UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

Integrating Sentinel-1/2 and machine learning models for mapping fruit tree species in heterogeneous landscapes of Limpopo

Name of student: Yingisani Winny Chabalala

Student number: 1251589

Name of supervisor: Elhadi Adam

A thesis submitted to the Faculty of Science, University of Witwatersrand, Johannesburg, in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Geography and Environmental Sciences

October 2024

Johannesburg

South Africa

Abstract

From ancient times to this century, Africa has relied chiefly on agriculture for survival. Crop type maps are crucial for agricultural management, sustainable farming systems, and realizing food security. Agronomists, agricultural extension officers, policymakers, and the government rely on crop type spatial distribution information to make informed decisions and optimize resource allocation for sustainable agricultural management. Attaining food security for all is an urgent need in Africa. However, the farming landscapes predominately comprise fragmented smallholder heterogeneous farms. The farming systems include intercropping and cultivating different crops that require different management strategies. This results in within-class spectral similarities and intra-spectral variability due to similar canopy structures and different phenologies, which complicates the application of remote sensing in crop type mapping. The free availability of Copernicus products such as Sentinel 1 and 2 have high temporal, spectral, and spatial resolution suitable for mapping smallholder agriculture. Thus, this research aimed to integrate Sentinel-1/2 and machine learning models for mapping fruit tree species in heterogeneous landscapes of Limpopo.

First, the research tested the applicability of sampling techniques and five mapping classifiers (i.e., Random Forest (RF), Support vector Machine (SVM), Adaptive Boosting (AdaBoost), Gradient Boosting (GB), and eXtreme Gradient Boosting (XGBoost) in mapping fruit trees and co-existing land use types. The original dataset was under-sampled randomly into two balanced datasets (i.e., Dataset 1 and Dataset 2) consisting of 100 and 150 sample points. Furthermore, the imbalanced ratio from the original dataset was reduced by applying different sampling strategies to extract four imbalanced datasets (i.e., at 40%, 50%, 60%, and 70%), which resulted in the formation of Dataset 3, Dataset 4, and Dataset 5, respectively. These samples, together with the original dataset (i.e., Dataset 7), were used as input to Sentinel-2 (S2) data using adaptive boosting (AdaBoost), gradient boosting (GB), random forest (RF), support vector machine (SVM), and eXtreme gradient boost (XGBoost) machine learning algorithms. The results showed that reducing the amount of imbalanced ratio by randomly under-sampling the original imbalanced dataset could increase the classification accuracy to 71% using the SVM classifier and 60% of the original dataset. Individually, the majority of the crop types were classified with an F1 score of between 60% and 100%.

Secondly, the research independently assessed the effectiveness of Sentinel-1 (S1) and Sentinel-2 (S2) data for fruit tree mapping using random forest (RF) and support vector machine (SVM) classifiers. Four models were tested using each sensor independently and

fusing both sensors. From the fused model, features were ranked using the RF mean decrease accuracy (MDA) and forward variable selection (FVS) to identify optimal spectral windows to classify fruit trees. The best fruit tree map with an overall accuracy (OA) of 0.91.6% with a kappa coefficient of 0.91% was produced using the RF MDA and FVS model and SVM classifier. The application of SVM to S1, S2, S2 selected variables and S1S2 fusion independently produced OA = 27.64, Kappa coefficient = 0.13%; OA= 87%, Kappa coefficient = 86.89%; OA = 69.33, Kappa coefficient = 69. %; A = 87.01%, Kappa coefficient = 87%, respectively. The green (B3), SWIR_2 (B10), and vertical horizontal (VH) polarization bands were identified as the optimal spectral features for S2 and S1 data, respectively.

The third part of the research identified the optimal growth window period in which fruit trees can be detected with high accuracy. Phenological metrics were extracted from 12 months (i.e., January to December) of Sentinel-2 (S2) data and were used to classify fruit trees using a random forest (RF) classifier in a Google Earth Engine environment. The results showed that fruit trees can be detected and mapped with high accuracy during winter months (i.e., April-July) with an overall accuracy (OA) of 84.89% and a kappa coefficient of 83%. The user accuracy ranged from 62 to 100%, while the producer accuracy ranged from 60 to 100%. The fruit trees were mostly differentiated from co-existing land use types using the short infrared and the red-edge bands.

The fourth part of the thesis attempted to increase fruit tree classification accuracy by classifying optimal Sentinel-2 images acquired during the fruit trees' critical growth stages using a Deep Neural Network (DNN) model. This was achieved by applying phenological metrics derived from Sentinel-2 images acquired during optimal crop-growing seasons (i.e., flowering, fruiting, harvesting). The DNN models were optimized by tuning the hyperparameters to achieve the best classification results. The DNN produced an OA of 86.96%, 88.64%, 86.76%, and 87.25% for April, May, June, and July images, respectively. The results indicate the DNN models were robust and stable across the selected fruit growth periods.

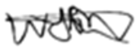
This research has shown that earth observation (EO) data such as Sentinel 1 and 2 can be used to map fruit trees in fragmented sub-tropical horticultural landscapes characterized by different environmental conditions and different crop cultivars operating under different management practices. The research results will assist agricultural stakeholders (i.e., farm managers, agronomists, agricultural extension officers, and policymakers) in allocating agricultural resources, devising effective agricultural management strategies, and attaining sustainable agriculture and food security.

Keywords: Data fusion, smallholder agriculture, horticulture, fruit-tree crops, phenology, fruit growth dynamics, optical data, SAR, Sentinel-2 data, Sentinel-1, imbalanced data; data sampling techniques, crop types mapping, Google Earth Engine, machine learning algorithms, deep neural network, precision agriculture.

Declaration 1 – Plagiarism

I, Yingisani Winny Chabalala, declare that this thesis titled "Integrating Sentinel-1/2 and Machine Learning Models for Mapping Fruit Tree Species in Heterogeneous Landscapes of Limpopo" and the work presented in it are my own. I confirm that:

1. This work was done while registered as a PhD candidate at the University of Witwatersrand in Johannesburg, South Africa,
2. The work reported in this thesis is my original research except where it is stated otherwise,
3. This work is being submitted to the University of Witwatersrand for examination for the first time and has not been submitted to any other university,
4. The literature cited in this research has been acknowledged and a list containing all sources of literature used has been provided at the end of this thesis.
5. The work conducted in collaboration with research assistants and other researchers has been acknowledged.



Signature:

Date: 17 October 2024

Declaration 2– Published papers

1. **Chabalala, Y.**; Adam, E and Ali, A.K. (2023). Exploring the effect of balanced and imbalanced multi-class distribution data and sampling techniques on fruit-tree crop classification using different machine learning classifiers and remote sensing data: *Journal of Geomatics*. 3 (1), 70-92. DOI: https://doi.org/10.3390/geomatics3010004
2. **Chabalala, Y.**; Adam, E and Ali, A.K. (2022). Machine Learning Classification of Fused Sentinel-1 and Sentinel-2 Image Data towards Mapping Fruit Plantations in Highly Heterogeneous Landscapes. *Journal of Remote Sensing*, 14 (11), 1-26. DOI: https://doi.org/10.3390/rs14112621
3. **Chabalala, Y.**; Adam, E, Adem Ali Khalid. (2023). Identifying the optimal phenological period for discriminating fruit tree crops using multi-temporal Sentinel-2 data and Google Earth Engine in a heterogeneous subtropical horticultural region. *South African Journal of Geomatics*, 14 (2), 262-283. DOI: https://doi.org/10.4314/sajg.v12i.2.10
4. **Chabalala, Yingisani.**; Adam, Elhadi. and Kganyago, Mahlatse. (2023). Mapping fruit tree dynamics using phenological metrics from optimal Sentinel-2 data and Deep Neural Network. *CABI for Agriculture and Bioscience* 4 (51), 1-20. DOI: https://doi.org/10.1186/s43170-023-00193-z

Dedication

I can do all this through Christ, who gives me strength. You never promised that things would be easy, but you promised to always be by my side. Thank you, Lord, for making this milestone possible; it really exceeded my expectations. Indeed, many are our plans, but only your plans will prevail. The delays were not denials but a realignment of my paths because you promised to make it happen when the time was right, and now it is my turn. To God be the glory.

I dedicate this degree to my late parents, Mr. Mkhatshani Phineas “Xtapura” Mathonsi and Mrs. Khumburisa N’wa Ben Mathonsi “Gov gov”. You both witnessed my previous accomplishments, and unfortunately, you could not live to witness this one, as God knew he got this covered. However, I know you are celebrating with me in heaven. May your soul continue to rest in perfect internal peace.

Acknowledgements

I thank the University of Witwatersrand and the University of South Africa for awarding me staff bursaries to study for this PhD. I will be forever grateful to the School of Geography, Archaeology, and Environmental Studies for granting me access to research facilities used to carry out the work required for this research.

To my supervisor, Professor Elhadi Adam, thank you so much for your support and supervision. Thank you for allowing me an opportunity to co-supervise your students. The experience I acquired through the supervision has helped me to grow my knowledge in the field of remote sensing and, opened opportunities & paved the way that led to where I am today. I would also like to thank Professor Adem Khalid Ali and Dr Colbert Jackson for their support in proofreading and editing my papers and final thesis. A special thank you to staff members and friends from the School of Geography, Archaeology, and Environmental Science, Professor Paida Mhangara, Professor Karim Sadr, Natanya Alexander, Thandizwe Nsimbi, Dr Simbarashe Jombo, Dr Mahlatse Khanyago, Dr Emmah Mandishona, and Dr Lenkwalo Thabeng. Your support has enabled me to pursue my passion in the field of agricultural remote sensing while I hold a full-time job at the school.

To my lovely husband, Mr. George Chabalala "Mchavi", thank you for your support and caring for our family while I worked on my studies for extended hours for over three years. The few words that you always uttered, when necessary, helped me to navigate this journey easily. To my kids, Vuyelo, Minkhenso, and Mpfuxelelo, I dedicate this degree to you. Your existence has motivated me to work hard, even when I felt like giving up. To my sister, Glory Mathonsi "Mbuyesi", through this journey, we encountered a lot of tragedies in our family, including the loss of our father, who was our pillar of strength. Through it all, you were always available to listen to my struggles and offer support. Sometimes, I felt like giving up as I was not coping mentally and physically. Thank you for being a consistent sister, a counsellor, and a praying partner. Thank you for encouraging me to continue my studies, even though it was sometimes unbearable. To my nephew, Ndzivalelo Mathebula, "boti Thevu", as my kids call you, thank you for caring for my kids and playing a brother role to the best of your ability. You took up the brotherly role without fail for two consecutive years while I worked on my studies. I am grateful for your support and love. I knew when you were born, through the bond we shared during your early life, that you were a blessing to the Mathonsi family.

Contents

Abstract……………………………………………………………………………………...i

Declaration 1 – Plagiarism..iv

Declaration 2– Published papers.. v

Acknowledgment. vi

Keywords...iii

Contents…………………………………………………………………………………..viii

List of Figures……………………………………………………………………….. ..xiii

List of tables……………………………………………………………………………..xviii

List of abbreviations…………………………………………………………….. ….xix

CHAPTER ONE…………………………………………………………………...1

General introduction……………………………………………………………………1

1.1. The importance of fruit tree agriculture 2

1.2. Mapping and monitoring fruit trees using Earth Observation data.................................... 2

1.3. Research problem statement 6

1.4. Research aim and objectives 7

1.5. Scope of the research 7

1.6. Research area……………………………………………………………………………...8

1.7. Thesis outline……………………………………………………………………………..9

CHAPTER TWO…………………………………………………………………………11

Exploring the effect of balanced and imbalanced multi-class distribution data and sampling techniques on fruit-tree crop classification using different machine learning classifiers and remote sensing………………………………………………………...11

Abstract…………………………………………………………………………………12

2.1. Introduction………………………………………………………………………...13

2.2. Materials and methods 15

2.2.1. Methods Overview.. 15

2.2.2. Research region... 16

2.3. Data collection and processing 17

2.3.1. Ground truth data 17

2.3.2. Data sampling………………………………………………………………………18

2.3.3. Remote sensing data acquisition and pre-processing.. 22

2.4. Machine learning classification algorithms 23

2.4.1. Random Forest (RF) 23

2.4.2. Support Vector Machine (SVM)... 23

2.4.3. Gradient Boosting (GB) 24

2.4.4. Adaptive Boosting (AdaBoost)... 24

2.4.5. eXtreme Gradient Boosting (XGBoost).. 25

2.5. Mapping accuracy assessment 25

2.6. Results……………………………………………………………………………...26

2.6.1. Spectral separability of fruit-tree crops and co-existing land-use types 26

2.6.2. Fruit-tree crop mapping using machine learning algorithms 27

2.7. The variable importance 28

2.8. The class accuracies 29

2.9. Statistical comparison of ML classifiers from seven datasets 33

2.10. Discussion……………………………………………………………………...36

2.10.1. The effect of sampling size on the performance of machine learning algorithms 36

2.10.2. The statistical comparison of the classifiers among and within the classifiers............ 38

2.10.3. Comparison of individual class accuracies 38

2.10.4. Importance of variables in mapping fruit-tree crops and co-existing land-use types .. 39

2.11. Conclusions……………………………………………………………………….40

CHAPTER THREE……………………………………………………………………...42

Machine Learning classification of fused sentinel-1 and sentinel-2 image data towards mapping fruit plantations in highly heterogeneous landscapes.. 42

Abstract………………………………………………………………………………….43

3.1. Introduction…………………………………………………………………………...44

3.2. Materials and Methods 48
3.2.1. Method Overview 48
3.2.2. Research region 49
3.2.3. Remote Sensing Data Acquisition and Pre-Processing 50
3.2.4. Data Fusion 53
3.2.5. Field Data Collection 54
3.3. Image classification 57
3.3.1.1. The random forest algorithm (RF) 57
3.3.1.2. Assessing the importance of variables 58
3.1.3.3. Forward variable selection 59
3.6. Mapping outputs 63
3.7. Classification accuracy 66
3.8. Class accuracies 68
3.9. Discussion 70
3.10. Conclusions 74
CHAPTER FOUR 75
Identifying the optimal phenological period for discriminating fruit tree crops using multi-temporal Sentinel-2 data and Google Earth Engine in a heterogeneous subtropical horticultural region 75
Abstract 76
4.1. Introduction 77
4.2. Materials and Methods 79
4.2.1. Research area 79
4.3. Earth observation data and pre-processing 80
4.3.1. Sentinel-2 data and pre-processing 80
4.4. Acquisition of ground control points 82
4.5. Feature extraction 84
4.6. Spectral separability of fruit tree crops 84
4.7. Training of the random forest classifier 84

4.8. Accuracy assessment.. 85

4.9. Statistical analysis to determine the reliability of the research results 85

4.10. Results……………………………………………………………………………….85

4.10.1. Spectral separability between fruit tree crops 85

4.10.2. Relative importance of variables.. 86

4.10.3. Accuracy assessment.. 87

4.10.4. Statistical analysis to determine the reliability of the research results 89

4.10.5. Spatial distribution of fruit tree crops 89

4.11. Discussion…………………………………………………………………………...93

4.12. Limitations…………………………………………………………………………..95

CHAPTER FIVE……………………………………………………………………...97

Mapping fruit tree dynamics using phenological metrics derived from optimal Sentinel-2 data and Deep Neural Network 97

Abstract……………………………………………………………………………………98

5.1. Introduction…………………………………………………………………………...99

5.3. Research area……………………………………………………………………...103

5.4. Data collection and processing 104

5.4.2. Spectral characteristics of fruit-tree crops and co-existing land use types 105

5.5. Earth observation data and pre-processing 107

5.5.1. Sentinel-2 data……………………………………………………………………..107

5.5.2. Image pre-processing... 108

5.5. Accuracy assessment 110

5.6. Results………………………………………………………………………………111

5.6.1. Overall accuracies of the DNN models 111

5.6.2. Performance of the classification models 113

5.6.3. Mapping outputs 114

5.7. Discussion…………………………………………………………………………...122

5.7.1. Overall classification accuracies... 122

5.7.3. The spatial distribution maps 124

5.8. Limitations………………………………………………………………………..125

5.9. Conclusions………………………………………………………………………125

CHAPTER SIX………………………………………………………………………127

Integrating Sentinel-1/2 and machine learning models for mapping fruit tree species in heterogeneous landscapes of Limpopo: A synthesis 127

6.1. Introduction……………………………………………………………………128

6.2. Exploring the effect of balanced and imbalanced multi-class distribution data and sampling techniques on fruit-tree crop classification using different machine learning classifiers and remote sensing data 129

6.3. To test the effectiveness of machine learning classification of fused Sentinel-1 and Sentinel-2 Image data towards mapping fruit plantations in highly heterogeneous landscape………………………………………………………………………………130

6.4. To identify the optimal phenological period for discriminating fruit tree crops using multi-temporal Sentinel-2 data and Google Earth Engine in a heterogeneous subtropical horticultural region………………………………………………………………………131

6.5. To assess the applicability of optimal Sentinel-2 data and Deep Neural Network in mapping fruit tree dynamics using phenological metrics. 133

6.6. Conclusions……………………………………………………………………133

6.7. Research limitations 134

During this research, the following limitations were encountered: 134

6.8. Future directions………………………………………………………………..135

References……………………………………………………………………………137

List of Figures

Figure 1.1. Map of the research site showing Levubu sub-tropical farms located in Collins Chabane and Makhado Local municipalities (light green) of the Limpopo Province (grey), South Africa…………………………………………………………………………...…7

Figure 2.1: Flowchart of the methodology developed in this research……………………...16

Figure 2.2: Map of the study region, depicting the Levubu subtropical farming region located in the Makhado and Collins Chabane Local Municipalities (lime green) of the Limpopo province (grey), South Africa………………………………………17

Figure 2.3: Location of fruit trees and the co-existing land-use types in the sub-tropical fruit farming area of Levubu. The background map represents a S2 RGB image obtained from European Space Agency…………………………………………..…..19

Figure 2.4: Spectral signature of fruit-tree crops, and co-existing land cover classes extracted from spectral Sentinel-2A algorithms image. The mapped classes are avocado (AV), banana (BN), bare soil (BS), guava (GV), macadamia nut (MN), mango (MG), pine tree (PT), built-up (BU), water body (WB), and woody vegetation (WV)…………………………………………………………………………..27

Figure 2.5: Comparison of the overall classification accuracies (a) and F1-Score (b) from GB, Ada Boost, RF, SVM, and XGBoost using seven different datasets. Dataset_1 (balanced undersampled at 100 points), Dataset_2 (balanced undersampled at 150 points), Dataset_3 (undersampled randomly at 40%), Dataset-4 (undersampled randomly at 50%), Dataset-5 (undersampled randomly at 60%), Dataset_6 (undersampled randomly at 70%), and Dataset-7 (imbalanced dataset)………....29

Figure 2.6: A comparative importance of explanatory variables used in permutation feature selection on seven datasets based on S2 spectral bands for Dataset 1 (balanced undersampled at 100 points), Dataset 2 (balanced undersampled at 150 points), Dataset 3 (undersampled randomly at 40%), Dataset 4 (undersampled randomly at 50%), Dataset 5 (undersampled randomly at 60%), Dataset 6 (undersampled randomly at 70%), and Dataset 7 (imbalanced dataset). The most influential bands are depicted by spikes indicating the highest feature importance scores. The color maroon, red, orange, yellow, lime green, green, light blue, blue, dark blue and purple represent Sentinel-2 bands, namely: Band 2 (Blue), Band 3 (Green), Band 4

(Red), Band 5 (Rededge_1), Band 6 (Rededge_2), Band 7 (Rededge_3), Band 8 (NIR), Band 8a (Rededge_4), Band 11 (SWIR_1) and Band 12 (SWIR_2)……..31

Figure 2.7: F1-scores rankings indicating the effect of sampling methods on class-specific accuracies obtained from GB, Ada Boost, RF, SVM, and XGBoost classifiers. The mapped classes are avocado (AV), banana (BN), bare soil (BS), built-up (BU), guava (GV), macadamia nut (MN), mango (MG), pine tree (PT), water body (WB), and woody vegetation (WV)…………………………………………………..32

Figure 2.8: User's (blue) and producer's (red) accuracy per class produced using GB, Ada Boost, XG-Boost, RF, and the SVM classifier. On each classifier, there is a trade-off between the user and producer accuracy across the datasets…………...33

Figure 2.9: The boxplot shows the mean (yellow line), the low and upper quantile (black whiskers) comparison of the model performance across datasets and among the models using students paired t-test, where (A) represents the gradient boosting; (B) AdaBoost; (C) random forest; (D) support vector machines; (E) eXtreme gradient boosting; while (F) depicts the comparison of model performance amongst models, respectively. From the x-axis, the 100, 150, 40, 50, 60, 70, and original dataset labels correspond to Dataset 1, Dataset 2, Dataset 3, Dataset 4, Dataset 5, Dataset 6, and Dataset 7, respectively. The green triangles represent the mean in relation to the overall performances of each classifier per dataset…………………………..35

Figure 3.1: Flowchart of the methodology implemented for mapping fruit trees and co-existing land-use types in Levubu, Limpopo, South Africa……………………………….49

Figure 3.2: Map of South African provinces showing the location of Levubu sub-tropical farming area in Sentinel-2 RGB image located in Makhado Local Municipality...50

Figure 3.3: Average spectral reflectance profiles extracted from Sentinel-2 data on fruit-tree crops and co-existing land-use types, reference data, and field photos. The mapped classes are avocado (AV), banana (BN), bare soil (BS), guava (GV), macadamia nut (MN), mango (MG), pine tree (PT), built-up (BU), water body (WB), and woody vegetation (WV)…………………………………………………….....57

Figure 3.4: The variable importance in the RF classification model indicates the mean decrease accuracy (MDA) for (a) selected variable models. A high MDA importance indicates that the predictor variable plays a crucial role in the classification. (b)

shows the mean decrease accuracy (MDA), indicating the relationship between S1S2 spectral bands and classes……………………………………………………62

Figure 3.5: Identification of the optimal variables based on RF MDA rankings produced using the forward variable selection (FVS). The arrow indicates the set of variables that have the lowest error rate……………………………………………………...63

Figure 3.6: Mapping fruit-tree crops and other existing land use types based on Sentinel-1 data ((A): RF, (B): SVM), Sentinel-2 ((C): RF, (D): SVM), Sentinel-2 selected variables ((E): RF, (F): SVM), Sentinel-1 and Sentinel-2 ((G): RF, (H): SVM), and Sentinel-1 and Sentinel-2 selected variables ((I): RF, (J): SVM)…………………………67

Figure 3.7: The overall accuracy (OA) obtained by RF (blue) and SVM (orange) classifiers using Sentinel1, Sentinel-2, Sentinel-2 selected variables, S1S2 fusion, and S1S2 selected variables in mapping tree crop species and other existing land uses…...68

Figure 4.1: Geographic location of Levubu subtropical farming area in Makhado and Collins Chabane Local Municipalities, Limpopo Province, South Africa……………….80

Figure 4.3: Spectral signatures of fruit tree crops at various growth stages extracted from Sentinel-2 images in 2019 (June and July) and 2020 (April and May): blue (B2), green (B3), red (B4), red-edge (B5, B6, B7, and B8A), near-infrared (B8), and shortwave-infrared (B11 and B12) bands………………………………………..86

Figure 4.4: Variable importance of bands in the best performing models used in mapping fruit trees using the Sentinel-2 dataset: (a) April, (b) May, (c) June, and (d) July……87

Figure 4.5: Comparison of the overall accuracy and kappa coefficients for the 12-month Sentinel-2 composites using a random forest classifier……………………......88

Figure 4.6: Producer and user accuracies for avocado, banana, bare soil, built-up, guava, macadamia nut, mango, pine tree, water body, and woody vegetation in Levubu, Limpopo Province, South Africa…………………………………………………..88

Figure 4.7: Areal coverage, in square kilometres, of fruit tree crops and co-existing land cover types in the study area. The areal calculations are based on the months of June 2019, July 2019, April 2020, and May 2020…………………………………....90

Figure 4.8: Classification maps of fruit tree crops with spatially explicit zoom-ins (a, b, and c), providing detailed spatial distributions from RF classifications in the research region for the best classification accuracy. They are based on four selected S2

monthly composite images captured in June (5A) and July 2019 (5B), and April (5C) and May 2020 (5D)…...……92

Figure 5.1: Flowchart showing the Methodology followed to identify fruit trees and other land use types using Sentinel-2 images and the DNN algorithm……………………103

Figure 5.2: Location of the research areas at the National (Top left corner) and District Municipality levels (Top right corner). Scene of the Sentinel-2 RGB image indicates the research area. The green (a) and light grey areas, (b) represent Limpopo Province and the Vhembe District Municipality, while the dark grey (c) represents the Makhado and Collins Chabane Local Municipalitie.....................104

Figure 5.3: The location distribution of fruit trees and other co-existing land-use types in the Levubu sub-tropical fruit farming area. The RGB bands represent S2 image downloaded from the European Space Agency………………………………...106

Figure 5.4: The variations in temporal spectral reflectance curves of within-season phenological stages of fruit trees and other land use types derived from Sentinel-2 multispectral images acquired over a different time range (April, May, June, and July). All fruit trees and their co-existing land use types are plotted together…...110

Figure 5.5: Comparison of the variation of model accuracy and loss function curves with increasing epochs on the prediction of the DNN models tested in different phenological stages of the fruit trees during April (Model 1), May (Model 2), June (Model 3: and July (model 4) ……………………………………………….……111

Figure 5.6: Relationship between the model accuracy and the loss function, showing the stability of the models across the tested 100 epochs…………………………....112

Figure 5.7: The User Accuracy and Producer Accuracy of the fruit trees and co-existing land use types in four growing months (April, May, June, and July) ……………….118

Figure 5.8: Discrimination results for fruit trees and co-existing land use types by DNN and Sentinel-2 images acquired in A April, B May, C June, and July……………….119

Figure 5.9: Fruit tree maps obtained using DNN with some zoomed spots showing the spatial structure of the classified fruit crops and co-existing land use types based on Sentinel-2 images across growing seasons in April, May, June, and July. The a–c depicts the zoomed areas of three sites with major classification differences obtained using the DNN classifier across the four months considered in this study

detected in July, while they were slightly detected in April to June image acquisitions…………………………………………………………………..121

List of tables

Table 2.1: Number of samples for fruit-tree crops and other existing land uses, the structure of training and testing data for each class used to map fruit-tree crops in Levubu subtropical farms...……………………………………………………..…..18

Table 3.1: Spectral configurations for Sentinel-1 and Sentinel-2 data……………………….52

Table 3.2: Confusion matrices for the models that produced the models that produced the highest overall accuracy using RF and SVM algorithms. The mapped classes are avocado (AV), banana (BN), bare soil (BS), guava (GV), macadamia nut (MN), mango (MG), pine tree (PT), built-up (BU), water body (WB), and woody vegetation (WV)………………………………………………………………65

Table 3.3: The performance of RF and SVM over four datasets (S1, S2, S2 selected variables, S1S2, S1S2 selected variables importance (VI) in terms of user's and producer's accuracies in percentages (%). The mapped classes are avocado (AV), banana (BN), bare soil (BS), guava (GV), macadamia nut (MN), mango (MG), pine tree (PT), built-up (BU), water body (WB), and woody vegetation (WV)…………..69

Table 4.1: Spectral configurations for the S2 ImageCollection used from GEE…………….81

Table 4.2: Number of Sentinel-2 images acquired for each monthly period for the Levubu subtropical farm region…………………………………………………………82

Table 4.3: Growth stages of fruit trees in the Levubu subtropical farming area (Note: The shades denote the flowering stage (Flo), fruiting stage (Fru), and harvesting stage (Harv)………………………………………………………………………83

Table 5.1: Phenology calendar of the major fruit trees in Levubu subtropical farms. The colour shading represents the flowering Flo fruiting (Fru) and harvesting (Harv)) stages…………………………………………………………………...107

Table 5.2: Spectral characteristics of Sentinel-2 sensor……………………………………108

Table 5.3: Sentinel-2 image and acquisition dates…………………………………………108

Table 5.4: Confusion matrices for the fruit trees and co-existing land use types in four growing months (April, May, June, and July)…………………………………………...115

List of abbreviations

α:	Alpha
AdaBoost	Addictive boosting
ADAM:	Adaptive Moment Estimation
AV:	Avocado
ANOVA:	Analysis of variance
ANN:	Artificial Neural networks
ATCOR:	ATmospheric CORrection
◦C:	Degree Celsius
CV:	Cross Validation
c:	Cost
CA:	Califonia
CNN:	Convolutional Neural Networks
dB:	Debicel
DEM:	Digital elevation model
DL:	Deep Learning
DNN:	Deep Neural Network
GB:	Gradient Boost
GPS:	Global positioning system
ESA:	European Space Agency
EO:	Earth Observation
FFS:	Forward feature selection
Fig:	Figure
FN:	False-Negatives
FP:	False-Positives
FVS:	Forward variable selection
GCPs:	Ground Control Points
GEE:	Google Earth Engine
GRD:	Ground range detected
ha:	Hectares
SRTM:	Shuttle Radar Topography Mission

S1S2:	Sentinel-1 Sentinel-2
HV:	Horizontal: Vertical
IW:	Interferometric Wide
Kappa:	Kappa coefficient
***k*-NN:**	*k*nearest neighbor
LULC:	Land use/land cover
m:	Metre
mm:	millimetre
MODIS:	Moderate Resolution Imaging Spectroradiometer
MDA:	Mean Decrease Accuracy
MDG:	Mean Decrease Gini
ML:	Machine Learning
MLR:	Multiple Linear Regression
***n*:**	Number
nm:	Nanometer
NIR:	Near Infrared
OA:	Overall Accuracy
OOB:	Out-of-bag
PA:	Producer Accuracy
RBF:	radial basis function
REP:	Red-edge position
ReLu:	Rectified linear unit
RF:	Random Forest
RMSprop:	Root Mean Squared Propagation
RS:	Remote Sensing
RUS:	Random undersampling
SAR:	Synthetic Aperture Radar
SDM:	Species Distribution Models
SDG:	Stochastic Gradient Descent
S1:	Sentinel-1
S2:	Sentinel-2
S2A:	Sentinel-2A

S2B:	Sentinel-2B
SMOTE:	Synthetic minority oversampling technique
SNAP:	Sentinels Application Platform
SPOT:	Satellite pour l'Observation de la Terre
SR:	Surface Reflectance
SVM:	Support vector machines
SWIR:	Short Wave Infrared
TN:	True negatives
TP:	True positives
TOA:	Top-Of-Atmosphere
UA:	User Accuracy
USA:	United States of America
UTM:	Universal Transverse Mercator
VI:	Variable Importance
VI:	Vegetation Indices
VV:	Vertical: Vertical
XGBoost:	Extreme Gradient Boost
XGBoost:	eXtreme gradient boosting
***y*:**	Gamma

CHAPTER ONE

General introduction

1.1. The importance of fruit tree agriculture

Fruits fight malnutrition and associated health problems by enhancing nutrition, promoting general health and well-being, and advancing food security (Jin *et al.*, 2019; Bey *et al.*, 2020). Fruit trees enable farmers to generate income and provide jobs at every stage of the value chain, from planting to harvesting to processing to marketing and distribution (Jamnadass *et al.*, 2011; Becker-reshef *et al.*, 2020). By boosting ecosystem resilience and offering a range of ecosystem services, fruit trees are essential to the sustainability of the environment (Kutya, 2013; Vogels *et al.*, 2020). Fruit trees contribute to biodiversity by offering a variety of animals, such as birds, insects, and mammals, homes and food sources (Vira *et al.*, 2015; Oeba and Illiassou, 2020). By cycling nutrients and accumulating organic matter, fruit trees' root systems enhance soil structure, reduce soil erosion, and increase soil fertility (Vira *et al.*, 2015). Trees—including fruit trees—absorb and store carbon dioxide from the atmosphere, thus helping to mitigate climate change (Gelaye and Getahun, 2024). By absorbing rainfall, lowering runoff, and encouraging groundwater recharge, fruit trees help regulate water resources and lessen the effects of floods and droughts (Vira *et al.*, 2015; Gelaye and Getahun, 2024). Fruit trees' canopies create favourable microclimates for plant growth and increase agricultural productivity by reducing wind speed, providing shade, and regulating temperature (Vira *et al.*, 2015; Gelaye and Getahun, 2024). Fruit tree integration into agroforestry systems improves land use efficiency, supports sustainable land management techniques, and improves ecosystem services (Deloitte, 2017). Therefore, encouraging fruit tree production and incorporating them into agricultural landscapes can help create food systems that are more sustainable, equitable, and resilient.

1.2. Mapping and monitoring fruit trees using Earth Observation data

Accurate mapping and monitoring of fruit tree distributions are necessary for efficient resource allocation and management (Kutya, 2013; Vogels *et al.*, 2020). Fruit tree mapping facilitates their distribution in agricultural environments—this knowledge helps farmers and agricultural extension services to maximize agricultural productivity by carefully considering where to plant fruit trees, choosing the best types, and putting into effect management techniques that are suited for the climate in the area (You and Dong, 2020; Wu *et al.*, 2020; Xiong *et al.*, 2022). The knowledge of fruit tree distribution enables the focused distribution of resources, including water, fertilizers, pesticides, and farming equipment. Increased resource efficiency minimizes waste and increases agricultural output by concentrating resources on regions with high concentrations of fruit trees or where they are most required (Wu *et al.*, 2020). Also, zoning laws, property tenure agreements, and conservation activities are among the land use planning considerations influenced by fruit tree distribution (Wu *et al.*, 2020; Xiong *et al.*, 2022). Planners can minimize conflicts and maximize land use efficiency by identifying suitable places for agroforestry systems, orchards, and other land uses by comprehending the spatial distribution of fruit trees (Mccarty *et al.,* 2017; Defourny *et al.,* 2019). Tracking variations in fruit tree distributions over time offers important insights into the dynamics of land cover, the fragmentation of habitats, and the health of ecosystems (Chabalala *et al*., 2023b; Chabalala *et al*., 2023c). Assessing the effects of shifting land uses, deforestation, and climate variability on fruit tree populations and related biodiversity requires the utilization of these data (Leroux *et al.,* 2019). To protect ecosystem services and preserve ecological balance, sensible management and conservation strategies can be created based on these discoveries (Leroux *et al.,* 2019, Chabalala *et al*., 2022). Accurate mapping of fruit tree distributions promotes market access and value chain growth by locating possible production areas, connecting producers to markets, and streamlining supply chain management (Ozdogan *et al.,* 2010). This makes it possible for farmers to reach markets, bargain for better terms, and enhance the value of their products through efforts in processing, packaging, and branding (Ozdogan *et al.,* 2010). Charting and tracking the spread of fruit trees aids in determining how vulnerable fruit tree ecosystems are to climate change and in creating risk-reduction plans (Tu *et al.*, 2019). To improve resilience and guarantee the long-term sustainability of fruit tree production systems, adaptive measures can be put into place by identifying locations that are more vulnerable to climate-related impacts, such as droughts, floods, and temperature extremes (Hadria *et al.*, 2010; Duveiller *et al*., 2013).

Conventional approaches to mapping fruit trees entail field surveys conducted on the ground, manual field observations, and community interviews (Wu *et al.*, 2020). Although these approaches have proven useful in the past, they have several drawbacks. To identify and map fruit trees, traditional methods frequently rely on ground surveys carried out by field personnel who physically visit areas (Geetha *et al.,* 2016; Leroux *et al.,* 2019). Although ground surveys offer comprehensive details about individual trees and their surrounds, they are costly, time-consuming, and labour-intensive, particularly when they cover vast areas (Chemura *et al*., 2015). In field observations, bias, inconsistency, and human error can occur during manual field surveys, which can result in inaccurate mapping data (Ouzemou *et al.*, 2018; Chabalala *et al*., 2020). Furthermore, fruit tree distributions may not be fully captured by field observations, particularly in isolated or difficult-to-reach places (Chemura *et al*., 2015; Ouzemou *et al*., 2018). Through community interviews about the distribution of fruit trees, methods of cultivation, and trends in land use might be subjective, lacking, or out of date, even if it is useful for comprehending historical trends and sociocultural factors (Geetha *et al.,* 2016). Furthermore, community interviews alone might not be able to supply the thorough spatial data required for efficient mapping and management (Xiong *et al.*, 2022). Additionally, maps produced by ground-based surveys are not as scalable or applicable because they can only cover a relatively small area (Xiong *et al.*, 2022). Seasonal variations, growth dynamics, and changes in land use are examples of temporal variability in fruit tree distributions that may be missed by traditional approaches (Chabalala *et al*., 2023c). Because manual field surveys are usually carried out at certain times, they can miss long-term patterns or slow changes in fruit tree populations over time (Chemura *et al*., 2015; Ouzemou *et al.*, 2018). Data integration and analysis are difficult because traditional approaches frequently produce heterogeneous datasets that are collected in different formats (You and Dong, 2020). Data gaps or inconsistencies may arise from combining data from community interviews, field observations, and ground surveys—this process takes a lot of work (Baghdadi *et al.,* 2009; Azzari *et al.,* 2021). Expenses associated with labor, transportation, equipment, and data administration can make traditional approaches expensive and resource intensive (Gourlay *et al.*, 2017).

Remote sensing techniques can greatly enhance fruit tree mapping and monitoring from a wider angle, which can offer important insights into numerous elements of agricultural landscapes, environmental sustainability, and fruit tree ecosystems (Gómez *et al.,* 2016). Fruit tree distributions over vast geographic regions may be effectively mapped using remote sensing (Song *et al.,* 2019; Becker-reshef *et al.,* 2020). High-resolution data can be captured using satellite images, aerial photography, and unmanned aerial vehicles (UAVs) (Mccarty *et al.,* 2017; Defourny *et al.,* 2019; Leroux *et al.,* 2019). This enables a precise analysis of the species composition, spatial patterns, and cover of fruit trees across various landscapes. Data from remote sensing can be gathered regularly to track changes in the temporal distribution of fruit trees over time (Ibrahim *et al.,* 2021; Chen *et al*., 2019). Time-series analysis provides valuable information for evaluating trends, recognizing patterns, and detecting seasonal variations, growth dynamics, and land cover changes (Jin *et al.*, 2019; Chabalala *et al*., 2023b; Chabalala *et al*., 2023c).

Furthermore, fruit tree species, health, and environmental conditions may all be thoroughly characterized thanks to the ability of remote sensing sensors to collect multispectral and hyperspectral data (Veloso *et al.,*2017). By reducing bias and human error in data gathering and processing, remote sensing-based mapping is frequently more objective and repeatable than traditional methods (Song *et al.,* 2019; Becker-reshef *et al.,* 2020). Using spectral signatures, remote sensing instruments, such as multispectral and hyperspectral imagers, may differentiate between fruit tree species and determine their current state of health (Tu *et al.,* 2020; Veloso *et al.*,2017). Fruit tree health can be monitored using sophisticated image processing techniques like vegetation indices and spectral unmixing to identify illnesses, stressors, and physiological changes (Preidl *et al.,* 2020). Fruit tree cover, canopy structure, biomass, and production may all be quantitatively analyzed based on remote sensing, which helps with yield forecasts, agricultural management techniques, and resource allocation choices (Karthikeyan *et al.,* 2020; Hao *et al.,* 2018). Adopting remote sensing technology makes it easier to monitor variables like soil moisture, temperature, precipitation, land use, and cover shifts that affect fruit tree ecosystems (Forkuor *et al.,* 2018). Assessing the susceptibility of fruit tree populations to climatic variability, land degradation, and habitat fragmentation is made easier by integrating remote sensing data with environmental models (Richard *et al.,* 2017; Vogels *et al.,* 2020).

Additionally, remote sensing helps evaluate the ecosystem services that fruit tree ecosystems offer, such as soil stabilization, carbon sequestration, biodiversity preservation, and water management (Vira *et al.*, 2015; Gelaye and Getahun, 2024). Comprehending the geographical dispersion and operational mechanisms of fruit tree ecosystems facilitates the assessment of their inputs to ecosystem services and the development of conservation plans (Hadria *et al.*, 2010; Duveiller *et al*., 2013). Interactive mapping, monitoring, and decision-making tools can be created by integrating remote sensing data with Geographic Information Systems (GIS) and Spatial Decision Support Systems (SDSS) (Baghdadi *et al.,* 2009; Csorba *et al.*, 2019). Using these tools, stakeholders can assess trends, prioritize interventions for the sustainable management of fruit tree resources, and visualize geographical patterns (You and Dong, 2020; Wu *et al.*, 2020; Xiong *et al.*, 2022). Lastly, by utilizing participatory methods and citizen science projects, remote sensing tools can involve local populations in mapping and monitoring fruit tree distributions. Giving communities access to remote sensing equipment and training improves their ability to track environmental changes, promote sustainable land use, and support conservation initiatives.

In summary, although conventional techniques possess advantages, remote sensing presents a potent and supplementary tool for more accurate, scalable, and efficient mapping of fruit tree populations. Stakeholders can improve their knowledge of fruit tree ecosystems, support sustainable management practices, encourage evidence-based decision-making regarding issues about agriculture, food security, and environmental sustainability, and support initiatives to conserve biodiversity, improve food security, and lessen the effects of climate change by utilizing the capabilities of remote sensing technologies.

1.3. Research problem statement

Attaining sustainable smallholder horticulture systems in the future depends on the availability of fruit tree maps. The use of remote sensing technology is crucial in providing spatial information required to develop horticulture-orientated management tools that increase fruit production and ensure that profitability is realised in the future. Accurate fruit tree mapping requires ground control points that will be used to train and validate RS models. However, the current ground-based methods used for data collection are expensive, labour-intensive and impractical over rugged terrains containing smallholder systems (Chabalala *et al*., 2022). The unavailability of fruit tree information that is up to date is a known worldwide problem, specifically for those systems located in developing countries (You et al., 2009). Based on existing literature, there is less research focusing on fruit tree mapping in smallholder agriculture, which threatens the efforts to innovate the industry and improve the livelihoods of farmers in fruit production (Pérez-Hoyos *et al.*, 2017).

There is incomplete spatial information regarding the distribution of smallholder fruit tree systems in Africa, which hinders the planning and implementation of policies that will improve agricultural sustainability and attain food security for all (Li *et al.,* 2019; Persello *et al.,* 2019). Furthermore, the lack of crop types spatial distribution impedes food security efforts and progress in agricultural remote sensing (Karthikeyan *et al.,* 2020). As a result, the African continent is still food insecure due to agricultural farming systems that are not properly managed (Jain *et al.,* 2013; Azzari *et al.,* 2021). Currently, smallholder farmers rely on farm boundaries created using Google Earth and outdated Nornamilsed Vegetation Index (NDVI) maps, which do not give detailed information about the current fruit tree growth conditions (Estes *et al*., 2022; Chabalala *et al*., 2022; Chabalala *et al*., 2023a).

Agronomists, agriculturalists, government and other stakeholders require fruit trees' spatial distribution maps to make evidence-based decisions regarding agricultural planning, resource management, the development of appropriate management strategies and formulation of agricultural policies for efficient resource management, biodiversity conservation, sustainable horticultural systems and socio-economic development. Thus, understanding the spatial distribution of fruit trees is a crucial precursor for the development and implementation of an RS-based fruit load monitoring system for smallholder orchard plantations. However, mapping and monitoring fruit tree species in heterogeneous landscapes is challenging due to fruit varieties and management strategies resulting in spectral overlap. Remote sensing (RS) can be used to develop efficient scalable solutions that will innovate the horticulture industry by optimising the production of good quality fruits cost-effectively. Therefore, the research aimed to map and monitor fruit tree species by integrating Sentinel-1 and Sentinel-2. The results of the research can assist agronomists, agriculturalists, policymakers, government and other stakeholders who are responsible for managing different agricultural activities.

1.4. Research aim and objectives

The aim of the research was to map fruit trees and co-existing land use types using multisource remote sensing (RS) data in conjunction with machine learning (ML) and deep learning (DL) techniques.

The objectives of the research were:

1. To explore the effect of balanced and imbalanced multi-class distribution data and sampling techniques on fruit-tree crop classification using different machine learning classifiers and remote sensing data.
2. To test the effectiveness of machine learning classification of fused Sentinel-1 and Sentinel-2 Image data towards mapping fruit plantations in highly heterogeneous landscapes.
3. To identify the optimal phenological period for discriminating fruit tree crops using multi-temporal Sentinel-2 data and Google Earth Engine in a heterogeneous subtropical horticultural region.
4. To assess the applicability of Sentinel-2 and Artificial Neural Networks in classifying dynamic phenological cycles in horticultural crops.

1.5. Scope of the research

This research used remotely sensed data to map and monitor fruit trees in Levubu sub-tropical farms in the Vhembe District Municipality in Limpopo Province of South Africa. The research evaluated the applicability of using synthetic aperture radar (SAR/radar) and multispectral (optical) data to discriminate common fruit trees in the research area. Furthermore, the research demonstrates the applicability of Sentinel-1 and Sentinel-2 data in mapping fruit trees and co-existing land use types in the research area. The research demonstrates the effectiveness of data sampling in modifying the data imbalance ratio on the training dataset. The results indicated that data sampling could be used to reduce the imbalanced ratio and attain high classification accuracy when robust machine learning classifiers such as RF, SVM, AdaBoost, GB, and XGBoost are used. Furthermore, the research demonstrates the utility of fusion optical and SAR data in mapping sub-tropical regions and the selection of optimal variables. The results showed that applying optimal variables from the fused dataset to RF and SVM can be used to map fruit trees and co-existing land use types with higher accuracy. The utility of using optical data acquired across the fruit full growth cycle can help identify the optimal window period in which fruit trees can be discriminated accurately using the RF classifier. The results showed that fruit trees can be discriminated with high accuracy using optical images acquired from April to July corresponding to fruit key growth stages. Lastly, the research mapped the fruit dynamics using phenological metrics extracted from optimal Sentinel-2 images using the Deep Neural Network. The results showed that deep learning models are stable to fruit growth stages grown under different management strategies and can be applied to increase classification accuracy. The results can be used to support decision-making and facilitate innovation in the horticulture industry.

1.6. Research area

The research site is located in Levubu sub-tropical farms in the northeastern part of South Africa in the Limpopo Province. The Levubu sub-tropical farms fall under Vhembe District Municipality within the boundaries of Collins Chabane and Thulamela Local Municipalities. The area covers the fruit farming production areas of 10, 000 hectares (Figure 1.1) (VDM, 2021). Spatially, the meridians of the research site are $23^{\circ}5^{0}31.25^{00}$ S to $23^{\circ}6^{0}45.03^{00}$ S, and the parallels are $30^{\circ}13^{0}40.88^{00}$ E to $30^{\circ}21^{0}26.03^{00}$ E. The sub-tropical farming area covers five villages, i.e., Levubu, Tshakuma, Makhuro, Valdezia, and Mulangaphuma (VDM, 2021). It has a total population of 207,000, with white people accounting for 10.9% (VDM, 2021). The dominant agriculture in the area is fruit production (i.e., banana, avocado, guava, macadamia nut, mango), although intercropped with cash crops and grains. These crops are rich in minerals, vitamins, proteins, carbohydrates, and beneficial oils, which contribute to a balanced diet.

