007. Comparison of petiole nitrate concentrations, SPAD chlorophyll readings, and QuickBird satellite imagery in detecting nitrogen status of potato canopies. *Field crops research*, *101*(1), pp.96-103.

Wu, Q., Zhang, Y., Zhao, Z., Xie, M. and Hou, D., 2023. Estimation of relative chlorophyll content in spring wheat based on multi-temporal UAV remote sensing. *Agronomy*, *13*(1), p.211.

Xue, J. and Su, B., 2017. Significant remote sensing vegetation indices: A review of developments and applications. *Journal of sensors*, *2017*(1). p.1353691.

Xue, L. and Yang, L., 2009. Deriving leaf chlorophyll content of green-leafy vegetables from hyperspectral reflectance. *ISPRS Journal of Photogrammetry and Remote Sensing*, *64*(1), pp.97-106.

Yang, H., Hu, Y., Zheng, Z., Qiao, Y., Zhang, K., Guo, T. and Chen, J., 2022. Estimation of potato chlorophyll content from UAV multispectral images with stacking ensemble algorithm. *Agronomy*, *12*(10), p.2318.

Yang, S., Hu, L., Wu, H., Ren, H., Qiao, H., Li, P. and Fan, W., 2021. Integration of crop growth model and random forest for winter wheat yield estimation from UAV hyperspectral imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, pp.6253-6269.

Yin, C., He, B., Quan, X. and Liao, Z., 2016. Chlorophyll content estimation in arid grasslands from Landsat-8 OLI data. *International Journal of Remote Sensing*, *37*(3), pp.615-632.

Yin, H., Huang, W., Li, F., Yang, H., Li, Y., Hu, Y. and Yu, K., 2023. Multi-temporal UAV Imaging-Based Mapping of Chlorophyll Content in Potato Crop. *PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, *91*(2), pp.91-106.

Zha, H., Miao, Y., Wang, T., Li, Y., Zhang, J., Sun, W., Feng, Z. and Kusnierek, K., 2020. Improving unmanned aerial vehicle remote sensing-based rice nitrogen nutrition index prediction with machine learning. *Remote Sensing*, *12*(2), p.215.

Zhai, W., Li, C., Cheng, Q., Ding, F. and Chen, Z., 2023. Exploring multisource feature fusion and stacking ensemble learning for accurate estimation of maize chlorophyll content using unmanned aerial vehicle remote sensing. *Remote Sensing*, *15*(13), p.3454.

Zhang, C. and Kovacs, J.M., 2012. The application of small unmanned aerial systems for precision agriculture: a review. *Precision agriculture*, *13*(6), pp.693-712.

Zhang, J., Han, W., Huang, L., Zhang, Z., Ma, Y. and Hu, Y., 2016. Leaf chlorophyll content estimation of winter wheat based on visible and near-infrared sensors. *Sensors*, *16*(4), p.437.

Zhang, L., Han, W., Niu, Y., Chavez, J.L., Shao, G. and Zhang, H., 2021. Evaluating the sensitivity of water stressed maize chlorophyll and structure based on UAV derived vegetation indices. *Computers and Electronics in Agriculture*, *185*, p.106174.

Zheng, T., Liu, N., Wu, L., Li, M., Sun, H., Zhang, Q. and Wu, J., 2018. Estimation of chlorophyll content in potato leaves based on spectral red edge position. *IFAC-PapersOnLine*, *51*(17), pp.602-606.

Zhou, X., Yang, L., Wang, W. and Chen, B., 2021. UAV Data as an Alternative to Field Sampling to Monitor Vineyards Using Machine Learning Based on UAV/Sentinel-2 Data Fusion. *Remote Sensing*, *13*(3), p.457.

Zhou, X., Zhang, J., Chen, D., Huang, Y., Kong, W., Yuan, L., Ye, H. and Huang, W., 2020. Assessment of leaf chlorophyll content models for winter wheat using Landsat-8 multispectral remote sensing data. *Remote Sensing*, *12*(16), p.2574.

Zhu, W., Sun, Z., Yang, T., Li, J., Peng, J., Zhu, K., Li, S., Gong, H., Lyu, Y., Li, B. and Liao, X., 2020. Estimating leaf chlorophyll content of crops via optimal unmanned aerial vehicle hyperspectral data at multi-scales. *Computers and Electronics in Agriculture*, *178*, p.105786.

Zillmann, E., Schönert, M., Lilienthal, H., Siegmann, B., Jarmer, T., Rosso, P. and Weichelt, T., 2015. Crop ground cover fraction and canopy chlorophyll content mapping using RapidEye imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *40*, pp.149-155.