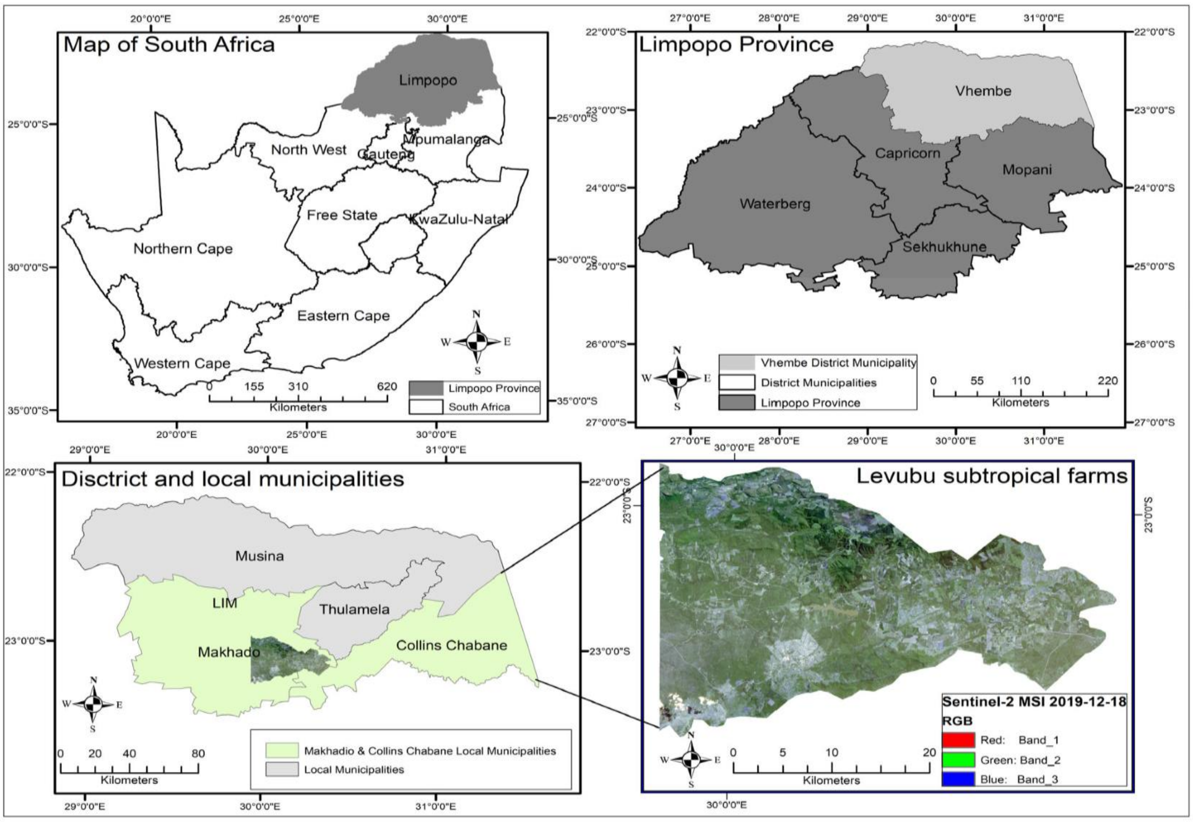


Figure 2.1. Map of the research site showing Levubu sub-tropical farms located in Collins Chabane and Makhado Local municipalities (light green) of the Limpopo Province *(grey), South Africa.*

Levubu's temperature ranges from 12-22ºC in winter and 16-45ºC in summer (VDM, 2021). The area encompasses a variety of water features, such as the Albasin dam, numerous perennial rivers (Livhuhu, Lotonyanda) and numerous non-perennial rivers (Mukhada *et al.,* 2021). the topography of the area is flat to undulating, with elevations ranging from 667 to 2221m (VDM, 2021). Induced by the Soutpansberg Mountains, the middle part of the mountain receives annual rainfall reaching 2000m at Entabeni, while the western part of the mountain receives 340mm during January and February. The soil is mostly red and very fertile, making it suitable for various sub-tropical fruit productions.

1.7. Thesis outline

The thesis comprises seven chapters, of which 2 to 5 have been submitted to peer-reviewed journals indexed by the Department of Higher Education and Training (DHET) and Directory of Open Access Journals (DOAJ). The thesis aim was achieved through four research papers that were published in peer-reviewed journals. Each objective was written as a standalone article that can be read independently and in conjunction with the other articles as the objectives are linked. Hence, the research is organized using the following chapters.

Chapter 1 is composed of a general introduction to the research. It introduces the current state of African agriculture, crop growth and crop phenological cycles and the application of remote sensing in crop types mapping. Furthermore, the chapter also outlined the objectives of this thesis.

Chapter 2 used data sampling techniques to examine the effect of imbalanced data and the sensitiveness of ML classifiers (i.e., RF, SVM, AdaBoost, GB, XGBoost) to detect and map fruit trees in a sub-tropical smallholder horticultural region. Six datasets (two balanced and three imbalanced were derived using RandomUnderSampler (RUS) and their classification performance was compared to the original imbalanced dataset. Furthermore, this chapter also identified the sampling strategy, optimal Sentinel-2 bands and the classifier that can be used to map fruit trees in the research area.

Chapter 3 applied fused optical and SAR data to map fruit trees and co-existing land use types. Six optimal bands were identified from the fused dataset and used to produce the final spatial distribution maps.

Chapter 4 used optical data, i.e., Sentinel-2 images and RF in a Google Earth Engine environment acquired over 12 months, to determine the optimal window period in which the fruit trees can be detected with high accuracy. The optimal spectral bands in mapping fruit trees were also identified.

Chapter 5 aimed to improve the fruit tree classification accuracy by utilizing phenological metrics derived from sentinel-2 images acquired during the key fruit tree growth period using a Deep Neural Network.

Chapter 6: This chapter synthesizes the key findings from the research objectives. Furthermore, this chapter provides the concluding remarks of the research and outlines the limitations of the research.

Chapter 7: This chapter concludes the major findings from the research objectives and recommends future research on the application of remote sensing in agriculture.

A single reference list of the literature cited in this research is provided at the end of the thesis, bringing the research thesis to an end.

CHAPTER TWO

Exploring the effect of balanced and imbalanced multi-class distribution data and sampling techniques on fruit-tree crop classification using different machine learning classifiers and remote sensing

This chapter is based on:

Chabalala, Y., Adam, E and Ali, A.K. (2023). Exploring the effect of balanced and imbalanced multi-class distribution data and sampling techniques on fruit-tree crop classification using different machine learning classifiers and remote sensing data. *Journal of Geomatics*. 3 (1), 70-92. DOI: https://doi.org/10.3390/geomatics3010004

Abstract

Fruit-tree crops generate food and income for local households and contribute to South Africa's gross domestic product. Timely and accurate phenotyping of fruit-tree crops is essential for innovating and achieving precision agriculture in the horticulture industry. Traditional methods for fruit-tree crop classification are time-consuming, costly, and often impossible to use for mapping heterogeneous horticulture systems. The application of remote sensing in smallholder agricultural landscapes is more promising. However, intercropping systems coupled with the presence of dispersed small agricultural fields that are characterized by common and uncommon crop types result in imbalanced samples, which may limit conventionally applied classification methods for phenotyping. This study assessed the influence of balanced and imbalanced multi-class distribution and data-sampling techniques on fruit-tree crop detection accuracy. Seven data samples were used as input to adaptive boosting (AdaBoost), gradient boosting (GB), random forest (RF), support vector machine (SVM), and eXtreme gradient boost (XGBoost) machine learning algorithms. A pixel-based approach was applied using Sentinel-2 (S2). The SVM algorithm produced the highest classification accuracy of 71%, compared with AdaBoost (67%), RF (65%), XGBoost (63%), and GB (62%), respectively. Individually, the majority of the crop types were classified with an F1 score of between 60% and 100%. In addition, the study assessed the effect of size and ratio of class imbalance in the training datasets on algorithms' sensitiveness and stability. The results show that the highest classification accuracy of 71% could be achieved from an imbalanced training dataset containing only 60% of the original dataset. The results also showed that S2 data could be successfully used to map fruit-tree crops and provide valuable information for subtropical crop management and precision agriculture in heterogeneous horticultural landscapes.

2.1. Introduction

Agriculture is regarded as the foundation of all civilization, and crops are crucial to human nutrition and societal stability (Zhong *et al.,* 2019). Due to the rising global human population, there is a need to closely monitor agricultural systems (Wang *et al.,* 2019). The human population of the world is experiencing a rapid increase—quadrupled in the last century, to *c*. 6.2 billion (Robert, 2002), and it is projected to be *c*. 8.7 billion by 2030 (United Nations, 2015), and 10 billion by 2050 (FAO, 2014). Therefore, food production needs to be doubled by 2050 (Foley *et al.,* 2011). Timely, efficient, and reliable agricultural data is of paramount importance to help monitor the demand for food, manage costs, and target agriculture-related policies (Santos *et al.,* 2019; Lehlou *et al.,* 2022). Additionally, monitoring agricultural production is critical in terms of climate change, loss of biodiversity, and natural disasters—e.g., drought and flood—which threaten food security at local and global scales (Shi *et al.,* 2021). The mapping of crops is key to the management of crops, estimation of yield, and food security (Hao *et al.*, 2018; Yang *et al.,* 2022).

Traditionally, information on the distribution of crops has been acquired through ground surveys and censuses (Wang *et al.,* 2019). However, field surveys are costly, outdated, cover small spatial extents, and are prone to human error (Gourlay *et al.,* 2017). Remote sensing (RS) technology, which has the advantages of monitoring extensive spatial extents and low cost, has substituted the traditional methods (Gourlay *et al.,* 2017). As RS data are advancing in terms of spatial and temporal resolutions, they are becoming increasingly important in generating crop-type maps (Wang *et al.,* 2019). Therefore, the need for timely, accurate, and reliable agricultural data has seen unprecedented growth in the economic importance of RS-based crop mapping over the last two decades (Lehlou *et al.,* 2022). The accessibility of satellite data by the public enabled many researchers to minimize ground surveys by creating low-cost crop-type maps—using features extracted from remotely sensed spectral differences in vegetation and other surfaces over time (Waldner *et al.,* 2015). Agricultural RS offers insights that will radically change the operation and management of agricultural systems (Preidl *et al.,* 2020). These efforts have been achieved through the successful application of machine learning (ML) algorithms, such as support vector machines (SVM), random forests (RF), and, increasingly, neural networks (Cai *et al*., 2018).

In addition to the number and quality of the training samples, algorithm performance may be affected by class imbalance (Devi *et al.*, 2019; Maxwell *et al.*, 2018). Class imbalance occurs when the number of reference samples varies among the classes, leading to an imbalanced training set (Devi *et al.,* 2019; Maxwell *et al.,* 2018). Most classifier learning algorithms that assume a fairly balanced distribution, such as *k*nearest neighbor (*k*-NN), support vector machine (SVM), and random forest (RF), are affected by imbalanced training data (Maxwell *et al.,* 2018). Imbalanced data may occur in a deliberative sample and are also expected in random sampling (Devi *et al.,* 2019; Maxwell *et al.,* 2018). In simple random sampling, the chance of choosing a class is related to the area covered by the class. Thus, relatively rare/minority classes will comprise smaller proportions of the training set, and vice versa. The original dataset was imbalanced despite efforts to add samples of fewer crop types— which compromises class-specific accuracy (Li *et al.,* 2018). The imbalanced data problem worsens in a multi-class skewed distribution because the standard classifiers are designed for binary classification and assume a well-balanced class distribution (Zhu *et al.,* 2017). Advanced boosting classifiers, such as gradient boosting (GB), adaptive boosting (AdaBoost), and eXtreme gradient boosting (XGBoost), are used to solve the limitations of standard classifiers.

Recently, the applicability of these boosting classifiers has been tested in crop type mapping (Zhu *et al.,* 2017; Prins and Van Niekerk, 2020) and land cover mapping (Waldner *et al.,* 2015). For example, Mashamba-Munghemezulu *et al.* (2021) tested the performance of the standard classifiers (RF, SVM) and boosting classifiers (GB, AdaBoost, XGBoost) in heterogeneous crop types mapping and reported a superior performance of 86.91% using XGBoost as compared to the RF and SVM. Similarly, (Bouras *et al.,* 2021) obtained the best accuracy metrics of 0.88% when forecasting cereal yield using the XGBoost classifier compared to SVM, RF, and multiple linear regression (MLR), whose accuracy was lower. The classification accuracies reported in these studies involved cereal, sugarcane, and wheat crops—they possess different canopy structures and spectral signatures. In this study, the crops were all fruit crops—namely, avocado, banana, guava, mango, and macadamia nut—which display almost similar spectral responses due to extreme heterogeneities. Such crop types are nearly impossible to discriminate using classical multi-spectral imagery analysis techniques (Li *et al.,* 2018). This is made worse if the landscape is rugged terrains, intercropping systems scattered in small patches, and different crop area coverage/levels of growth with varying health statuses (Tu *et al.,* 2020). Additionally, the differences in management practices and soil fertility result in spectral similarities among crops, leading to uneven distribution of crop types (Feyisa *et al*., 2019).

Thus, different sampling techniques have been designed to solve the imbalanced class problem and ensure generalized data characteristics for class separability (Naboureh *et al.,* 2020; Walder *et al.,* 2019). Such techniques include the modification of the classifier and data sampling (Azadbakht *et al.,* 2018). The applicability of data sampling techniques, such as random oversampling (ROS) using the synthetic minority oversampling technique (SMOTE) and random undersampling (RUS), have been demonstrated in land cover mapping (Azadbakht *et al.,* 2018; Walder *et al.,* 2019). These techniques provide land cover, crop, and soil mapping solutions by modifying the training dataset (Ghaseminik *et al.,* 2021). Although they produce acceptable classification accuracy, they fail to produce acceptable individual class accuracy (Azadbakht *et al.,* 2018; Walder *et al.,* 2019). Most of these techniques overfit the model or discard valuable samples required to determine the decision boundary between the classes (Ghaseminik *et al.,* 2021; Taghizadeh-Mehrjardi *et al.,* 2020; Chen *et al.,* 2020). In all classification schemes, the datasets were split into test and training data before running the models. Running models before splitting datasets can allow identical samples to be present in both the test and training data, leading to the model's overfitting the training data—the sampling techniques were only applied to training data to avoid model overfitting (Li *et al.,* 2018; Zhu *et al.,* 2017).

Accurate and timely mapping of crop types and having reliable information about the cultivation pattern/area play a key role in sustainable agriculture management. High resolution, national-scale maps of agricultural land are needed to develop strategies for future sustainable agriculture. Thus, this study applied Sentinel-2 multispectral data and five widely-used ML classifiers—which were used for comparison—i.e., SVM, RF, an extreme gradient boosting machine called XGBoost, AdaBoost, and GBBoost, to map fruit crops and co-existing land use types in Levubu area, Limpopo province, South Africa, using imbalanced datasets.

The organization of the paper is as follows: Section 2 describes the materials and methods used in this research. Section 3 depicts the results of the paper, while the results of the research are discussed in Section 4. Finally, the conclusion of the paper is presented in Section 5.

2.2. Materials and methods

2.2.1. Methods Overview

The potential of data sampling using Sentinel-2 multispectral data for crop type classification was studied by selecting the Levubu sub-tropical farming area as a research site. The Sentinel-2 image was pre-processed using SEN2COR in the ESA Snap platform to remove atmospheric effects, and five machine learning classifiers (i.e., GB, Ada Boost, RF, SVM, and XGB) were used to classify the processed image. Finally, the effectiveness of the classifier in classifying the monthly images was evaluated by computing different classification metrics, that is, the overall accuracy (OA), F1 score, user's accuracy (UA), and producer's accuracy (PA). A flowchart of the methodology followed in this research is presented in Figure 2.1.

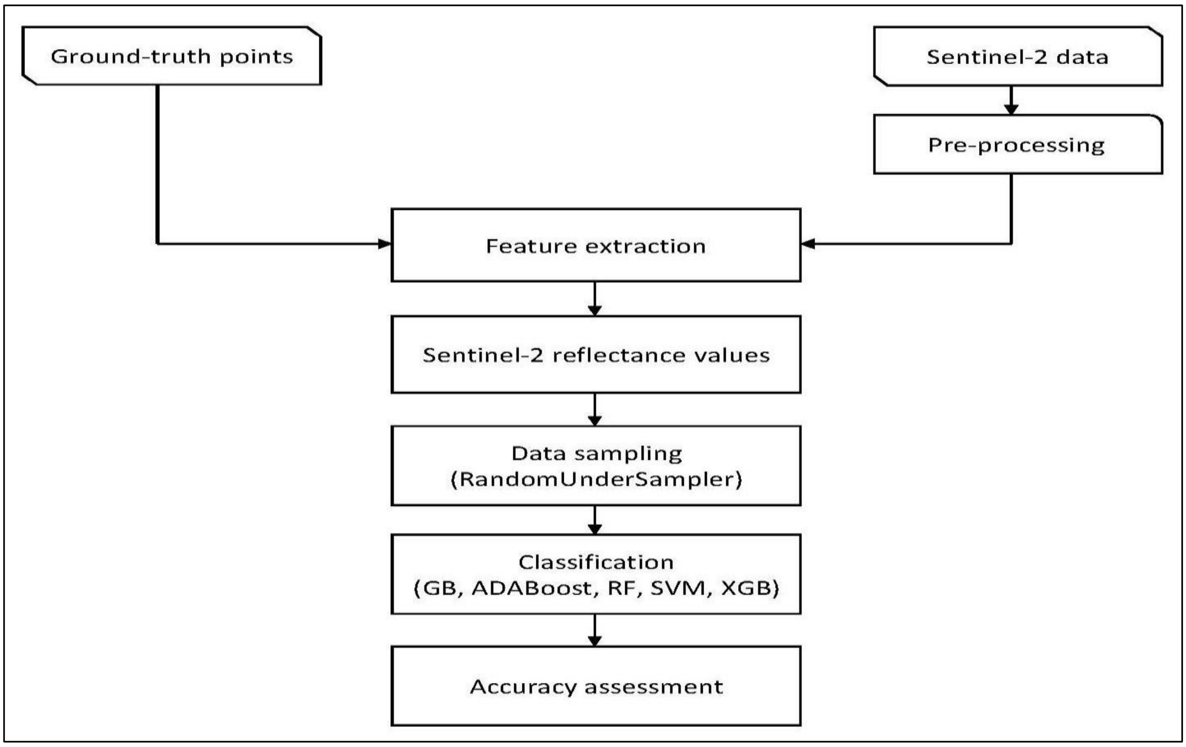


Figure 2.1: *Flowchart of the methodology developed in this research.*

2.2.2. Research region

This research was conducted in Levubu, a subtropical fruit-tree crop farming region located at 23◦4′60.00″ S and 30◦16′60.00″ E in the Vhembe District, Limpopo, South Africa (Figure 2.2). The region experiences a warm subtropical climate with annual temperatures ranging from ~16 to 22 ◦C (Maponya and Mpandeli, 2012; Chabalala *et al.*, 2022). The annual rainfall ranges from ~200 to 1500 mm; however, it varies substantially owing to the geographic effect of the Soutpansberg mountain range (Louw and Flandorp, 2017; Weier *et al.*, 2018). Regions windward and east of the mountain range receive 200 to 400 mm, while regions leeward and west of the mountain range receive around 1000 mm (Mukwada *et al.*, 2021). The rainfall of the region is seasonal, with distinct wet and dry seasons ranging from October to March and April to September, respectively (Fraser, 2008).

Levubu, a smallholder agricultural region in Limpopo, South Africa, has diverse agroclimatic conditions that offer endless possibilities for commercial horticulture farming (DAFF, 2012). The dominant tree crop types are the macadamia nut, avocado (Hass, Maluma Hass, Pinkerton, and Fuerte), banana, guava, mango, and pine tree. In addition to the crop area, the region is characterized by other land-use types, such as water bodies, built-up areas, and woody vegetation. Most crops grown in the Levubu area are subtropical, and only a few citrus crops, such as oranges, are grown. Tree crops dominate the site, and small cultivated crops are grown depending on the season. Levubu crops are primarily rain-fed; however, during the dry season, crops are irrigated. Most of the tree crops bear fruit seasonally, and the flowering and harvesting time differs with each crop type. Many crops, including the macadamia nut, banana, mango, and avocado, are exported to international markets (Department of Agriculture, Forestry and Fisheries) (Chen *et al.,* 2016).

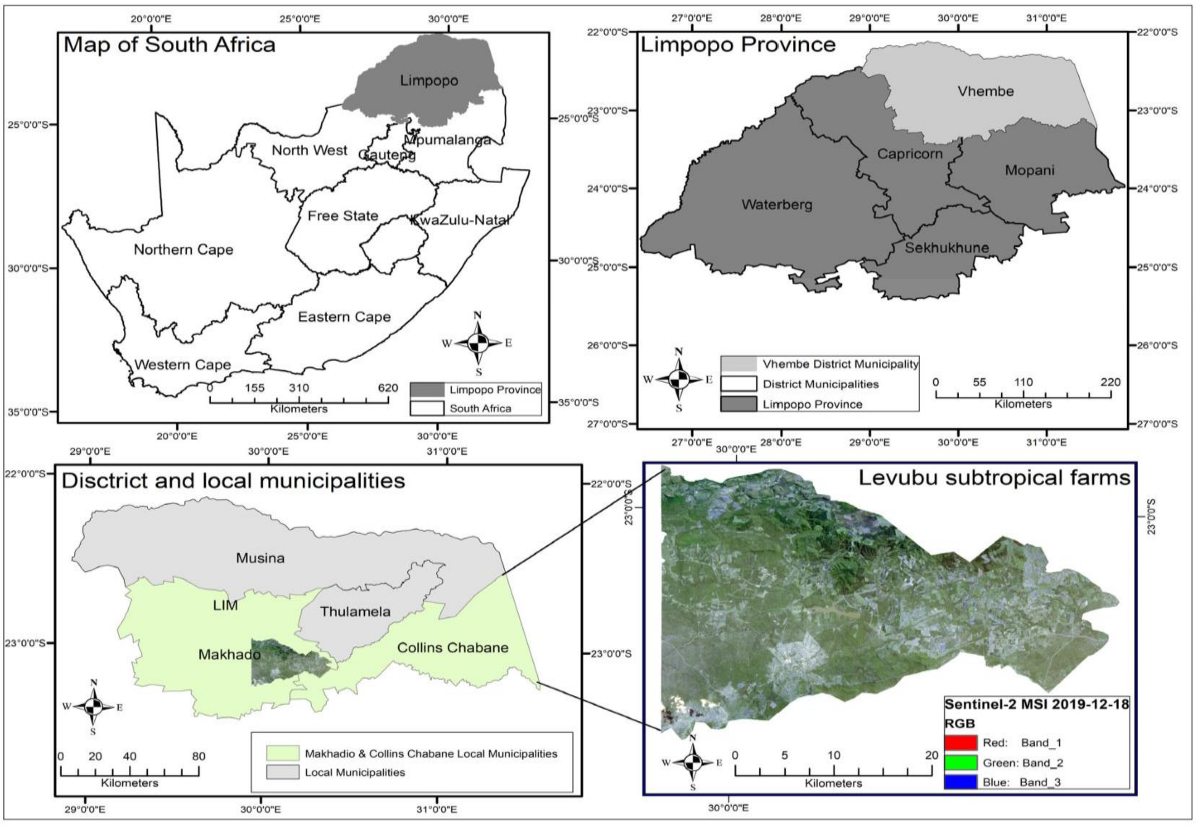


Figure 2.2: *Map of the study region, depicting the Levubu subtropical farming region located in the Makhado and Collins Chabane Local Municipalities (lime green) of the Limpopo province (grey), South Africa.*

2.3. Data collection and processing

2.3.1. Ground truth data

The ground truth data used for calibrating and validating fruit-tree crop maps and applying ML algorithms in this study were collected using field surveys conducted on 28 and 29 December 2019 and 2 and 3 January 2020. A Garmin eTrex 20 Global Positioning System, with a positional accuracy of 5 m, was used to capture the geographic locations of dominant fruit-tree crops and co-existing land cover classes, namely avocado, banana, bare soil, built-up, macadamia nut, mango, pine tree, water body, and woody vegetation. A combined total number of n = 304 training samples was collected.

The collected field samples were used as a guide for visual comparison, identifying and labelling samples corresponding to the tree crop species and other co-existing land-use classes. A single-pixel sampling approach was applied when digitizing the training samples to select pure pixels and minimize the effect of landscape heterogeneity and mixed spectral signatures on classification accuracy (Brownlee, 2020). A purposeful sampling method was used to increase the sample number and to consider the proportion of different types of fruit trees in the study area. The additional training samples were digitized using Google Earth Pro and ArcGIS version 10.6.1. The proportion of the farm was considered in generating the number and distribution of the samples, as depicted in Figure 2.2. The structure of the reference data is presented in Table 2.1, while the spatial distribution of the samples is depicted in Figure 2.3.

Table 2.1: *Number of samples for fruit-tree crops and other existing land uses, the structure of training and testing data for each class used to map fruit-tree crops in Levubu subtropical farms.*

Tree Species and Other Land Use Classes	**Reference Data**		**Total**
	Training	**Testing**	
Avocado	109	47	156
Banana	181	77	258
Bare soil	120	52	172
Built-up	122	52	174
Guava	154	66	220
Macadamia nut	113	48	161
Mango	160	68	228
Pine tree	117	50	167
Waterbody	128	53	181
Woody vegetation	126	54	180
Total sample size = 1897			

2.3.2. Data sampling

Data sampling is crucial in supervised land cover mapping based on imbalanced data (Quan *et al.*, 2021). Previous research has dealt with class imbalance problems using data or algorithm-level methods (Khaldoon *et al.,* 2020). The data-level methods modify the distribution and degree of imbalance of classes in the training dataset to fit the configurations of standard classifiers, while the algorithm-level methods modify the classifiers (Khaldoon *et al.,* 2020; Waldner *et al.,* 2019). Two data modification methods applied in this research are undersampling and oversampling (Chen *et al.,* 2020). The former balances data by removing points from the majority class, while the latter duplicates points from the minority class (Waldner *et al.,* 2019; Wang *et al.,* 2020). Based on the existing remote sensing literature, much attention has been given to improving the accuracy of minority classes (Azadbakht *et al.,* 2018). However, this study is built on the assumption that all crops are equally crucial in small-scale farming and must be recognized to allow efficient farm management practices (Waldner *et al.,* 2019). Hence, the study employed undersampling techniques to retain the in-situ data and prevent the loss of crucial samples from the imbalanced dataset (Krawczyk, 2016). This was achieved by applying a RandomUnderSampler, a non-heuristic approach (naïve sampling) that selects samples from the majority class to balance the class distribution (Waldner *et al.,* 2019).

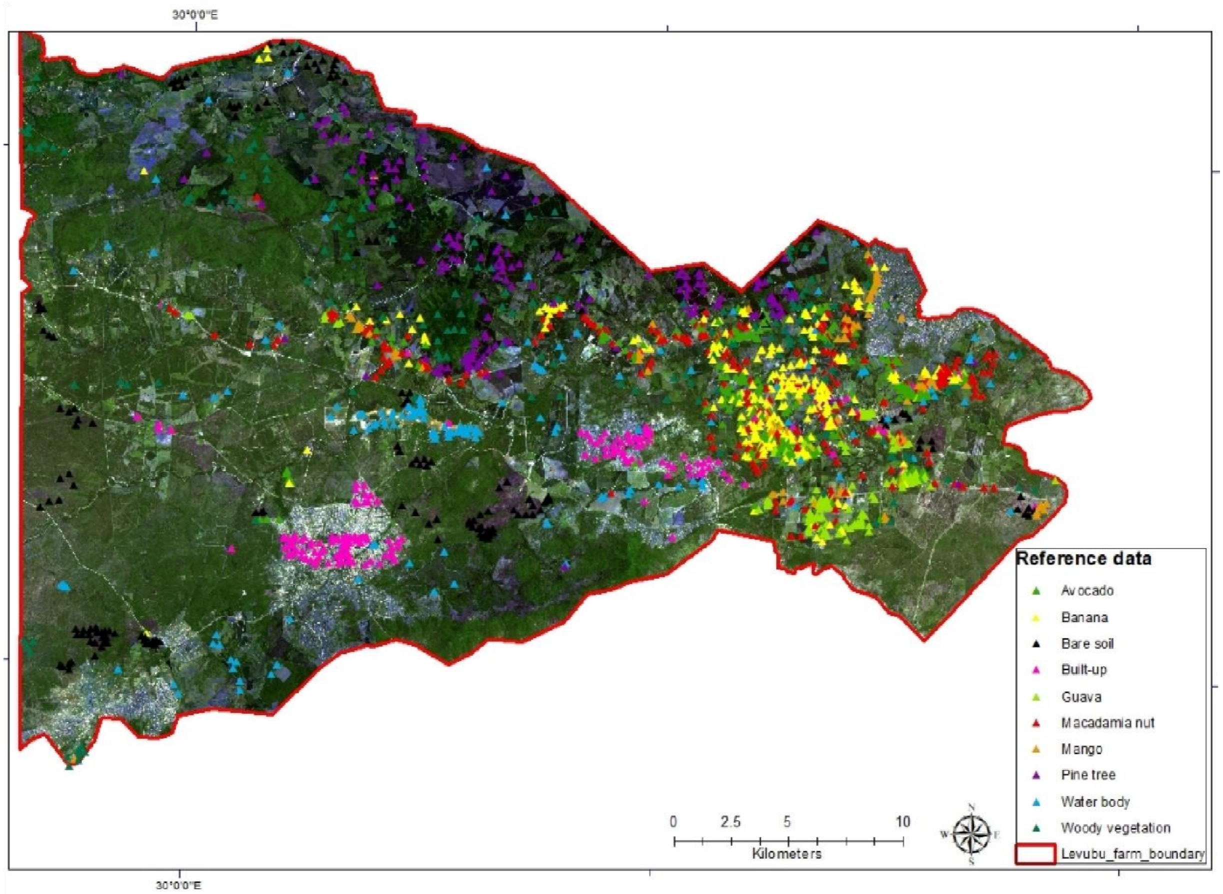


Figure 2.3: *Location of fruit trees and the co-existing land-use types in the sub-tropical fruit farming area of Levubu. The background map represents a S2 RGB image obtained from the European Space Agency.*

In this study, the collected in-situ data is presented as Dataset 7 in Table 2.1. and consists of an imbalanced class ratio of 2:3, whereby the common (i.e., majority) species had more samples than the less common (i.e., minority) species. Six different sampling strategies were applied to generate new datasets from the original dataset, consisting of a decreased imbalanced ratio and balanced class distribution (Table 2.1). Firstly, two balanced datasets (i.e., Dataset 1 and Dataset 2) were extracted, consisting of 100 and 150 balanced points.

Secondly, the imbalanced ratio on the original dataset was decreased by extracting four datasets using sampling strategies based on percentages (i.e., 40%, 50%, 60%, and 70%) and probability sampling. These strategies resulted in the formation of Dataset 3, Dataset 4, Dataset 5, and Dataset 6, respectively.

The sampling procedures were carried out using Imbalanced-Learn, a Python toolbox designed to handle problems associated with imbalanced datasets (Kiyohara, 2015). This toolbox consists of different libraries, where Scikit-Learn and RandomUnderSampler were used to sample the datasets in this study. Seven experiments of training datasets, including the original dataset, were tested to select the best model for mapping fruit-tree crops and co-existing land-use types in our study regions. The dataset was randomly divided into 70% training and 30% validation datasets for each class. All datasets were cross-validated to detect the behaviour of different classifiers on different samplings.

Table 2.2: List of fruit-tree crops and co-existing land-use types and the descriptive statistics of the undersampled and original datasets. The # stands for the number of the datasets used in this research.

Class Name	**Abbreviation**	**Balanced Datasets (#1 and 2)**		**Percentages Undersampled Datasets (#3–6)**				**Unsampled Imbalanced Dataset (#7)**
		Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5	Dataset 6	Dataset 7
Avocado	AV	100	150	62	78	94	109	156
Banana	BN	100	150	103	52	154	181	258
Bare soil	BS	100	150	69	86	103	120	172
Guava	GV	100	150	70	87	104	121	174
Macadamia nut	MN	100	150	88	110	132	154	220
Mango	MG	100	150	64	81	97	113	161
Pine tree	PT	100	150	91	114	137	160	228
Built-up	BU	100	150	69	86	100	117	167
Waterbody	WB	100	150	72	91	109	127	181
Woody vegetation	WV	100	150	72	90	108	126	180
Total number of instances		1000	1500	760	875	1138	1328	1897

2.3.3. Remote sensing data acquisition and pre-processing

Sentinel-2A (S2A) and Sentinel-2B (S2B) are a constellation of two polar-orbiting sensors placed in the same sun-synchronous orbit with the same sensor configurations. Both are equipped with a multi-spectral imaging (MSI) sensor. The imagery acquired from the two sensors is freely available from the European Space Agency (ESA) Copernicus Open Access Hub website (https://scihub.copernicus.eu/). The MSI sensor consists of 13 spectral bands ranging from visible and near-infrared (VNIR) to the shortwave infrared (SWIR) region (Table 2.2) (Djamai and Fernandes, 2018). The image has a wide-swath width of 290 km with 13 spectral bands consisting of three spatial resolutions (10 m, 20 m, 60 m). The 10 m spatial resolution bands (Blue: B2, Green: B3, Red: B4, NIR: B8) are centered at 490 nm, 560 nm, 665 nm, and 842 nm, and the 20 m spatial resolutions ban_1: B5, nm, and 842 nm, and the 20 m spatial resolutions bands (red edge_1: B5, red edge_2: B6, red edge_3: B7, Narrow NIR: B8A, SWIR_1: B11, SWIR: B12) are located at 705 nm, 740 nm, 783 nm, 865 nm, 1,610 nm, and 2,190 nm. The 60 m spatial resolution bands (coastal aerosol: B1, water vapour: B9, SWIR-cirrus: B10) are located at 443 nm, 945 nm, and 1, 375 nm, respectively (Table 2.1). Additionally, it consists of two spectral bands that were systematically positioned in the red-edge position (REP). These bands are sensitive to chlorophyll content and are essential for crop mapping (Louw and Flandorp, 2017). The temporal resolution between the two sensors is five days, allowing for continuous crop mapping throughout the growing season (Saini and Ghosh, 2018).

Pre-processing of S2 data involves atmospheric corrections, geometric rectification, and radiometric calibration. In this study, the pre-processing was executed using a level 2A Prototype processor (SEN2COR) version 2.5.5 processing module in the ESA SNAP toolbox (Louw and Flandorp, 2017). Specifically, the atmospheric/topographic correction for satellite imagery (ATCOR) algorithm within the SEN2COR toolbox was applied to correct the image (Saini and Ghosh, 2018; Chabalala *et al.*, 2022). All of the S2 band images were resampled to a 10 m spatial resolution using a nearest neighbour resampling method. Red edge_2: B6, red edge_3: B7, Narrow NIR: B8A, SWIR_1: B11, SWIR: B12) are located at 705 nm, 740 nm, 783 nm, 865 nm, 1,610 nm, and 2,190 nm. The 60 m spatial resolution bands (coastal aerosol: B1, water vapour: B9, SWIR-cirrus: B10) are located at 443 nm, 945 nm, and 1, 375 nm, respectively (Table 2.1). Additionally, it consists of two spectral bands that were systematically positioned in the red-edge position (REP). These bands are sensitive to chlorophyll content and are essential for crop mapping (Louw and Flandorp, 2017). The temporal resolution between the two sensors is five days, allowing for continuous crop mapping throughout the growing season (Saini and Ghosh, 2018).

2.4. Machine learning classification algorithms

Selecting an appropriate classifier for a classification problem remains challenging because the sensitivity of the classifiers during the learning stage varies per dataset, owing to topography, land use, and class distribution (Saini and Ghosh, 2018). The algorithms used in this study were selected based on their simplicity, robustness, and popularity in remote sensing of biophysical parameters. Therefore, it is imperative to elucidate their performance and consistency with previous research using the same acquisition condition but in a different environment. The study compares ML algorithms for mapping fruit-tree crops and co-existing land-use types in smallholder agriculture using balanced and imbalanced data. The classification was implemented with 70/30 split samples for training/testing. The analysis was carried out using Python Jupiter Notebook Version 3.9 (https://jupyter.org/).

2.4.1. Random Forest (RF)

The RF classifier is an ensemble tree-based classifier that randomly selects features from the training sample data and reduces tree correlations (Breiman, 2001). It uses bootstrap aggregating to build models using two parameters, ntre*e* and *mtry*, interactively and independently by randomly selecting samples from the training dataset to make a final prediction (Breiman, 2001). All trees are aggregated to reach a final prediction. Thus, applying diverse decision trees grown using different random subsets reduces bias and prevents overfitting in the model (Breiman, 2001). The bootstrap aggregating (bagging) method is robust against model fitting and assists in obtaining a stable model (Kganyago *et al.*, 2021).

2.4.2. Support Vector Machine (SVM)

The SVMs are discriminative classifiers based on a statistical learning framework (Vapnick, 1995). The SVM uses a margin-based classifier to identify optimal linear hyperplanes with high-significance class prediction using a kernel function (Vapnick, 1995). The SVM consists of kernels that model various classification problems as a hybrid classifier. Examples include the linear, polynomial, sigmoid, and radial basis function (RBF) kernels (Chabalala *et al.*, 2020). In this study, the SVM was applied using the function in Equation [2.1], which was implemented using the RBF kernel in Equation [2.1], consists of two tuning parameters, termed the "gamma" (y) and "cost" (c) (Saini and Ghosh, 2018).

$$\boldsymbol{f(x) = sign\left(\sum_{i=1}^{n} a_i y_i K(x, xi) + b\right)}, \tag{2.1.}$$

where αi illustrates the Lagrange multiplier and K (x, xi) represents the kernel fun.

2.4.3. Gradient Boosting (GB)

The GB classifier is an ensemble classifier that improves final prediction by combining residuals from prior weak models (Friedman, 2001). A gradient descent algorithm is used to sequentially train individual classifiers using equally weighted training data during the training process. The data are further re-calculated during the boosting process (Friedman, 2001). At the same time, the loss function is minimized by sequentially and iteratively adding new models to improve the base learners from the weak classifiers in an additive manner until the training data set is predicted perfectly (Woodruff, 2017).

2.4.4. Adaptive Boosting (AdaBoost)

AdaBoost is a meta-algorithm that can be used with other classifiers to improve classification accuracy (Freund, 1997). It treats observations equally during optimization using stochastic GB machines (Barrow and Crone, 2016]. The boosting process uses training data to build short decision trees in succession (Cao *et al.*, 2012). The performances of the trees are measured to determine the weight allocation during the next iteration (Peng *et al.*, 2017).

The AdaBoost classifier is initiated by allocating weight on the training dataset based on how hard or easy the observations can be predicted. More weight is allocated to complex observations, while less is given to observations that can be easily predicted $(u_1, v_1), \ldots, (u_n, v_n)$; $u_i \in U$, $v_i \in \{-1,+1\}$. At the start of the boosting process, all models are assigned the same distribution of $D_1(i) = 1/N$, $i = 1, \ldots, N$, where N represents the total number of observations (Sun *et al.,* 2020). The models are sequentially trained using the distribution (D_t), and each model updates the weights of training instances of weak learners (Patil and Burkpalli, 2021). The new weights are calculated, and the distribution D_t: $D_{t+1}(i)$ is updated according to function 2. A final prediction is made based on the weighted vote accuracy of the trees on the training data (Cao *et al.,* 2012).

$$\frac{D_{t}(i) exp\left(-a_t y_t T_t(u_t)\right)}{C_t} \quad (2.2)$$

2.4.5. eXtreme Gradient Boosting (XGBoost)

The XGBoost classifier has gained popularity in data science because it uses a gradient descent from the trees constructed parallel to build an optimal model using the majority rule (Woodruff, 2017). The XGBoost was developed to overcome limitations associated with previous boosting algorithms (i.e., GB and AdaBoost) (Saini and Ghosh, 2018). It uses a regularized technique to control model overfitting while maximizing the model accuracy by sequentially adding models to correct errors from the existing models (Mashaba-Munghemezulu *et al.,* 2021; Rumora *et al.,* 2020). Furthermore, the classifier is an advanced gradient boosting that is computationally efficient and uses a cluster machine to continuously train large models while boosting fitted models on newly generated data (Brownlee, 2021).

The XGBoost contains different tuning hyper-parameters that are decided on using a grid search, making it robust and suitable for classifying imbalanced multi-class data from complex heterogeneous landscapes (Mashaba-Munghemezulu et al., 2021; Rumora et al., 2020).

In this study, the XGBoost classification was performed using the XGBoost package with an objective function of "multi. softprob" to model the multi-class classification problem. The calibration of the models was carried out using stratified 10-fold cross-validation (CV). The CV is used to cater for the variance and enforce class distribution in imbalanced multi-class datasets while giving the dataset an equal chance to be returned to the testing set at fold iteration (Heydari and Mountrakis, 2018). The learning rate was set to 0.01 with a base score of 0.5.

2.5. Mapping accuracy assessment

The prediction performance of the ML algorithms (GB, AdaBoost, SVM, RF, and XGBoost) in mapping fruit-tree species and co-existing land use types using S2 was evaluated using 30% of the training samples from each of the seven models used in this study. A confusion matrix was used to calculate the threshold metrics widely used in imbalanced classification problems (Vanino *et al.,* 2018). The OA, user's accuracy (UA), producer's accuracy (PA), and F-measures (F1-score) were used to evaluate the classifiers' performance Taghizadeh-Mehrjardi *et al.,* 2020; Chen *et al.,* 2021). The equations below were applied for the assessment.

$$OA = \frac{Sd}{n} \times 100\% \quad (2.3)$$

$$UA = \frac{Xij}{Xj} \times 100\% \quad (2.4)$$

$$PA = \frac{Xij}{Xi} \times 100\% \quad (2.5)$$

$$F_{score} = 2 \times \frac{UA \times PA}{UA + PA} \quad (2.6)$$

where S_d demonstrates the samples correctly classified; n is the total number of validated samples; X_{ij} represents the observation in row i of column j; X_i is the marginal total of row i, and Xj denotes the marginal of the total in column j.

The accuracy of the classifiers was then compared using a student's paired t-test (Dietterich, 1998). The paired t-test was applied with an alpha level of 5% ($\alpha = 0.05$), using the classifiers' OA to assess whether significant differences exist across the sampled datasets (Brownlee, 2016). Boxplots were also created for visual interpretation to complement the t-test results.

2.6. Results

2.6.1. Spectral separability of fruit-tree crops and co-existing land-use types

The mean spectral reflectance of the dominant fruit tree crops and co-existing land cover types identified during the field surveys are shown in Figure 2.4. Compared to the other nine classes, the built-up class has the highest reflectance values in the visible spectrum (Blue: 490 nm; Green: 560 nm; red edge_1:705 nm). The spectral reflectance values superficially on red edge_2, red edge_3, red edge_4, and NIR are higher for guava (GV), woody vegetation (WV), avocado (AV), macadamia nut (MN), mango (MG), banana (BN), and pine tree (PT) than for bare soil (BS) and water body (WB) (Figure 2.4). The GV has the highest spectral values in the REP among the fruit-tree crops, followed by the MN, MG, BN, and PT. For the non-crops, the built-up areas have the highest spectral values, followed by bare soil, and as expected, the waterbody has the lowest spectral values because of its high absorption.

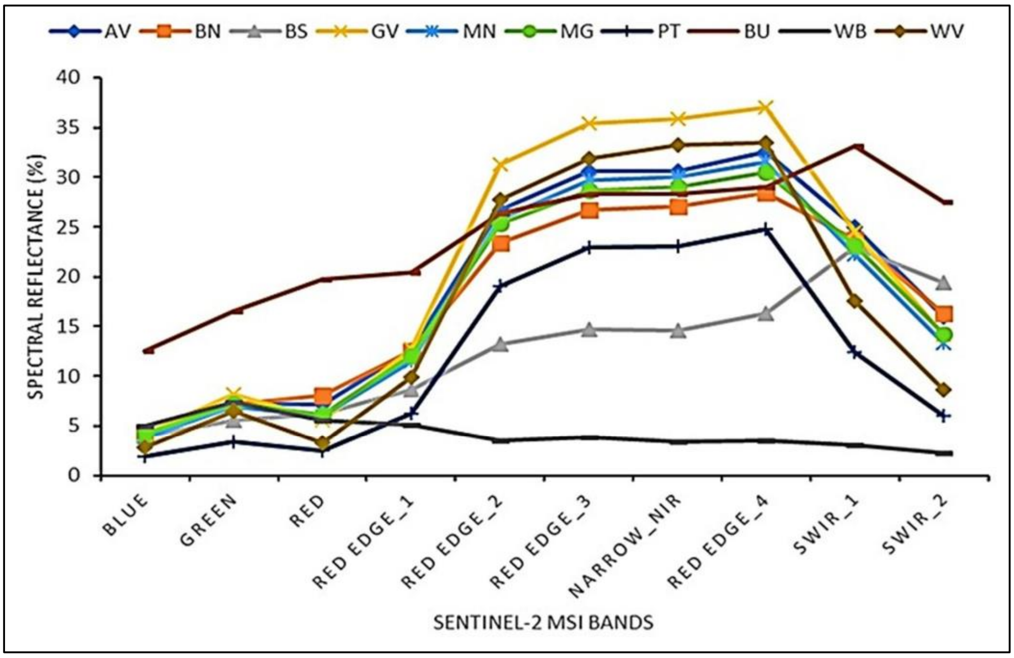


Figure 2.4: *Spectral signature of fruit-tree crops, and co-existing land cover classes extracted from spectral Sentinel-2A algorithms image. The mapped classes are avocado (AV), banana (BN), bare soil (BS), guava (GV), macadamia nut (MN), mango (MG), pine tree (PT), built-up (BU), water body (WB), and woody vegetation (WV).*

2.6.2. Fruit-tree crop mapping using machine learning algorithms

This study used seven data sampling scenarios and single date S2 imagery for fruit tree crop mapping by applying the various ML algorithms (RF, SVM, GBoost, AdaBoost, and XGBoost) (Figure 2.5a). The mapping results in Figure 2.5b show the effect of sampling techniques on the performance of the algorithms. The SVM classifier produced 71% and 69% OA at 150 balanced samples (Dataset_2) and 60% undersampling (Dataset_6). Next, the RF model recorded an OA of 65% at 100 balanced undersampled (Dataset_1) and 50% undersampling strategy (Dataset 4). The GB and XGBoost algorithms demonstrate the poorest performance for the data balanced at 100 sample points (Dataset_1) and sampled at 40% (Dataset_3) but recorded a slight improvement when each class has 150 balanced sample points (Dataset_2), scoring 50% and 70% for the undersampling strategies, respectively. Conversely, the performance of the AdaBoost classifier was stable across all datasets except at 40% undersampling (Dataset_3), in which the other classifiers also recorded accuracy scores of below 65%.

2.7. The variable importance

Variable selection is an important process that eliminates redundant information and increases the classifier's performance (Quan *et al.,* 2021). The permutation feature selection was used to retrieve important variable(s) from all possible parameter combinations (Quan *et al.,* 2021). The results show that the S2 bands' contribution and strength in discriminating classes vary across datasets, suggesting that the classifiers were sensitive to the sampling methods (Figure 2.6). For the 100 balanced samples, the SWIR_2 (B12: 2190 nm), SWIR_1 (B11: 1610 nm), and red (B4: 665 nm) are shown as the most optimal variables. For 150 balanced data points (Dataset_2), Red-Edge_2 (B6: 740 nm), green (B3: 560 nm), and SWIR_1 (B11: 1610 nm) were most important for discriminating the classes. A sampling at the 40% strategy, SWIR_2 (B12: 2190 nm), Red (B4: 665 nm), and RedEdge_2 (B6:740) produced the highest MDA scores, while at 50% sampling, the SWIR_1 (B11:1610), SWIR_2 (B12: 2190 nm) and Red-Edge_2 (B6: 740 nm) were the top contributing bands. At 60%, the SWIR_1 (B11: 1610 nm), SWIR_2 (B12: 2190 nm), and red (B4: 665 nm) were recorded as the best variables, while, at 70% sampling, the Red-Edge_2 (B6: 740 nm), red (B4: 665 nm) and SWIR_2 (B12: 2,190 nm) were the best variables. The original dataset displayed the red (B4), SWIR_1 (B11: 1610 nm), SWIR_2 (B12: 2190 nm), and Red-Edge_2 (B6: 740 nm) to be the bands crucial for accurate mapping.

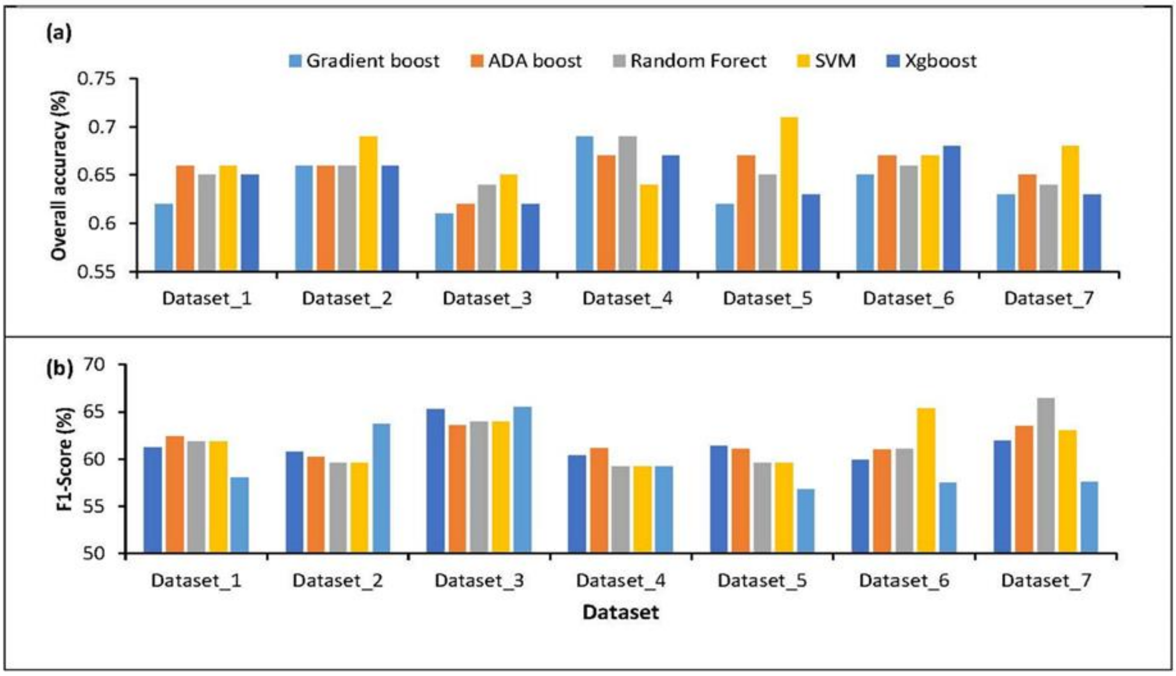


Figure 2.5: *Comparison of the overall classification accuracies (a) and F1-Score (b) from GB, Ada Boost, RF, SVM, and XGBoost using seven different datasets. Dataset_1 (balanced undersampled at 100 points), Dataset_2 (balanced undersampled at 150 points), Dataset_3 (undersampled randomly at 40%), Dataset-4 (undersampled randomly at 50%), Dataset-5 (undersampled randomly at 60%), Dataset_6 (undersampled randomly at 70%), and Dataset-7 (imbalanced dataset).*

2.8. The class accuracies

In Figure 2.7, the F1 scores for avocado, water body, and woody vegetation are above 60% for all ML classifiers across all datasets. Figure 2.8 shows that individual classes' UA and PA values were inconsistent across the datasets. The PA values for individual classes ranged between 17.78% (pine tree), using balanced data with 100 observations, to 100% (waterbody), using Dataset_2 (resampled at 40%) and the gradient boost classifier. At the same time, the UA ranged from 17.78% (pine tree, Dataset_2) to 100% for the water body class when applying the GB algorithm. The AdaBoost produced a PA of 23.33% for the mango class using Dataset_2 and 100% (waterbody) using Dataset_3. The UA accuracies ranged between 19.23% for macadamia nut, using Dataset_2, and 100% for waterbody. The PA values for individual classes ranged from 0 (pine tree) to 100 (avocado, waterbody) for XGBoost. The UA accuracies ranged from 0% (pine tree) for Datasets 3, 5, and 6 to 100% for the avocado and waterbody classes for Dataset_3,5 and 6. The pine tree class recorded the lowest user accuracy on account of the low PA produced using the XGBoost classifier. The RF recorded PA values ranged between 15.38% for the mango class using Dataset_4 and 100% for the waterbody class using Dataset_5. The UA ranged between 32% for the pine tree class using Dataset_4 and 100% for the avocado and water body class using Dataset_1, 3, and 4, respectively., using Dataset_1 and Dataset_3 and the AdaBoost classifier. The SVM produced PA values ranging between 15.38% for the mango class using Dataset_4 and 100% for the waterbody class using Dataset_5.

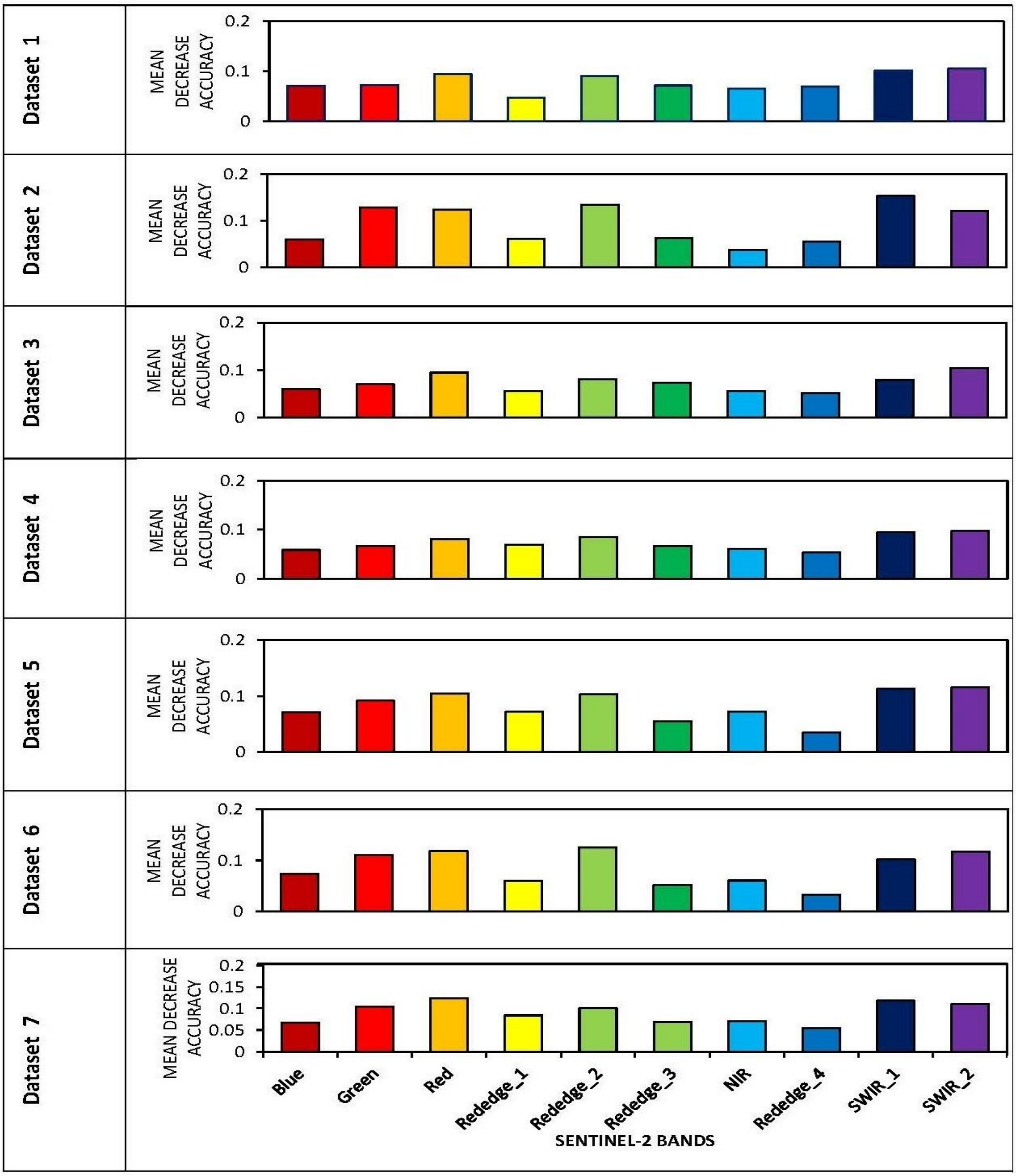


***Figure 2.6**: A comparative importance of explanatory variables used in permutation feature selection on seven datasets based on S2 spectral bands for Dataset 1 (balanced undersampled at 100 points), Dataset 2 (balanced undersampled at 150 points), Dataset 3 (undersampled randomly at 40%), Dataset 4 (undersampled randomly at 50%), Dataset 5 (undersampled randomly at 60%), Dataset 6 (undersampled randomly at 70%), and Dataset 7 (imbalanced dataset). The most influential bands are depicted by spikes indicating the highest feature importance scores. The colour maroon, red, orange, yellow, lime green, green, light blue, blue, dark blue and purple represent Sentinel-2 bands, namely: Band 2 (Blue), Band 3 (Green), Band 4 (Red), Band 5 (Rededge_1), Band 6 (Rededge_2), Band 7 (Rededge_3), Band 8 (NIR), Band 8a (Rededge_4), Band 11 (SWIR_1) and Band 12 (SWIR_2).*

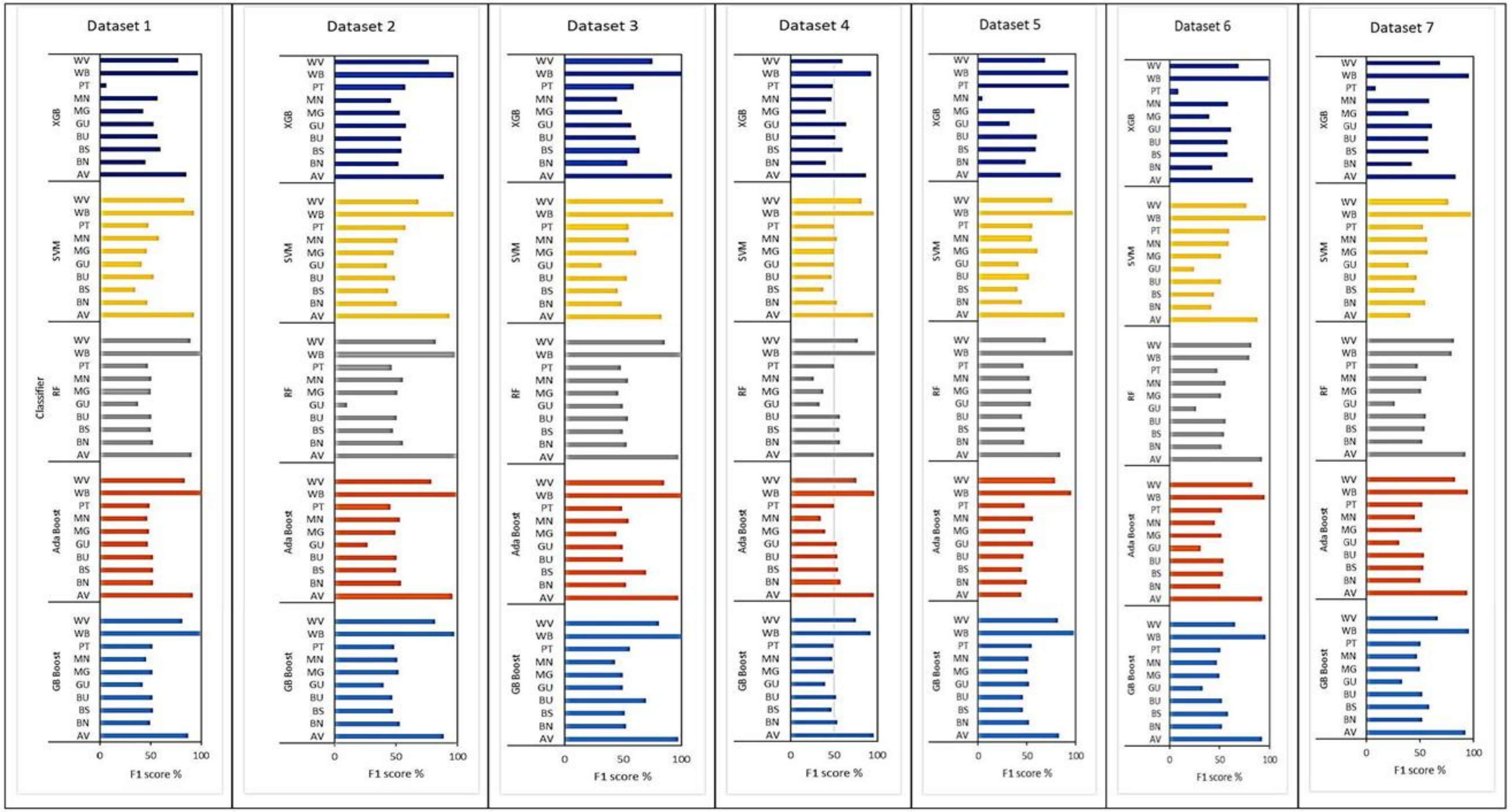


Figure 2.7: *F1-scores rankings indicating the effect of sampling methods on class-specific accuracies obtained from GB, Ada Boost, RF, SVM, and XGBoost classifiers. The mapped classes are avocado (AV), banana (BN), bare soil (BS), built-up (BU), guava (GV), macadamia* nut (MN), *mango (MG), pine tree (PT), water body (WB), and woody vegetation (WV).*

Figure 2.8: *User's (blue) and producer's (red) accuracy per class produced using GB, Ada Boost, XG-Boost, RF, and the SVM classifier. On each classifier, there is a trade-off between the user and producer accuracy across the datasets.*

2.9. Statistical comparison of ML classifiers from seven datasets

A comparison of the performance of the classifiers across the sampling techniques for the seven datasets shows that the Ada Boost and RF algorithms are sensitive (Figure 2.9). The overall mean ranged from 2.37 for RF and XGB ($p = 0.001$) to 5.42 for SVM ($p = 0.008$). The results indicate that the performance of the models across the datasets is statistically different and significant ($p = 0.001$). However, the GB, SVM, and XGBoost algorithms have the same mean and p of 0.292 and 0.780, respectively. The prediction of the fruit-tree crops with their co-existing land uses using different classifiers was significant across the datasets.

Furthermore, the prediction was more statistically significant using the GB and XGBoost than for AdaBoost, RF, and SVM.

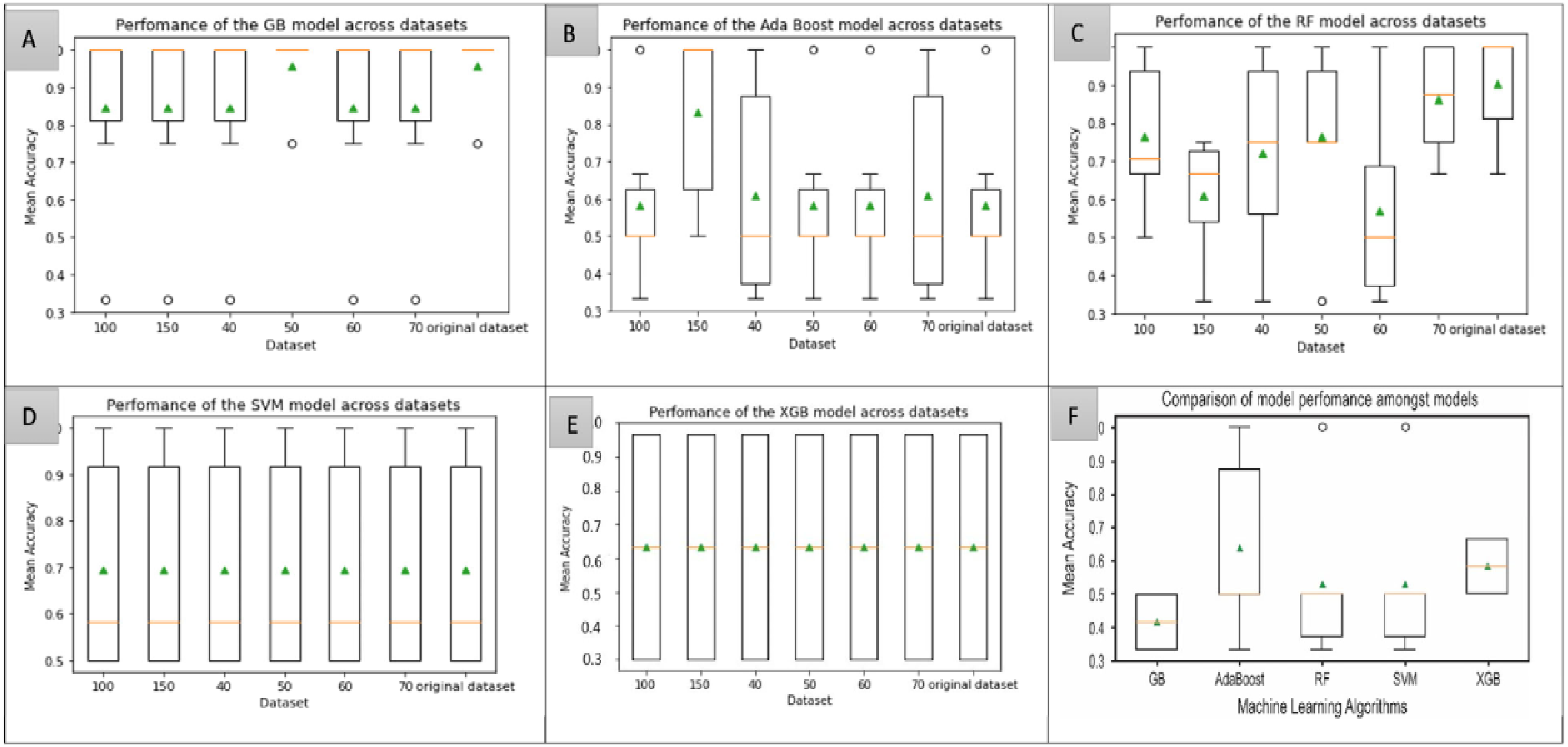


Figure 2.9: *The boxplot shows the mean (yellow line), the low and upper quantile (black whiskers) comparison of the model performance across datasets and among the models using students paired t-test, where (A) represents the gradient boosting; (B) AdaBoost; (C) random forest; (D) support vector machines; (E) eXtreme gradient boosting; while (F) depicts the comparison of model performance amongst models, respectively. From the x-axis, the 100, 150, 40, 50, 60, 70, and original dataset labels correspond to Dataset 1, Dataset 2, Dataset 3, Dataset 4, Dataset 5, Dataset 6, and Dataset 7, respectively. The green triangles represent the mean in relation to the overall performances of each classifier per dataset.*

2.10. Discussion

2.10.1. The effect of sampling size on the performance of machine learning algorithms

Timely and accurate image classification from earth observation data involves ongoing efforts in various global agricultural systems based on distinctive climatic regions and different user needs, advances in mapping approaches, and easier access to remote sensing data. Spatially explicit information at a fine scale is required for site-specific management to optimize production and maximize returns. Thus, continuous mapping of smallholder agricultural systems is required for proper agricultural management, informing decisions, and developing policies to ensure sustainable food systems and economies in rural communities. This study evaluated the influence of sampling techniques on mapping fruit-tree crops in a subtropical agricultural region, namely Levubu, South Africa, using a pixel-based approach with Sentinel-2 data.

The results demonstrate that applying effective sampling techniques and S2 data could provide spatially explicit information to improve the mapping of fruit-tree crops in a heterogeneous horticultural environment while reducing the need for extensive data collection. The fruit trees and co-existing land types in the study area were classified at an OA of 71% when applying the SVM algorithm, and the imbalanced dataset was sampled at a 60% level (i.e., Dataset_6). Recent studies by (Rumora *et al.,* 2020; Zhou *et al.,* 2017) also reported the SVM as superior to other ML algorithms in discriminating crop species. The SVM uses spectrally weighted kernels to eliminate redundant information, select optimal variables based on their relative importance, and improve classification accuracy (Brownlee, 2020; Rumora *et al.,* 2020). This advantage corroborates findings in previous crop type classification studies (Heydari and Mountrakis, 2018; Grabska *et al.,* 2020; Forkuor *et al.,* 2018). The ability of the SVM to achieve good classification results using small training samples should be considered when mapping landscapes with terrain characteristics that hinder the collection of balanced samples (Belgiu *et al.,* 2018).

The superiority of the boosting classifiers (GB, AdaBoost, XGBoost) to the benchmark classifiers (RF, SVM) has been reported in previous studies (Grabska *et al.,* 2020; Gašparović and Dobrinić, 2020). The boosting algorithms have also been applied in crop mapping to overcome the configurations of the standard classifiers (Mashaba-Munghemezulu *et al.,* 2021; Saini and Ghosh, 2018). Studies by (Bouras *et al.,* 2021; Saini and Ghosh, 2021) also compared different ML algorithms for crop classification. They concluded that the XGBoost was the best classifier over SVM, RF, and GB. However, our results showed that the XGBoost algorithm performed poorly compared to SVM and RF for our study area. These results are consistent with findings from recent similar studies (Mashaba-Munghemezulu *et al.*, 2021; Rumora *et al.*, 2020). The performance of the GB classification showed higher accuracy when applied to a 50% sampled dataset (Dataset_4) but produced poor results for the other sampled datasets. This may be attributed to the sensitivity of the GB algorithm to hyperparameter selection, which was not evaluated in this study (Rumora *et al.*, 2020). In all tested experiments, the results of the AdaBoost classifier were similar to the other classifiers, which contradicts the findings by (Baumann *et al.*, 2018). In their study, the AdaBoost was more accurate in classifying the canola crops than the SVM.

The sensitivity of the classifiers differed based on the degree of class overlap, class imbalance, and lack of density in minority classes (Zhou *et al.*, 2015). The effect of sampling on classifiers' performance was noticeable in imbalanced training samples. The GB, AdaBoost, and XGBoost were mostly affected compared to RF and SVM. The GB was affected by small balanced and imbalanced training samples (Dataset 1 and Dataset 3), while the overall accuracy of the AdaBoost, RF, and XGBoost decreased on small imbalanced training samples (i.e., Dataset 3). The performance of the XGBoost in this research was not stable on imbalanced samples, which contradicts the findings by Ayyagari (2020). The RF and SVM overall accuracy was over 65% on selected imbalanced training samples (i.e., Dataset 3, Dataset 5, Dataset 6, and Dataset 7). The performance of the RF was comparable on balanced samples. The differences in overall accuracy were significant with the increase in imbalanced training, which is similar to the report by (Sun *et al.*, 2020). The SVM was least sensitive to imbalanced training samples, performed well in detecting overlapping instances, and showed the most accurate results even on small training samples. However, the results were slightly different from RF (Devi *et al.*, 2019; Sun *et al.*, 2020). These findings are consistent with the results by (Waldner *et al.*, 2018). We caution, however, that this conclusion is based on a small study area with a limited number of reference data. We suggest that the findings of this study should receive considerable additional investigation with other and bigger numbers of samples before it is accepted as a substitute for reliable results.

2.10.2. The statistical comparison of the classifiers among and within the classifiers

The five classifiers recorded significant differences ($p = 0.001$), suggesting that the sampling techniques had an effect on the classification results and can be used to improve crop discrimination in smallholder agriculture. In contrast, the mean differences in overall accuracies of the classifiers within the classifiers were statistically insignificant for GB, SVM, and XGBoost across the datasets ($p = 0.780$). Although the highest classification accuracy was obtained using the SVM model, the classifier's statistical comparison within the datasets was not statistically significant, corroborating the findings by (Noi and Kappas,2017; Yousefi *et al.,* 2021), who stated that the SVM classifier was reliable and able to perform well in classification using limited training samples. The SVM classifier was robust and produced stable accuracies across the various balanced and unbalanced training set samples (Zhou *et al.*, 2017; Sun *et al.,* 2020). Conversely, the RF and Ada Boost performances differed across the various datasets and were statistically significant (Ustuner *et al.,* 2016). The models were robust and independent of the statistical distribution of training data, corroborating the statistical significance obtained using Student's paired *t*-test.

2.10.3. Comparison of individual class accuracies

The discrimination between classes was assessed using the F1 scores, UA, and PA. As inferred from the F1 scores, it is evident that the performance of the classifiers is optimized when the data is balanced with a low imbalanced class ratio in the training data (Heydari and Mountrakis, 2018). Although there was spectral overlapping among the mango, macadamia nut, and pine tree, the separability of the mango class from other classes was enhanced as a result of its particular structural and textural properties (Maldonado and López, 2014). In addition, it has a loose canopy that displays unique tonal changes. However, these classes recorded low accuracies due to the differences in stem elongation, flowering, and fruiting, suggesting that the image acquisition was not optimal for these crops (Hong-xia *et al.,* 2020). In some instances, spectral confusion can be exacerbated by factors, such as landscape heterogeneity, variations in agronomical practices, cropping calendars, and the pixel-based approach applied in this study (Hong-xia *et al.,* 2020). Consistent with findings by (Bouras *et al.*, 2021; Saini and Ghosh, 2021), the best class-specific metrics in this study were achieved using GB, AdaBoost, RF, and XGBoost. The results indicate that the avocado class was the most accurately classified crop across the datasets. The co-existing land uses (i.e., water bodies and woody vegetation) have the highest classification accuracy except for the bare soil and pine tree classes. The performance of the classifiers was consistent and comparable in detecting the built-up class due to its high scattering mechanism, which was not affected by the balanced and imbalanced ratio of the datasets (Richard *et al.,* 2017). The lowest F1-score value was achieved using the SVM classifier, which is similar to the (Peng *et al.,* 2017) findings. Their study achieved the highest F1 score using the RF algorithm when mapping the winter wheat crop in an urban agricultural region in Jiangsu Province, China. The crop spectral similarities affected the classification algorithms as they lowered the classification results on some datasets (Rumora *et al.,* 2020). High misclassifications among the macadamia nut, mango, and pine tree were observed from the accuracy metrics resulting from their small field size and spectral similarities. The difficulties in mapping macadamia trees are attributed to their hedgerow structures, as reported in other horticultural studies (Hing-xia et al., 2020; Richard *et al*., 2017).

The pixel-based classification failed to capture the spectral heterogeneity in macadamia nuts and mango trees owing to intercropping, structural overlapping, and mixed spectral signatures (Prins and Van Niekerk, 2020; Sivasankar *et al.,* 2018]. The findings corroborate previous research that cited class overlapping and spectral similarities as the main factors hindering classification accuracy (Sivasankar *et al.,* 2018; Johansen *et al.,* 2020). The results further demonstrate the effectiveness of S2 data in discriminating horticultural crops with mixed classes and high spatial heterogeneity (Chabalala *et al.,* 2022; Sivasankar *et al.,* 2018). Overall, stable class accuracies were obtained using Datasets 3 and Dataset 4, indicating the robustness of the classifiers in handling imbalanced data.

2.10.4. Importance of variables in mapping fruit-tree crops and co-existing land-use types

Measuring the importance of the explanatory variables concerning the modelled biophysical parameters has been fully explored in the existing literature (Saini and Ghosh, 2018; Kganyago *et al.,* 2021). In addition, variable importance is used to investigate the structure and the parameters that are used to train the model to understand their significance or a set of critical variables used to derive predictions (Kganyago *et al.,* 2021). The sensitivity of the S2 bands across different training datasets was evaluated using permutation feature selection. The comparative assessment of the derived feature importance (Figure 2.6) indicates a shifting pattern in the importance of S2 bands across different sampling techniques. The prominence of the S2 red band (B4: 665 nm), RedEdge_2 (B6: 740 nm), and the SWIR bands (B11,12: 1610–2190 nm) are comparable and more significant for class separability, regardless of the dataset used in this study. The results are consistent with the findings by (Saini and Ghosh, 2018). The performance of RedEdge-1 (B5: 705 nm) is similar when data is balanced, sampled at 60%, 70%, and the original imbalanced dataset. The blue band (B2: 490 nm) scored high at 100 equal splits (balanced), sampling at 60% and 70%. The additional S2 RedEgde_2 band (B6: 740 nm) improved the classification results in the research area and was among the top three crucial variables. As expected and reported in previous studies, the results highlighted the effect of biophysical parameters on various electromagnetic spectra (Kganyago *et al.,* 2021; Gutierrez-Coarite *et al.,* 2018). These findings are similar to previous research on crop type mapping using Sentinel-2 data (Grabska *et al*., 2020; Forkuor *et al.,* 2018). The S2 shortwave region is characterized by increased spectral variability sensitive to water content on leaf canopies (Gašparović and Dobrinić, 2020). The impact of phenology on the SWIR reflectance enhanced the crop classification, as noted in recent research (Darvishzadeh *et al.*, 2019). In contrast, the S2 visible spectrum and REP were found to be sensitive to chlorophyll content and leaf area index (LAI) in a study by (Luo *et al.,* 2021). The S2 10m spatial resolution reduces landscape heterogeneity and creates opportunities for crop type mapping in complex smallholder systems (Liu *et al.*, 2018; Cui *et al.*, 2019).

2.11. Conclusions

The influence of data sampling techniques and class imbalance for mapping fruit-tree crops and co-existing land uses was evaluated using five machine learning algorithms (GB, AdaBoost, RF, SVM, and XGBoost). The SVM algorithm outperformed the other algorithms with imbalanced data sampled at 60% and a balanced dataset with 150 instances per class. The overall accuracies ranged between 69% for 150 balanced datasets and 71% for imbalanced data sampled at a 60% level. Compared to GB, AdaBoost, RF, and XGBoost, the results show the superiority of the SVM classifier and Sentinel-2 data in mapping fruit tree crops in a heterogeneous horticultural landscape.

It is concluded:

- Data sampling and selecting appropriate classification algorithms are essential for accurately mapping fruit trees in a horticultural environment characterized by complex and heterogeneous landscapes.
- Sentinel-2 offers similar classification accuracy and can be used for crop type inventories; these reduce the need for extensive data collection.
- The Sentinel-2 Red-Edge_2, SWIR_2, SWIR_1 (B11), and red (B4) bands are the most crucial predictor variables for crop classification using all datasets.
- The S2 Red-Edge bands are centred in the biomass region and contribute more to biomass studies.
- The best overall accuracy was achieved using the SVM and the dataset sampled at 60% (i.e., Dataset 7), while the class accuracies were stable when the dataset was sampled at 40% and 50%, respectively (i.e., Dataset 3 and Dataset 4).

CHAPTER THREE

Machine Learning classification of fused sentinel-1 and sentinel-2 image data towards mapping fruit plantations in highly heterogeneous landscapes

This chapter is based on:

Chabalala, Y., Adam, E and Ali, A.K. (2022). Machine Learning Classification of Fused Sentinel-1 and Sentinel-2 Image Data towards Mapping Fruit Plantations in Highly Heterogenous Landscapes. *Journal of Remote Sensing*, 14, 1-26. DOI: https://doi.org/10.3390/rs14112621

Abstract

Mapping smallholder fruit plantations using optical data is challenging due to morphological landscape heterogeneity and crop types having overlapping spectral signatures. Furthermore, cloud covers limit the use of optical sensing, especially in subtropical climates where they are persistent. This research assessed the effectiveness of Sentinel-1 (S1) and Sentinel-2 (S2) data for mapping fruit trees and co-existing land-use types by using support vector machine (SVM) and random forest (RF) classifiers independently. These classifiers were also applied to fused data from the two sensors. Feature ranks were extracted using the RF mean decrease accuracy (MDA) and forward variable selection (FVS) to identify optimal spectral windows to classify fruit trees. Based on RF MDA and FVS, the SVM classifier resulted in relatively high classification accuracy with overall accuracy (OA) = 0.91.6% and kappa coefficient = 0.91% when applied to the fused satellite data. Application of SVM to S1, S2, S2 selected variables and S1S2 fusion independently produced OA = 27.64, Kappa coefficient = 0.13%; OA= 87%, Kappa coefficient = 86.89%; OA = 69.33, Kappa coefficient = 69. %; OA = 87.01%, Kappa coefficient = 87%, respectively. Results also indicated that the optimal spectral bands for fruit tree mapping are green (B3) and SWIR_2 (B12) for S2, whereas for S1, the vertical-horizontal (VH) polarization band. Including the textural metrics from the VV channel improved crop discrimination and co-existing land use cover types. The fusion approach proved robust and well-suited for accurate smallholder fruit plantation mapping.

3.1. Introduction

Timely and precise spatial information on crop types is crucial for managing and monitoring agricultural activities, especially in areas with increasing populations and rapidly growing food security demands (Crommelinck and Höfle, 2016). Accurate location monitoring of agricultural land at a crop types level plays an essential role in precision agriculture and achieving a better understanding of farm management practices, forecasting crop yields, and facilitating an enhanced response to food insecurity (Abdel-Rahman and Ahmed, 2008; Brinkhoff *et al.,* 2020). Promoting good agricultural practices and achieving sustainable agricultural systems depend on detailed spatial information of the current cropping systems on small-field parcels, especially in South Africa (Useya and Chen, 2019). The fruit-tree industry is a pillar that provides livelihoods, creates employment opportunities, and reduces poverty and food insecurity by generating income for local agriculturally based livelihoods while contributing to sub-tropical rural economic development (Mukwada *et al.,* 2021; Abbott *et al.,* 2021). Production of fruit-tree crops is increasing worldwide, and export sales have become more competitive (Amare *et al.,* 2019). Thus, there is a need to develop effective management to improve fruit-tree crop production and advance the competitive standing of small-scale farmers. One of the essential requirements for precise crop management is providing decision support tools that offer up-to-date spatial information on the tree crop systems and dynamics to maximize production and returns (Johansen *et* al., 2019; Danylo *et al*., 2021).

Historically, crop types were conventionally identified and mapped using field-based methods such as ground surveys and agricultural statistics, as they work well at small local scales (Ge *et al.,* 2016; Kolecka *et al.,* 2018; Sinha *et al.,* 2021). Although this method provides accurate spatial data about crop type distribution, many studies have shown the method is costly, time-consuming, and labour-intensive and is not feasible for large-scale crop mapping (Meroni *et al.,* 2021). As such, the agricultural industry has undergone a digital revolution characterized by technologies such as artificial intelligence, remote sensing and geographical information systems specifically designed to advance the agricultural sector (Wolfert *et al.,* 2017; Segarra *et al.,* 2020). These technologies are promising tools for innovating agricultural operations, fostering the adoption of precision agriculture measures, and enhancing production (Kellengere Shankarnarayan and Ramakrishna, 2020; Sishodia *et al*., 2020).

Remote sensing has shown potential to overcome the abovementioned limitations as it provides accurate crop information and timely and cost-effective farming opportunities needed for site-specific crop management (Böhler *et al.*, 2020). It is a tool used in conjunction with machine learning algorithms, big data analytics and artificial intelligence to extract useful information on fruit species regardless of the varieties (Yandún Narváez *et al.*, 2016). The unprecedented availability of high-resolution spectral, spatial and temporal resolution images has advanced the use of remote sensing in crop mapping over various scales and geographic locations (Maimaitijiang *et al*., 2020).

The application of different satellite technologies for decision-making and precision agriculture in horticulture has been shown in various studies (Sinha *et al.,* 2021; Bai *et al.,* 2016; Bai *et al.,* 2019). Among the various satellite types, MODIS data has been widely tested in crop mapping at regional and global scales due to its global coverage (Gilbertson and Van Niekerk). However, the efficacy of using MODIS data in crop type mapping is limited to commercial farming (Qader *et al.,* 2021). Its solutions for precision farming practices are challenging to envisage and implement due to MODIS spatial resolution, resulting in mixed pixels and deteriorated classification accuracy when applied to heterogeneous terrains where smallholder farmers operate (Aguilar *et al*., 2018).

Remote sensing data with intermediate spatial resolution < 30 m (Landsat 8) and high spatial resolution < 10 m, such as Quickbird, Worldview, SPOT, and Sentinel-2 data, are used to overcome the problem of spectral mixing and improve crop classification in agricultural systems characterized by small fields (Hong-xia *et al.,* 2020). The applicability of high spatial resolution data has been tested in producing accurate site-specific geospatial information for precision agriculture due to their resolutions being smaller than smallholder fields (Lebourgeois *et al.,* 2017; Veloso *et al.,* 2017; Rao *et al.,* 2021). The regional and localized applications of high-resolution data include cropland mapping (Neigh *et al.,* 2018), identification of irrigated crops (Zheng *et al.,* 2015), and tree crop condition mapping (Chemura *et al.,* 2015). However, the applicability of the above datasets for designing precise sustainable farming solutions in smallholder farming systems in sub-Saharan Africa is still limited due to various dimensions such as lack of modern hardware and network infrastructure, software, technological and technical skill gaps, and lack of access to funding (Aguilar *et al*., 2018).

Optical data is weather-dependent, relies on the sun’s energy, and is mainly affected by clouds, thus hampering their applicability in crop mapping in subtropical regions such as South Africa (Torbick *et al.,* 2017). The synthetic aperture radar (SAR) data from the Copernicus program has unlimited availability and is a powerful tool that can be used to overcome the impact of optical data unavailability, as radar systems lack illumination and can penetrate cloud covers (Steinhausen *et al.,* 2018; Yang *et al.,* 2021). The SAR sensors can differentiate the crops' geometric and structural features due to their ability to capture textural information consisting of pixel intensities characterized by spatial variations that can detect subtle structural plant traits (Maimaitijiang *et al.,* 2020). The radar systems are sensitive to the geometric and dielectric properties of the crops, enabling the systems to transmit emissions and backscatter from the objects, rendering them viable for vegetation mapping (Yang *et al.,* 2021). However, the lack of spectral features and sensitivity to the surface parameters (e.g., vegetation, soil moisture, soil roughness) limit their applicability in crop mapping (Faye *et al.,* 2020). Conversely, the leaf structure and water content influence the reflected energy from optical sensors, rendering them effective for crop mapping (Moumni and Lahrouni, 2021).

In recent studies, the hypothesis of data fusion in increasing crop mapping accuracy by capturing vegetation structural information representations from SAR and optical has been supported (Ren *et al.,* 2022; Cai *et al.,* 2019). Data fusion compensates for the limitations of single-source sensors by combining different sensor spectral, spatial and temporal characteristics, offering broader information content of the target object (Ren *et al.,* 2022; Orynbaikyzy *et al.,* 2019). The growing research on crop types mapping reflects the diverse nature of agricultural landscape locally, regionally, and globally and its site-specific challenges such as spectral similarities and intra-class variability, crop diversity, environmental and weather conditions, and farming system (Ren *et al.,* 2022; Orynbaikyzy *et al.,* 2019; Wardlow *et al.,* 2007; Ni *et al.,* 2021). Different approaches and earth observation data have been tested to improve crop classification. The combination of Sentinel-1 and Sentinel-2 is well-investigated in various countries such as Belgium, Australia, India, and China (Brinkhoff *et al.,* 2020; Van Tricht *et al.,* 2018; Ni *et al.,* 2020). These studies confirm that SAR data improve crop type classification accuracies and overcome the limitations of cloud cover. This is due to SAR data characteristics and signals sensitivity to plant canopies' physical and structural properties, which complements optical information from multispectral sensors such as Sentinel-2 (Ni *et al.,* 2020; Blickensdörfer *et al.,* 2022). However, the potential for mapping smallholder farming systems, particularly in African countries, has not been thoroughly tested and remains challenging due to the high spatial and temporal dynamics of the agricultural landscapes. For example, in rural areas of South Africa, the farmland depicts a small-scale but very diverse mix of crops such as maize, fruit trees and vegetables. Fields are relatively small and often separated by woody vegetation or plantation; the growing conditions and mixed crop phenology and canopies' physical structure vary enormously (Lakshmanan and Mathiyazhaga, 2021; Gené-Mola *et al.,* 2020). The optical sensors' observations show high spatial and temporal crop spectral variability, and therefore, they may not be adequate to map crop types in small farms accurately (Hong-xia *et al.,* 2020; Steinhausen *et al*., 2018).

Accurate remote sensing depends not only on the data characteristics and the site specifications but also on the classification methods. The selection of classifiers in image classification depends on the size of the crop field, landscape complexities, management strategies, and the diversity of the crops (Conrad *et al.,* 2019). Machine learning-based classifiers such as support vector machines (SVM), random forest (RF), decision trees, and k_nearest neighbours (*k*-NN) have been used to create thematic maps from satellite images for different fruit-tree crops with reasonable accuracies (Conrad *et al.,* 2019; Ouzemou *et al*., 2018; Saini and Ghosh, 2021; Prins and Van Niekerk, 2021).

Levubu is one of the sub-tropical farming areas that makes Limpopo Province a leading fruit export in South Africa (Mukwada *et al.,* 2021). Smallholder farming systems (defined here as farms under 2 ha) play a significant role in global food production and are the backbone of South Africa's economy (Useya and Chen, 2019). These agricultural systems create employment opportunities that generate income and ensure food security for the local agricultural-based livelihoods (Mukwada *et al.,* 2021). The Levubu farms are predominantly planted with fruit tree crops such as macadamia nut, avocado, banana, guava, mango, and pine trees. The co-existing land-use types include water bodies, built-up areas, and woody vegetation. Hence, mapping these fruit trees with higher accuracy is necessary for precise agricultural management and improving food production. Despite being the most prominent fruit exporter in Limpopo Province, the management of these fruit tree crops relies on ground-based surveys that are time-consuming and often subjected to human error (Hong-xia *et al*., 2020).

The European Space Agency (ESA) provides SAR and multispectral data with a high spatial resolution, giving ample opportunities for crop phenotyping in smallholder agriculture in sub-tropical regions (Sarzynski *et al.,* 2020). Sentinel-1 is suitable for smallholder crop mapping due to different polarization channels (VH and VV), which interact with crops differently, providing increased data content to support decision-making in a cost-effective manner (Qadir and Mondal, 2020). The potential for integrating the Sentinels products has been tested for mapping crop growth conditions, leaf area index, and crop type patterns in various landscapes (Meroni *et al.,* 2021; Van Tricht *et al.,* 2018; Kraaijvanger *et al.,* 2019). However, the Sentinels products' integrated use has not been tested in mapping smallholder fruit plantations in South Africa despite their free availability, consisting of suitable spatial and spectral resolutions that offer unlimited capabilities for mapping at the landscape level. There is limited research in South Africa that has integrated the SAR and optical data from Sentinel satellite sensors to map smallholder fruit-tree crops due to complex image processing and complex sensor interactions with target parameters (Torbick *et al.*, 2017). Therefore, the study aims to address this problem by investigating the integrated use of Sentinel-1 and Sentinel-2 data in fruit-tree crop mapping using machine learning classifiers. The objectives of the research are:

- To evaluate the effectiveness of RF and SVM in discriminating tree crops from each sensor separately;
- To examine if the application of RF and SVM classifiers on fused data can improve the crop mapping accuracy at the landscape level;
- Assess the use of selected important variables in improving the mapping accuracy of fruit tree crops at the landscape level.

3.2. Materials and Methods

3.2.1. Method Overview

The proposed methodology shown in Figure 3.1 consists of the following steps: image pre-processing, converting Sentinel-1 (S1) image intensities (VV and VH) into decibels (dB), resampling six bands of Sentinel-2 data (i.e., rededge_1(B5), rededge_2(B6), rededge_3(B7), narrow NIR(B8A), SWIR_1(B11) and SWIR(B12)) were resampled from 20 m to 10 m spatial resolution using the Nearest Neighbor algorithm. After that, the resampled bands were combined with the 10 m resolution bands (i.e., Blue: B2, Red: B3, Green: B4, NIR: B8), and the resulting image composite consisted of 10 spectral bands. The image classification process involved the creation of an image composite using S1 and S2, variable selection on the fused image (S1S2). The ground truth data were used to extract the image reflectance using four imageries (S1, S2, S1S2, S1S2 variable selection) and classify them using random forest (RF) and support vector machine (SVM). The accuracy assessment was carried out on the classified maps to evaluate the image that produced the best fruit trees and co-existing land-uses map.

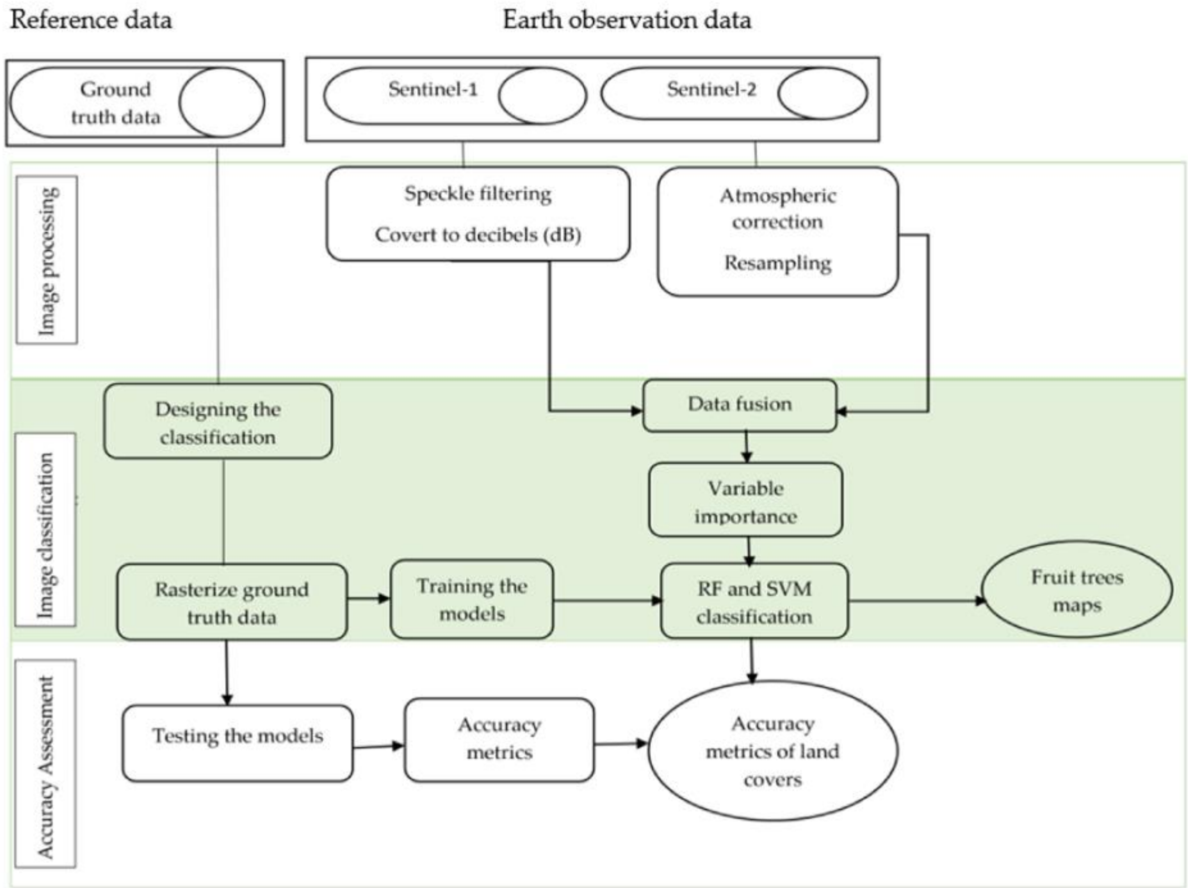


Figure 3.1: *Flowchart of the methodology implemented for mapping fruit trees and co-existing land-use types in Levubu, Limpopo, South Africa.*

3.2.2. Research region

The research region (10,000 hectares) covers the Levubu sub-tropical fruit-tree crops farming area in Vhembe District Municipality, Limpopo, South Africa (Figure 3.2). The study area is geographically located at latitudes $23^{\circ}5^{0}31.25^{00}$ S to $23^{\circ}6^{0}45.03^{00}$ S and longitudes $30^{\circ}13^{0}40.88^{00}$ E to $30^{\circ}21^{0}26.03^{00}$ E. The subtropical farming area covers Tshakuma village, Mulangampuma, Levubu, Mukhuro, Valdezia and Ratondo. Levubu has a population of 207,000, of which white people account for 10.9% (VDM, 2021).

The area's temperature is warm and wet from December to February, ranging from 16 to 40 °C, and the cool, dry season from May to August, ranging from 12 to 22 °C. The environmental, ecological, and biophysical environments are optimal for growing subtropical tree crops.

The Levubu area has two major rivers, namely the Luvuvhu (the main river) and the Lotonyanda River (Mukhada *et al.,* 2021). The Soutpanberg influences the annual rainfall distribution, reaching 2000 mm at Entabeni, the middle part of the Soutpanberg and 340 mm in the western part, peaking up during January and February. The precipitation season of the area hinders the acquisition of good-quality optical data, which presents an ideal region to examine the potential for integrating SAR and optical data.

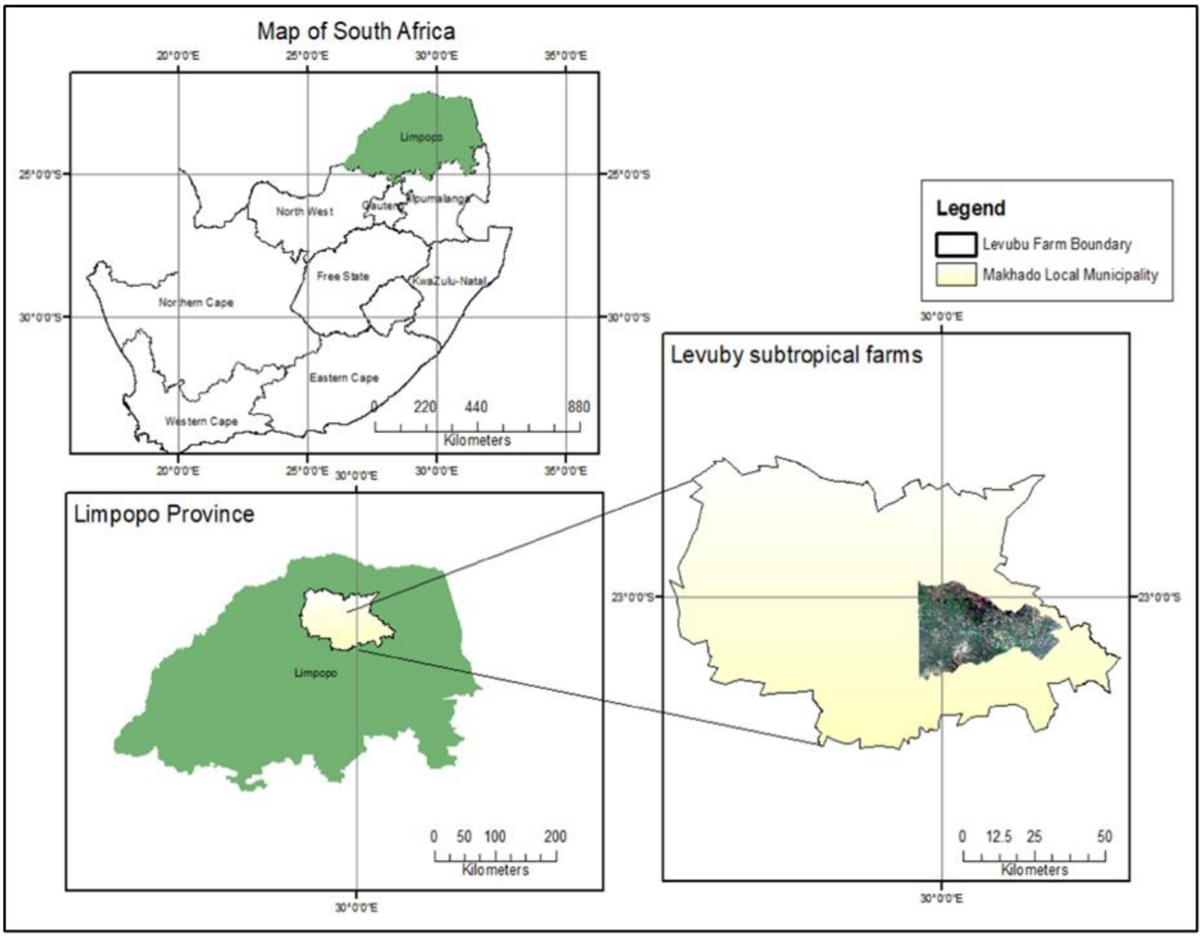


Figure 3.2*: Map of South African provinces showing the location of Levubu sub-tropical farming area in Sentinel-2 RGB image located in Makhado Local Municipality*.

3.2.3. Remote Sensing Data Acquisition and Pre-Processing

The optical Sentinel-2A (S2) Level-1C and Interferometric Wide (IW) level-1 GRD (ground range detected) SAR Sentinel-1A (S1) were used in this study. The data were freely downloaded from the European Space Agency Portal, the Sentinel Scientific Hub (https://scihub.copernicus.eu/ (accessed on 16 January 2021). Both S1 and S2 images were acquired on 28 December 2019, which coincided with the data collection. The sentinel mission aims to provide data continuity to MSI from the SPOT mission and a continued supply of observations to be used in services such as land-cover mapping, food security, and forest monitoring, to name a few (Berger *et al*., 2012).

Sentinel-2 (S2) is a Multispectral Instrument (MSI) launched on 23 June 2015, with a wide- swath of 290 km providing data at 10 m to 60 m spatial resolution in 13 spectral bands ranging from visible to shortwave infrared regions with a revisit time of 5 days (ESA, 2015). S2 contains three spectral bands strategically positioned in the red-edge position, which is crucial for improving crop condition mapping and retrieval of chlorophyll content (Clevers and Gitelson, 2013).

S2 image was atmospherically corrected using a level 2A Prototype processor (SEN2COR) version 2.5.5, as stipulated on the Sentinel-2 processor for users (Louis *et al.*, 2016). All bands were resampled to 10 m resolution, and the corrected spectral bands were used in the analysis. The spectral configurations of the used sensors are presented in Table 3.1.

Table 3.1: *Spectral configurations for Sentinel-1 and Sentinel-2 data.*

Sentinel 2 MSI					**Sentinel-1 SAR**	**Wavelength**
Band	**Name**	**Centre**	**Width**	**Resolution(m)**	**Name**	**Centre**
2	Blue	490	65	10	VH polarization	C-band
3	Green	560	35	10	VV polarization	C-band
4	Red	665	30	10		
5	Red edge	705	15	20		
6	Red edge	740	15	20		
7	Red edge	783	20	20		
8	NIR	842	115	10		
8a	Narrow NIR	865	20	20		
11	SWIR	1610	30	20		
12	SWIR	2190	180	20		

The Sentinel-1A (S1) is a synthetic aperture radar (SAR) satellite, launched on 3 April 2014. The data comprises two polar-orbiting instruments performing radar imaging through the C-band at 5.405 GHz, operating in four imaging channels (Hass and Ban, 2017). The sensor has a wide swath of 400 km, different spectral resolutions of down to 5 m and concise temporal resolution, enabling rapid data collection. Sentinel-1 is in dual-polarization mode, and its data layers include the vertical transmit and horizontal polarized received (VH) and vertical transmit and vertical polarized received (VV) operating at 1 dB (3Σ) radiometric accuracy, sensitive to vegetation dynamics and crucial for crop mapping (Vreugdenhil *et al*., 2018).

Sentinel-1 was atmospherically corrected using the Sentinel-1 Kit Module in ESA SNAP and a python script configured to read the digital elevation model (DEM) and the radiative transfer lookup tables during image processing using Anaconda (python 2.7) (Louis *et al.*, 2016). The image corrections include geometric correction, which is an essential step in quantitative analysis. In addition, we conducted radiometric calibration and speckle filtering to remove the sensor distortions and noise. The image speckles were reduced using a 5 × 5 m adaptive Lee filter commonly used for SAR data and have been proven to produce the best filtering results, especially in areas with high heterogeneity (Moumni and Lahrouni, 2021). The image was terrain corrected using a Range Doppler STRM digital elevation model (DEM) 1Sec HGT (auto download) as an input and a pixel spacing of 10.0 m × 10.0 m with a line spacing of 10.0 m.

The analysis and interpretation of the Sentinel-1 SAR GRD product were enhanced by converting the processed image into Beta sigma naught (σ^0), and the VH and VV units were converted into decibels (dB) based on Equation (3.1).

$$X = 10 \log_{10} x \tag{3.1}$$

where x illustrates the pixel value.

3.2.4. Data Fusion

Integrating Sentinel-1 and Sentinel-2 is used to improve the textual, structural, and spatial resolution for better crop type discrimination (Orynbaikyzy *et al.*, 2019). In this study, data fusion was performed using the Sentinel toolbox from SeNTinel Application Platform (SNAP) software at a pixel level using band combination. The images were first co-registered before the synergy. Sentinel-1 data were used as a master product since its geolocation accuracy was high after range Doppler terrain correction.

The nearest neighbour model was used to resample and produce a fused image with 10 m pixel resolution using the Universal Transverse Mercator (UTM) at zone 36 South (Zhang *et al.,* 2014; Abubakar *et al.,* 2020). Sentinel-1 was used as master and Sentinel-2 as a slave, and the fused image was segmented using the collocation layer to aggregate the pixels with similar values.

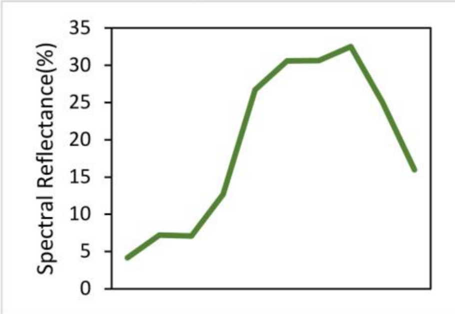
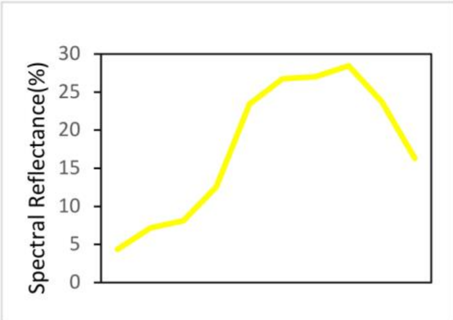
3.2.5. Field Data Collection

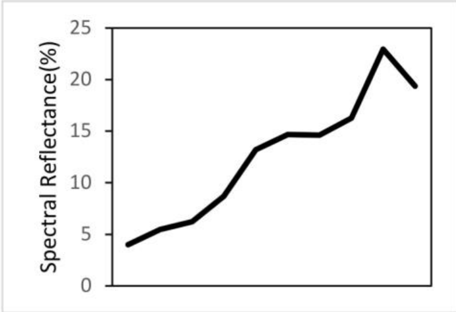
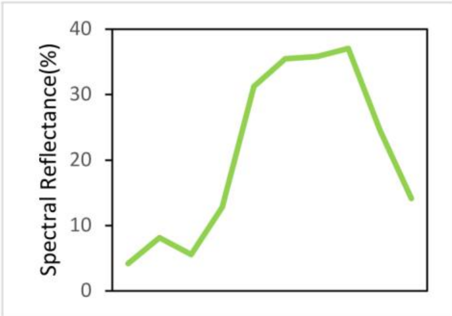
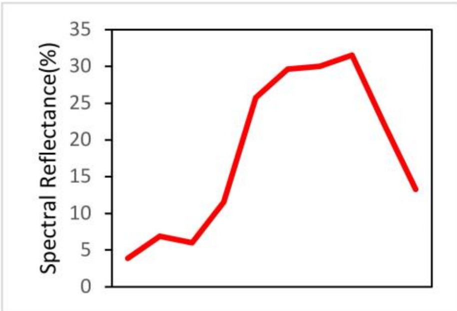
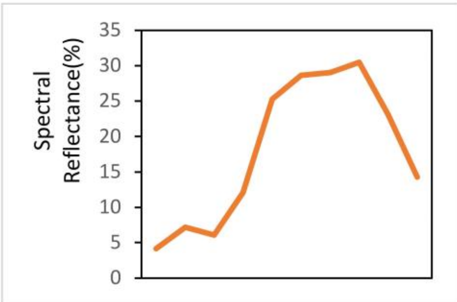
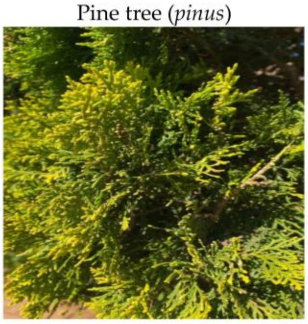
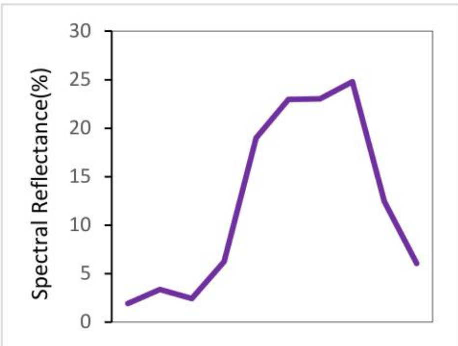
The training data were collected on 28 and 29 December 2019, as well as on 2 and 3 January 2020. The common fruit-tree crops were identified through extensive field surveys and consulting Google Earth Pro. Where there was accessibility, a roadside sampling method was used to capture points (Waldner *et al.,* 2019). A handheld Global Position System (GPS), Garmin eTrex 20X, with a 5 m positional accuracy, was used to capture the coordinates systems of the samples at the centre of the farms. At each point, metadata were captured, and a note was made on mixed cropping systems to ensure that these fields were avoided during digitizing. Using a 10 m distance to account for Sentinel-2 spatial resolution. A total number of 304 (n = 304) ground-truth points were captured from the dominant land cover classes; these include avocado (n = 49), banana (n = 53), built-up (n = 7), guava (n = 12), macadamia nut (n = 95), mango (n = 18), pine tree (n = 22), water bodies (n = 4) and woody vegetation (n = 10). These points (n = 304) were then used to guide digitizing additional points where the field is big enough using informative observation on ArcGIS and Google Earth (Fowler et al., 2020).

3.2.6. Assessing Spatial Autocorrelation

The ground truth data were evaluated for spatial autocorrelation, which produced a Moran's Index of 1.004 with a Z-score of −24.76, indicating a dispersed pattern (Ahmed *et al.,* 2021). Furthermore, the points were spatially separated using the "Spatially Rarerely Occurrence Data for Species Distribution Models (SDM) (reduce spatial autocorrelation)" function available in SDMPro Toolbox in ArcGIS Pro version 2.4. A Moran's Index of 0.09 with a Z-score of 0.68 was obtained, which indicates a strong random pattern among the points (Abdulhafedh, 2017).

The retained data were partitioned into 70% (387) train data and a 30% (176) validation set using the SDM Random Select Points function to prevent bias in the classification model (Hansch *et al.,* 2017). Furthermore, purposeful sampling was employed to digitize additional samples for the underrepresented classes (i.e., bare soil, build-up, guava). The summary statistics of the most common fruit-tree crops and other land-use types identified in the field with their average reflectance, field photos, and reference data are presented in Figure 3.3.

Class	Spectral	Training	Testing	Total
Avocado (*Persia americana Mills*) 		45	18	63
Banana *(Musa acuminate)* 		45	19	64

Bare soil 			22	15	37
Guava (*Psidium*) 			27	14	41
Macadamia nut (*Macadamia*) 			77	32	109
Mango (*Mangifera indica*) 			31	13	44
Pine tree (*pinus*) 			38	15	53

The average spectral profiles for common fruit-tree crops and co-existing land-use types.

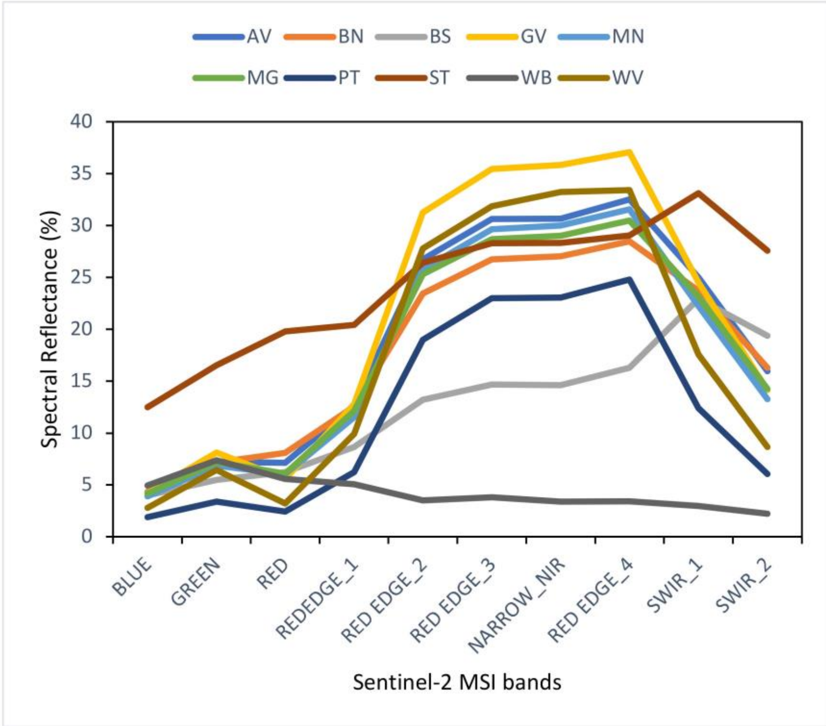


Figure 3.3: *Average spectral reflectance profiles extracted from Sentinel-2 data on fruit-tree crops and co-existing land-use types, reference data, and field photos. The mapped classes are avocado (AV), banana (BN), bare soil (BS), guava (GV), macadamia nut (MN), mango (MG), pine tree (PT), built-up (BU), water body (WB), and woody vegetation (WV).*

3.3. Image classification

3.3.1. Machine learning classification algorithms

Two supervised machine learning algorithms, random forest (RF) and support vector machines (SVM), were evaluated in classifying the tree crop types using four dataset configurations at a pixel-based level. The models were tuned and trained using the "caret package" in R studio, statistical software, and programming language version 4.0 (Kuhn, 2014; Feyisa *et al.*, 2020).

3.3.1.1. The random forest algorithm (RF)

Random Forest (RF) is an ensemble machine learning algorithm that uses bagging to improve predictions by reducing the variance introduced by Breiman (Breiman, 2001). In this study, the RF model was optimized using two primary parameters, namely *mtry* and *ntree*. The former denotes the number of predictors tested at each tree node, and the latter illustrates the number of decision trees ran at each iteration, respectively (Tatsumi *et al.*, 2015). The classification accuracy was improved by optimizing the number of grown trees (*ntree)* and variables at each tree split *(mtry)* based on the OOB error rate using a 10-fold cross-validation method and a grid search (Zhang *et al.,* 2020). The optimal trees were searched at 500 intervals, using *ntree* parameters ranging from 500 to 10,000 (Tatsumi *et al.,* 2015). On the other hand, the optimal *mtry* was searched using the square root of the total number of explanatory variables, as it has been proven to provide satisfactory results (Conrad *et al*., 2019; Abdullah *et al*., 2019).

The RF model measures the importance of each explanatory variable applied in the classification using the OOB samples (unbiased estimates) and permutes the variables to calculate the mean decrease accuracy (MDA) and mean decrease Gini (MDG) (Mashaba-Munghemezulu *et al.,* 2021). The RF also builds trees using randomly replaced bootstrapped samples comprising two-thirds of the training dataset, decreasing the variance in classification errors (Genuer *et al.,* 2015; Richard *et al.,* 2017). The one-third of the training data excluded during the bootstrapped sample is referred to here as the out-of-bag (OOB) error rate. The OOB is used to internally evaluate the performance of the RF classification during the cross-validation process (Yu *et al.,* 2018). The model with the lowest OOB error is considered the best data fit for the RF model (Useya and Chen, 2019). The equation used to define the OOB error in a classification framework is as follows (Genuer *et al.,* 2015).

$$errOOB = \{1/n \text{ Card } \{i \in \{1, \ldots, n\} \mid y_i \, 6{=}\hat{y}_i\} \quad (3.2)$$

where $\hat{y}_i$ illustrates the predicted values aggregated by the number of trees t where (x_i, y_i) belongs to the *OOB* sample associated with it.

The RF algorithm has been commonly used to create crop type maps, as it has been shown to produce better accuracies (Prins and Van Niekerk, 2021). The classifier is robust and suitable for heterogeneous landscapes and can handle the problem of noise and correlations that are intrinsically inherited in remote sensing data (Genuer *et al.,* 2015). The crop types were classified using datasets generated from Sentinel-1 and Sentinel-2.

3.3.1.2. Assessing the importance of variables

The application of noisy bands has been reported to significantly impact classification conducted at the pixel level (Zhao *et al.,* 2020). Selecting a set of optimal variables is essential to properly characterize objects and improve classification accuracy (Cui *et al.,* 2020). In this study, the RF algorithm’s out-of-bag (OOB) strategy was used to measure and rank the importance of variables from the training data using myImagePlorR in R studio from the data fusion model (Breiman, 2001). The RF model quantifies the variable importance (VI) using a mean decrease in accuracy (MDA) index, which ranks the importance of independent variables in predicting the dependent variable using a random permutation-based score (Breiman, 2001;

Chan and Paelinckx, 2008). The variables with lower MDA values indicate less association with the dependent variable, while the variables with higher MDA values are considered the most significant independent variables in classification (Richard *et al*., 2017; Vuolo *et al*., 2018; Schmidt et al., 2016; Zhang et al., 2020).

The variable importance (VI) produced by each model differed significantly between the four dataset configurations. The spikes show the significance of each band per class in the band importance graphs. Standard bands for different datasets were identified for mapping the same crop types, but band importance inconsistencies amongst classes were observed (Zhao *et al.,* 2013). The selected optimal bands were used to form the model used to create the final crop maps.

3.1.3.3. Forward variable selection

A forward variable selection (FVS) is used to select relevant features that will yield lower misclassification error rates and use them to build multiple RF models (Adam *et al.,* 2012). FVS is based on an RF mean decrease accuracy (MDA) metric that ranks the variable importance of the remote sensing dataset (Breiman, 2001; Hu *et al.,* 2017; Loggenberg *et al.,* 2018). Statistically, the objective of variable selection is to assist diagnostics and interpretation, as well as shorten the data processing time (Genuer *et al.,* 2015). The data fusion comprised 12 spectral bands; hence, the FVS method was used to select a set of optimal variables to improve the overall accuracy. Six variables were selected and applied to the fused model using the training data to eliminate model overfitting and used thereafter to classify crops using Random Forest (RF) (Morin et al., 2019; Wang et al., 2016).

3.1.4. The support vector machines algorithm (SVM)

The support vector machine (SVM) is a non-parametric supervised learning algorithm based on spectrally weighted kernel functions (Vapnik, 1995; Vapnik, 2000). The method uses a hyperplane to separate classes, which minimizes misclassifications (Cui et al., 2020).

The SVM consists of four types of kernels different types of kernels. However, only four kernels are commonly used to classify remote sensing data, which are linear, radial basis function (RBF), sigmoid, and polynomial kernels (Zhang *et al.,* 2019). In this study, the SVM classifier was trained using the radial basis function (RBF) kernel in the Equation below:

$$K(u,v) = \exp{-\gamma \times |u-v|^2} \qquad (3.3)$$

where γ denotes the kernel parameter in the RBF kernel.

In the SVM model, two parameters are used to tune the RBF kernel and select the best parameters, which are "C" cost and "gamma" (γ) (Haas and Ban, 2017). The gamma and cost were optimized using 10-fold cross-validation to select the optimal C and γ parameters. The C parameter allows a trade-off between model complexity and training error. The optimal hyperplanes with sensitive spectral information are extracted and used to increase discrimination and classification accuracy (Wang *et al.,* 2019). Using a small C value increases the training errors while the large C value overfits the model (Kranjcˇic´ *et al.,* 2019). The gamma (γ) is defined as a free parameter of the RBF; when the γ value is large, it indicates a small variance, which leads to a high-bias model (Kranjcˇic´ *et al.,* 2019).

The SVM has proven robust on imbalanced samples and is computationally fast (Vluymans, 2019; Maldonado and Barbosa, 2016). The RBF is commonly used in remote sensing existing literature and has been proven to produce high overall accuracy compared to the other kernels, which depend on image resolution (Mashaba-Munghemezulu *et al.,* 2021). The RF-based selected variables were again applied to the SVM S1S2 variable selection model and used to classify crops.

3.4. Accuracy Assessment and Model Validation

Accuracy assessment was used to check how well the locations of the mapped crops correspond to the actual earth surface locations (Olofsson *et al.,* 2014). The method tests whether the sampling design and data analysis methods best fit the data (Olofsson *et al.,* 2014). Further, it validates the classified map based on location-specific data using the 30% test dataset. Previous studies have cited the confusion matrix as a widely used method for assessing the accuracy of thematic maps classified using reference data (Gómez *et al.,* 2016). The confusion matrixes were constructed, and the OOB error was used to identify the actual and misclassified class (Adam *et al.,* 2012). Further, different confusion matrix metrics such as the overall accuracy (OA), user accuracy (UA), producer accuracy (PA) and kappa coefficient (kappa) were computed to aid the explanations through Equations (3.4) and (3.5) (Olofsson *et al.,* 2014; Taghizadeh-Mehrjardi *et al.,* 2020).

$$\left.\begin{aligned} OA &= \frac{TP+TN}{TP+TN+FN+FP} \\ Precision &= \frac{TP}{TP+FP} \\ Recall &= \frac{TP}{TP+FN} \end{aligned}\right\} \qquad (3.4.)$$

$$K - index = 1 - \frac{1 - \mathrm{OA}}{1 - Pe} \qquad (3.5)$$

where *TP*, *TN*, are the samples predicted as true positives and true negatives, while *FP*, *FN* are samples predicted as false positives and false negatives. At the same time, P_e indicates the probability of chance agreement.

3.5. Results

3.5.1. The variable importance

The variables selected as the most optimal for the data fusion model are shown in Figure 3.4a. In order of merit, these are the green (B3:560 nm), red (B4:665), rededge_1 (B5:705 nm), SWIR_1 (B11:1610 nm), SWIR_2 (B12:2190 nm) and the VH polarization. The significance of the SAR and optical bands in classifying classes was evaluated using the RF's variable importance and mean decrease accuracy (MDA) consisting of spikes showing the most optimal bands, as presented in Figure 3.4b. The contribution of each band is highly noticeable when data dimensionality is reduced.

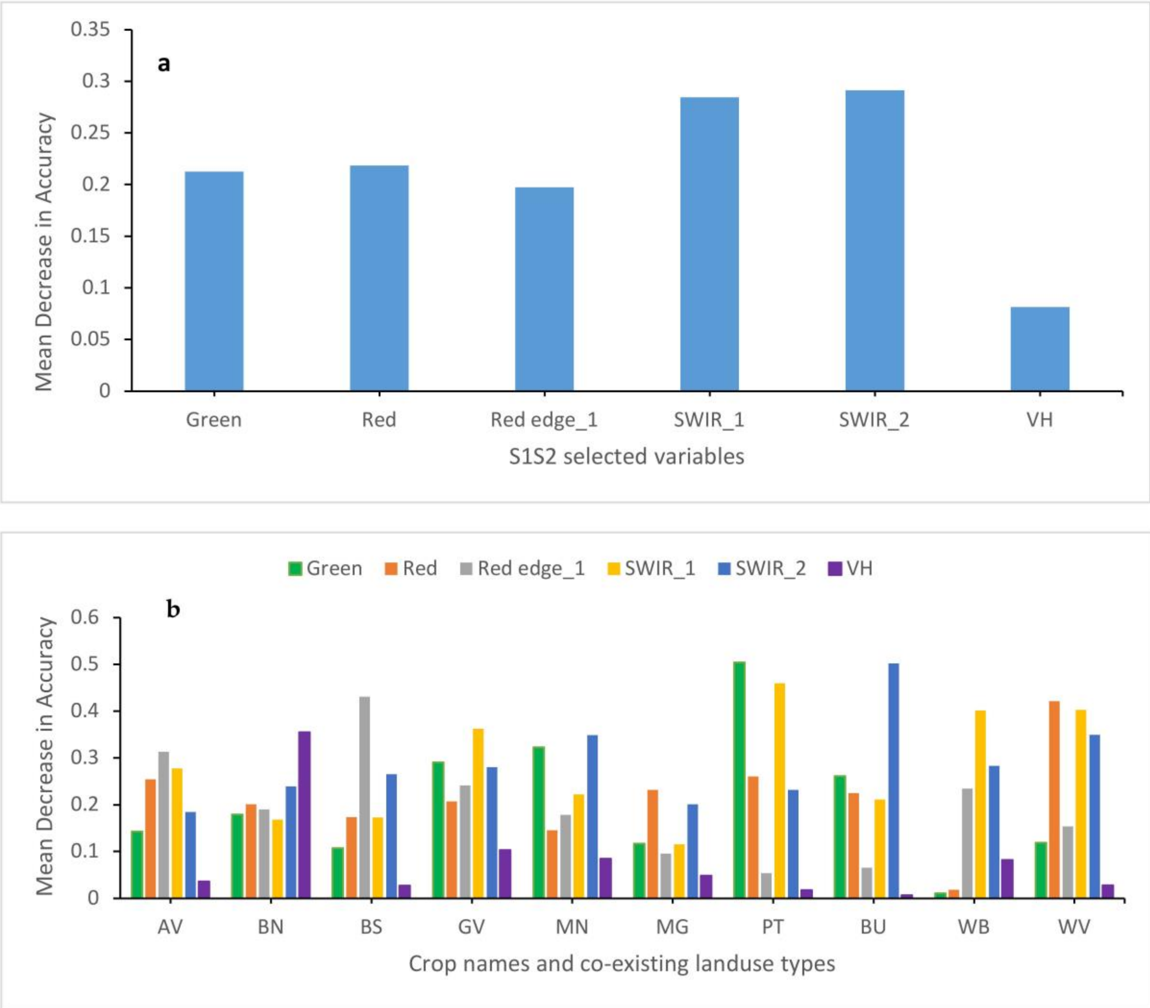


Figure 3.4: *The variable importance in the RF classification model indicates the mean decrease accuracy (MDA) for (a) selected variable models. A high MDA importance indicates that the predictor variable plays a crucial role in the classification. (b) shows the mean decrease accuracy (MDA), indicating the relationship between S1S2 spectral bands and classes.*

The S2 Green band (B3) located in the visible region played a major role in differentiating between built-up areas, pine trees, guava, and mango. The red band (B4) located at 650–680 wavelength played a significant role in distinguishing woody vegetation, guava and avocado, followed by the green band, which detected the pine trees, built-up areas, guava, and mango, better. The SWIR_2 (B12:2190) was imperative in discriminating woody vegetation, pine trees, bare soil, and water bodies. Similarly, the built-up areas, pine trees and bare soil were better identified using the RedEdge_1 (B5:665). The SWIR_1 (B11:1610) played a significant role in distinguishing water bodies, pine trees, guava, avocado, and woody vegetation. For S1, the results highlight a vast contribution of VH polarization to discriminating the banana class.

3.5.2. The variable selection using the forward feature selection (FFS) method

All fused Sentinel-1 and Sentinel-2 bands (*n* = 12) produced an OOB error of 14.29%. A forward feature selection (FFS) was used to select a set of optimal feature combinations that could lead to high mapping accuracy while mitigating high dimension and model over-fitting (Meyer *et al.*, 2018). The FFS was carried out based on mean decrease accuracy (MDA) calculated during the training of the RF model. The MDA measure filters the explanatory variables based on their prediction strength (Kuhn and Johnson, 2013). The results indicate that the lowest OOB error is achievable with six variables, as the error rate increases from seven variables upwards (Figure 3.5).

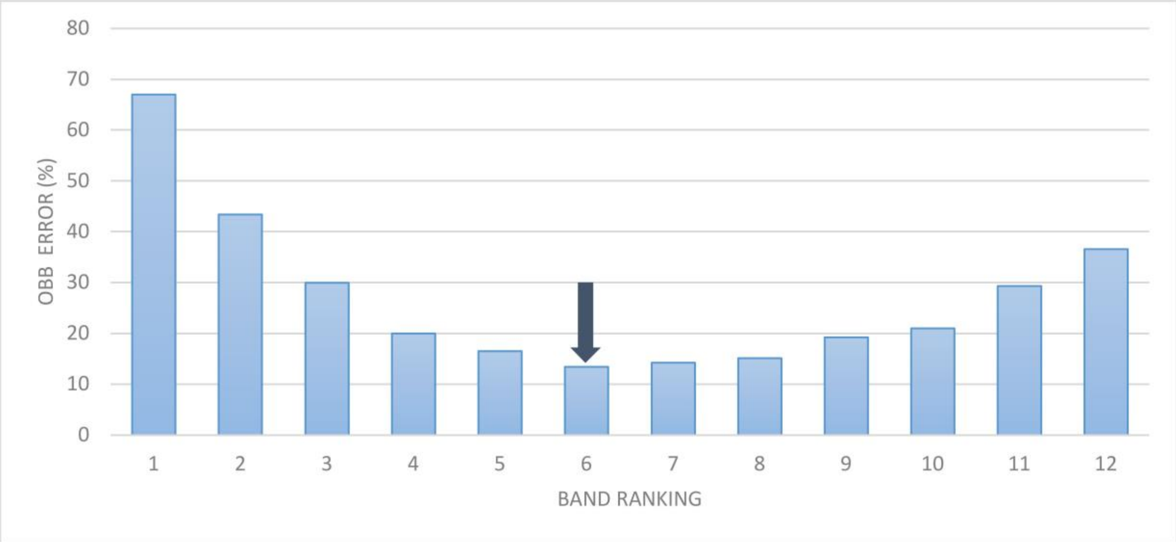


Figure 3.5: *Identification of the optimal variables based on RF MDA rankings produced using the forward variable selection (FVS). The arrow indicates the set of variables that have the lowest error rate.*

3.6. Mapping outputs

The patterns of RF and SVM in mapping tree crop species and other existing land use types based on Sentinel-1 SAR data. Figure 3.6A, B) shows that the woody vegetation is the dominant land use in the radar scenes. The other land use types were not detected accurately, making it difficult to interpret the spatial distribution patterns.

The optical data classification shows a heterogeneous area of bananas mixed with guava. Most of the classes are incorrectly classified by optical scene compared to the radar scenes. The RF map presented in Figure 3.6C, D display the spatial distribution of classes, showing misclassifications between avocado, bare soil and built-up areas. As a result, the bananas were misclassified as guavas and mangoes. The highest spectral confusions were reported between guava, banana, avocado, built-up regions, mango, and macadamia nut trees. Conversely, in Figure 3.6D and Table 3.2B, the SVM map displayed less spectral confusion between mangoes and bananas, macadamia nuts and mangoes.

As depicted in Figure 3.7E, F, Tables 3.1 and 2C, the RF maps show the spectral confusion between avocados and mangoes. The avocado crop overlaps with banana, mango, and macadamia nuts, as shown in Table 3.1. Some mangoes were classified as macadamia nuts, while some macadamia nut trees were classified as guavas. The SVM classifier recorded misclassifications between these classes (Figure 3.7F).

The addition of radar data to the optical data reduced the image speckles. The SVM map displayed spectral confusion for most classes except avocados, mangoes, and bananas (Table 2.2D). The RF map shows less spectral, and a shift of spectral confusion towards the avocado and the guava, as well as the mangoes and the banana class, was noticed in Figures 3.6G and 3.7B. As depicted in Table 3.1, the mango crop overlaps with avocado and macadamia classes. Hence, most of the areas of avocadoes were classified as mangoes, while a few macadamia nut trees were misclassified as guava in Figure 3.6H.

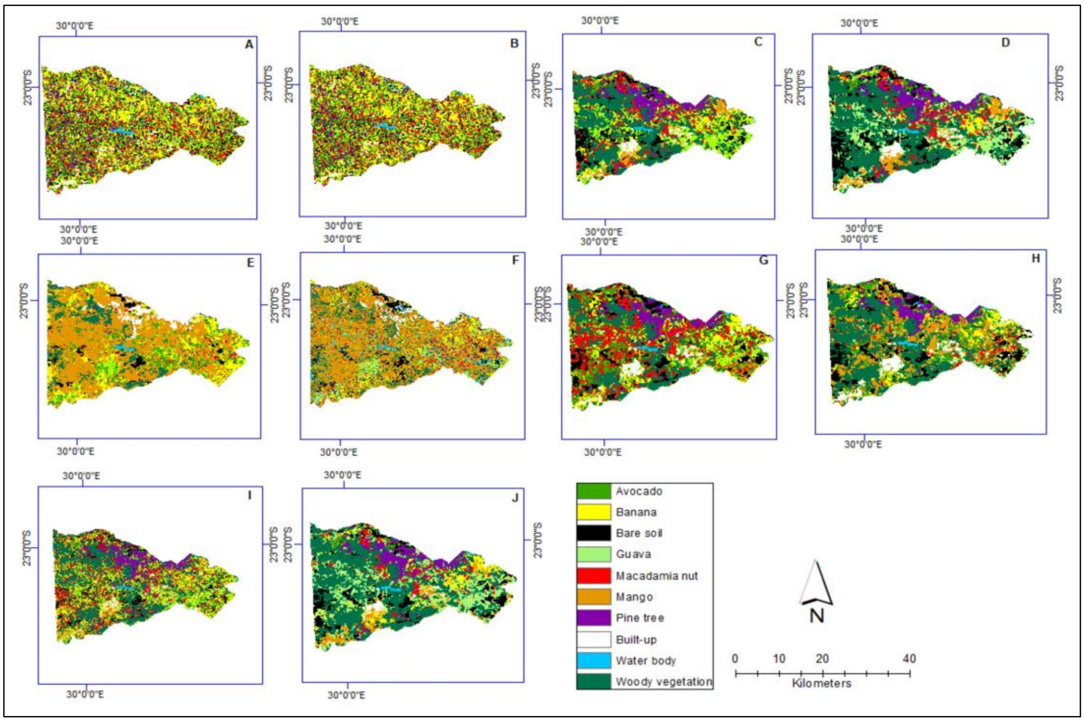


Figure 3.6: *Mapping fruit-tree crops and other existing land use types based on Sentinel-1 data ((A): RF, (B): SVM), Sentinel-2 ((C): RF, (D): SVM), Sentinel-2 selected variables ((E): RF, (F): SVM), Sentinel-1 and Sentinel-2 ((G): RF, (H): SVM), and Sentinel-1 and Sentinel-2 selected variables ((I): RF, (J): SVM).*

Table 3.2: *Confusion matrices for the models that produced the models that produced the highest overall accuracy using RF and SVM algorithms. The mapped classes are avocado (AV), banana (BN), bare soil (BS), guava (GV), macadamia nut (MN), mango (MG), pine tree (PT), built-up (BU), water body (WB), and woody vegetation (WV).*

(A) RF-S1

Class	AV	BN	BS	GV	MN	MG	PT	ST	WB	WV	Total
AV	4	2	1	2	2	2	1	2	0	2	18
BN	2	5	3	2	2	3	0	1	0	1	19
BS	1	1	4	2	1	1	2	1	1	1	15
GV	2	1	3	4	1	2	1	0	0	0	14
MN	4	8	3	4	4	2	2	2	1	1	31
MG	2	1	0	2	2	4	1	1	0	0	13
PT	1	0	0	2	1	2	8	1	0	0	15
BU	2	2	1	0	0	2	1	4	0	1	13
WB	1	0	0	0	0	0	0	2	12	0	15
WV	2	2	1	0	0	1	1	1	2	13	23
Total	21	22	16	18	13	19	17	15	16	19	176

OA: 29.84%

kappa: 0.23%

(B) SVM = S2

Class	AV	BN	BS	GV	MN	MG	PT	ST	WB	WV	Total
AV	16	0	0	0	1	1	0	0	0	0	18
BN	0	16	0	0	1	1	0	0	0	1	19
BS	0	0	14	1	0	0	0	0	0	0	15
GV	1	1	0	11	0	1	0	0	0	0	14
MN	0	1	1	0	28	1	0	0	0	0	31
MG	0	0	0	1	1	11	0	0	0	0	13
PT	0	1	0	0	1	0	13	0	0	0	15
BU	2	0	1	1	0	0	0	10	0	0	13
WB	1	0	0	0	1	0	0	0	13	0	15
WV	0	1	0	0	0	0	0	0	0	22	23
Total	20	20	16	14	33	15	13	10	13	23	176

OA: 87%

kappa: 86.89%

(C) RF-S2 SELECTED

Class	AV	BN	BS	GV	MN	MG	PT	ST	WB	WV	Total
AV	15	0	0	1	1	1	0	0	0	0	18
BN	0	16	0	0	1	1	0	0	0	1	19
BS	0	0	12	2	1	1	0	0	0	0	15
GV	0	1	0	7	5	1	0	0	0	0	14
MN	0	2	2	4	18	5	0	0	0	0	31
MG	0	1	0	0	4	7	1	0	0	0	13
PT	0	1	0	0	2	1	9	1	0	0	15
BU	2	0	0	0	0	1	1	9	0	0	13
WB	0	0	0	0	0	0	0	3	12	0	15
WV	0	1	0	1	0	4	0	0	0	17	23
Total	17	22	14	15	32	22	11	13	12	18	176

OA = 69.33%

Kappa = 69%

(D) SVM-S1S2

Class	AV	BN	BS	GV	MN	MG	PT	ST	WB	WV	Total
AV	16	0	0	0	1	1	0	0	0	0	18
BN	0	17	0	0	0	1	0	0	0	1	19
BS	0	0	13	0	1	1	0	0	0	0	15
GV	1	1	0	11	0	1	0	0	0	0	14

MN	0	0	1	1	28	1	0	0	0	0	31
MG	0	0	0	2	1	10	0	0	0	0	13
PT	0	1	0	0	1	1	12	0	0	0	15
BU	1	0	0	0	0	0	0	12	0	0	13
WB	1	0	0	0	1	0	0	0	13	0	15
WV	0	1	0	0	0	0	0	0	0	22	23
Total	19	20	14	14	33	16	12	12	13	23	176
OA = 87.01%											
Kappa = 0.87											

(E) SVM-S1S2 SELECTED											
Class	AV	BN	BS	GV	MN	MG	PT	ST	WB	WV	Total
AV	17	0	0	0	0	1	0	0	0	0	18
BN	0	17	0	0	0	1	0	0	0	1	19
BS	0	0	13	0	1	1	0	0	0	0	15
GV	1	1	0	11	0	1	0	0	0	0	14
MN	0	0	0	1	29	1	0	0	0	0	31
MG	0	0	0	0	1	12	0	0	0	0	13
PT	0	1	0	0	1	0	13	0	0	0	15
BU	1	0	0	0	0	0	0	12	0	0	13
WB	0	0	0	0	0	0	0	0	15	0	15
WV	0	1	0	0	0	0	0	0	0	22	23
Total	19	20	13	12	32	17	13	12	15	23	176
OA = 91.63%											
Kappa = 0.91%											

Accurate class separation was achieved based on selected six optical variables using 1 SAR and 1 Optical scene. There is a reduction in class confusion in macadamia nut and avocado areas (Table 2.2), while a noticeable amount of mango pixels was still misclassified (Figure 3.6G**),** as well as some avocados being misclassified as guavas. The best classification model was based on the SVM algorithm (Figure 3.6H). The area covered by the map with the highest classification accuracy shows the woody vegetation as the largest class, covering 36.3% of the south, northeast, and past southeast of the study area. The Guava crop is the second largest, covering 16.80% and dominates in the eastern part of the study area compared to a water body and built-up, covering 0.8% and 2.6%, respectively.

3.7. Classification accuracy

In this study, the classification accuracy was assessed using testing data. Five different models comprising different variables, namely S1 (n = 2) (VV, VH); S2 (n = 10) (B3: Blue, B3: green, B4: red, B5: red edge_1, B6: red edge_2, B7: red edge_3, B8: NIR, B8A: Narrow NIR, B11: SWIR_1, B12: SWIR_2); S2 selected variables (n = 6)(B2: blue, B3:green, B4: red, B5: red edge_1, B11:SWIR_1S1S2, SWIR_2:B12); S1S2 (n = 12) (B2: blue, B3: green, B4: red, B5: red edge_1, B6: red edge_2, B7: red edge_3, B8: NIR, B8A: Narrow NIR, B11: SWIR_1, B12: SWIR_2, VV, VH), and S1S2 selected variables (n = 06) (B3: green, B4: red, B5: Red Rdge_1, B11: SWIR_1, B12: SWIR_2, VH), were used and classified using RF and SVM. The S1S2 selected variable model using the SVM classifier obtained the highest classification accuracy compared to the RF classifier for the other models (S1, S2, S2 selected variables and S1S2), as shown in Figure 3.8. The S1 produced an overall classification accuracy of 28.04% and 27.64%, using the RF and SVM model. The S2 model had an overall classification accuracy of 81.96% and 87% for RF and SVM, while the S2 selected variables produced an overall accuracy of 70,04% and 69.33% using the RF and SVM, respectively. A steady accuracy was observed from the data fusion approach, wherein an overall classification of 81.93% and 87.01% was achieved. The combination of selected optimal variables produced an increased accuracy of 86.50 and 91.6% for RF and SVM, respectively, amounting to a 4% improvement over S2 data. A bar plot showing the highest classification accuracies from both models is presented in Figure 3.5.

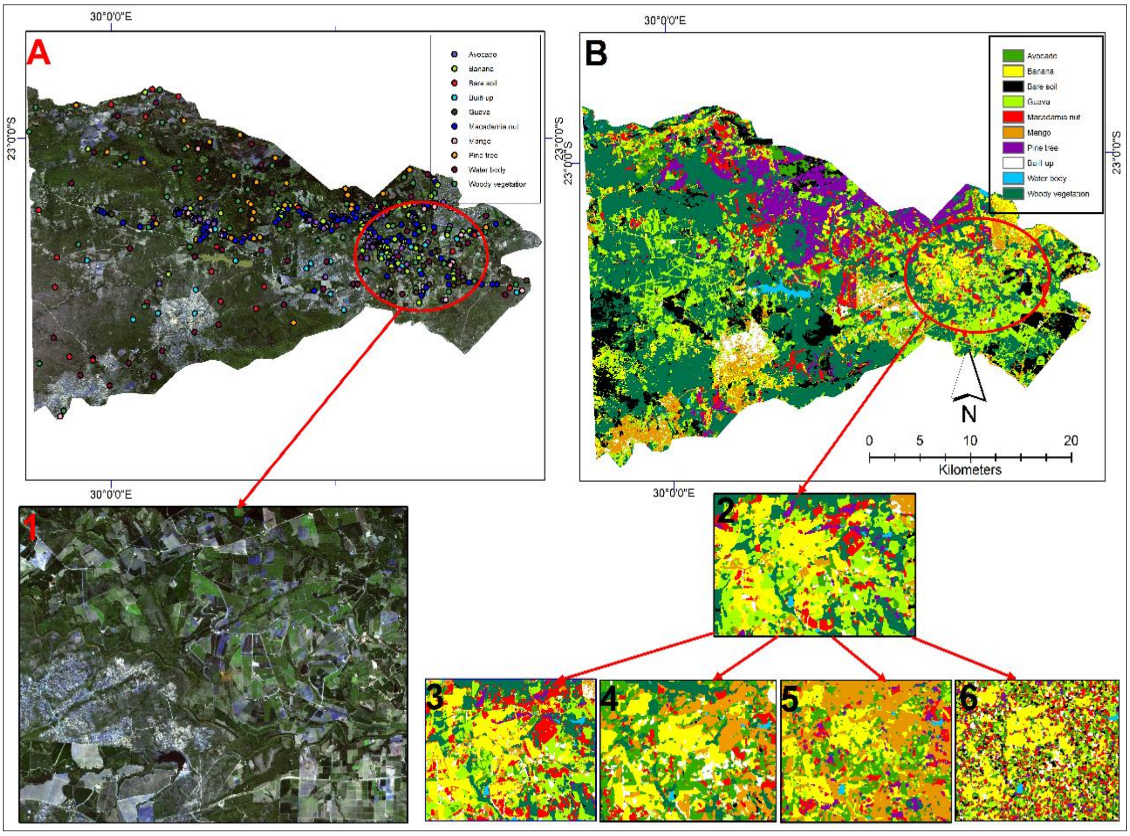


Figure 3.7: *The spatial distribution of fruit-tree crops and co-existing land use types for the Levubu subtropical farming area using S1S2 selected variables and SVM (B). (A) and zoom-in in (1) shows the unclassified Sentinel-2 RGB image, including the collected ground truth data. (2–6) show the classified zoom-ins of the same area using different models. In order of merit, these are S1S2 selected variables, S1S2, S2, S2 selected variables and S1.*

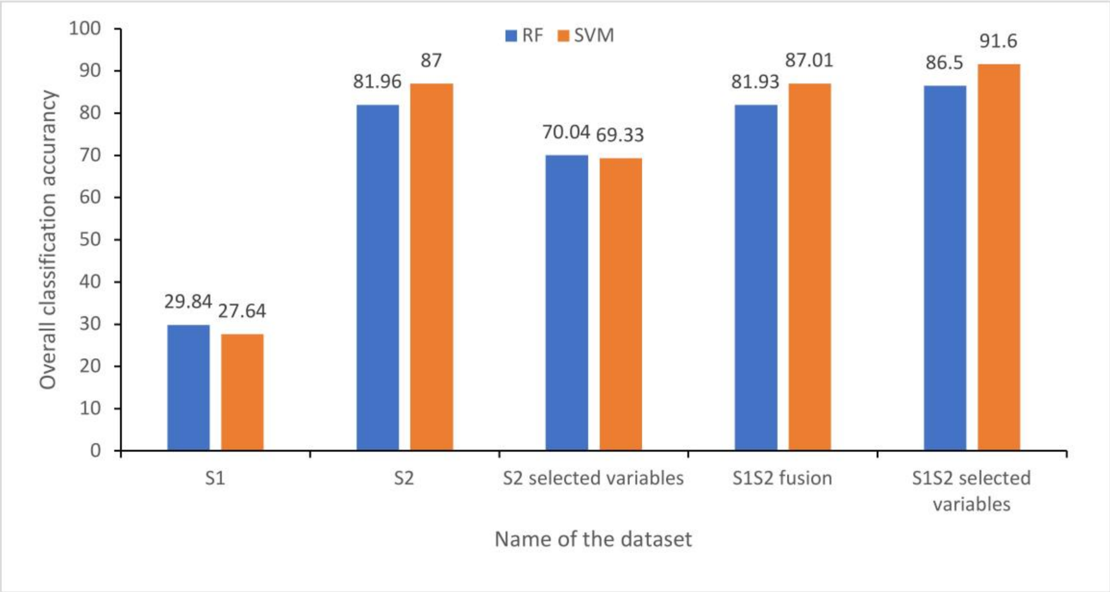


Figure 3.7: *The overall accuracy (OA) obtained by RF (blue) and SVM (orange) classifiers using Sentinel1, Sentinel-2, Sentinel-2 selected variables, S1S2 fusion, and S1S2 selected variables in mapping tree crop species and other existing land uses.*

3.8. Class accuracies

The classification accuracies of the four datasets at a class-wise level were evaluated using the test dataset. The user's accuracy (UA) and producer's accuracy (PA) in Table 3.3 show the overall accuracies obtained by each model using RF and SVM. The producer's and user's accuracies give crucial insight into how each model contributed to detecting classes.

Table 3.3: *The performance of RF and SVM over four datasets (S1, S2, S2 selected variables, S1S2, S1S2 selected variables importance (VI) in terms of user's and producer's accuracies in percentages (%). The mapped classes are avocado (AV), banana (BN), bare soil (BS), guava (GV), macadamia nut (MN), mango (MG), pine tree (PT), built-up (BU), water body (WB), and woody vegetation (WV).*

	(a) RF User's Accuracy (%)					(b) RF Producer's accuracy (%)				
Class	**S1**	**S2**	**S2 VI**	**S1S2**	**S1S2 VI**	**S1**	**S2**	**S2 VI**	**S1S2**	**S1S2 VI**
AV	18.57	87.32	44.16	71.08	93.75	19.12	77.50	49.28	73.75	75.00
BN	40.00	75.71	62.5	78.75	88.06	28.57	76.81	67.90	91.30	85.51
BS	18.18	84.48	82.35	89.09	90.74	17.39	89.09	91.30	89.09	89.06
GV	19.05	80.00	60.87	86.59	79.79	24.00	81.93	66.67	84.52	90.36
MN	15.28	75.58	50.00	74.67	85.91	16.92	79.27	45.83	68.29	86.84
MG	23.73	74.19	80.00	65.96	66.67	29.17	47.92	61.54	64.58	75.56
PT	27.69	92.06	50.00	96.72	93.75	35.29	95.08	84.31	96.72	98.36
BU	34.88	79.22	82.14	93.33	90.63	27.78	89.71	82.14	82.35	86.57
WB	93.88	98.11	96.00	91.53	96.30	86.79	95.63	96.00	100.00	96.30
WV	100.00	99.06	100.00	100.00	100.00	17.65	93.15	100.00	87.67	94.52
	(a) SVM User's accuracy (%)					(b) SVM Producer's accuracy (%)				
Class	**S1**	**S2**	**S2 VI**	**S1S2**	**S1S2 VI**	**S1**	**S2**	**S2 VI**	**S1S2**	**S1S2 VI**
AV	17.74	38.71	44.87	83.75	94.67	32.35	39.71	50.72	83.75	94.67
BN	38.96	67.65	64.71	85.14	88.73	38.96	59.74	67.90	91.30	88.33
BS	23.08	84.91	87.76	94.64	94.55	13.04	96.53	93.48	93.36	90.00
GV	23.89	50.67	84.31	93.6	87.78	50.00	76.00	76.79	91.67	96.3
MN	14.58	52.63	55.10	72.94	94.59	11.48	45.45	64.29	75.61	95.18
MG	0.00	57.58	52.00	80.00	85.11	0.00	39.58	54.17	66.67	92.11
PT	20.83	85.11	74.07	95.08	93.00	29.41	78.43	76.79	95.08	88.89
BU	29.09	74.24	91.11	93.85	95.00	15.69	90.74	80.39	89.71	98.36
WB	34.61	100.00	94.12	94.74	97.36	34.91	97.26	88.89	94.52	85.07
WV	0.00	100.00	100.00	100.00	100.00	0.00	79.33	74.07	100.00	98.63

The produced UA's and PA's accuracies ranged from 17.74 to 100%. The sentinel-1 SAR data discriminated the water bodies and woody vegetation with an overall accuracy of 80% and above, while the other crops were below 40% for both classifiers. Optical data (S2) increased the UA's and PA's accuracies to above 70% for all classes using the RF. Conversely, the UA's and PA's accuracy scores were below 60% for avocado, guava, pine tree and mango using the SVM classifier. However, a decrease in UA's and PA's scores was recorded for the mango and pine tree classes. The model S2 selected variables recorded an increase of 5.81% in the UA for the mango class. The UA values for the built-up and woody vegetation classes increased by 2.98% and 0.40%. The PA values of the bare soil, macadamia nut, mango and water body increased by 2.21%, 12.41%,1 7.85% and 4%, respectively.

Using the selected optical variables from data fusion, the mango class recorded a 20–30% accuracy increase compared to the other models. Though the SVM proved superior to the RF model, the individual class accuracies for avocado, guava, mango, and macadamia nut are comparable. The RF improved the producer and user accuracy margins of areas identified as guava, pine tree, built-up, water bodies and woody vegetation by over 80%, including the mango class, which was challenging to detect in other models.

3.9. Discussion

The study evaluated the utility of using RF and SVM models using Sentinel-1 (SAR) and Sentinel-2 Optical, their synergy to map fruit-tree crop species and co-existing land-use types in a heterogeneous smallholder landscape. The study further tested the utility of feature selection by investigating the relevance of the Sentine-1 and Sentinel-2 bands by calculating the variable importance using an RF algorithm to eliminate data dimensionality and increase classification accuracy (Loggenberg *et al.*, 2018). The free availability of SAR and multispectral data ESA Copernicus with suitable spectral, temporal, and spatial resolution similar to SPOT-5 provides promising opportunities for mapping smallholder agriculture at the landscape level (Mandanici *et al*., 2016).

The results showed that Sentinel-1 data could not be used alone, as it could not differentiate between the studied fruit-tree crops and other existing land uses due to a limited number of channels (Zeyada *et al.*, 2016). As expected, the classification based on the SAR image produced an overall accuracy of 28.04% using RF and 27.64% using the SVM, respectively. Similar to the findings by Brinkhoff *et al*. (2020), the VH polarization was optimal in distinguishing the fruit tree crops due to its sensitivity to the structure and geometrical features of crops. The application of Sentinel-1 data alone resulted in false cropping patterns using both classifiers (RF and SVM). The findings concur with previous studies that mapping tree crops using single date images produced unsatisfactory results due to similarities in physiological and morphological characteristics that commonly exist in fruit-tree species

(Conrad *et al.,* 2014) Preidl *et al.,* 2020). It is even more challenging to obtain precise information when mapping crops at the landscape level (Cui *et al.*, 2020). The reason why the Sentinel-1 SAR data report confused cropping patterns on all land cover classes, except the banana crop, is yet to be studied.

The Sentinel-2 data performed well in discriminating crops since they consist of spectral information, which is crucial in crop discrimination (Sun *et al.,* 2019). In contrast, the NIR is less critical to all models in discriminating tree crops, as Heckel *et al*. (2020) reported. The NIR band was noisy and less beneficial in crop discrimination (Zhao *et al.,* 2020). The individual class accuracy results showed some difficulties on the Sentinel optical data in discriminating banana, guava, and mango within built-up areas due to spectral overlapping, suggesting a high level of crop dependence on RS data spatial resolution to be classified accurately. Hence, the success of crop mapping relies on the careful selection of remote sensing data with appropriate spectral and spatial characteristics to improve classification accuracy (Conrad *et al*., 2014).

The use of 12 spectral bands yielded less accuracy than the selected six optimal bands due to noise contamination (Conrad *et al.,* 2019). Similar to the existing literature on mapping fruit-tree parameters, the highest classification accuracy of 91.60% was achieved from the combination of selected variables from SAR and optical data, much better than results derived from high-resolution SAR and optical images (Steinhausen *et al.,* 2018; Sun *et al.,* 2019; Clerici *et al.,* 2017). Through the use of the SVM model, the classification accuracy improved by 4%, showcasing the importance of using SAR-optical fusion in areas where landscape heterogeneity and climatic conditions challenge remote sensing data acquisition (Preidl *et al.,* 2020). Data fusion is a promising approach in remote sensing that provides smallholder farmers with spatially explicit information while offering practical opportunities to achieve precision agriculture at the landscape level.

As reported in previous studies (Mercier *et al.,* 2020; Tavares *et al.,* 2019), including Sentinel-1 backscatter, improved crop separability and overall classification accuracy due to its sensitivity to biomass and tree canopies. The class accuracy was improved for crops and other land-use types when radar data were integrated with optical data. The effect of variable selection for enhancing class accuracy was high with regard to mango crops, where a 20–30% increase was recorded, which agrees with the findings by Akbari *et al.* (2020). The RF improved the producer and user accuracy margins of areas identified as guava, pine tree, built-up, water bodies and woody vegetation by over 80%, including the mango class, which was challenging to detect in other models. Heterogeneous classes such as avocado, guava, mango, and macadamia nut were distinguished correctly from the other crops using the variable selection. Although the SVM proved superior to the RF model in terms of overall accuracy, the implemented RF model using the selected variables was robust and informative.

The class-specific variable selection showed that the green band (B3:560 nm) was the most influential spectral band in discriminating pine trees, mango, and guava. The Sentinel1 VH band was effective in detecting the banana crop, which corresponds with previous studies (Talema and Hailu, 2020). Sentinel-2 green band (B3: 560 nm) was crucial for mapping pine trees, mango, and guava. The red band was essential for detecting avocado and woody vegetation, while the red Edge_1 (B5: 705 nm) distinguished avocado and bare soil. The SWIR bands were critical in discriminating built-up areas, banana, mango, guava, pine tree, water bodies and woody vegetation. The model detected all crops well, including the small patches of mango and bananas hidden within built-up areas. Similar to Mercier *et al.* (2020), the data fusion model was a better discriminator for mango patches within built-up regions that could not be detected using S1 and S2 alone. The visible bands (green (B3:560 nm) and red (B4: 665 nm), SWIR_2(B12: 2.190 nm), SWIR_1 (B11: 1.610 nm) and the rededge_1 (B5: 705 nm) were the most sensitive to chlorophyll content and leaf structure and contributed highly in distinguishing crops (Punalekar *et al.*, 2018). Using selected pivotal bands from the data fusion produced less spectral confusion and was suitable for mapping fruit-tree crops in a heterogeneous environment. The criticality of the SWIR bands in fruit-tree crop mapping was highlighted (Orynbaikyzy *et al.*, 2019). The maps depicting tree crops and co-existing land-use types show that the integrated approach discriminates subtle areas, which was not the case in optical and SAR models alone (Muro *et al.*, 2016). The selection of optimal variables proved to be a well-suited mapping approach for the study area, which agrees with the findings of Akbari *et al*. (2020) and Kyere *et al*. (2020)

The comparison of Sentinel-2 and data fusion showed good agreement in this study. As Haas and Ban (2017) reported, including the SAR data, particularly the VH polarization, increased the detection of smaller crop aggregations, while the optical data overestimated the crop distribution in the study area. The Sentinel-2 method missed some small patches of mango and guava crops (Figure 3.6C, 3.6D). The data fusion model accurately captured the cropping dynamics in heterogeneous smallholder agriculture. The visual assessment of the final maps generated using the selected optimal variables showed enhanced interpretability (Orynbaikyzy *et al.*, 2019) on SVM products while some noise in homogenous maps produced by the RF algorithm, similar to (Tatsumi *et al.*, 2015) findings. The data distinguished crops better due to the swath width and high spatial resolution, which account for crop patterns in small field sizes for heterogeneous areas. The noise depicted on maps produced using each dataset separately was reduced by the fused model. The radar data is sensitive to vegetation structure and texture, and it contributed much to discriminating banana, mango, and guava found within build-up areas, which was difficult to determine using single-date images.

The contribution of SAR and optical data in crop mapping of smallholder agriculture in an African setting was ascertained. Further, the potential of S1S2 spectral discriminating power was analyzed. Although the results were not significantly increased, the model stability on fused data increased, demonstrating data fusion's potential to enhance smallholder crop mapping. The combination of backscatter from Sentinel-1 (SAR) at C-band and spectral features from Sentinel-2 (optical) have proven to be a better approach to discovering complex distribution patterns of crops, including classes that are difficult to map using a single image. Integrating S1S2 provided spectral and textural information to better detect different tree crops and co-existing land-use types in a heterogeneous landscape (Clerici *et al.*, 2017). To ascertain this, we further investigated the effect of variable selection on S2 data. The model produced an overall accuracy of 70.04% and 69.33% using RF and SVM, respectively. Although the levels of class accuracies were low on the S1S2 model compared to the S2 model, the results of this study proved that incorporating the SAR data enhanced the discrimination of the crop types. A clear trade-off was observed between S2 and data fusion, where the highest classification results were obtained using data fusion information. Due to the Sentinel-2 low geometric resolution, the crops from small farm parcels that belong to the same family were not accurately classified (Conrad *et al*., 2014).

The approach provides opportunities to map sub-tropical smallholder agriculture with mixed cropping systems for field management, precision agriculture, and decision-making by agronomists and the government. However, the method was time-consuming and required advanced remote sensing and GIS skills and a high-processing computer to accomplish the task. The study was limited to a single date image from each Sentinel platform to simplify the mapping process. The launch of Sentinel-1B with five days' temporal resolution allows the application of SAR data in mapping heterogeneous landscapes. Although Sentinel-1 failed to detect crop types, it is still helpful to integrate its data with optical data as it offers opportunities to enhance distribution patterns in heterogeneous landscapes.

Therefore, the study recommends testing this model on multi-time series data to improve crop detection using crop temporal behaviours from different phenological stages. Due to various cropping calendars, it is advisable to use multi-time series images to assess different vegetative stages and identify the optimum window relevant for accurate crop mapping for precision agriculture (PA) (Zhao *et al.*, 2020). Further, the study recommends selecting a few optimal polarimetric images to reduce high data dimensionality and increase classification accuracy. Discriminating all crops with high accuracy, including minor crops, is essential. Hence, the study suggests an image-fusing approach based on time-series data to identify temporal windows crucial for accurate crop mapping.

3.10. Conclusions

Overall, the use of single-source data (S1) proved unsuitable for mapping crops due to the lack of spectral characteristics. Although the optical data (S2) performed better than S1, it could not discriminate crops with spectral similarities (i.e., mango, macadamia, pine tree). Combining SAR (S1) and optical (S2) data into machine learning classifiers can be used to map fruit-tree crops in a sub-tropical heterogeneous landscape experiencing frequent cloud coverage. The results show that the SVM classifier provides the best classification results for fruit-tree crops and co-existing land use types by utilizing six key variables, of which five were S2 bands (i.e., green (B3), red (B5), rededge_1 (B5), SWIR_1 (B11), SWIR_2 (B12)) and one S1 channel (the VH polarization). The multi-source approach provides timely information for evidence-based decision-making.

CHAPTER FOUR

Identifying the optimal phenological period for discriminating fruit tree crops using multi-temporal Sentinel-2 data and Google Earth Engine in a heterogeneous subtropical horticultural region

This chapter is based on:

Chabalala, Y.; Adam, E and Ali, K.A. (2023). Identifying the optimal phenological period for discriminating fruit tree crops using multi-temporal Sentinel-2 data and Google Earth Engine in a heterogeneous subtropical horticultural region. *South African Journal of Geomatics*,12 262-283. DOI: https://doi.org/10.4314/sajg.v12i.2.10

Abstract

The accurate and appropriate monitoring of the spatial distribution of fruit tree crops is crucial for crop management and yield forecasting. Owing to inter- and intra-farm fragmentation and overlapping phenological cycles, classifying fruit tree crops in subtropical agriculture using single-date images is challenging. Therefore, this research aimed to identify the optimal temporal window in which the crucial phenological stages can be used to classify fruit tree crops in Levubu, Limpopo province, using a random forest (RF) classifier. Phenological metrics were extracted from 12-month Multispectral Instrument (MSI) images from Sentinel-2 (S2). The RF classification algorithm attained an overall accuracy of 84.89% and a kappa coefficient of 83%. The user accuracy ranged from 62 to 100%, while the producer accuracy ranged from 60 to 100%. An analysis of variance was used to assess whether the overall accuracies among the S2 monthly composites were statistically significant. The results showed distinct spectral differences between fruit trees. In April, classification differences were observed during the harvesting and senescence of the mango and macadamia nut crops. In May, differences were observed during the senescence stage of the macadamia nut, mango, and guava crops. In June and July, there were distinct spectral differences during the peak flowering stage of the avocado, macadamia nut, and mango crops, as well as in the fruiting stage of the banana crops. Followed by the red-edge bands, the shortwave infrared bands were significant in differentiating between the respective fruit tree crops. The results of this research provide evidence-based information that can assist farm managers and horticulturists in making informed decisions. This is critical in achieving effective agricultural management and in ensuring the sustainability of local horticultural systems.

4.1. Introduction

Mapping crop types is an important activity that provides evidence-based information for guiding agricultural practices and policies for food security (Jin *et al.,* 2019). Accurate and timely maps showing the spatial distribution of crop types provide valuable information for monitoring crops, estimating yields, scheduling agronomic practices, and detecting crop anomalies (Remelgado *et al.,* 2020). Moreover, crop type maps can determine the ecological suitability for extending the existing crop production area and for increasing the volume of harvested products (Singh *et al.,* 2017). Crop type mapping using remote sensing is an efficient method for assessing agricultural production. Satellite sensors such as the Sentinel-2 (S2) Multispectral Imager (MSI) and the Landsat 8 Operational Land Imager (OLI) have medium and high spatial and temporal resolutions that provide information suitable for crop modelling (Neigh *et al.,* 2018). Owing to the spectral similarity among specific tree species during their phenological stage, the mapping of crop types using single-date images may not be sufficient (Mercier *et al.,* 2020). Fruit tree species are the only crop types that have a stable canopy cover and a similar leaf structure (Asgarian *et al.,* 2016). This results in overlapping spectral signatures, which impair the accuracy of the classification process (Neigh *et al.,* 2018; Nabil *et al.,* 2022). For example, Chabalala *et al.* (2022) applied a random forest (RF) classifier to an S2 single-date image and reported misclassifications of mango and avocado crops which were attributed to overlapping spectral reflectance. Therefore, multi-temporal data based on key crop growth and phenology stages can serve as a promising approach to spectrally distinguish between tree species and improve the accuracy in classifying tree species (Peña-Barragán *et al.,* 2011).

Several studies have used multi-temporal remote sensing information and different classifiers to predict vegetation greenness across key crop growth stages in various agricultural landscapes (Luo *et al.,* 2021; Shelestov *et al.,* 2017). Nabil *et al.* (2022) applied an RF classifier to S2 time series data to map fruit tree crops in Egypt and obtained an overall classification accuracy of 96%, while Busetto *et al.* (2019) analysed temporal changes in rice crops in Senegal (West Africa) using Moderate Resolution Imaging Spectroradiometer (MODIS) time series data and the PhenoRice algorithm. The latter researchers obtained an overall accuracy of 81% and 65% for dry and wet seasons, respectively. Pena *et al.* (2017) used Landsat-8 time series data to classify fruit tree crops in Chile, and accurate results were achieved using linear discriminant analysis during the crop green-up stage. Zurita-Milla *et al.* (2017) obtained an overall accuracy of 80% when identifying crop types in smallholder agriculture in sub-Saharan Africa using Worldview-2 time series data and an RF classifier. Similarly, Zhang *et al.* (2018)

improved crop-type classification accuracy in Ontario, Canada, using an RF classifier and Satellite pour l'Observation de la Terre (SPOT-5) time-series data. However, most of these studies were conducted at regional and country levels (Kumari *et al.,* 2021; Wang *et al*., 2021; Kraaijvanger *et al.,* 2019). Becau**se** of the challenges regarding the suitability and consistency of these mapping approaches, they cannot be easily replicated in different contexts, such as in highly fragmented African agricultural environments (You and Dong, 2020).

Farming systems have site-specific characteristics that influence the spatial and spectral information of crops—this determines how different crop types can be accurately distinguished from one another. Studies such as Luo *et al.* (2021), Shelestov *et al. (*2017), and You and Dong (2020) used remote sensing data at Google Earth Engine (GEE) to map crop types. However, these studies were conducted in areas such as China and Ukraine, which have site-specific characteristics that differ from those of the Levubu subtropical farms. Subtropical agriculture in Levubu is characterised by highly fragmented farms growing clonal fruit trees, with minimal seasonal growth dynamics associated with overlapping phenological cycles (Neigh *et al.,* 2018).

The mapping of fruit trees in smallholder agriculture is complex, owing to agronomic factors such as wide-ranging cropping patterns, numerous crop varieties, different crop-growth calendars, and high crop densities (Nabil *et al.,* 2022; Wang *et al.,* 2021). Smallholder agriculture is characterised by diverse and patchy fragmented landscapes, a variety of management practices, and varying agro-climatic conditions (Potgieter *et al.,* 2021). The spectral behaviour of fruit tree crops in such landscapes presents challenges in crop observation when optical sensors with coarse or medium spatial resolutions are used. These images may not correspond with the ideal period optimal for inter-class separation (Pena *et al.,* 2017; Shelestov *et al.,* 2017; Lamour *et al.,* 2019).

Identifying optimal growth periods and the phenological traits of fruit-tree crops using high temporal resolution remote sensing data can be applied to develop a robust method for identifying fruit-tree species (Mercier *et al.,* 2020). Copernicus (European Space Agency) offers optical constellation missions, such as S2, with high spatial and temporal resolutions suitable for accurately identifying crop types (Lamour *et al.,* 2019; Gao *et al.,* 2020). The mission provides opportunities for smallholder farmers to adopt sustainable farming technologies by accelerating smart and effective crop mapping (Shelestov *et al.,* 2017). Furthermore, it facilitates the development of operational mapping tools that will encourage the continued use of earth observation data in monitoring crop types in smallholder landscapes in Africa at the local, regional, and national levels (Arias *et al.*, 2020; Veloso *et al.*, 2017). However, the effective use of S2 time series data for the detection of crop types is still limited, especially in smallholder subtropical agriculture (Veloso *et al*., 2017).

The objective of this research was to identify the optimal phenological window in which fruit trees can be effectively classified in a highly heterogeneous horticultural landscape in Levubu, Limpopo province, South Africa, using phenological metrics derived from multi-temporal S2 data and an RF classifier. This method provided evidence-based information that would enable efficient horticultural monitoring systems and sustainable agricultural management (Usha and Singh, 2013).

4.2. Materials and Methods

4.2.1. Research area

Levubu is in Limpopo Province, northeastern South Africa (Figure 4.1), and the fruit-farming area covers *c*. 10,000 hectares (Fraser, 2008). The elevation varies from 677 to 2221 m, with extensive flat and fertile areas in the plains to the southeast of the Makhado local municipal area. The sub-tropical regions have temperatures ranging from 16 to 40°C during summer (i.e., from December to February) and 12 to 22°C in winter (i.e., from May to August). The area has an average annual rainfall of 340 mm. The research area has diverse agricultural landscapes and agro-meteorological conditions that influence the performance and management of fruit tree crops. The fruit trees commonly planted in the area are avocado, banana, guava, mango, and macadamia nut.

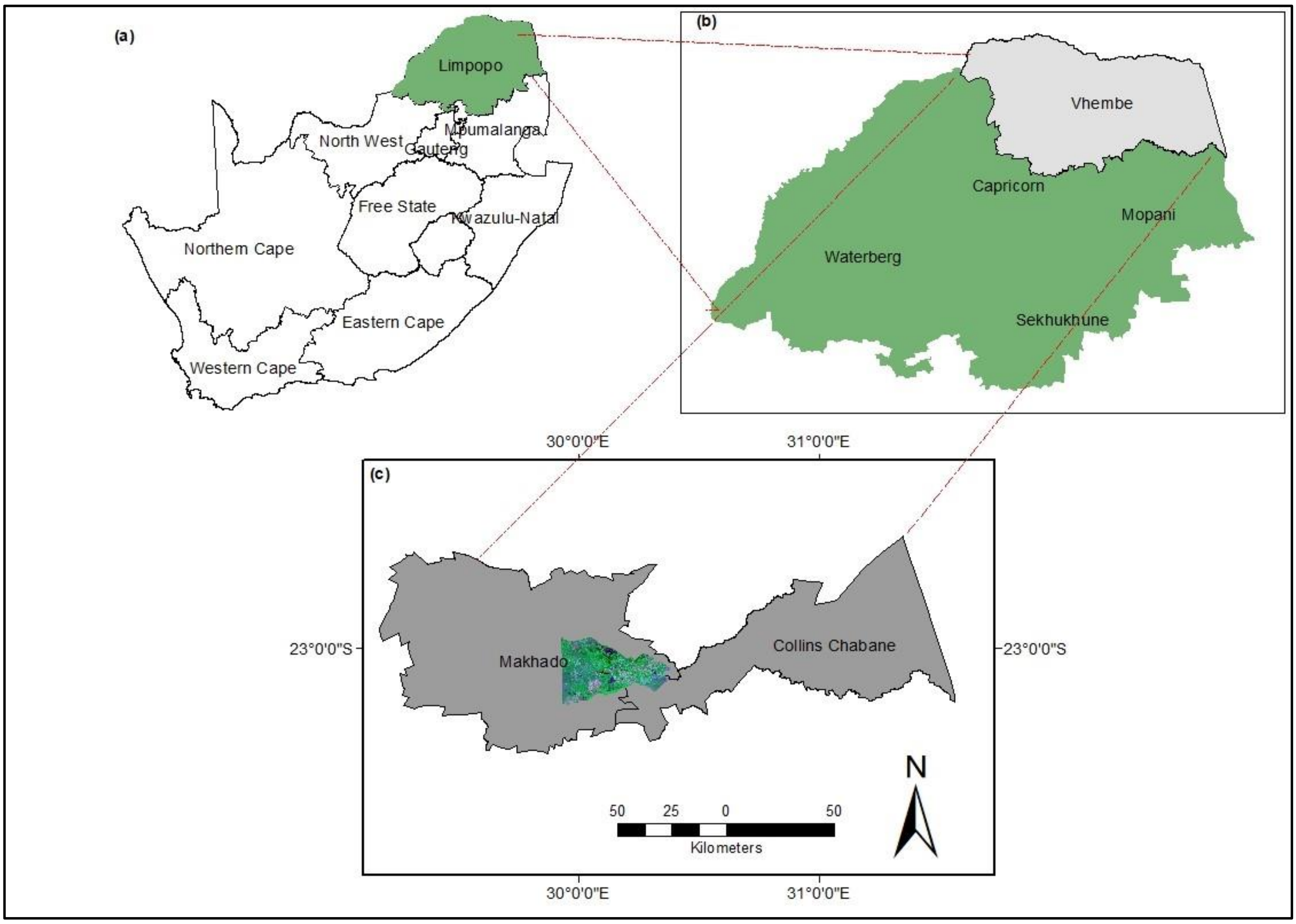


Figure 4.1: *Geographic location of Levubu subtropical farming area in Makhado and Collins Chabane Local Municipalities, Limpopo Province, South Africa.*

4.3. Earth observation data and pre-processing

4.3.1. Sentinel-2 data and pre-processing

This study used Copernicus/S2 SR ImageCollection, and all available data from June 2019 to May 2020 were sourced. The Copernicus/S2 SR is a Level-2A Surface Reflectance ImageCollection that is atmospherically corrected and made freely available for public consumption on GEE (https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR) (European Space Agency, 2023). The ImageCollection has 11 bands, namely, four visible bands (B1, B2, B3, B4)), three red-edge bands (B5, B6, B7), two near-infrared bands (B8, B8A), and two shortwave bands (B11, B12) (European Space Agency, 2023). Integrating spectral bands and vegetation indices allows for a means to determine redundancy and correlation, which in turn affects the classification accuracy (Peña-barragán *et al.*, 2011). Hence, the research used only the spectral data acquired at 10 and 20m (i.e., blue (B2), green (B3), red (B4), red-edge (B5, B6, B7, and B8A), near-infrared (B8), and shortwave-infrared (B11 and B12) bands). Fewer images were available for the summer months owing to the dense cloud cover during that season. Table 4.1 shows the spatial and spectral properties of the Sentinel-2 (S2) data. The optimal temporal window for accurate tree crop and co-existing land cover mapping was identified using a median composite method, which synthesises the pixels of the available time series images (Luo *et al.,* 2021). Owing to the variation in cloud cover, the monthly composites included different numbers of images for each month. They were created using all images acquired with a maximum filter of five percent (5%) of cloud cover each month.

Table 4.1: *Spectral configurations for the S2 ImageCollection used from GEE.*

S2 MSI				
Band	Description	Pixel Size (m)	Wavelength Centre	Wavelength Width
B1	Coastal aerosol	60	443	20
B2	Blue	10	490	65
B3	Green	10	560	35
B4	Red	10	665	30
B5	Red-edge 1	20	705	15
B6	Red-edge 2	20	740	15
B7	Red-edge 3	20	783	20
B8	Near-infrared (NIR)	10	842	115
B8A	Red-edge 4 (NIR narrow)	20	865	20
B9	Water vapour	60	945	20
B11	SWIR	20	1375	30
B12	SWIR	20	2190	180
QA60	Cloud mask	60		

The ImageCollection used in this study to map the fruit tree crops and the surrounding land cover types was S2 SR. It was applied to the data acquired for the agronomic year, June 2019 - May 2020 (i.e., 12 months). A total of 18 S2 images (Table 4.2) were collected monthly, equating to 216 images for the study period. As commonly observed in sub-tropical regions, the monthly composites were affected by clouds and shadow clouds. The S2 cloud masking was computed using the QA60 band that combines the cirrus cloud and dense cloud masks (Jin *et al.*, 2019). The QA60 band has a spatial resolution of 60m and is provided with S2 SR data in the GEE platform (Claverie *et al.*, 2018; Kolecka *et al.*, 2018).

Table 4.2: *Number of Sentinel-2 images acquired for each monthly period for the Levubu sub-tropical farm region.*

S2 monthly composites	
Month	Number of images
January	18
February	18
March	18
April	18
May	18
June	18
July	18
August	18
September	18
October	18
November	18
December	18

4.4. Acquisition of ground control points

Coinciding with the acquisition of image data for the respective crop growing seasons (Table 4.2), the field survey to map common fruit trees and other land cover types was conducted in December 2019, January, and April 2020. A Global Positioning System (eTrex® 20x GPS Receiver; Garmin, Olathe, KS, USA), with a positional accuracy of five metres (5 m) was used to record the longitudinal and latitudinal locations of the fruit-tree crops common in the area. A total number of 304 (n=304) ground-truth points were mapped for the dominant land-cover classes, including avocado (n=49), banana (n=53), built-up area (n=7), guava (n=12), macadamia nut (n=95), mango (n=18), pine tree (n=22), water body (n=4) and woody vegetation (n=44). The collected (n = 304) GPS points were used to guide the digitalization of additional points in cases where field sizes were big enough; this was achieved by using informative observations on Google Earth Pro and ArcGIS (Chabalala *et al*., 2022).

Field surveys conducted in smallholder areas are challenging, and some of the collected data were imbalanced owing to fragmented landscapes and the lack of access to some parts of the study area (Ren *et al.*, 2022). Because the unproportional distribution of the field data could affect classification accuracy (Maxwell *et al.*, 2018), field-mapped points (n=304) were used to guide the digitalisation of additional points in the data-scarce regions to account for this bias (Ni *et al.*, 2021). This approach resulted in 1894 points, which were then split into training and validation sets by using the ratio of 70:30. This approach has proved to be robust in mapping fruit trees with varied phenological growth characteristics (You and Dong, 2020). The growth stages of fruit tree crops across the Levubu subtropical farming region are presented in Table 4.3.

Table 4.3: *Growth stages of fruit trees in the Levubu subtropical farming area (Note: The shades denote the flowering stage (Flo), fruiting stage (Fru), and harvesting stage (Harv).*

Class	Varieties	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
Avocado	Hass	Fru	Fru	Fru	Fru	Harv	Harv			Flo	Flo	Flo	Fru
	Fuerte	Fru	Harv	Harv						Flo	Flo	Flo	Fru
	Pinkerton	Fru	Fru	Fru	Harv	Harv			Flo	Flo	Flo	Fru	Fru
	Ryan	Flo	Flo	Fru	Fru	Fru	Fru	Harv	Harv				Flo
Banana		Harv	Harv	Flo	Flo	Flo	Fru	Fru	Fru	Fru	Harv	Harv	Harv
Guava		Fru	Fru	Fru	Harv							Flo	Flo
Macadamia nut	Beaumont 695	Fru	Fru	Fru	Harv	Harv			Flo	Flo	Fru	Fru	Fru
	Nelmak 2	Fru	Harv	Harv			Flo	Flo	Fru	Fru	Fru	Fru	Fru
	A4	Fru	Harv	Harv			Flo	Flo	Fru	Fru	Fru	Fru	Fru
	814	Fru	Harv	Harv			Flo	Flo	Fru	Fru	Fru	Fru	Fru
	816	Fru	Harv	Harv			Flo	Flo	Fru	Fru	Fru	Fru	Fru
	344	Fru	Harv	Harv			Flo	Flo	Fru	Fru	Fru	Fru	Fru
Mango	Tommy Atkins	Harv						Flo	Flo	Fru	Fru	Fru	Harv

4.5. Feature extraction

Optimal selection of remote sensing data is crucial for the efficient mapping of crop types. In this research, the optimal phenological window was identified by developing 12 monthly composite-based models (i.e., January to December). In a Google Earth Engine (GEE) environment, the collected ground-truth points (GCPs) were overlaid on the monthly composites and used to extract the multi-temporal phenological metrics (i.e., Flo (flowering), Fru (fruiting), and Harv (harvesting) used to train and classify the fruit trees.

4.6. Spectral separability of fruit tree crops

The first step in the classification of tree species is to determine whether the classes are spectrally different. For this research, the spectral information was extracted from the Sentinel-2 images using the point data collected from the field. The assessment of the spectral properties of the sample training classes and their separability over others was conducted using spectral reflectance curves.

4.7. Training of the random forest classifier

The random forest (RF) classifier is an ensemble tree-based classifier developed in 2001 (Breiman, 2001). It combines random sub-control methods and integrated learning theory, making it suitable for processing heterogeneous data with high dimensionality (Zhi *et al.,* 2022). The random forest (RF) classifier has been extensively used because of its high accuracy, robustness, feature importance, and scalability (Tatsumi *et al.,* 2015; Saini and Ghosh, 2018; Chabalala *et al.,* 2023). Also, the RF classifier is less sensitive to noise and correlations that are intrinsically inherited in remote sensing data and reduces overfitting by averaging multiple decision trees. As such, it is suitable for mapping crop types in heterogeneous landscapes (Tatsumi *et al.,* 2015; Saini and Ghosh, 2018; Chabalala *et al.,* 2023). Two parameters were adjusted when training the RF model in the GEE environment: 1) the *ntree*, which was set to a default value of 500. It determines the total number of binary trees that can be used to build an RF model. An increase in the number of trees results in an increase in the cost calculation; 2) the *mtry* parameter was set as a square root of the total number of the input variables (Kraaijvanger *et al.,* 2019). About 67% of the samples from the training dataset were regarded as in-bag, while the remaining one-third (i.e., 33%) were termed out-of-bag (OOB) samples (Kraaijvanger *et al.*, 2019). The performance of the RF classification was examined and optimised using a 10-fold cross-validation (Tatsumi *et al*., 2015).

4.8. Accuracy assessment

Accuracy assessment is crucial when analysing remote sensing data (Banko, 2016). An accuracy assessment tests the effectiveness of the applied methodology in predicting the data (Olofsson *et al.*, 2014). For this study, the accuracy of the classified images was assessed using 30% (validation data) of the reference data. A confusion matrix was used to calculate the three metrics (i.e., the overall classification accuracy (OA), the user accuracy (UA), the producer accuracy (PA), and the kappa coefficients (κ), all of which were used to interpret the results of the models (Congalton, 1991; Olofsson *et al*., 2014).

4.9. Statistical analysis to determine the reliability of the research results

The reliability of the research results was tested by comparing results from a one-way analysis of variance (ANOVA) and the student's paired t-test, conducted at a 0.5% confidence level (Gutierrez-Coarite *et al*., 2018; Bagheri, 2019).

4.10. Results

4.10.1. Spectral separability between fruit tree crops

The mean spectral reflectance curves of the training data were extracted from the S2 data for the respective months and plotted with their standard deviations (Figure 4.3). Generally, all fruit tree classes exhibited considerable spectral overlaps. In April, only the green band (B3) separated macadamia from the other four classes beyond one standard deviation of uncertainty. In May, the green band (B3) differentiated avocado trees from the other classes, while the shortwave-infrared (B11 and B12) bands separated guava from the remaining four classes beyond one standard deviation of uncertainty. In the months of June and July, the shortwave-infrared (B11 and B12) bands again differentiated guava from the avocado, banana, macadamia nut, and mango trees from the remaining four classes. The lowest spectral reflectance values were recorded by the green band (B3) for macadamia nut in April and avocado in May.

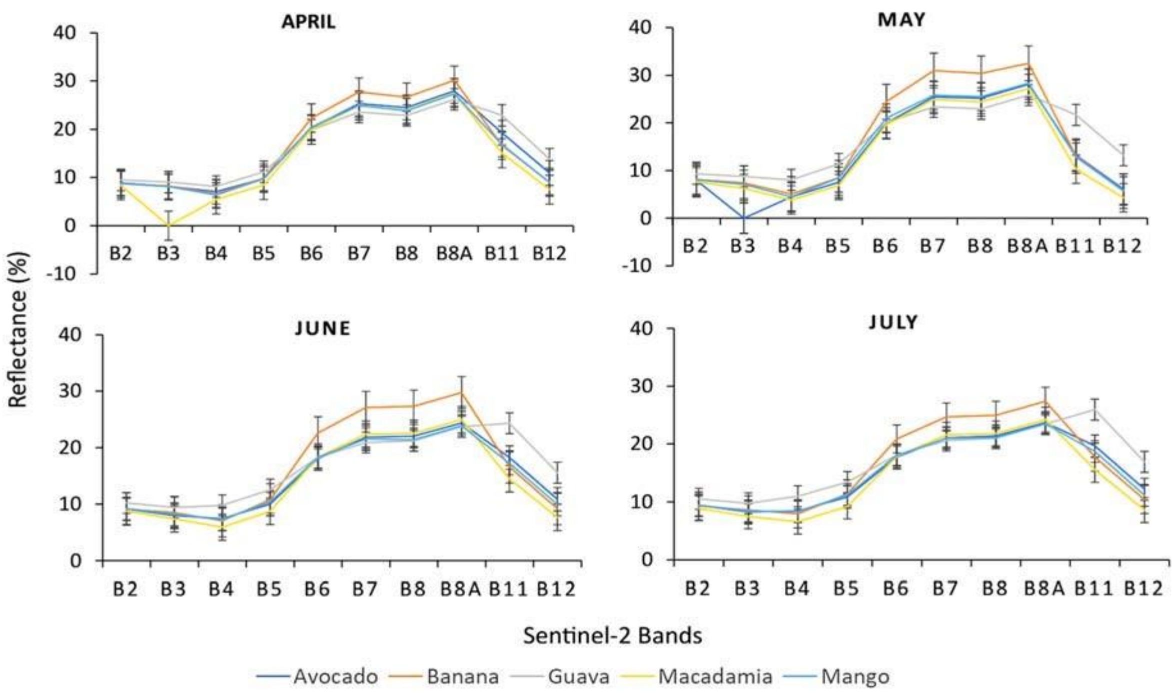


Figure 4.3: *Spectral signatures of fruit tree crops at various growth stages extracted from Sentinel-2 images in 2019 (June and July) and 2020 (April and May): blue (B2), green (B3), red (B4), red-edge (B5, B6, B7, and B8A), near-infrared (B8), and shortwave-infrared (B11 and B12) bands.*

4.10.2. Relative importance of variables

The Sentinel-2 (S2) data acquired from April to July—during the peak of crop productivity—were crucial in differentiating between fruit trees in Levubu. Figure 4.4 shows the performance of predictors applied in modelling fruit-tree crops using the variable importance tool in the random forest (RF) classifier for the winter months (i.e., April-July — the months that produced the highest overall accuracy (OA). The red band (B4: 665 nm) ranked the highest for the months of May - July, followed by the green band in April. During the peak of the flowering (August-December) and fruiting (January - March) stages, the fruit trees differed distinctly from the surrounding land use/land cover (LULC) types, i.e., built-up area, pine tree, other woody vegetation, bare soil, and water body. The blue band (B2: 490 nm) was a strong predictor in May, followed by the shortwave-infrared band (B11:1610 nm) in May and June. The shortwave-infrared band (B11:1610 nm) was a strong predictor in April, while the shortwave-infrared (B11: 1610 nm) and the red-edge (B5:705 nm) bands were strong in April, May, June, and July. Overall, the results showcase the importance of the visible bands (i.e., B2:490 nm, B3:560 nm, and B4:665 nm), the shortwave-infrared bands (i.e., B11:1610 nm and B12:2190 nm), and the red-edge bands (i.e., B5:705 nm and B7:740 nm) in differentiating between the fruit tree crops.

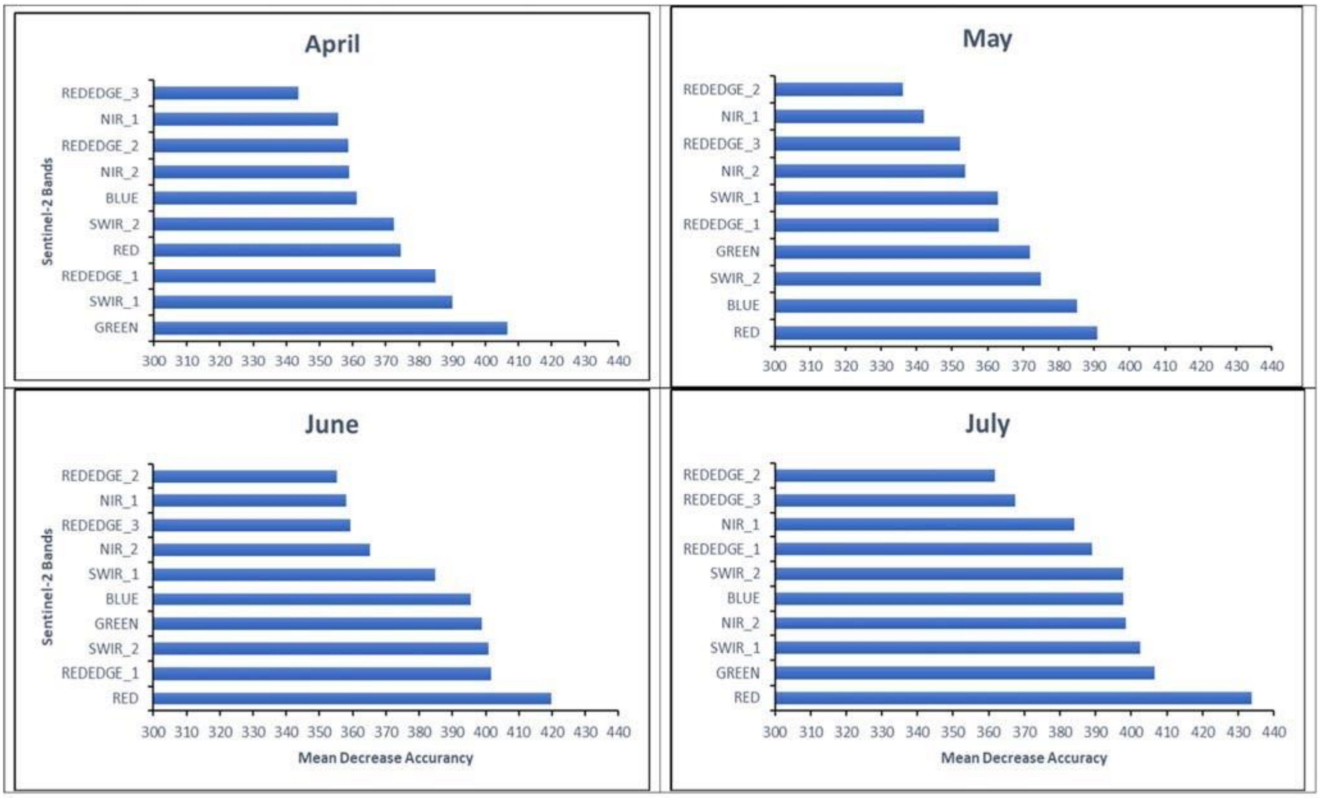


Figure 4.4: *Variable importance of bands in the best performing models used in mapping fruit trees using the Sentinel-2 dataset: (a) April, (b) May, (c) June, and (d) July.*

4.10.3. Accuracy assessment

The highest classification accuracy was obtained for the winter Sentinel-2 (S2) images (i.e., April (81.16%), May (83.57%), June (81.87%), and July (84.89%)). The lowest classification accuracy was recorded for summer and autumn (i.e., August to March, with overall accuracies (OA) ranging from 70.07 to 78.82% and the kappa coefficient ranged from 67 to 83%. The OA and kappa coefficients (κ) for the 12-month S2 composites are presented in Figure 4.5.

Figure 4.6 presents the overall classifications and the accuracy statistics. The results show that the major land cover types are distinct from the fruit trees, with user accuracies (UA) of 70 - 100%. The differentiation between the fruit trees was fair in that the user (UA) and producer accuracies (PA) were between 60 and 100%. The most successful differentiation between the tree crop's ability was achieved during the winter season (in order of merit in July, June, May, and April), while it was at its lowest level in August in the case of the avocado crop, where the UA and PA were 65% and 58.43%, respectively. The avocado crop recorded a similar UA of 69.93% and a PA of 70.86% in October. In September, the banana crop recorded a UA and a PA of 62.5% and 74.53%, respectively. In January and February, the banana crop recorded a UA of 69.9% and a PA of 68.50%, while in March, the UA and PA declined to

62.90% and 68.42%, respectively. In January and February, the guava crop reported a UA of 65%. However, the PA was higher at 75%.

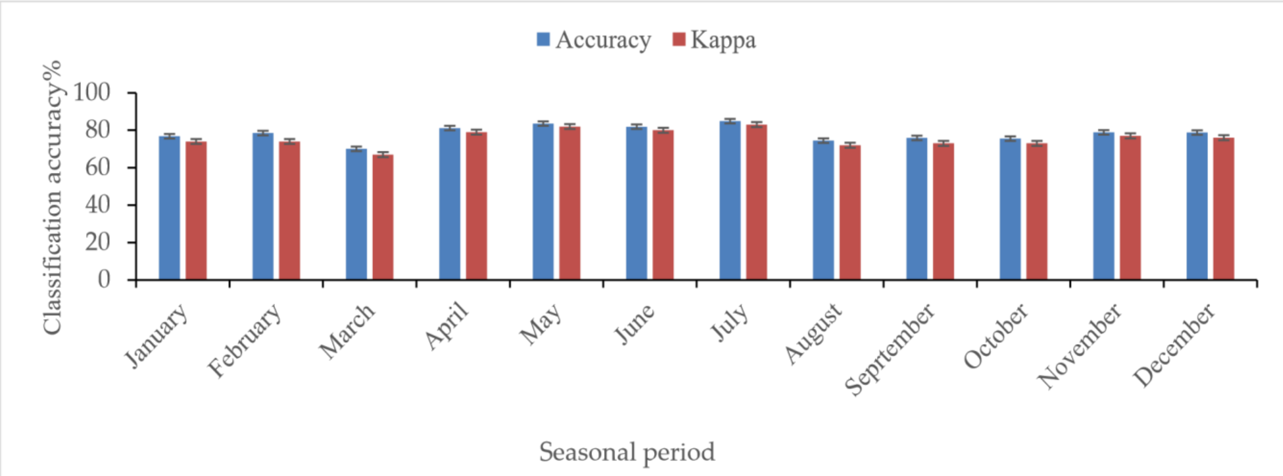


***Figure 4.5:** Comparison of the overall accuracy and kappa coefficients for the 12-month Sentinel-2 composites using a random forest classifier.*

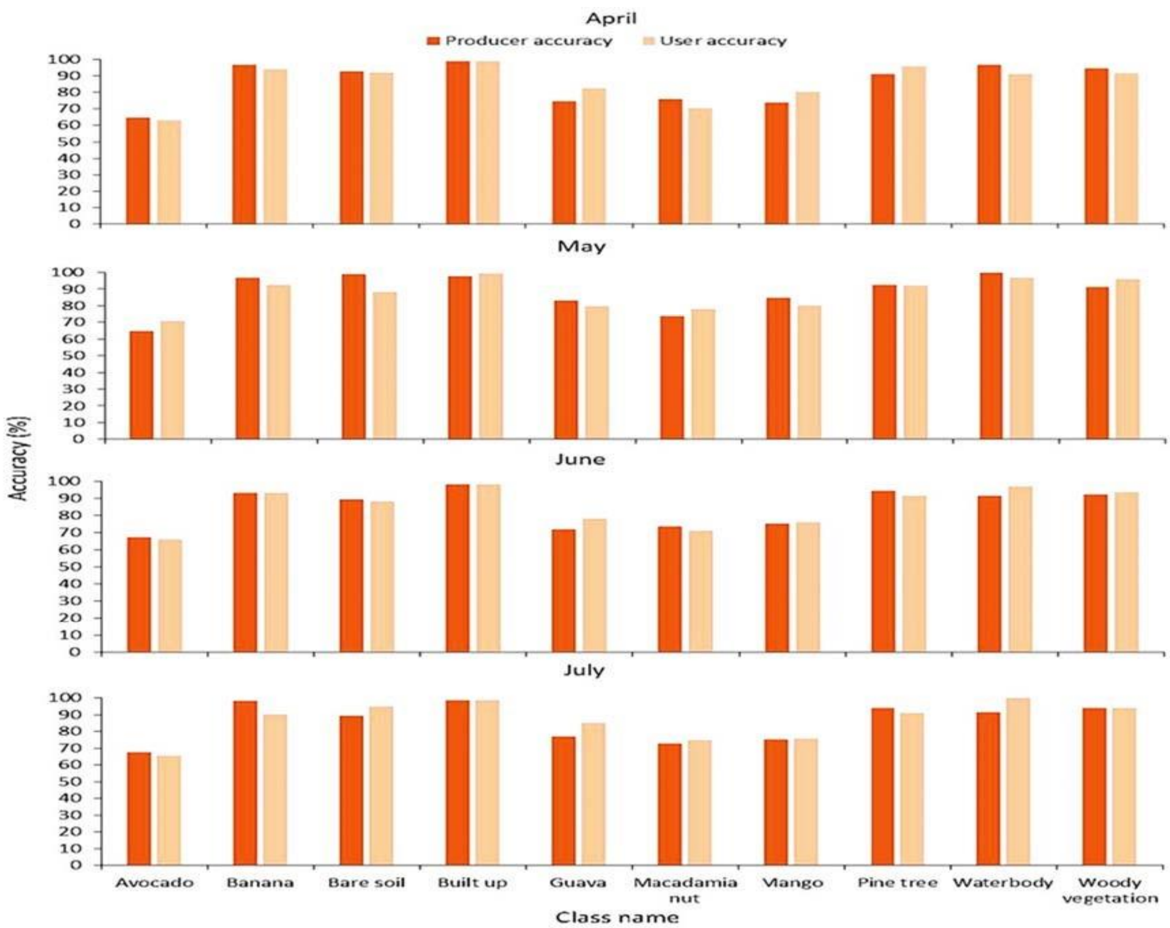


Figure 4.6:** Producer and user accuracies for avocado, banana, bare soil, built-up, guava, macadamia nut, mango, pine tree, water body, and woody vegetation in Levubu, Limpopo Province, South Africa.**

The predictions for mango for March, June, and August were low, with the UA ranging from 64 - 69% and the PA from 62 - 84%. The other land cover types (i.e., bare soil, built-up area, pine tree, other woody vegetation, and water body) had UAs of over 70% throughout the year, except for March, where a UA of 60.19% and a PA of 68.18% were obtained for the bare soil and pine tree classes, respectively.

4.10.4. Statistical analysis to determine the reliability of the research results

The changes in the RF accuracies were related to S2 monthly composites using ANOVA and the Student's t-test. The tests, conducted at a 95% confidence interval, produced a P-value of 0.170 and 0.174, greater than the confidence level of 0.05%, indicating an insignificant difference in the RF classifications across the twelve months.

4.10.5. Spatial distribution of fruit tree crops

Classified winter images show the predicted fruit tree crops and various land cover types within the study area (Figure 4.7). In the April image (Figure 4.7 and Figure 4.8A), the bare soil class covered 20.55 % of the study area, followed by the woody vegetation class with an area coverage of 23.40%. In the same month, the avocado and macadamia nut classes covered approximately 11.15% and 16.01% of the study area, respectively. In May (Figure 4.7 and Figure 4.8B), woody vegetation was the dominant class, covering 22.20% of the study area. This was followed by the macadamia nut, avocado, built-up area, and banana classes, which recorded coverage of 10.90%, 16.60%, 7.05%, and 6.32%, respectively. Both guava and mango had more areal coverage than the avocado crop. In June (Figure 4.7 and Figure 4.8C), bare soil covered 30.29% of the study area and dominated the southeastern and southwestern parts of the research area, while woody vegetation covered 18.14%. Macadamia nut, built-up area, and avocado covered 11.01%, 11.83%, and 13.52%, respectively, of the area. All classes were separable in July (Figure 4.7 and Figure 4.8D) — where within-class misclassifications were low. The area coverage of bare soil increased to 32.29% and continued to dominate in the southeastern and southwestern parts of the research area. The woody vegetation was concentrated in the southern and northern parts of the study area, covering 14.05% of the area. Avocado and macadamia nuts dominated the study area in the eastern, central, and north-western parts, covering 12.88% and 11.09%, respectively. Overall, the results demonstrate the potential of S2 multi-temporal data for mapping fruit trees in a highly fragmented heterogeneous horticultural landscape.

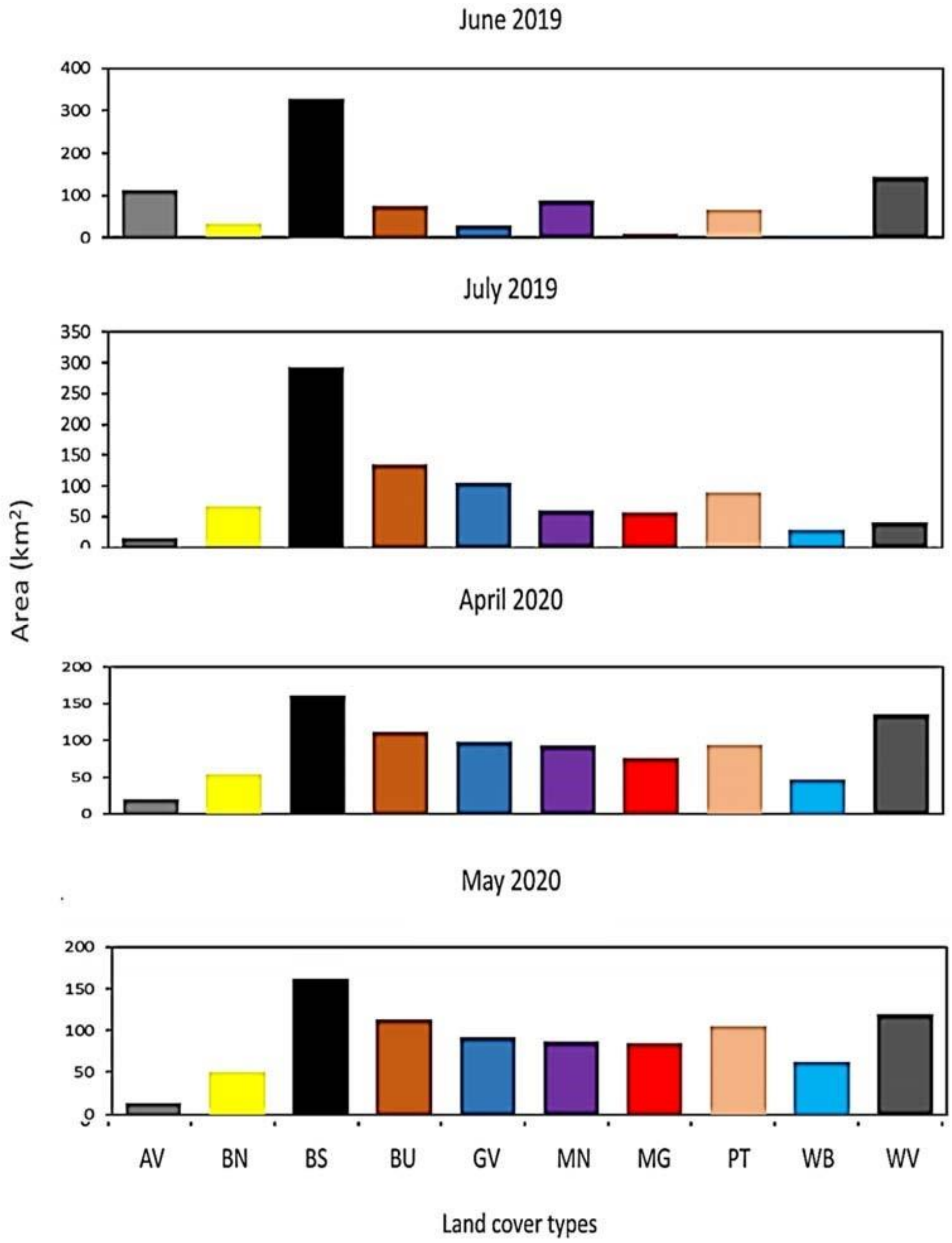
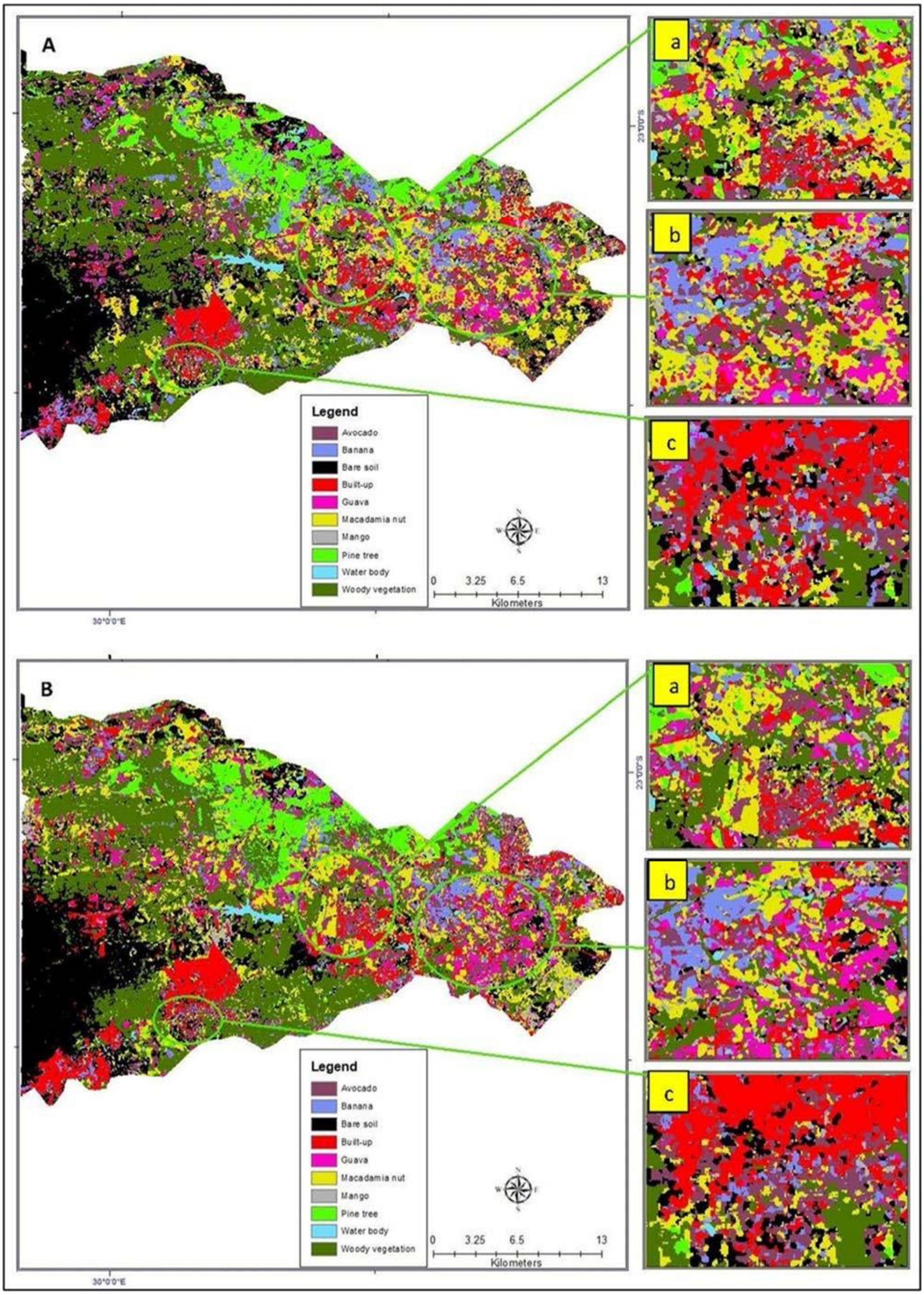


Figure 4.7: *Areal coverage, in square kilometres, of fruit tree crops and co-existing land cover types in the study area. The areal calculations are based on the months of June 2019, July 2019, April 2020, and May 2020.*


A
a
b
c
Legend
Avocado
Banana
Bare soil
Built-up
Guava
Macadamia nut
Mango
Pine tree
Water body
Woody vegetation
0 3.25 6.5 13
Kilometers
30°0'0"E
23°0'0"S
B
a
b
c
Legend
Avocado
Banana
Bare soil
Built-up
Guava
Macadamia nut
Mango
Pine tree
Water body
Woody vegetation
0 3.25 6.5 13
Kilometers
30°0'0"E
23°0'0"S

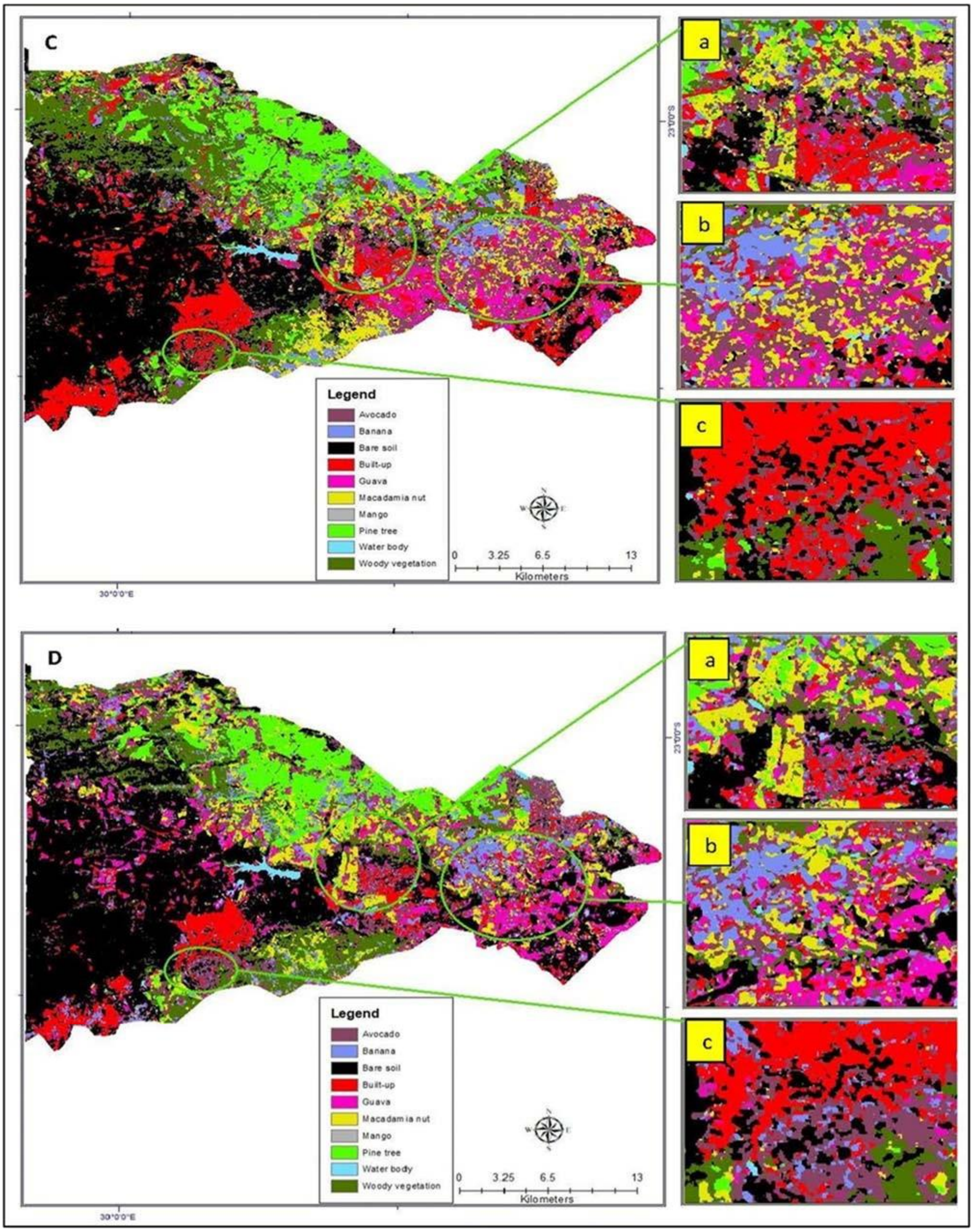


Figure 4.8: *Classification maps of fruit tree crops with spatially explicit zoom-ins (a, b, and c), providing detailed spatial distributions from RF classifications in the research region for the best classification accuracy. They are based on four selected S2 monthly composite images captured in June (5A), July 2019 (5B), April (5C) and May 2020 (5D).*

4.11. Discussion

The significance of this study was to identify the optimal temporal window in which crucial phenological stages can be used to effectively classify fruit tree crops in a highly fragmented heterogeneous horticultural landscape. The classification was performed using a Sentinel-2 (S2) dataset and a random forest (RF) classifier in a Google Earth Engine (GEE) environment. Furthermore, multi-temporal images allowed for the identification of an optimal phenology stage in which different fruit trees could be accurately differentiated from one another (Santos *et al*., 2019).

The spectral reflectance of the crops was determined by a myriad of factors (e.g., fruit varieties, different management, and mixed cropping systems, etc.). Different crop varieties of the same family can furthermore have different growth stages (i.e., flowering, fruiting, or harvesting). As such, they exhibited different biochemical traits (Ren *et al.,* 2022). Figure 4.3 shows that within one standard deviation, all fruit tree classes showed varying but significant degrees of spectral overlap. The best four performing models (i.e., April, May, June, and July models) revealed that beyond one standard deviation from the mean, only the green band (B3) separated macadamia from avocado, banana, guava, and mango. In May, again, the green band (B3) separated avocado from banana, guava, macadamia, and mango trees, while the shortwave-infrared (B11 and B12) bands differentiated guava trees from the other fruit tree classes. In the months of June and July also, beyond one standard deviation of uncertainty, the shortwave-infrared (B11 and B12) bands differentiated guava from avocado, banana, macadamia nut, and mango trees. However, an advantage of machine learning classifiers, such as RF, is that they are suited to extreme cases of binary classification, and as such, lead to highly accurate classifications (Kuter, 2021). The major objective in selecting variables is to choose and identify those of high importance, which makes the associated models simpler and faster to fit and predict. Overall, the red band outperformed the other bands in the months of May, June, and July. The greatest contribution of the red band in differentiating between the respective fruit trees in the study area was in July (Figure 4.4). The best-performing bands in April were the green (B3), shortwave-infrared (B11), red-edge (B5), red (B4), and shortwave-infrared (B12) bands. The bands that performed dismally in the four best-performing models were the near-infrared band in May and the red-edge bands (i.e., B7 and B6) in the months of May and April, respectively. The near-infrared (B8) band performed well in July. Generally, the visible bands performed well in differentiating between the fruit tree crops in the study area.

The flowering of the fruit trees was detected mostly by the red (B4: 665 nm) and green (B3:560 nm) bands from the visible light spectrum, as these bands show correlations with the photosynthetic components (i.e., pigment, water, and chlorophyll) of the fruit trees (Pena *et al.*, 2017). Comparable with the findings of Peña and Brenning (2015), the green band was saturated at the end of the vegetative stage as a result of the high chlorophyll content and contributed again in June and July, when the mango and macadamia nut crops started flowering, thus indicating that this band performs well at the green-up stage. Previous work has demonstrated that the near-infrared bands are influential in differentiating between the crop types (Lakshmanan and Mathiyazhagan, 2019). In this study, the near-infrared (B8A:865 nm) spectral band performed well and contributed substantially to the classification of crop types during the flowering stage. The highly predictive power of the near-infrared (B8A:865 nm) band is attributed to its sensitivity to the vegetation metrics (i.e., chlorophyll, LAI, and nitrogen), which are in turn related to crop growth and productivity (Tu *et al.*, 2019). A reduced differential ability in the near-infrared region was apparent at the end of the fruiting stage and the start of the harvesting stage when crop classification favours clear differences in relation to the amount of foliage of the various crops (Pena *et al.*, 2017). The observed multi-series signatures of the S2 bands concurred with those observed by Mercier *et al.* (2020). A peak variation in spectral signatures was recorded for the red-edge bands during the flowering and fruiting stages (Figure 4.8). Feng *et al.* (2019) and Figueroa-Figueroa *et al.* (2020) recently reported that the inclusion of shortwave-infrared and red-edge bands had a significant effect in distinguishing between the fruit tree crops. Their results concur with previous findings by Delloye *et al.* (2018). Comparable to the report by Luo *et al.* (2021), the red-edge (B5:705 nm) band proved to be a strong predictor during the senescent and fruiting stages because of its sensitivity to chlorophyll levels. The high temporal resolution of S2, with its three additional spectral bands located in the red-edge position (REP), allowed for the retrieval of the vegetation parameters such as chlorophyll, leaf nitrogen content, and leaf area index (LAI) (Chabalala *et al.*, 2020). The shortwave-infrared bands (B11:1610 nm, B12:2190 nm) could detect the variation in the water content and foliar chlorophyll of the fruit tree crops (Peña and Brenning, 2015). The results are supported by Ren *et al.* (2022), who highlighted the importance of shortwave-infrared bands in distinguishing between crop types.

The RF classification of S2 images (i.e., from June 2019 -May 2020) classified fruit trees with overall accuracies (OA) ranging from 67 - 84.80% and a kappa coefficient (κ) ranging from 67 - 83%. The winter images reported the highest accuracies in their classification results (i.e., OA greater than 80%). The results concur with previous crop-type mapping research findings by Blickensdörfer *et al.* (2022), who obtained the highest accuracies in their classification results of 80% using winter images. Research by Luo *et al.* (2021) achieved similar OA results of 89.75% when these researchers mapped crop types in China using S2 time series data at a regional level. The winter months (i.e., April - July) coincided with peak greenness (i.e., the fruiting period of most crops) (Luo *et al.,* 2021). The user accuracy (UA) ranged from 60 to 100%, while the producer accuracy (PA) ranged from 58 to 100%. Owing to spectral similarities emanating from overlapping phenological stages, the UA and PA accuracies for banana, mango, and macadamia nut classes were low from January to March 2020 (Tu *et al.,* 2019; Kyere *et al.,* 2020). It is worth noting that these were crop growth periods during which Levubu experienced a high volume of rainfall (i.e., in 2020). The existence of double-row planting systems and a mixture of young *versus* mature macadamia and banana trees made it challenging to accurately detect such trees. Furthermore, since the banana crop is asynchronous, its growth stages often overlap with those of the avocado, mango, guava, and macadamia nut trees (Hebbar *et al.,* 2014; Gao *et al.,* 2020). The findings are consistent with observations by Hebbar *et al.* (2014), who reported poor classification accuracy in orchard plantations owing to mixed spectral reflectance emanating from the simultaneous presence of young and mature plantations. For example, the avocado and mango crops reported higher accuracy levels from their larger canopies and leaf structures. In contrast, the macadamia nut and banana classes produced lower classification accuracies for most months. This can be attributed to their canopy structures and the shape of their leaves, which expose the soil background and result in mixed spectral signatures (Rajah *et al.,* 2017). Overall, the class accuracies of the fruit tree crop and the co-existing land cover types differed because of the differences in the spectral properties exhibited by the crops during their growth cycles (Ge *et al*., 2016).

The results of this study proved that S2 images are sufficient for the accurate mapping of fruit trees in the fragmented horticultural landscapes of South Africa. The classification maps produced in this research could assist agronomists, horticulturists, and farm managers in devising precise management strategies to build sustainable horticultural food systems (Yi *et al.,* 2020).

4.12. Limitations

Tracking crop growth by using multi-series data was limited by the prevailing cloud cover at the time of the research. The unavailability of quality images limited the accuracy during the summer months. The monitoring of the banana green-up and senescent stages was missed because of crop mismanagement and the presence of clouds in the S2 multi-temporal data from December to March. The problem was aggravated by the cloud and cloud shadow detection algorithms, which proved to be unreliable. Future research should focus on applying more robust models, such as deep learning and selecting images acquired during optimal seasons, to untangle spectral overlapping and similarities and improve classification accuracy.

4.13. Conclusion

This research leveraged the power of the spatial and temporal resolution of S2 data analyzed in a Google Earth Engine (GEE) environment and an RF classifier in mapping fruit trees. The results led to the conclusion that optimizing the data acquisition period and the potential of temporal resolution in remote sensing offers considerable promise for accurately identifying fruit trees. The optimum period for establishing the type of fruit trees was during the early and middle growth stages (i.e., during the flowering and fruiting stages). This study has proved the Sentinel-2 (S2) sensor to be a useful tool for modelling and monitoring fruit trees at the landscape level. The results provide evidence-based information that could allow for efficient horticultural monitoring systems and sustainable agricultural management, particularly in countries or regions where specific crop databases are not available. The approach used in this study could be applied as a framework for areas with similar fruit tree types, phenologies, and landscape characteristics.

CHAPTER FIVE

Mapping fruit tree dynamics using phenological metrics derived from optimal Sentinel-2 data and Deep Neural Network

This chapter is based on:

Chabalala, Y.; Adam, E. and Kganyago, M. (2023). Mapping fruit tree dynamics using phenological metrics derived from optimal Sentinel-2 data and Deep Neural Network. *CABI for Agriculture and Bioscience*, 4, 1-20. DOI: https://doi.org/10.1186/s43170-023-00193-z

Abstract

Accurate and up-to-date crop-type maps are essential for efficient management and well-informed decision-making, allowing accurate planning and execution of agricultural operations in the horticultural sector. The assessment of crop-related traits, such as the spatiotemporal variability of phenology, can improve decision-making. The study aimed to extract phenological information from Sentinel-2 data to identify and distinguish between fruit trees and co-existing land use types on subtropical farms in Levubu, South Africa. However, the heterogeneity and complexity of the study area—composed of smallholder mixed cropping systems with overlapping spectra—constituted an obstacle to the application of optical pixel-based classification using machine learning (ML) classifiers. Given the socio-economic importance of fruit tree crops, the research sought to map the phenological dynamics of these crops using deep neural network (DNN) and optical Sentinel-2 data. The models were optimized to determine the best hyperparameters to achieve the best classification results. The classification results showed the maximum overall accuracies of 86.96%, 88.64%, 86.76%, and 87.25% for the April, May, June, and July images, respectively. The results demonstrate the potential of temporal phenological optical-based data in mapping fruit tree crops under different management systems. The availability of remotely sensed data with high spatial and spectral resolutions makes it possible to use deep learning models to support decision-making in agriculture. This creates new possibilities for deep learning to revolutionize and facilitate innovation within smart horticulture.

5.1. Introduction

The phenological stages of vegetation act as important indicators in monitoring vegetation growth and evaluating how climate change may affect vegetation, among other functions (Pan *et al*., 2021). Phenology and related studies may be as old as civilization itself—farmers settled in particular places and carried out certain agricultural tasks, including sowing seeds, tending crops, and harvesting, at certain times of the year (Pan *et al*., 2021).

Historical interest in phenology was sparked by a desire to understand how farming evolved and how it related to the climate (Chabalala *et al*., 2020). Phenological dynamics are influenced by local environmental interactions, genetic factors, seasons, and agronomic management of the growing environment, which results in phenology variation (Xie and Niculescu, 2022; Chabalala *et al*., 2023a). Mapping crop types using phenological information acquired during crop key growth stages offers additional possibilities for tracking crop growth changes (Pan *et al*., 2021; Aitelkadi *et al*., 2021). The information about crop phenological events and their spatiotemporal variability can assist in improving the quality of crops and yields through the implementation of appropriate and sustainable crop management practices (Pan *et al*., 2021; Elders *et al*., 2022; Yedage *et al*., 2013). Therefore, accurate mapping of crops' phenological dynamics is essential in crop production estimation, especially given the current climate uncertainty in sub-tropical Africa (Pan *et al*., 2021). Accurate crop yield prediction assists decision-makers and farmers in planning harvests, storage, and preparing for food import or export in the event of either shortage or surplus (Kordi and Yousefi 2022a).

Previous fruit-tree inventories were based on conventional field surveys, which are costly, time-consuming, and impractical over vast areas, lack spatial variability, and are often subject to farmers' reporting biases (Chabalala *et al*., 2020, 2023a; Xie and Niculescu, 2022). Conversely, remote sensing can be applied to monitor crop phenology at a landscape level in a timely and effective manner (Chabalala *et al*., 2023b). The application of phenological metrics for crop classification is mostly conducted using stacked time-series data (Pan *et al*., 2021; Xie and Niculescu, 2022). (Singh *et al*., 2022) used phenology metrics derived from Landsat for sugarcane crop mapping in North India and obtained an overall mapping accuracy of 84.5%. A study by Feng *et al*. (2023) used phenology metrics derived from Sentinel-2 to map rice, maize, and soybean for crop type mapping in Fujin, China and achieved an overall accuracy of 97.14%. However, existing approaches used for extracting temporal features lack the adaptability to handle sub-tropical farming systems with complex vegetation dynamics characterized by intra-class variability, interclass similarity, and persistent clouds resulting in disparate temporal patterns (Pan *et al*., 2021; Zhong *et al*., 2019). Furthermore, stacked images have high data dimensionality and are highly computational (Pan *et al*., 2021). Other studies applied vegetation indices (VIs) derived from time-series data, which has been proven to produce high classification accuracy in vegetation and crop mapping (Chen *et al*., 2019). Although the approach captures important traits related to vegetation health and growth. The vegetation indices (VI) are effective for crop types with distinct temporal spectral characteristics and remain challenged by the spectral complexities emanating from diverse cropping patterns in morphologically heterogeneous landscapes with similar spectra (Chabalala *et al*., 2023b). However, the classification accuracy depends on the number of images used to create the time-series product (Pan *et al*., 2021). Vegetation indices are created using specific spectral features, which ignores other bands that might be crucial in the overall classification model.

While crop phenology mapping approaches are well established, their applicability is subject to uncertainties in smallholder agriculture due to agronomic factors (Aitelkadi *et al*., 2021). Furthermore, it is challenging to map fruit tree crops in sub-tropical smallholder regions due to the limited availability of cloud-free observations (Yin, 2023). Although many studies have mapped horticulture crops and heterogeneous landscapes (Yin 2023). Sub-tropical smallholder agriculture in South Africa is planted in small plots (< 1 ha) and is characterized by multiple scales of mixed, irregular, and intercropping systems resulting in landscape heterogeneities—crops with different textures, shapes, sizes, colours, and morphological features, and sharing spectral and canopy similarities (Biffi *et al*., 2021; Ukwuoma *et al*., 2022). Also, intra-class variation occurs because of fruit-tree mutations (Gao and Zhang 2021; Bal and Kayaalp 2021). Furthermore, the farming systems follow different growing calendars and management strategies, resulting in within-farm heterogeneity and fruit trees with similar growth profiles and morphological features (appearances, shapes, and colour) (Elders *et al*., 2022). This results in a multi-form classification problem that is difficult to solve using single-date images as the available observations might not be sensitive to specific crop growth stages (Gao and Zhang 2021; Kordi and Yousefi 2022b). The management strategies for horticultural crop cultivation differ from those for common crops (Elders *et al*., 2022). Therefore, the classification models that are suitable for common crops tested in these regions cannot be transferred to other regions as they are sensor, region, and crop-specific (Villa *et al*., 2015).

According to existing literature, two classification approaches exist, i.e., Machine Learning (ML) and Deep Learning (DL). Machine learning (ML) algorithms such as random forest (RF), k-nearest neighbours (*k*-NN), and support vector machines (SVM) have been applied in various plant studies (Prins and Niekerk 2020; Chabalala *et al*., 2022; Schreier *et al*., 2021). Although remarkable results were achieved in classifying certain crop types, difficulties in distinguishing between crop types were nevertheless reported. Machine learning (ML) approaches are generative and require predetermined classes, which render them unsuitable for identifying heterogeneous crop types with high crop variation, occlusion, and overlapping spectra (Kestur *et al*., 2018). Machine Learning classifiers such as SVM rely on features that are not designed for multi-temporal data, making them unable to model the inherent feature variation in time-series data (Pan *et al*., 2021). Deep Learning (DL) has opened new horizons for innovation and the introduction of novel approaches within the agricultural industry (Vasconez *et al*., 2020; Southworth and Muir, 2021). As such, the extraction of information from nonlinear objects in complex farming systems is now possible through DL (Elders *et al*., 2022; Vasconez *et al*., 2020). Deep learning (DL) models are versatile tools that assimilate heterogeneous big data and solve the complex problem of fruit tree classification accuracy (Elders et al., 2022). Thus far, different DL models, such as Convolutional Neural Networks (CNN), YOLOv5, and RetinaNet, have been used to distinguish and classify crop types (Biffi *et al*., 2021; Ukwuoma *et al*., 2022; Cai *et al*., 2018; Xiong *et al*., 2022). Ismail and Malik (2021) compared five DL models (DenseNet, EfficientNet, NASNet, MobileNetV2, and ResNet) for grading apples and bananas and obtained a recognition rate of 98.6% and 99.2% for bananas and apples, respectively, using the EfficientNet model. The study by Xiong *et al*. (2022) evaluated five deep learning models (VGC16, AlexNet, InceptionV3, MobileNetV2, and ResNet) for classifying date fruit types and obtained a classification accuracy of 99% based on the MobileVetV2 model. When mapping wheat, maize, squash, and sunflower in China, (Xiong *et al*., 2022) found the Unet model to be superior to RF, SVM, Extreme Gradient Boosting (XGB,) and deplabv3 +. However, most of these studies concentrated on large-scale crops such as maize, rice, wheat, apple, citrus, and olives, grown in commercial farms with no overlapping spectra (Elders *et al*., 2022; Li *et al*., 2022; Mashonganyika *et al*., 2021). Each crop has a microclimate that is dependent on plant development, and such traits differ tremendously under climatic conditions (Kumar *et al*., 2021). Therefore, the performance of these mapping approaches tested on single crops might be limited if transferred to landscape locations with smallholder management practices (Ukwuoma *et al*., 2022; Zhang *et al*., 2021).

The horticulture industry has a crucial role to play in advancing the economy, ensuring food and nutritional security in South Africa, and creating job opportunities for residents in Levubu (Vasconez *et al*., 2020). Despite this, the horticulture industry in South Africa is still faced with challenges, including a lack of innovation, outdated agricultural practices, and insufficient technical skills. As a result, decision-making by farmers in Levubu relies on fruit tree inventories compiled manually, which is expensive, time-consuming, and prone to human errors. There are no fruit tree datasets or maps of the spatial distribution of fruit trees, and only farm boundaries ascertained by means of Google Earth are available—and this is for only a small number of farming systems. Research on improving the mapping approach is crucial to obtaining robust results and thus developing appropriate management strategies.

Although phenology metrics and time series data have been widely used to map fruit trees (Pan *et al*., 2021; Feng *et al*., 2023). Crop type mapping using time-series data in the context of smallholder agriculture is still challenging due to the lack of remote sensing data during the current analyzed season, mostly caused by clouds in sub-tropical regions (Pan *et al*., 2021). Thus, the development of a remote sensing-based fruit tree mapping model requires the selection of optimal images and adequate training data (Gallo *et al*., 2023). The study by Chabalala *et al*., 2023b identified the optimal window period to map fruit tree crops during their critical phenological stages in Levubu, using Sentinel-2 monthly composites and Random Forest. Their study revealed that fruit trees can be optimally mapped with an accuracy of 85% using images acquired from April to July. However, image composites have high data dimensionality, which has been reported to limit the performance of the ML models (Pan *et al*., 2021).

Therefore, the research aimed to overcome the inherent problem of data dimensionality in time series data and the inability of the ML classifiers to handle heterogeneous information. This research applied phenological metrics derived from Sentinel-2 images acquired during optimal crop-growing seasons (i.e., flowering, fruiting, harvesting) to map fruit trees in Levubu, South Africa, using a Deep Neural Network (DNN) model. These months correspond to clear conditions with no cloud interference in Levubu, which will enable the characterization of the seasonal patterns of fruit trees based on their temporal phenological profiles according to their seasonal events (Singh et al., 2019). This approach can provide reliable results for a challenging and highly heterogeneous agricultural region like Levubu, which is characterized by fragmented small-sized fields with mixed cropping systems. The results of this research can help innovate and improve fruit management in smallholder horticulture systems.

5.2. Materials and methods Methodological framework

The research tested the applicability of Sentinel-2 time series images acquired during the optimal growing season using a Deep Neural Network (DNN) to distinguish fruit trees and surrounding land use types on sub-tropical horticultural farms in Levubu, South Africa. Figure 5.1 shows the methodology followed in this study.

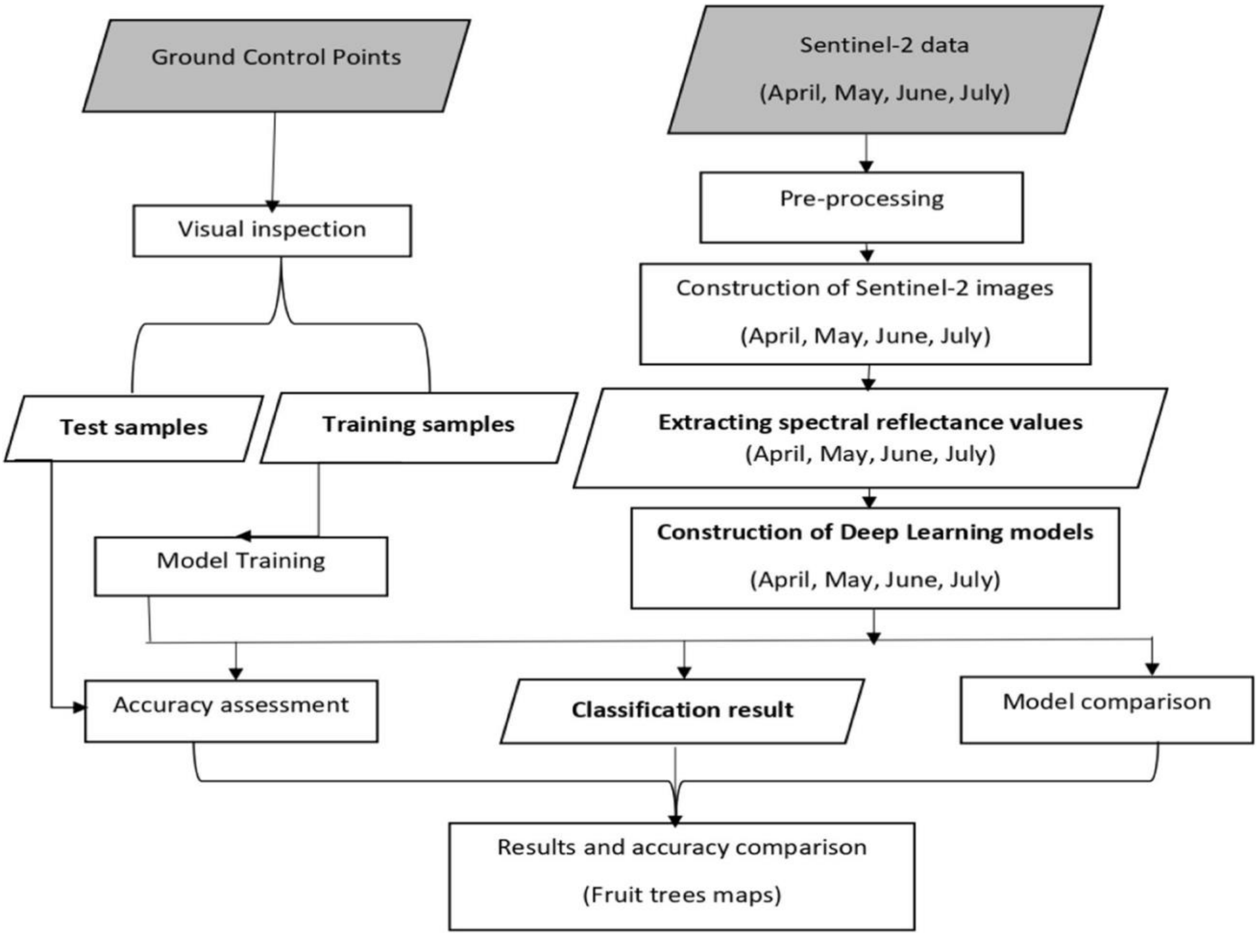


Figure 5.1: *Flowchart showing the Methodology followed to identify fruit trees and other land use types using Sentinel-2 images and the DNN algorithm.*

5.3. Research area

The research area, Levubu sub-tropical farms, is located in the Northern Limpopo Province (Figure 5.2), which is very productive in horticultural farming in South Africa. The area has a total land coverage of approximately 10, 000 hectares, and much of the land is over the Soutpansberg Mountains, located at 775 m above sea level (Chabalala *et al*., 2023a). The area is characterized by a warm climate and an average rainfall of 1000 mm accompanied by persistent cloud cover along the mountains, especially during the summer seasons (September to March), which coincide with the beginning of the flowering and fruiting seasons. Cloud cover is a bit better in winter months (i.e., April to August). Hence, the images acquired during this period will be used in this research. The Levubu farming area is composed of fragmented smallholder horticultural farming systems. The average agricultural field sizes are smaller than 1 hectare (ha) planted with fruit tree crops (i.e., avocado, banana, guava, mango, macadamia nut) and surrounding land use types (i.e., bare soil, build-up, pine trees, water body, and woody vegetation) (Chabalala *et al*., 2022; Chabalala *et al*., 2023a). The fruits are used for local, national, and global consumption and contribute to seasonal employment generation to the local surrounding communities and Gross domestic product in South Africa.

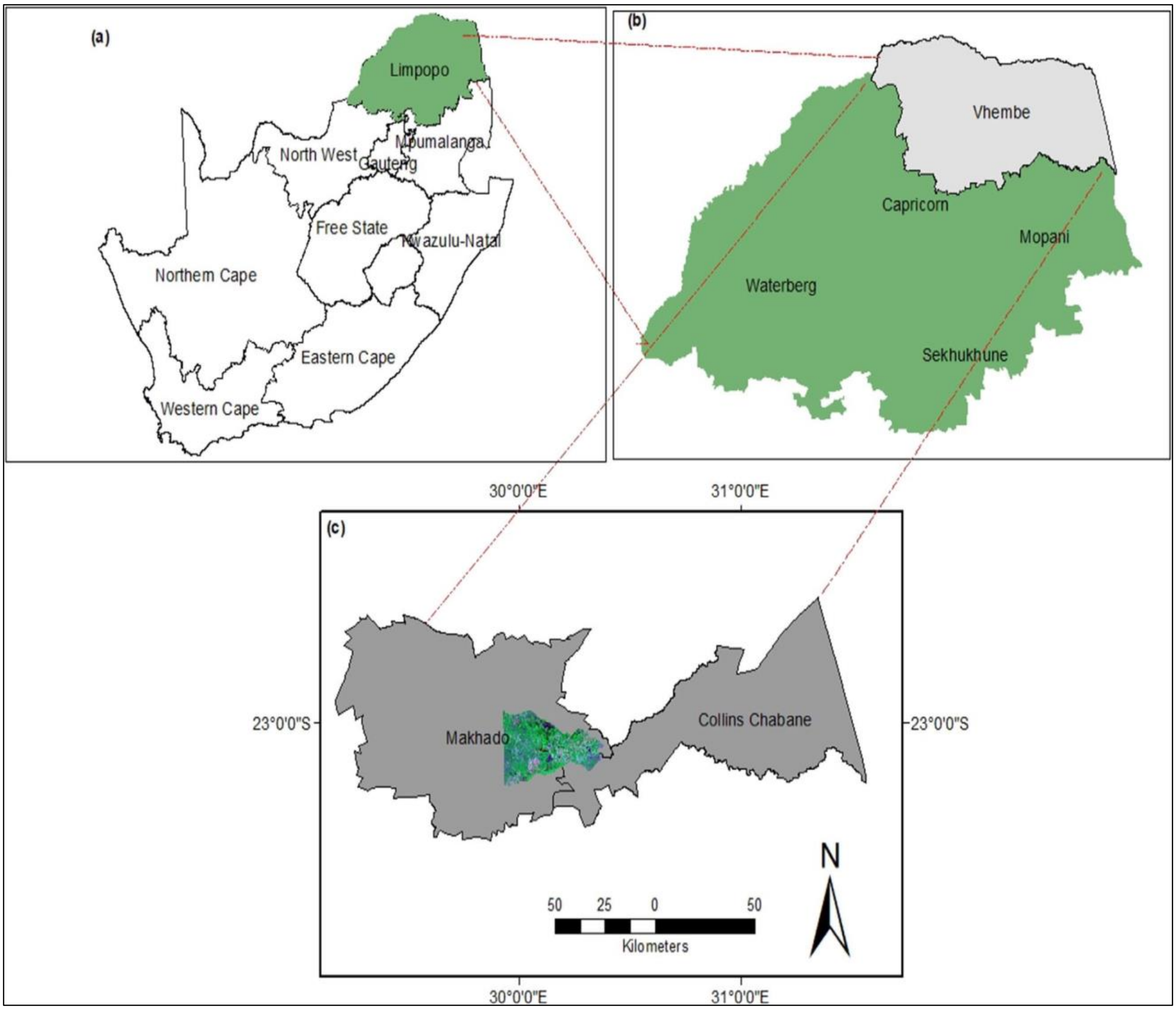


Figure 5.2: *Location of the research areas at the National (Top left corner) and District Municipality levels (Top right corner). The scene of the Sentinel-2 RGB image indicates the research area. The green (a) and light grey areas, (b) represent Limpopo Province and the Vhembe District Municipality, while the dark grey c represents the Makhado and Collins Chabane Local Municipalities.*

5.4. Data collection and processing

This section explains the datasets used in this study (in-situ and earth observation) and how earth observation data was pre-processed before analyzing the data. Furthermore, it shows the behavioural changes of crops in different growing seasons. Additionally, it discusses the classification algorithm applied in this study and how it was implemented in Jupyter Notebook in a Python environment. The concluding sub-subsection explains the assessment of the classification results.

5.4.1. In-situ data collection

The in-situ data were collected during field campaigns conducted in December 2019, January 2020, and April 2020. A handheld Garmin eTrex 20X global positioning system (GPS) with an accuracy of a sub-meter was used to record the geographic locations (latitude and longitude). A total number of 304 (n = 304) ground control points were captured from the dominant land cover classes; these include avocado (n = 49), banana (n = 53), built-up (n = 7), guava (n = 12), macadamia nut (n = 95), mango (n = 18), pine tree (n = 22), water bodies (n = 4) and woody vegetation (n = 10) (Figure 5.3). These points (n = 304) were then used to guide digitizing additional points using the visual interpretation of Sentinel-2 Images on ArcMap (ArcGIS® v. 10.6; ESRI, Redlands, CA, USA) and Google Earth Pro (Chabalala *et al*., 2020, 2023a). The ground data were subdivided into a 70% training set which was used to train and fine-tune the hyperparameters of the classification models, while the remaining 30% testing set was used to validate the classification results.

5.4.2. Spectral characteristics of fruit-tree crops and co-existing land use types

Efficient orchard management is dependent on crop phenology. The farming systems are intensive and intercropped, making it challenging to map using single-date images (Chabalala *et al*., 2023b). Hence, for this research, phenological metrics were extracted from four optimal Sentinel-2 images acquired during key phenological stages of the fruit trees (i.e., flowering, fruiting, and senescence) (Table 5.1). This information was utilized to account for spectral confusion that mostly occurs while mapping crop types using single-date images in heterogeneous landscapes**.** The spectral reflectance data for each image was extracted using the collected in-situ data using ArcGIS version 10.5. The average spectral characteristics curves of the fruit trees and other land use types considered in the research are presented in Figure 5.4. The spectral curves were extracted from 10 Sentinel-2 multispectral bands using the collected ground truth data.

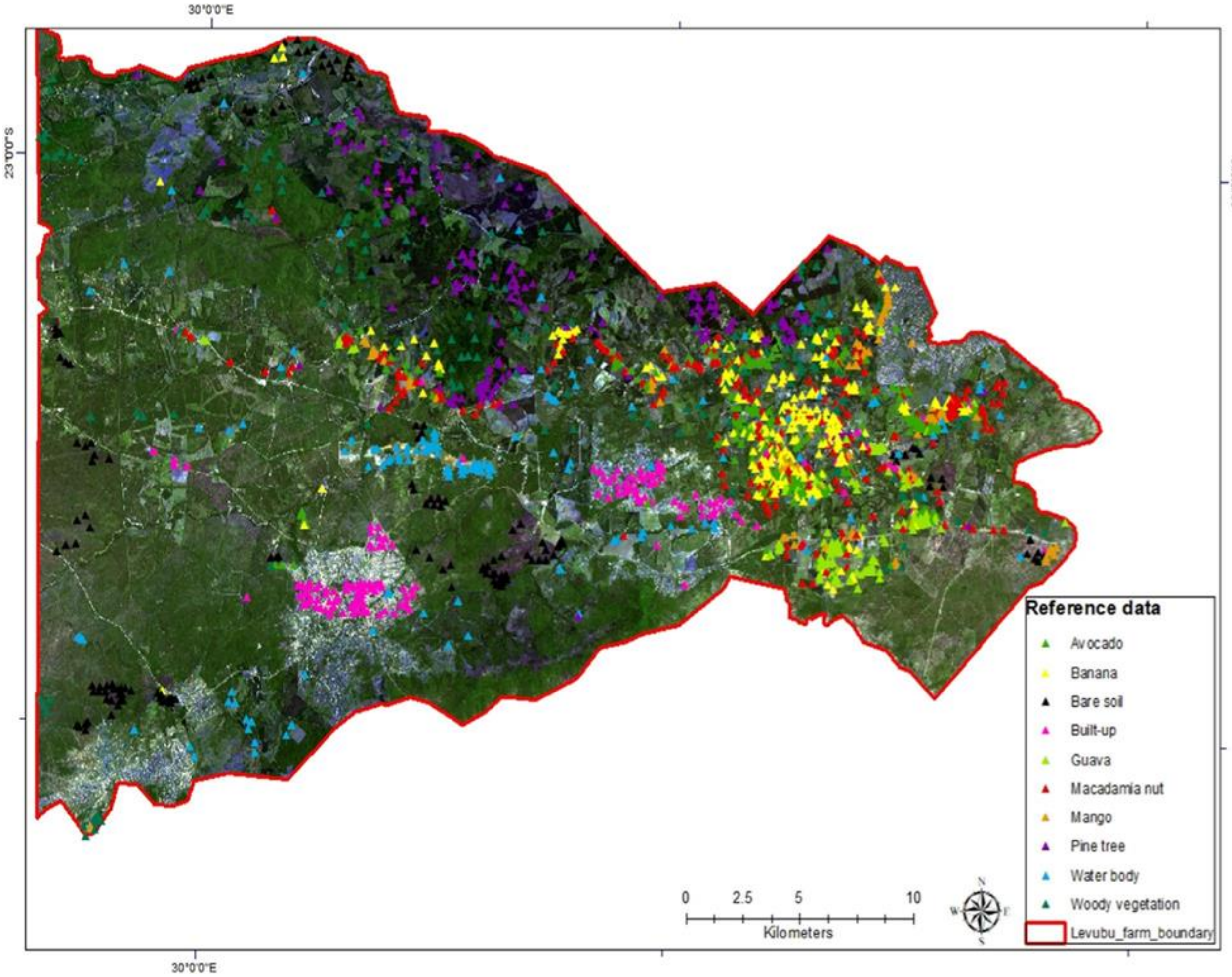


Figure 5.3: *The location distribution of fruit trees and other co-existing land-use types in the Levubu sub-tropical fruit farming area. The RGB bands represent S2 image downloaded from the European Space Agency.*

Table 5.1: *Phenology calendar of the major fruit trees in Levubu subtropical farms. The colour shading represents the flowering Flo fruiting (Fru) and harvesting (Harv)) stages.*

Class	Varieties	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
Avocado	Hass	Fru	Fru	Fru	Fru	Harv	Harv			Flo	Flo	Flo	Fru
	Fuerte	Fru	Harv	Harv						Flo	Flo	Flo	Fru
	Pinkerton	Fru	Fru	Fru	Harv	Harv			Flo	Flo	Flo	Fru	Fru
	Ryan	Flo	Flo	Fru	Fru	Fru	Fru	Harv	Harv				Flo
Banana		Harv	Harv	Flo	Flo	Flo	Fru	Fru	Fru	Fru	Harv	Harv	Harv
Guava		Fru	Fru	Fru	Harv							Flo	Flo
Macadamia nut	Beaumont 695	Fru	Fru	Fru	Harv	Harv			Flo	Flo	Fru	Fru	Fru
	Nelmak 2	Fru	Harv	Harv			Flo	Flo	Fru	Fru	Fru	Fru	Fru
	A4	Fru	Harv	Harv			Flo	Flo	Fru	Fru	Fru	Fru	Fru
	814	Fru	Harv	Harv			Flo	Flo	Fru	Fru	Fru	Fru	Fru
	816	Fru	Harv	Harv			Flo	Flo	Fru	Fru	Fru	Fru	Fru
	344	Fru	Harv	Harv			Flo	Flo	Fru	Fru	Fru	Fru	Fru
Mango	Tommy Atkins	Harv						Flo	Flo	Fru	Fru	Fru	Harv

5.5. Earth observation data and pre-processing

5.5.1. Sentinel-2 data

The Sentinel-2 (S2) images were downloaded from the Copernicus website (https://scihubcopernicus.eu/), which is owned by the European Space Agency (ESA). S2 consists of two satellites (Sentinel-2A and Sentinel-2B) with a temporal-spatial resolution of 5–6 days (Darvishzadeh *et al*., 2019). The data have 13 spectral bands with spatial resolutions of 10, 20, and 60 that are in different wavelengths ranging from 443 to 2190 nm and consist of three additional red edge bands with a central wavelength of 705 nm (Band 5), 710 nm (Band 6), and 783 nm (Band 7), respectively (Table 5.2). The visible and red edge bands have been proven to have capabilities to detect vegetation foliar properties due to their sensitivity to leaf chlorophyll and water content (Darvishzadeh *et al*., 2019). In this research, four Sentinel-2 time series images optimally acquired from April to July, acquired in 2019/2020, corresponding to key fruit tree phenological stages (i.e., flowering, fruiting, and harvesting) were used (Table 5.1 and Table 5.3). The utilization of optimal images acquired during critical crop growth stages has the capability to reveal the within-crop spectral similarities for accurate spectral separability of a wide range of crop types (Zhang *et al*., 2018).

Table 5.2: *Spectral characteristics of Sentinel-2 sensor*

Band	Description	Spatial resolution	Wavelength centre	Wavelength width
B2	Blue	10	490	65
B3	Green	10	560	35
B4	Red	10	665	30
B5	Red-Edge 1	20	705	15
B6	Red-Edge 2	20	740	15
B7	Red-Edge 3	20	783	20
B8	Near-infrared (NIR)	10	842	115
B8A	Red-Edge 4	20	865	20
B11	SWIR 1	20	1375	30
B12	SWIR 2	20	2190	180

Table 5.3: *Sentinel-2 image and acquisition dates*

Images	Acquisition dates	Day of Year
April	2020-04-06	97
May	2020-05-26	147
June	2019-06-11	163
July	2020-07-05	187

5.5.2. Image pre-processing

The selected Sentinel-2 images used in the research were atmospherically and geometrically corrected using the SEN2COR processor in the ESA SNAP software to obtain the reflectance data in Level-2A Top-Of-Atmosphere (TOA) format (Darvishzadeh *et al*., 2019). During pre-processing, the selected Sentinel-2 images were resampled to 10 m spatial resolution, and the remaining ten (10) bands (i.e., band 2, band 3, band 4, band 5, band 6, band 7, band 8, band 9, band 11, and band 12) were stacked together using ArcGIS version 10.6 to form the four Sentinel-2 images used in this research. The corrected images were classified using a Deep Neural Network (DNN).

5.4. Deep learning models

A Deep Neural Network (DNN) algorithm that uses multiple artificial neural networks and has the capability to model complex non-linear objects (Hu *et al*., 2016). DL can (a) extract features directly from the dataset, (b) use a deep network to learn hierarchical features, and (c) automate predictions, making it more robust and generalized than traditional ML classifiers (Biffi *et al*., 2021; Ukwuoma *et al*., 2022). DL methods have high complexities and are designed to solve complex problems with non-linear transformations (Biffi *et al*., 2021; Bargiel, 2017).

The study employed Keras Python 3.6.13 libraries on top of Tensorflow 2.4.1 using an interactive Jupyter Notebook to implement the DNN model. Typically, the DNN model architecture is determined using four parameters, namely, the number of neurons, the number of layers, the learning algorithm, and the activation function. For the study, three layers/neurons, one input layer consisting of 10 classes, two hidden layers, and one output layer consisting of 10 classes. The layers were added sequentially to identify the best network architecture—all neurons were activated to achieve the best model performance using the ReLu and SoftMax activation functions (Tian *et al*., 2019a). The transfer learning approaches were applied by fine-tuning the model using different image batch sizes of 10, 50, and 100. The model performance was optimized by applying the glorum_uniform, which draws samples from a uniform distribution (Amani *et al*., 2020). Furthermore, other optimization functions such as ADAM, RMSprop, and SDG were tested with a learning rate of 0.1 and 1 to identify the differences between in-situ data and network outputs (Loss Function) (Ukwuoma *et al*., 2022; Li *et al*., 2022). Different iterations with 50 and 100 epochs were tested to identify the model that would be more sensitive to fruit trees and co-existing land use types and give the best classification accuracy (Amani *et al*., 2020). The dropout rate was set to 0,0.3, and a 'max' mode was used for early stopping when the model accuracy stopped improving (Amani et al., 2020). The trained models were assessed using the test dataset. The loss and accuracy of the training and validation data were recorded, and the optimum overall accuracy was computed at the last batch/epoch (Xiong *et al.*, 2022).

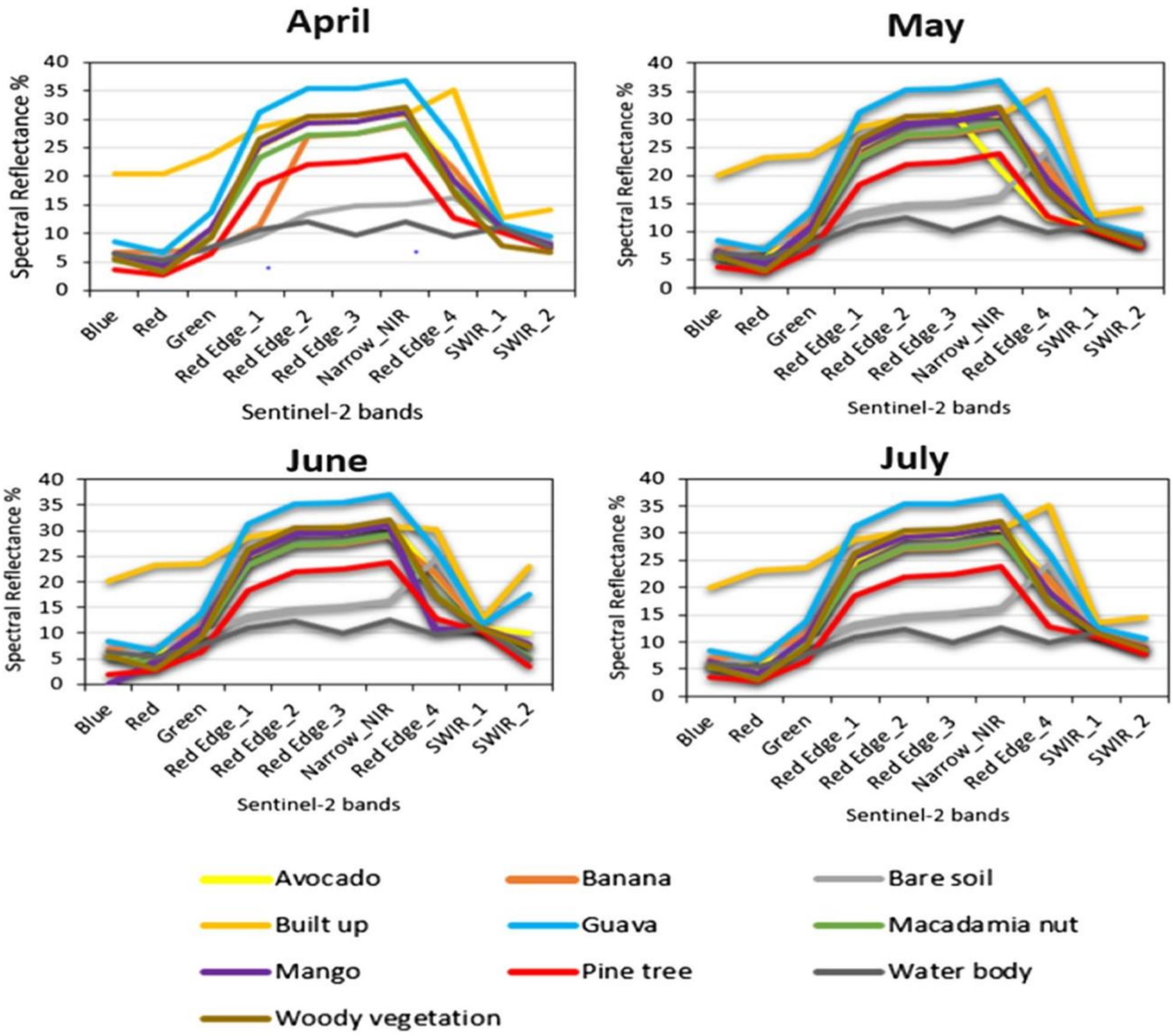


Figure 5.4: *The variations in temporal spectral reflectance curves of within-season phenological stages of fruit trees and other land use types derived from Sentinel-2 multispectral images acquired over a different time range (April, May, June, and July). All fruit trees and their co-existing land use types are plotted together.*

5.5. Accuracy assessment

Accuracy assessment is the process of evaluating the classification results as a way of assessing the reliability of the classification approach proposed in research (Amani *et al*., 2020). The results of this research were assessed using a confusion matrix which was calculated using test samples (30% of the dataset) (Amani *et al*., 2020). The performance of the DNN model in classifying the optimal images was assessed by computing the performance evaluation metrics user's accuracy (UA), producer's accuracy (PA), and overall accuracy (OA), which were computed from the confusion matrix of the four DNN models applied in the study.

5.6. Results

5.6.1. Overall accuracies of the DNN models

Figure 5.5 presents the accuracy and loss curves on the validation and training datasets of Model 1 (April image), Model 2 (May image), Model 3 (June image), and Model 4 (July image). The curves show the relationships between model accuracy and loss function, indicating the performance and stability of the models across the datasets were tested using 100 epochs. In all models, the learning curves are smooth, suggesting that the models were robust in learning the data and had the capacity to produce promising results. A direct correlation is observable between the training and testing data.

In April, the loss function of the DNN model is around 40 epochs and stabilizes around 60 epochs. In May, the loss started at epoch 40 and saturated around 57 epochs. The loss started at 5 epochs and saturated early at around 49 in June. A similar trend was observed in July, where a loss was experienced at three epochs and reached a maximum performance at around 67 epochs. In terms of accuracy, the May model achieved a superior performance of 87% compared with the other models, indicating that the recall and overall precision performance of the DNN model in May were better than in the other months (April, June, and July).

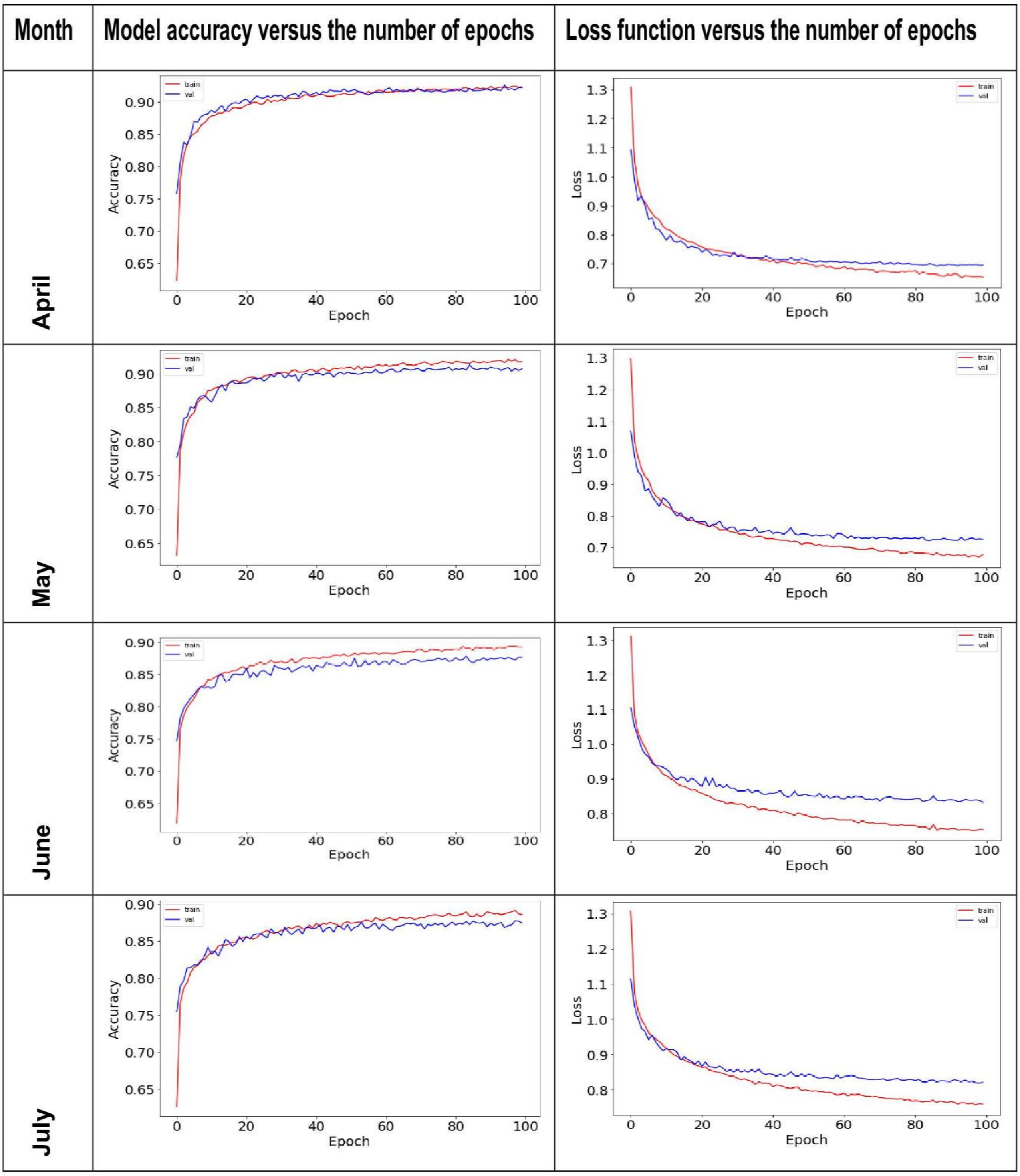


Figure 5.5: *Comparison of the variation of model accuracy and loss function curves with increasing epochs on the prediction of the DNN models tested in different phenological stages of the fruit trees during April (Model 1), May (Model 2), June (Model 3: and July (model 4).*

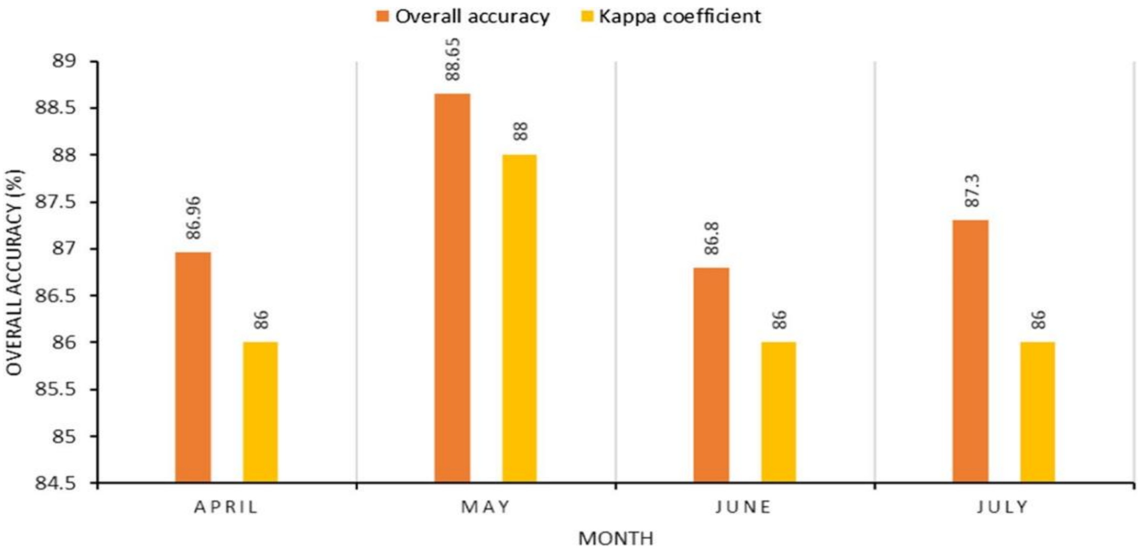


Figure 5.6: *Relationship between the model accuracy and the loss function, showing the stability of the models across the tested 100 epochs.*

5.6.2. Performance of the classification models

The overall classification accuracies of the models followed a similar order of 88.64%, 87.25%, 86.96 and 87.25% (Figure 5.6). The kappa coefficient values ranged from 86% to 88% for May to July, respectively. The May image was decisive and made an important contribution towards identifying the fruit trees and co-existing land use types.

Figure 5.7 shows the validation results of each fruit-tree species and co-existing land use types. These figures show that better class accuracies were achieved using the images from late autumn (May) and late winter (July).

In April (Figure 5.7A), the producer accuracy (PA) ranged from 64.76 to 98.66% for the avocado (AV) and woody vegetation (WV), respectively. The other crops (i.e., banana (BN), bare soil (BS), build-up (BU), pine tree (PT), macadamia nut (MN), and water body (WB) had a PA accuracy of 90% and above. The user accuracy (UA) ranged from 62.96% for the avocado class to 98.66% for the BU class.

In May (Figure 5.7B), the PA ranged from 64.95 to 100%, with the AV class recording the lowest while the WB class had the highest. An increase was observed for all classes except for the BU, MN, and WB. The PA for the BU class decreased by 1.23%. The UA increased for all classes, ranging from 70.79 to 99.37%.

In June (Figure 5.7C), the PA ranged from 67,39 to 98.01%. The UA decreased from 83.01% to 72% for the GV class, while for the mango (MG) and BN classes, it decreased from 84.75 to 75,23% and 96.83 to 93.24%, respectively. The UA ranged from 65.96 to 96.88 for the WB class. There was a decrease in UA values recorded for AV, BU, GV, MN, PT, WB, and MN.

In the July image (Figure 5.7D), the PA ranged from 68 to 99%, while the UA ranged from 65 to 100%. The PA value for the avocado crop was 67.33%, indicating that 32.67% of pixels attributed to avocado were misclassified. The UA values ranged from 65 to 100%, showing an increase for the BN, GV, MN, MG, and WB.

5.6.3. Mapping outputs

A DNN model was applied to four Sentinel-2 monthly images (April, May, June, and July) and used to distinguish the presence of fruit trees and surrounding land use types in subtropical farms in Levubu, South Africa.

Figure 5.8 shows pixel-based classification maps produced using four optimal Sentinel-2 images acquired in April to July and classified using the DNN algorithm. The proposed approach made it possible to distinguish and classify fruit trees and co-existing land use types. As depicted in 6A, the pine trees are preferentially found in high-elevation areas on the northern side of the research area. During this month, the avocado crop overlapped with the macadamia nut (MN), guava (GV), and mango (MG) crops, resulting in high spectral confusion between the classes (Table 5.4A). This could have been attributed to the fact that some of the avocado cultivars (Hass, Ryan) are fruiting, while the (Fuerte cultivar) is in the senescence stage and the AV (Pinkerton cultivar) is in the harvesting stage. These phenological stages coincide with the senescence stage of the mango crop and some macadamia nuts (A4,814,816,344), while the Beaumont 695 and Nelmark 2 cultivars are in their harvesting stages (Table 5.1).

Table 5.4: *Confusion matrices for the fruit trees and co-existing land use types in four growing months (April, May, June, and July).*

(A) DNN APRIL										
Class	AV	BN	BS	BU	GV	MN	MG	PT	WB	WV	Total
AV	68	3	1	1	11	9	12	0	0	3	108
BN	3	141	1	0	2	0	3	0	0	0	150
BS	2	1	79	1	0	1	0	0	0	2	86
BU	0	0	0	147	1	0	0	0	1	0	149
GV	6	1	1	0	56	0	4	0	0	0	68
MN	12	0	0	0	0	76	13	6	0	1	108
MG	9	0	0	0	4	11	105	1	0	1	131
PT	1	0	0	0	0	2	0	135	0	3	141
WB	0	0	1	0	0	0	1	0	30	1	33
WV	4	0	2	0	1	1	4	6	0	197	215
Total	105	146	85	149	75	100	142	148	31	208	1189
OA: 86.96%											
kappa: 86%											

(B) DNN MAY											
Class	AV	BN	BS	BU	GV	MN	MG	PT	WB	WV	Total
AV	63	2	0	1	3	12	7	1	0	0	89
BN	6	122	0	0	1	2	1	0	0	0	132
BS	1	1	82	2	3	0	0	3	0	1	93
BU	0	0	0	158	1	0	0	0	0	0	159
GV	11	0	0	0	59	1	2	0	0	1	74
MN	10	1	0	0	0	88	7	2	0	5	113
MG	4	0	0	0	2	11	100	2	0	6	125
PT	0	0	1	0	2	4	1	152	0	5	165
WB	0	0	0	0	0	0	0	0	32	1	33
WV	2	0	0	1	0	1	0	1	0	198	206
Total	97	126	83	162	71	119	118	164	32	21	1189
OA: 88.65%											
kappa: 88%											

(C) DNN JUNE											
Class	AV	BN	BS	BU	GV	MN	MG	PT	WB	WV	Total
AV	51	1	0	1	11	6	9	0	0	3	94
BN	0	138	2	1	0	2	4	0	0	0	148
BS	0	3	75	1	2	0	0	2	2	0	85
BU	0	0	1	148	2	0	0	0	0	0	151
GV	8	1	0	0	54	0	4	0	0	2	72
MN	16	1	0	0	0	78	6	5	0	4	110
MG	4	3	0	0	5	11	82	0	0	3	108
PT	0	0	4	0	0	5	1	152	0	4	166
WB	0	0	0	0	0	0	0	0	32	1	33
WV	1	0	2	0	1	4	3	2	1	208	222
Total	92	148	82	151	78	106	109	161	35	225	1189
OA = 86.80%											
Kappa = 0.86%											

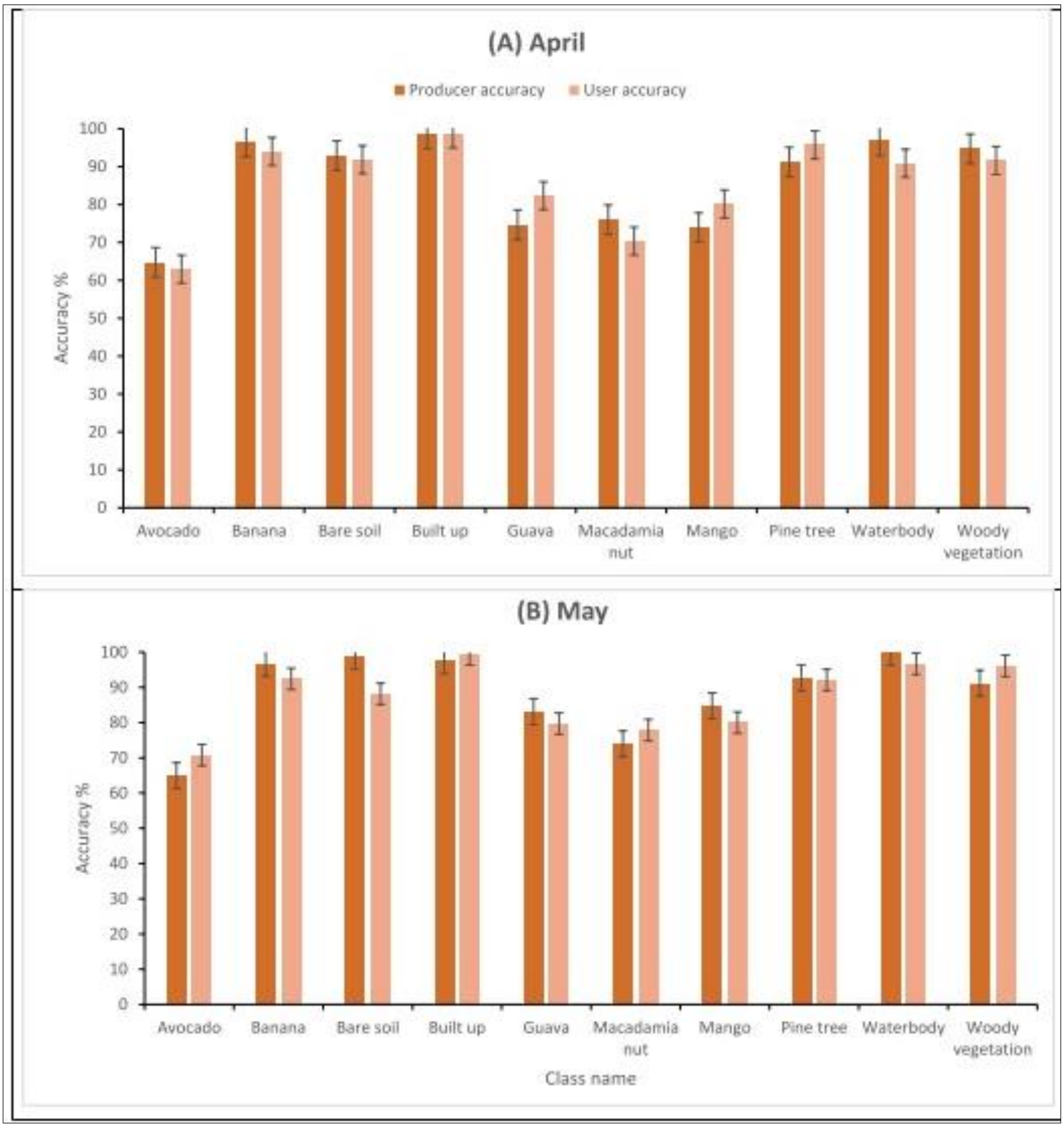
(D) DNN JULY											
Class	AV	BN	BS	BU	GV	MN	MG	PT	WB	WV	Total
AV	68	2	0	1	10	9	11	0	0	3	104
BN	3	108	3	1	3	1	0	0	1	0	120
BS	2	0	73	0	0	0	0	2	0	0	77
BU	1	0	1	161	0	0	0	0	0	0	163
GV	3	0	3	0	50	2	1	0	0	0	59
MN	11	0	2	0	0	91	15	2	0	1	125
MG	9	0	0	0	2	13	90	0	0	5	119
PT	1	0	0	0	0	7	1	142	0	5	156
WB	0	0	0	0	0	0	0	0	33	0	33
WV	3	0	0	0	0	2	2	5	2	219	233
Total	101	110	82	163	65	128	120	151	36	233	1189
OA = 87.30%											
Kappa = 86%											

The land use classes are, namely, *AV* avocado, *BN* banana, *BS* bare soil, *BU* built-up, *GV* guava, *MN* macadamia nut, *MG* mango, *PT* pine tree, *WB* water body, and *WV* woody vegetation.

In Figure 5.8B, the classification during May showed misclassification between banana (BN), MN, and MG, although the spatial distribution is similar to April. However, it is evident in Table 5.4B that 11, 10, and 6 pixels belonging to the avocado crop were misclassified as GV, MN, and BN, respectively. Furthermore, pixels 12 and 11 belonging to the mango crop were misclassified as avocado and woody vegetation (WV), while 6, 5, and 5 pixels of the WV class were misclassified as mango, guava, and macadamia nut, respectively. In Figure 5.5, it was observed that there are spectral similarities between these classes due to overlapping phenological stages. The spectral similarities emanated from the senescence stage of the mango crop, which coincides with the senescence stage of the majority of the MN varieties (A4, 814, 816, 344) and that of the avocado, specifically the Fuerte cultivar.

The image acquired in June (Figure 5.8C) was able to detect the banana and mango crops located within the built-up areas on the southern side of the study area, which decreased the spectral confusion between the crops (Table 5.4C). However, spectral confusion was observed within the avocado, mango, and guava crops. The AV (Hass) cultivar was in the harvesting stage, and the AV (Fuerte, Pinkerton) cultivars together with GV, MN (Beaumont 695), and MG (Keitt) were in their senescence stages, while the AV (Ryan) variety, the BN and MN (Nelmark 2, A4, 814, 816, 344) and MG (Sabre cultivar) were in flowering stages.

Figure 5.8D shows the spatial distribution of the fruit trees in July. It was observed from the confusion matrix that 11, 10, and 9 pixels corresponding to the avocado crop were misclassified as mango, guava, and macadamia nut, respectively (Table 5.4D). During this month, all crops were in their fruiting and flowering stages except for the MN (Keith cultivar), the GV, and the AV (Hass, Fuerte) cultivars. Overall, all fruit trees were accurately detected during the wet season (i.e., June and July) as they correspond to the flowering and fruiting seasons, which are easily detected by the optical sensor due to the high presence of chlorophyll and water contents in leaves.



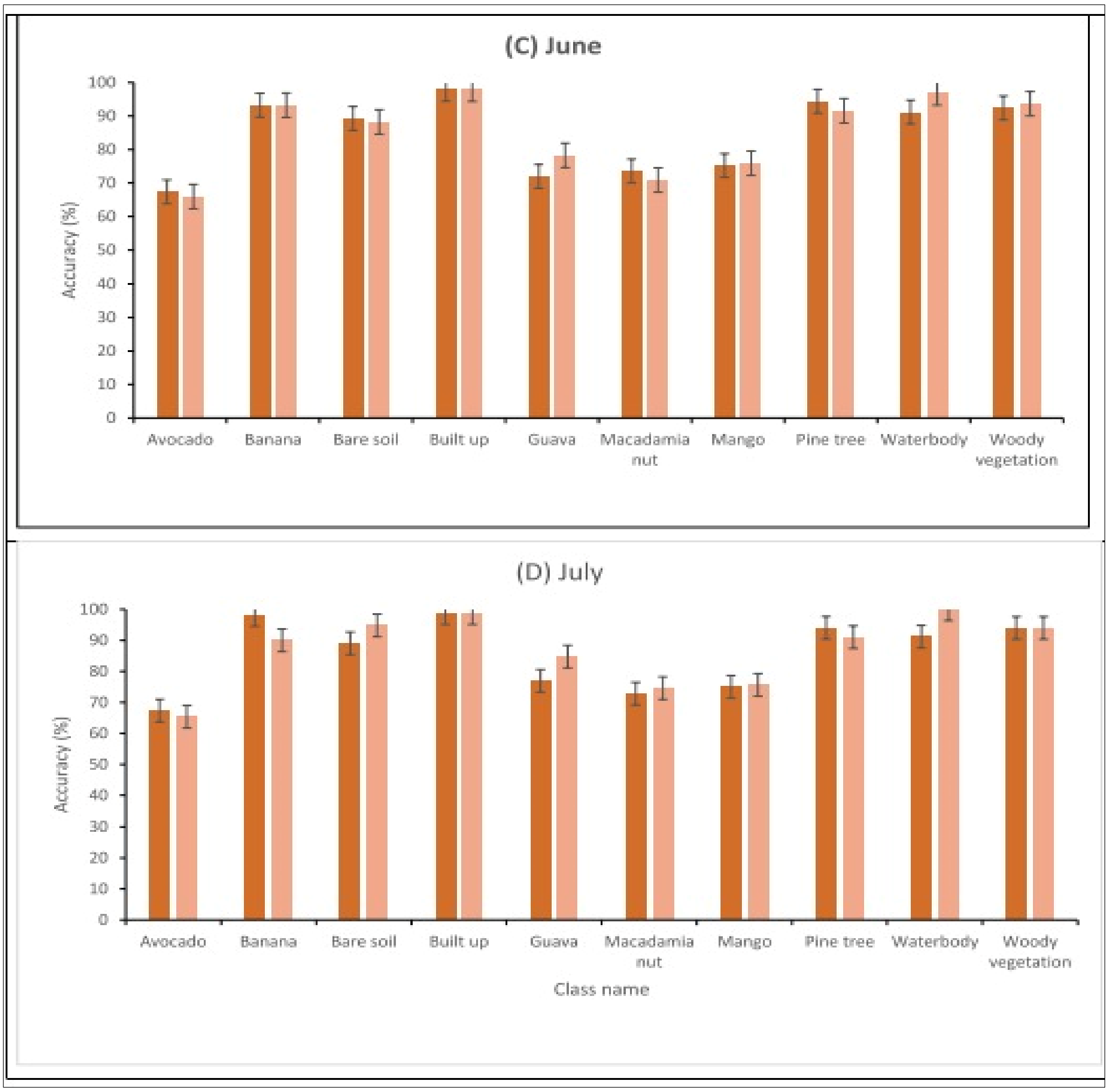


Figure 5.7: *The User Accuracy and Producer Accuracy of the fruit trees and co-existing land use types in four growing months (April, May, June, and July).*

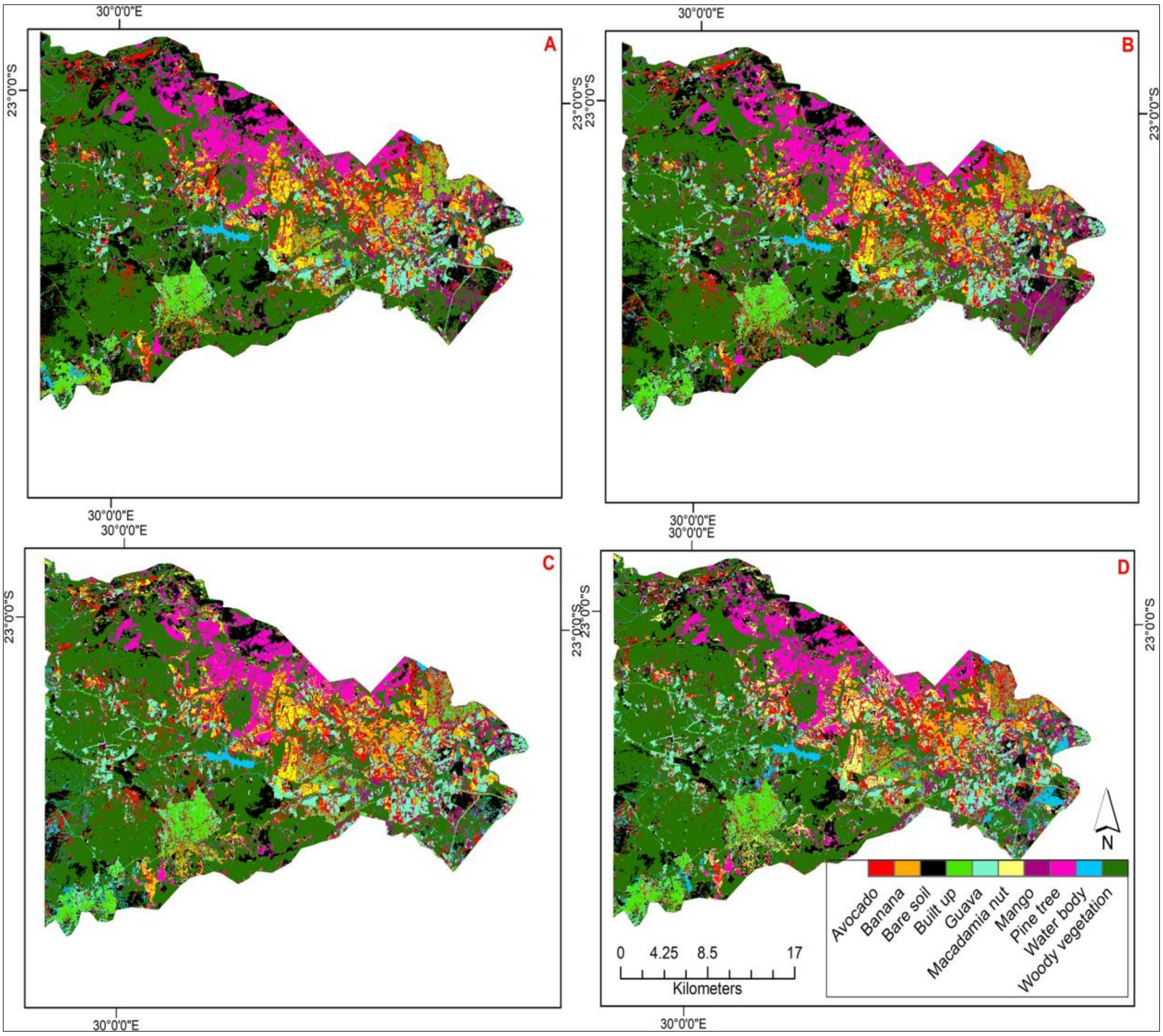


Figure 5.8: *Discrimination results for fruit trees and co-existing land use types by DNN and Sentinel-2 images acquired in A April, B May, C June, and July.*

Figure 5.9 shows the visual enhancement through the provision of zoomed-in maps of the regions with classification irregularities (Figure 5.9: inset a, b, and c). The first inset region (Figure 5.9, inset (a) is located in the northwestern part of Levubu, where mountains and irregular patches of pine trees exist. The approach applied made it possible to distinguish pine trees from woody vegetation. The dominant fruit trees were avocado and macadamia nut. Similarly, the applied approach was able to classify the fruit trees in the central north of Levubu [Figure 5.9, May, inset (b)]. The guava, avocado, mango, and pine trees have dominated this region. Furthermore, the approach was also effective in classifying fruit and co-existing land use areas in the northeastern part of Levubu [Figure 5.9, June, inset (c)]. This region is dominated by banana, guava, and avocado.

In Figure 5.9, inset (a), located in the Southeastern part of the study area, shows the differences in the detection of the guava crop across the four months. However, in April, a few pixels of the guava crop were detected by the model. In May, June, and July, more pixels of the guava crop were detected, as depicted on inset b-c, however, some of the bare soils and avocado and woody vegetation classes. The central part of the study area, depicted by inset (b), shows the spectral confusion between the guava, avocado, and woody vegetation in April and May, while in June and July, the same areas are correctly classified as guava, avocado and macadamia nut. On inset (c), some of the pixels of the woody vegetation were misclassified as guava and built-up areas during June. However, the water features present around those areas were detected in July, while they were slightly detected in April to June image acquisitions.

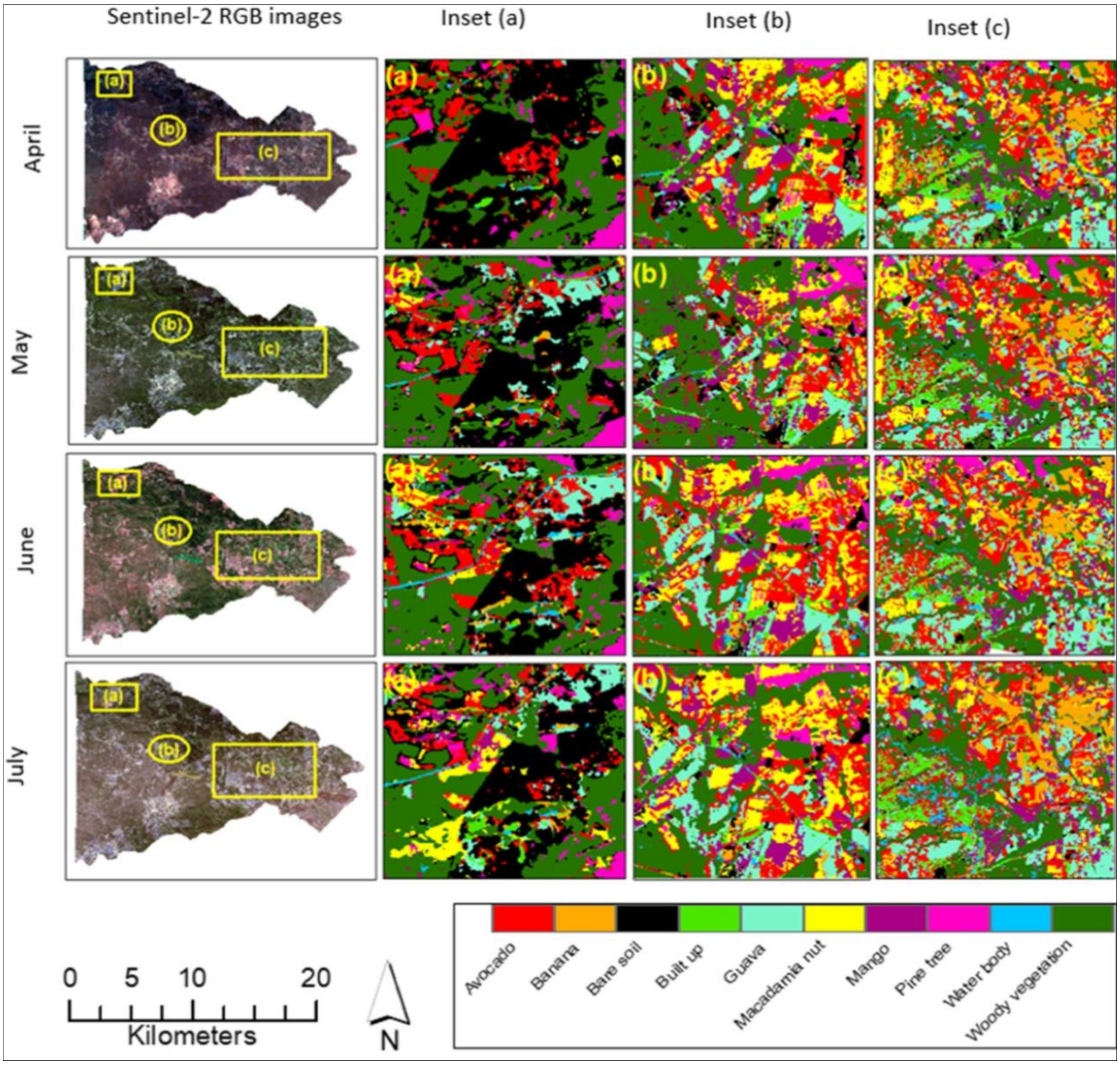


Figure 5.9: *Fruit tree maps obtained using DNN with some zoomed spots showing the spatial structure of the classified fruit crops and co-existing land use types based on Sentinel-2 images across growing seasons in April, May, June, and July. The a–c depicts the zoomed areas of three sites with major classification differences obtained using the DNN classifier across the four months considered in this study detected in July, while they were slightly detected in April to June image acquisitions.*

5.7. Discussion

5.7.1. Overall classification accuracies

Fragmented smallholder landscapes dominated by fruit trees with similar spectral characteristics present difficulties in terms of the performance of machine learning classifiers. This situation can be overcome through the application of DL models and new-generation sensors with a high temporal and spatial resolution, such as Sentinel-2 (Lanaras *et al*., 2018). The research constituted an investigation into the utility of using Sentinel-2 (S2) optimal multispectral data and phenological information integrated with DNN for mapping heterogeneous fruit trees and co-existing land use types in Levubu, South Africa. The performance of S2 multi-temporal images across crop growth season was compared using the overall accuracy, user, and producer accuracies. The approach produced satisfactory results, and superior results of 88.64% OA were achieved for the month of May. As shown by the results in Figure 5.6, the model performance was not affected by the image acquisition gaps from April to July, corroborating findings by Zhou *et al*., 2022. In their research, an OA of 91% was achieved using the ANN model, constituting an 8% improvement on the RF results. The accuracy obtained in the research is comparable with the study by Pena *et al*. (2017), who obtained 85.56% when classifying crop types using DL and Sentinel-2-time series data. In this research, the spectral features extracted in late winter (July) contributed 0.87%, whereas those extracted in early autumn (April) and early winter (June) contributed 86.96% and 86.8% to the classification model. Although the research produced various accuracies, an OA of 80% and above was achieved in both months. This suggests that the optimal period for mapping fruit trees in Levubu is autumn to winter (i.e., April to July). Similar to observations in Paris *et al*. (2020), the OA was low during the early seasons of crops, but an increase in OA was achieved as the seasons progressed. Identifying the optimal window period has proven to be an alternative to the utilization of all-time series images present in the year. The results are comparable with studies that used image composites from time series data (Kordi and Yousefi, 2022b; Vuolo *et al*., 2018).

5.7.2. Comparison of spatiotemporal accuracies of individual classes

In terms of individual class accuracies, better results were achieved for specific crops (Figure 5.7 and Table 5.4 (A–D)). The spectral differences with respect to fruit trees were large in late autumn and late winter, as most of the fruit trees were in the vegetative stage (flowering and fruiting). The early seasons are characterized by low vegetation cover and high reflectance originating from the soil background, and they provide little contrast between crops and their spectral differences (Vuolo *et al*., 2018). In May, the banana and mango crops begin production and seed development, while macadamia nut and mango begin to senesce. The structural differences were evident in May, with the banana and avocado crops (Ryan) cultivar in fruit development while the macadamia nut (Beaumont 695) cultivar, avocado (Pinkerton and Hass) cultivars, and guava crops were in the senescence and harvest stages. Similar to what was reported by Vuolo *et al*. (2018), low accuracies were obtained using the image acquired in June. The low accuracies for the avocado crops in the tested months are attributed to the sensitivity of the DNN model and the robustness of the applied classification approach to subtle changes.

In contrast, the differences in the spectral reflectance of types of fruits were more prominent later in the growing season due to green vegetation, increased moisture content, and canopy structure (McNairn *et al*., 2009). Similarly to our findings, the research by McNairn *et al*. (2009) reported the significance of late growing season optical acquisitions for crop discrimination. Applying spectral features at different times increased the distinction between fruit trees and co-existing land use types (Amani *et al*., 2020).

The change in the fruit tree canopies was not detected in April but was indistinctly captured in May. This emphasizes the need for phenological information captured by Sentinel-2 (S2) data due to its high-frequency interval. The temporal resolution of the S2 data makes it possible to detect these variations at the fruit-tree level. Corroborating findings by Zhou *et al*., 2022, the accuracy for the mango crop was lower than for the other fruit trees, which was expected because of the S2 image resolution. In June and July, the banana and avocado crops (Ryan cultivar) are in the vegetative stage with maximum biomass, as they are fruiting, while the biomass of avocado and guava crops is declining, as they are in the senescence and harvesting stages. During these stages, the sparse canopy correlates with less vitality in guava crops, while dense canopies relate to high vitality in pine trees (Bai *et al*., 2016). For example, the optical image acquired in May permitted the best classification of the banana and the guava crops, as they were at the greenup and full canopy stage, which correlates with maximum photosynthesis (Schirrmann *et al*., 2016). At the senescence stage, the leaves have low water content, leaf area index, and dry biomass (Bai *et al*., 2016). The results indicate the phenology stage in which each fruit tree type is best distinguished according to the crop calendar, making optimization of the imagery collection schedule possible (Paris *et al*., 2020). The crop classification in winter (May–July) detected the within-class variations emanating from fruit-trees structure, local calendars, and different management strategies (Paris *et al*., 2020).

Although satisfactory results were achieved for all four months, the UA and PA for the avocado crop were low. This might be attributed to differences between avocado varieties, exacerbated by mixed pixels arising from spectral similarities among the different fruit trees grown in the area, as cited in previous research on fruit trees (Chabalala *et al*., 2023a, 2023b, 2022). Furthermore, prioritizing the image acquisition based on the phenological stage increases the chances of the crops being spectrally distinguished with high accuracy (Pan *et al*., 2021). Figure 5.4 shows a high spectral overlap between the land use classes at different months, which challenged the classification accuracy. This was evident on the classification map (Figure 5.8) and the confusion matrix (Table 5.4 (A–D)), where misclassifications were observed. Overall, spatial and accuracy variations were observed between fruit trees and within cultivars, which further complicates the mapping process.

5.7.3. The spatial distribution maps

The seasonality of fruit trees and co-existing land use types were analyzed by producing spatial distribution maps for a period of 4 months covering different crucial growth stages. From Figure 5.8, a shifting pattern in the spatial distribution of the mapped fruit trees and co-existing land use types is apparent. The April and May images underestimated the avocado crop, while the same crop was overestimated in early and late winter (June and July). In late autumn (May), reasonable spatial patterns were generated that concurred with the researcher's knowledge.

On the northeastern side, the optimal time to separate from other fruit trees appears to be late May, when vegetation reflectance dominates the optical signals. The low classification results for avocados were associated with April and May acquisitions, suggesting that for increased accuracy, the mid to late-season optical acquisitions must be included in the classification. Unlike ML, DL models are consistent with regard to accuracy, as witnessed in this research, rendering them suitable for large-scale mapping (Pan *et al*., 2021).

The pigmentation, structure, and water content of crops change during their phenological stages from flowering to harvesting. The optical data are sensitive to differences in vegetation canopies (Tain *et al*., 2019b). Furthermore, the clustering densities of fruit trees vary within phenological stages (Tain *et al*., 2019b)]. During the fruiting stage, fruits have different sizes and colours and have sparse distribution (Saedi and Khosravi, 2020). The volume of fruit trees is large during the flowering period, the leaf colour changes and the clustering density decreases after harvesting due to pruning and other management strategies (Saedi and Khosravi, 2020). The effectiveness of the spectral reflectance decreases as fruits mature and senesce before harvest based on fruit variety and farmers' management strategies (Asgarian *et al*., 2016). Depending on the aim of the farmers' assessments, the zoom-in maps become useful for other agricultural operations such as crop change monitoring, irrigation scheduling, and fruit disease scouting. For agricultural management and food security, the classification maps produced through the research reported on in this article could support the development of appropriate management strategies that will increase food production.

5.8. Limitations

Deep Learning is computationally extensive and requires a large amount of labelled ground truth data and cadastral datasets to train the network; however, these datasets are very scant in Africa. The developed mapping tool can be transferred to other areas with similar geographic landscapes. However, model uncertainties will always exist because the overall model accuracy is pixel-specific and geographically specific and often varies per training data set.

5.9. Conclusions

The research developed a DNN-based mapping tool to improve the classification of fruit trees and co-existing land use types in a smallholder horticulture system in Levubu, Limpopo Province. The developed mapping tool identified the optimal phenological stages at which the fruit trees peak production and assessed how the DNN accuracy behaves across seasons. It was discovered that the developed mapping approach is robust and resilient to variations in climate, farming management strategies, and crop growth conditions as it considers both phenological stages and the growing period of fruit trees. Given the complex cropping systems in Levubu sub-tropical sites, individual classes were well classified. Regardless of the complexity of the cropping systems in Levubu, the mapped ten classes were distinguished with an overall accuracy of 89% and a kappa coefficient of 88%.

The findings suggest that selecting optimal image acquisition is important in discriminating crop types. The accessibility of optical imagery with high spatial resolution and temporal coverage is suitable for per-field crop type mapping and continuous refinement of spectral–temporal profiles for common crop types in different agronomic regions. This is expected to improve the classification accuracy of crop-type maps using these profiles. The research highlights the importance of Sentinel-2 data and deep learning models in mapping fruit trees and coexisting land use types in smallholder food systems for enhanced decision-making and the development of resilient food systems that will enable the attainment of food security. The application of phenological information derived from optimal S2 images showed high potential for fruit tree crop classification in fragmented sub-tropical regions. The results can be used as a valuable tool to develop sustainable horticultural management practices for the regulation of fruit tree crops in complex agricultural landscapes under different environmental conditions.

CHAPTER SIX

Integrating Sentinel-1/2 and machine learning models for mapping fruit tree species in heterogeneous landscapes of Limpopo: A synthesis

6.1. Introduction

The uncertainties of the spatial distribution of crop types have been identified as a major factor contributing to inadequate crop management and food insecurity (Nabil *et al.*, 2020; Elders *et al.*, 2022). As a result, the African continent is still food insecure, even in this era of unprecedented amounts of remote sensing (RS) data that can be applied to devise products that can provide solutions (Tariq *et al.*, 2022). The availability of accurate spatial distribution of crop types is important for effective crop research and for developing appropriate management strategies in the future (Chabalala *et al*., 2022).

Crop types in African agriculture have been mapped using field-based methods, which are laborious, time-consuming, subjected to human error, inefficient and strenuous in hard-to-reach areas (mountainous and heterogeneous landscapes) (Chabalala *et al*., 2023). Thus, RS products are used to collect vast amounts of crop data over various spatial coverages in a timely and efficient manner (Chabalala *et al*., 2022). However, African agriculture is mostly small-scale farming systems, characterised by intercropping from various cultivars with different management strategies, which results in complex overlapping crop canopies that are difficult to detect and monitor (Chabalala *et al*., 2023b). Mapping crop types in small-scale systems is challenging; hence, there is no universal mapping approach due to differences in crop type environmental factors and management strategies (crop calendars, irrigation scheduling, etc.) (Abubakar et al., 2020; Conrad *et al*., 2020).

Crop type mapping was previously done using field-based methods, which are time-consuming and impractical over large areas (Minaei *et al.*, 2022; Alabi *et al.*, 2022). The introduction of very high-resolution (VHR) data has been cited as suitable for mapping small-scale food systems in previous crop types (Kraaijvanger *et al.*, 2019; Roth *et al.*, 2019). However, the data is commercialized, which hinders its application in Africa due to a lack of funds. However, its usage was revolutionalised by technological advances through synthetic aperture radar (SAR) and unmanned aerial vehicles (UAVs). SAR and UAVs have been used to map fruit trees (Csillik *et al.*, 2018; Zhao *et al.*, 2023). However, it is worth noting that regardless of the spatial resolutions and mapping methods, challenges are still anticipated in each data type when mapping crop types in small-scale farming systems. Due to a lack of research on African agriculture, very few methods have been tested; thus, there is a need for the development of remote sensing-based mapping methods that are transferable to landscapes with similar characteristics. Although the research was conducted at one study site, Levubu subtropical farms in Limpopo Province. The training and testing samples used to map the fruit trees and co-existing land use types differ per objective depending on the requirements of the methods and classifiers.

In this thesis, fruit trees and co-existing land use types were mapped using different datasets in conjunction with Machine learning and deep learning techniques.

The objectives of the research were:

1. Exploring the effect of balanced and imbalanced multi-class distribution data and sampling techniques on fruit-tree crop classification using different machine learning classifiers and remote sensing data
2. To test the effectiveness of machine learning classification of fused Sentinel-1 and Sentinel-2 Image data towards mapping fruit plantations in highly heterogeneous landscapes.
3. To identify the optimal phenological period for discriminating fruit tree crops using multi-temporal Sentinel-2 data and Google Earth Engine in a heterogeneous subtropical horticultural region.
4. To assess the applicability of Sentinel-2 and Artificial Neural Networks in classifying dynamic phenological cycles in horticultural crops.

6.2. Exploring the effect of balanced and imbalanced multi-class distribution data and sampling techniques on fruit-tree crop classification using different machine learning classifiers and remote sensing data

The prevalence of heterogeneous landscapes in fragile and insecure regions often results in shortages of field samples needed to train the classifier (Qader *et al.*, 2021). The smallholder agricultural landscapes are characterized by intercropping systems and scattered small agricultural fields that consist of common and uncommon crop types, resulting in imbalanced samples that limit the performance of conventional classifiers (Chabalala *et al*., 2022). Therefore, data sampling can be used as an alternative approach for land cover classification conducted using imbalanced data (Quan *et al*., 2021). Based on existing literature, more emphasis has been given to improving the accuracy of minority or rare crops and the most applied approaches are oversampling and undersampling (Waldner *et al*., 2019). However, in smallholder agriculture, all crops are equally important. For efficient farm management, this objective aimed to assess the influence of balanced and imbalanced multi-class distribution by undersampling the original datasets into 6 datasets, i.e., 2 balanced and 4 imbalanced and comparing them to the original imbalanced dataset. The seven datasets were used as input into five machine learning classifiers, which are random forest (RF), support vector machine (SVM), eXtreme gradient boost (XGBoost), gradient boosting (GB), and adaptive boosting (AdaBoost) and used to discriminate fruit trees species. The classification was conducted at the pixel-based level of Sentinel-2 (S2) data.

The results indicated that the highest classification accuracy of 71 % can be achieved using the SVM algorithm classifier as compared to AdaBoost (67%), RF (65%), XGBoost (63%), and GB (62%), respectively. The class accuracy recorded an F1 score ranging from 60% to 100%, respectively. The study further assessed how the sensitiveness and stability of the algorithms are affected by the size and ratio of class imbalance in the training datasets. It was discovered that the selected ML classifiers are not affected by class imbalance, as the highest classification accuracy of 71% was achieved using an imbalanced training dataset of 60% of the original dataset. The research further assessed the importance of the Sentinel-2 bands across the seven datasets using permutation feature selection. The performance classifiers (i.e., RF, SVM, Ada Boost, GB, XGBoost) were different across the datasets and were significant (P = 0.001) using the XGB, and GB, RF, SVM and AdaBoost. The results showed that the RF and AdaBoost are sensitive to sampling techniques, while the other classifiers (SVM, GB, XGBoost) are not affected by data sampling. A shifting pattern was observed in the importance of Sentinel-2 bands across the sampling techniques. The red band (B4: 665 nm) was ranked as the most contributing explanatory variable on the classification model. This was followed by the RedEdge_2 (B6: 740 nm) and the SWIR bands (B11, B12: 1610-2190 nm), as they are sensitive to chlorophyll and moisture content.

6.3. To test the effectiveness of machine learning classification of fused Sentinel-1 and Sentinel-2 Image data towards mapping fruit plantations in highly heterogeneous landscapes

Mapping and monitoring spatial distribution changes of crop types provide a reliable source of information that supports agricultural policies and helps attain environmental sustainability (Valcarce-Diñeiro *et al.*, 2019). Accurate discrimination of crop types using single-date optical images is hindered by the inability to capture the temporal behaviour of crops during crucial crop growth cycles due to the persistence of clouds, making it difficult to discriminate fruit trees due to the poor quality of the available images (Xiong *et al.*, 2022; Singh *et al.*, 2021). In such cases, using single-date images is not reliable for crop identification (Chabalala *et al.*, 2022). Crop type mapping using optical data is challenging due to limited data availability during crucial crop growth stages due to persistent cloud cover in tropical and sub-tropical climates (Gao *et al.*, 2022). Synthetic aperture radar (SAR) data is useful in discriminating and identifying crop types at different growth stages. The high resolution of the SAR data provides information suitable for area-specific detailed analysis. Therefore, this objective assessed the effectiveness of fusing Sentinel-1(S1) and Sentinel-2 (S2) data for mapping fruit trees and co-existing land-use types by using two machine learning (ML) classifiers, i.e., random forest (RF) and support vector machine (SVM). The mapping process was conducted in four stages. The first classification process was performed independently using single-date S1 and S2 datasets. The second classification approach was based on data fusion of S1 and S2 data. The third stage involved the identification of optimal variables using the RF mean decrease accuracy (MDA) and forward variable selection (FVS). The final classification was based on optimally selected variables.

The results showed that the SVM classifier could be used to map the fruit trees and co-existing land-use types with an overall accuracy (OA) = 0.91.6% and kappa coefficient = 0.91% when applied to the combination (fused) S1 and S2 satellite data. Application of SVM to S1, S2, S2 selected variables and S1S2 fusion independently produced OA = 27.64, Kappa coefficient = 0.13%; OA= 87%, Kappa coefficient = 86.89%; OA = 69.33, Kappa coefficient = 69. %; OA = 87.01%, Kappa coefficient = 87%, respectively. It was observed that the optimal spectral bands for fruit tree mapping are green (B3: 560) and SWIR_2 (B12: 2190) on S2 data, whereas the vertical-horizontal (VH) polarization band was identified as the optimal band for S1 data. The green, red, red-edge_1, SWIR_1, SWIR_2, and VH were selected as the optimal variables for the data fusion approach. The results proved that including the textural metrics from the S1 VV channel improved discrimination of fruit trees and co-existing land use cover types. The fusion approach proved robust and appropriate for mapping smallholder fruit plantations.

6.4. To identify the optimal phenological period for discriminating fruit tree crops using multi-temporal Sentinel-2 data and Google Earth Engine in a heterogeneous subtropical horticultural region.

Single-date images have been widely used to map crop phenology because they are user-friendly and time-efficient (Li *et al.*, 2022). However, fruit tree phenology at the landscape level comprises inter-season changes in crop development stages such as greening, flowering, fruiting (peak of growing season), and senescence (Feng *et al.*, 2023). Thus, the single-date approach captures crop information at their specific growth stage and fails to discriminate crop spectral similarities emanating from different growth seasons (Southworth and Muir, 2021; Feng *et al.,* 2023). This necessitates the use of phenological information from multi-temporal data that will correlate to the crop's crucial phenological cycles in Africa's fragmented and heterogeneous agricultural landscapes (Singh *et al.*, 2021). Crop type mapping based on multi-temporal data can depict spectral similarity at a specific phenological stage, which can change enormously in another phenology stage, resulting in spectral confusion (Htitiou *et al.*, 2019). Obtaining accurate multi-year spatial distribution information of crops combined with the corresponding agricultural production data is of great significance to optimal agricultural production management in the future. However, the use of multi-series data is highly computational and data redundancy (Singh *et al.*, 2021). Thus, selecting the optimal window period in which remote sensing data can be acquired is a critical step in successful orchard classification.

In light of the above challenges, this objective was aimed at identifying the optimal temporal window in which the crucial phenological stages can be used to classify fruit tree crops in Levubu, Limpopo province, using a random forest (RF) classifier in a Google Earth Engine (GEE) environment. Phenological metrics were extracted from Sentinel-2 12-monthly composite images that were created using images available each month (i.e. 18 images per month) that were searched using a criteria of 5% cloud coverage. The RF classifier was used to classify the fruit trees and co-existing land use types and attained the highest overall accuracy of 84.89% with a kappa coefficient of 83% during July. The user accuracy on class accuracies ranged from 62 to 100%, while 60 to 100% was attained on producer accuracy. The analysis of variance overall accuracies among the S2 monthly composites showed statistical significance, indicating that the spectral differences between fruit trees differed. The mango and macadamia nut crops exhibited spectral differences during the harvesting and senescence in April. Spectral differences were observed for the macadamia nut, mango, and guava crops in May, while in June and July, distinct spectral differences were observed in avocado, macadamia nut, and mango crops during the peak flowering stage, as well as the banana crop that was in the fruiting stage. The contribution of the S2 in the classification process was assessed using the RF Mean Decrease Accuracy (MDA). The results showed that the Red-Edge bands (B5, B6, B7) and the shortwave infrared bands (B11, B12) were significant in discriminating the fruit tree crops. The results provide evidence-based information that can assist farm managers and horticulturists in making informed decisions, which is crucial in achieving effective agricultural management and sustainability of horticultural systems.

6.5. To assess the applicability of optimal Sentinel-2 data and Deep Neural Network in mapping fruit tree dynamics using phenological metrics.

Agricultural productivity is maximized by optimizing agricultural activities such as irrigation scheduling and application of pesticides. The notion is true when it comes to crop type mapping. Selecting optimal remote sensing that corresponds to critical crop growth stages is crucial in crop type mapping using optical data that is affected by the persistence of cloud coverage that camouflages critical information that optimises crop classification (Chabalala *et al*., 2023). This objective aimed to enhance the detection of fruit trees by using optimal Sentinel-2 images acquired during the optimal crop growth window period (i.e., April to July) as identified by Chabalala *et al*. (2023a) using Deep Neural Network. This was achieved by extracting phenological information from Sentinel-2 data to distinguish between fruit trees and co-existing land use types on subtropical farms in Levubu, South Africa. The hyperparameters of the DNN models were optimized to achieve the best classification accuracy. The overall accuracies of the models were 86.96%, 88.64%, 86.76%, and 87.25% for the April, May, June, and July images, respectively. The results demonstrate the suitability of temporal phenological optical-based data and DNN in mapping fruit tree crops under different management systems.

6.6. Conclusions

This research aimed to map fruit tree species and co-existing land use types in Limpopo's heterogeneous landscapes in South Africa by integration of Sentinel-1 and Sentinel-2 data using machine learning (ML) and deep learning (DL) models. Five ML classifiers, random forest (RF), Support Vector Machine (SVM), Gradient Boosting (GB), Adaptive Boosting (AdaBoost) and eXtreme Gradient Boost (XGBoost), deep neural networks (DNN), were used.

The research objectives lead to the following conclusions:

- Reducing the imbalanced ratio through data sampling can improve the classification accuracy of both minor and majority classes. The results showed that the RF and AdaBoost are sensitive to balanced and imbalanced data samples, while SVM, GB, and XGBoost are not affected by sample distribution. The most crucial bands in the classification of fruit trees and co-existing land use types are the red (B4: 665 nm), RedEdge_2 (B6: 740 nm), and the SWIR bands (B11, B12: 1610-2190 nm), respectively.

- Fusing Sentinel-1 and Sentinel-2 enhanced the mapping of fruit plantations in the heterogeneous horticultural landscape of Levubu. Selecting optimal variables that include textual information from Sentinel-1 and five variables from Sentinel-2 performed better than the single image mapping approaches.
- Classification based on images acquired during the fruit trees' full growth cycle helped identify the optimal window period in which fruit trees can be distinguished from the co-existing land use types with high accuracy. Images acquired in April, May, June, and July were identified as optimal images for detecting fruit trees at the research site.
- The application of phenological metrics derived from optimal Sentinel-2 images acquired during crucial fruit tree growth stages and the DNN model improved the classification accuracy compared to those produced by ML classifiers for the same periods.

In summary, the study showed the applicability of various techniques and mapping approaches in mapping fruit trees and co-existing land use types in heterogeneous smallholder agriculture. The single-date optical data proved inadequate for mapping fruit tree species in fragmented landscapes. This research provides mapping approaches such as data fusion, identification of optimal images and data sampling that can be used to accurately capture fruit metrics in mixed cropping systems at 10m resolution for smart orchard management in South Africa and other horticultural areas around the world with similar climatic conditions.

6.7. Research limitations

During this research, the following limitations were encountered**:**

- The Levubu smallholder horticulture is composed of fruit trees from different cultivars with different calendars and phenological cycles and operates under different management strategies, resulting in spectral similarities that affect the prediction models.
- The Sentinel-2 data was chosen because it is freely available and has high spatial, temporal, and spectral resolution suitable for smallholder agricultural applications (Chapter 3). The Sentinel-2 data has a spatial resolution of 10m, suitable for smallholder fruit crop mapping. However, the spectral resolution of the sensor was unable to detect

the subtle spectral dynamics of fruit trees from co-existing land use types in the research region.

- Regardless of the dominance of multispectral data in mapping crop types, the use of single-date multispectral images in mapping fruit trees and co-existing land use types is faced with considerable challenges in sub-tropical regions due to consistent cloud coverage with results in a lack of quality data during the critical growth seasons of fruit trees.

6.8. Future directions

The following recommendations can be implemented to solve the challenges.

- With recent advancements in remote sensing data, future research in horticulture should focus on incorporating biophysical and biochemical traits of fruit trees per cultivar to enhance the spectral separability for accurate fruit tree maps.
- Future research should focus on developing mapping approaches that are transferable and compatible with advanced, cost-effective RS instruments such as drones, UAVs and SAR.
- Due to the prevalence of complex farming systems in Africa, data-driven interventions must consider the most appropriate ways to effect change to understand how different fruit tree species respond to climate change and extreme weather conditions. Evidence-based interventions can be facilitated by knowing where fruit tree crops are planted. This information will optimize inputs through AI and Remote Sensing. Hence, space agencies should be established in Africa dedicated to capturing consistent spatial and RS data across scales.
- Investigate how various fruit tree species respond differently to changes in temperature, precipitation patterns, and extreme weather events. This could involve studying physiological responses, such as changes in growth rates, phenology, fruiting patterns, and stress tolerance mechanisms.
- Explore the genetic diversity within different fruit tree species and how this diversity influences their ability to adapt to changing environmental conditions to identify traits associated with climate resilience and the potential for selective breeding or genetic modification to enhance adaptive capacity.
- Examine the interactions between fruit tree species and other components of their ecosystems, including pests, diseases, and soil microorganisms. Understanding how

these interactions are affected by climate change and extreme weather events can help predict future ecosystem dynamics and inform management strategies.

- Investigate how responses to climate change vary geographically, both within and across different fruit tree species ranges. This research could involve spatial modelling and analysis to identify hotspots of vulnerability or resilience and assess the potential for range shifts or habitat fragmentation under different climate scenarios.

References

Abbott, P. and Checco, A. (2021). Smart Farming in Sub-Saharan Africa: Challenges and Opportunities. *Sensornets*, 159–164.

Abdel-Rahman, E.M. and Ahmed, F.B. (2008). The application of remote sensing techniques to sugarcane (Saccharum spp. hybrid) production: A review of the literature. *International Journal of Remote Sensing*, 29, 3753–3767.

Abdulhafedh, A.A. (2017). A novel hybrid method for measuring the spatial autocorrelation of vehicular crashes: Combining Moran's Index and Getis-Ord Gi* statistic. Open *Journal of Civil Engineering*, 7, 208–221.

Abdullah, A.Y.M., Masrur, A., Gani Adnan, M.S., Baky, M., Al, A., Hassan, Q.K. and Dewan, A. (2019). Spatio-temporal patterns of land use/land cover change in the heterogeneous coastal region of Bangladesh between 1990 and 2017. *Remote Sensing*, 11, 790.

Abubakar, G. A., Wang, K., Shahtahamssebi, A., Xue, X., Belete, M., Gudo, A. J. A., Shuka, K. A. M. and Gan, M. (2020). Mapping Maize Fields by Using Multi-Temporal Sentinel-1A and Sentinel-2A Images in Makarfi, Northern Nigeria, Africa. *Sustainability,* 12, 1–18.

Adam, E., Mutanga, O. and Rugege, D. (2010). Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: A review. *Wetlands Ecology and Management*, 18, 281–296.

Adam, E.M., Mutanga, O., Rugege, D. and Ismail, R. (2012). Discriminating the papyrus vegetation (Cyperus papyrus L.) and its co-existent species using random forest and hyperspectral data resampled to HYMAP. *International Journal of Remote Sensing*, 33, 552–569.

Adam, E. and Mutanga, O. (2009). Spectral discrimination of papyrus vegetation (Cyperus papyrus L.) in swamp wetlands using field spectrometry. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64, 612–620.

Adjorlolo, C., Mutanga, O., Cho, M. A. and Ismail, R. (2012). Challenges and opportunities in the use of remote sensing for C-3 and C-4 grass species discrimination and mapping. *African Journal of Range & Forage Science*, 29, 47–61.

Aguilar, R., Zurita-Milla, R., Izquierdo-Verdiguier, E., and de By, R.A. (2018). A cloud-based multi-temporal ensemble classifier to map smallholder farming systems. *Remote Sensing*, 10, 729.

Ahmed, N., Atzberger, C. and Zewdie, W. (2021). Species Distribution Modelling performance and its implication for Sentinel-2-based prediction of invasive Prosopis juliflora in lower Awash River basin, Ethiopia. *Ecological Processes*,10, 1–16.

Aitelkadi, K., Outmghoust, H., Laarab, S., Moumayiz, K. and Sebari, I. (2021). Detection and counting of fruit trees from RGB UAV Images by convolutional neural networks approach. *Advances in Science, Technology and Engineering Systems Journal*, 6, 887–93.

Akbari, E., Boloorani, A.D., Samany, N.N., Hamzeh, S., Soufizadeh, S. and Pignatti, S. (2020). Crop mapping using random forest and particle swarm optimisation based on multi-temporal Sentinel-2. *Remote Sensing*, 12, 1449.

Alabi, T. R., Adewopo, J., Duke, O. P. and Kumar, P. L. (2022). Banana Mapping in Heterogenous Smallholder Farming Systems Using High-Resolution Remote Sensing Imagery and Machine Learning Models with Implications for Banana Bunchy Top Disease Surveillance. *Remote Sensing*. 14, 1-22.

Amani, M., Kakooei, M., Moghimi, A., Ghorbanian, A., Ranjgar, B., Mahdavi, S., Davidson, A., Fisette, T., Rollin, P., Brisco, B. and Mohammadzadeh, A. (2020). Application of Google Earth Engine cloud computing platform, sentinel imagery, and neural networks for crop mapping in Canada. *Remote Sensing*, 12, 3561.

Amare, M., Mariara, J., Oostendorp, R. and Pradhan, M. (2019). The impact of smallholder farmers' participation in avocado export markets on the labour market, farm yields, sales prices, and incomes in Kenya. *Land Use Policy*, 88, 104168.

Arias, M., Campo-Bescós, M.Á. and Álvarez-Mozos, J. (2020). Crop classification based on temporal signatures of Sentinel-1 observations over Navarre province, Spain. *Remote Sensing*, 12, 1-29.

Asgarian, A., Soffianian, A. and Pourmanafi, S. (2016). Crop type mapping in a highly fragmented and heterogeneous agricultural landscape: A case of central Iran using multi-temporal Landsat 8 imagery. *Computers and Electronics in Agriculture*, 127, 31–540.

Ayyagari, M.R. (2020). Classification of Imbalanced Datasets using One-Class SVM, k-Nearest Neighbors and CART Algorithm. *International Journal of Advanced Computer Science and Applications*, 11, 1–5.

Azadbakht, M., Fraser, C. and Khoshelham, K. (2018). Synergy of sampling techniques and ensemble classifiers for classification of urban environments using full-waveform LiDAR data. *International Journal of Applied Earth Observation and Geoinformation*, 73, 277–291.

Azzari, G., Jain, S., Jeffries, G., Kilic, T. and Murray, S. (2021). Understanding the requirements for surveys to support satellite-based crop type mapping: Evidence from **S**ub-**S**aharan Africa. *Remote Sensing*, 13, 1-27.

Baghdadi, N., Boyer, N., Todoroff, P., El, M. and Bégué, A. (2009). Remote Sensing of Environment Potential of SAR sensors TerraSAR-X, ASAR / ENVISAT and PALSAR / ALOS for monitoring sugarcane crops on Reunion Island. *Remote Sensing of Environment*, 113, 1724–1738.

Bagheri, N. (2019). Application of aerial remote sensing technology for detection of fire blight-infected pear trees. *Computer and Electronics in Agriculture*, 168, 105147.

Bai, G., Ge, Y., Hussain, W., Baenziger, P.S. and Graef, G. (2016). A multi-sensor system for high throughput field phenotyping in soybean and wheat breeding. *Computer and Electronics in Agriculture*, 128, 181–192.

Bai, T., Zhang, N., Mercatoris, B. and Chen, Y. (2019). Improving jujube fruit tree yield estimation at the field scale by assimilating a single Landsat remotely sensed LAI into the WOFOST model. *Remote Sensing*, 11, 1119.

Bal, F. and Kayaalp, F. (2021), Review of machine learning and deep learning models in agriculture. *International Advanced Researchers and Engineering Journal*, 5, 309–323.

Banko, G. (1998). A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data and of Methods, Including Remote Sensing Data in Forest Inventory. Available at: https://core.ac.uk/reader/33897040 (accessed on 16 November 2022).

Bargiel, D. (2017). A new method for crop classification combining time series of radar images and crop phenology information. *Remote Sensing of Environment*, 198, 369–383.

Barrett, B., Nitze, I., Green, S. and Cawkwell, F. (2015). Support vector machines in remote sensing: A review. *International Journal of Applied Earth Observation and Geoinformation*, 34, 66–75.

Barrow, D.K. and Crone, S.F. (2016). A comparison of AdaBoost algorithms for time series forecast combination. *International Journal of Forecasting*, 32, 1103–1119.

Baumann, M.; Levers, C.; Macchi, L.; Bluhm, H.; Waske, B.; Gasparri, N.I.; Kuemmerle, T. (2018). Mapping continuous fields of tree and shrub cover across the Gran Chaco using Landsat 8 and Sentinel-1 data. *Remote Sensing of Environment*, 216, 201–211.

Becker-reshef, I., Justice, C., Barker, B., Humber, M., Rembold, F., Bonifacio, R., Zappacosta, M., Budde, M., Magadzire, T., Shitote, C., Pound, J., Constantino, A., Nakalembe, C., Mwangi, K., Sobue, S., Newby, T., Whitcraft, A., Jarvis, I. and Verdin, J. (2020). Strengthening agricultural decisions in countries at risk of food insecurity: The GEOGLAM Crop Monitor for Early Warning. *Remote Sensing of Environment*, 237, 1-19.

Belgiu, M. and Csillik, O. (2018). Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sensing of Environment*, 204, 509–523.

Berger, M., Moreno, J., Johannessen, J.A., Levelt, P.F. and Hanssen, R.F. (2012). ESA's sentinel missions in support of Earth system science. *Remote Sensing of Environment*, 120, 84–90.

Bey, A., Jetimane, J., Nogueira, S., Ribeiro, N., Sitoe, A. and Meyfroidt, P. (2020). Remote Sensing of Environment Mapping smallholder and large-scale cropland dynamics with a flexible classification system and pixel-based composites in an emerging frontier of Mozambique. *Remote Sensing of Environment*. 239, 111611.

Biffi, L.J., Mitishita, E., Liesenberg, V., Santos, A.A., Gonçalves, D.N. and Estrabis, N.V. (2020). ATSS deep learning-based approach to detect. Apple Fruits. *Remote Sensing*, 13, 1-22.

Blickensdörfer, L., Schwieder, M., Pflugmacher, D., Nendel, C., Erasmi, S. and Hostert, P. (2022). Mapping of crop types and crop sequences with combined time series of Sentinel-1, Sentinel-2 and Landsat 8 data for Germany. *Remote Sensing of Environment*, 269, 112831.

Böhler, J.E., Schaepman, M.E. and Kneubühler, M. (2020). Crop separability from individual and combined airborne imaging spectroscopy and UAV multispectral data. *Remote Sensing*, 12, 1256.

Bouras, E., Jarlan, L., Er-Raki, S., Balaghi, R., Amazirh, A., Richard, B. and Khabba, S. (2021). Cereal yield forecasting with satellite drought-based indices, weather data and regional climate indices using machine learning in Morocco. *Remote Sensing*, 13, 3101.

Breiman, L. Random forests. Machine Learning 2001, 45, 5–32. Available at: https://link.springer.com/article/10.1023/A:1010933404324 (accessed on 16 October 2022).

Brinkhoff, J., Vardanega, J. and Robson, A.J. (2020). Land cover classification of nine perennial crops using sentinel-1 and -2 data. *Remote Sensing*, 12, 96.

Brownlee, J. (2020). Imbalanced classification with python. In machine learning mastery; python: Wilmington, DE, USA, 463.

Brownlee, J. (2021). Gradient boosted trees with xgboost and scikitlLearn; Packet Publishing: Birmingham, UK, 148, 148–162.

Brownlee, J. Master machine learning algorithms: Discover how they work and implement them from scratch. In machine learning mastery; python: Wilmington, DE, USA, 2016; 1–163. Available at: http://machinelearningmastery.com/a-tourof-machine-learning-algorithms/(accessed on 14 September 2022).

Busetto, L., Zwart, S. J. and Boschetti, M. (2019). Analysing spatial–temporal changes in rice cultivation practices in the Senegal River Valley using MODIS time-series and the PhenoRice algorithm. *International Journal of Applied Earth Observation and*

Geoinformation, 75, 15–28.
Cai, Y., Guan, K., Peng, J., Wang, S., Seifert, C., Wardlow, B. and Li, Z. (2018). A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. *Remote Sensing of Environment*, 210, 35–47.
Cai, Y., Lin, H. and Zhang, M. (2019). Mapping paddy rice by the object-based random forest method using time series Sentinel-1/Sentinel-2 data. *Advances in Space Research*, 64, 2233–2244.
Cao, J., Kwong, S. and Wang, R. (2012). A noise-detection based AdaBoost algorithm for mislabelled data. *Pattern Recognition*, 45, 4451– 4465.
Central Statistical Agency. Agricultural Sample Survey 2010/2011. Available online: https://www.scirp.org/reference/referencespapers?referenceid=2611368 (accessed on 18 November 2022)
Chabalala, Y., Adam, E., and Ali, K.A. (2022). Machine Learning Classification of Fused Sentinel-1 and Sentinel-2 Image Data Towards Mapping Fruit Plantations in Highly Heterogeneous Landscapes. *Remote Sensing*, 14, 2621.
Chabalala, Y., Adam, E., and Ali, K.A. (2023a). Exploring the Effect of Balanced and Imbalanced Multi-Class Distribution Data and Sampling Techniques on Fruit-Tree Crop Classification Using Different Machine Learning Classifiers. *Geomatics*, 3, 70-92.
Chabalala Y., Adam E., and Ali, K.A. (2023b). Identifying the optimal phenological period for discriminating subtropical fruit tree crops using multi-temporal sentinel-2 data and Google Earth Engine. *South African Journal of Geomatics*, 12:262–83.
Chabalala Y., Adam E., and Ali, K.A. (2023c). Mapping fruit tree dynamics using phenological metrics from optimal Sentinel-2 data and Deep Neural Network. *CABI for Agriculture and Bioscience*, 4, 1-20.
Chabalala, Y., Adam, E., Oumar, Z. and Ramoelo, A. (2020). Exploiting the capabilities of Sentinel-2 and RapidEye for predicting grass nitrogen across different grass communities in a protected area. *Applied Geomatics*, 12, 379–395.
Chan, J.C.W. and Paelinckx, D. (2008). Evaluation of random forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment*, 112, 2999–3011.
Chemura, A., van Duren, I., and van Leeuwen, L. M. (2015). Determination of the age of oil palm from crown projection area detected from WorldView-2 multispectral remote sensing data: The case of Ejisu-Juaben district, Ghana. *ISPRS Journal of Photogrammetry and Remote Sensing*, 100, 118–127.
Chen, B., Jin, Y. and Brown, P. (2017). Classifying fruit-tree crops by landsat-8 time series. *IEEE 1st International Symposium on Geoscience and Remote Sensing, GRSS-CHILE*, 19-22.
Chen, B., Jin, Y. and Brown, P. (2019). Automatic mapping of planting year for tree crops with Landsat satellite time series stacks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 151, 176–188.
Chen, R.C., Dewi, C., Huang, S. and Caraka, R.E. (2020). Selecting critical features for data classification based on machine learning methods. *Journal of Big Data*, 7, 52.
Chen, Y., Hou, J., Huang, C., Zhang, Y. and Li, X. (2021). Mapping maize area in heterogeneous agricultural landscape with multi-temporal sentinel-1 and sentinel-2 images based on random forest. *Remote Sensing*, 13, 2988.
Chen, Y., Song, X., Wang, S., Huang, J. and Mansaray, L.R. (2016). Impacts of spatial heterogeneity on crop area mapping in Canada using MODIS data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 119, 451–461.
Claverie, M., Ju, J., Masek, J. G., Dungan, J. L., Vermote, E. F., Roger, J. C., Skakun, S. V.

and Justice, C. (2018). The Harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote Sensing of Environment*, 219, 145–161.

Clerici, N., Calderón, C.A.V. and Posada, J.M. (2017). Fusion of Sentinel-1A and Sentinel- 2A data for land cover mapping: A case study in the lower Magdalena region, Colombia. *Journal of Maps*, 13, 718–726.

Clevers, J.G.P.W. 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, 344–351.

Congalton, R. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37, 35–46.

Conrad, C., Dech, S., Dubovyk, O., Fritsch, S., Klein, D., Löw, F., Schorcht, G. and Zeidler, J. (2014). Derivation of temporal windows for accurate crop discrimination in heterogeneous croplands of Uzbekistan using multitemporal RapidEye images. *Computer and Electronics in Agriculture*, 103, 63–74.

Conrad, Y., Zhong, X., Hu, L.W. and Zhang, L. (2019). A robust spectral-spatial approach to identifying heterogeneous crops using remote sensing imagery with high spectral and spatial resolutions. *Remote Sensing of Environment*, 239, 111605.

Crommelinck, S. and Höfle, B. (2016). Simulating an autonomously operating low-cost static terrestrial LiDAR for multitemporal maize crop height measurements. *Remote Sensing*, 8, 205.

Csillik, O., Cherbini, J., Johnson, R., Lyons, A. and Kelly, M. (2018). Identification of citrus trees from unmanned aerial vehicle imagery using convolutional neural networks. *Drones*, 2, 1–16.

Csorba, A., Szegi, T., Fodor, H., Bukombe, B., Uwiragiye, Y., Xavier, F. and Michéli, E. (2019). Characterization of rice agriculture in the Southern Province of Rwanda by means of microwave remote sensing. *Physics and Chemistry of the Earth*, 112, 58–65.

Cui, B., Zhao, Q., Huang, W., Song, X., Ye, H. and Zhou, X. (2019). Leaf chlorophyll content retrieval of wheat by simulated RapidEye, Sentinel-2 and EnMAP data. *Journal of Integrative Agriculture*, 18, 1230–1245.

Cui, J., Zhang, X., Wang, W. and Wang, L. (2020). Integration of optical and SAR remote sensing images for crop-type mapping based on a novel object-oriented feature selection method. *International Journal of Agricultural and Biological Engineering*, 13, 178–190.

DAFF. Department of Agriculture, Forestry and Fisheries. (2012),1–44. Available at: https://www.daff.gov.za/docs/ AMCP/MaizeMVCP2011.pdf. AMCP/MaizeMVCP2011.pdf. (accessed on 18 October 2022).

Danylo, O., Pirker, J., Lemoine, G., Ceccherini, G., See, L., McCallum, I., Hadi; Kraxner, F., Achard, F. and Fritz, S. (2021). A map of the extent and year of detection of oil palm plantations in Indonesia, Malaysia, and Thailand. *Scientific Data*, 8, 4–11.

Darvishzadeh, R., Skidmore, A., Abdullah, H., Cherenet, E., Ali, A., Wang, T., Nieuwenhuis, W., Heurich, M., Vrieling, A. and O'Connor, B. (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, 58–70.

De Oliveira Santos, C.L.M., Lamparelli, R.A.C., Figueiredo, G.K.D.A., Dupuy, S., Boury, J., Luciano, A.C.S., da Silva Torres, R. and le Maire, G. (2019). Classification of crops, pastures, and tree plantations along the season with multi-sensor image time series in a subtropical agricultural region. *Remote Sensing*, 11, 334.

Defourny, P., Bontemps, S., Bellemans, N., Cara, C., Dedieu, G., Guzzonato, E., Hagolle, O., Inglada, J., Nicola, L., Rabaute, T., Savinaud, M., Udroiu, C., Valero, S., Bégué, A., Dejoux, J., El, A., Ezzahar, J., Kussul, N., Labbassi, K., Lebourgeois, V., Miao, Z., Newby, T., Nyamugama, A., Salh, N., Shelestov, A., Simonneaux, V., Sibiry, P., Traore, S. S. and Koetz, B. (2019). Near real-time agriculture monitoring at national scale at parcel resolution: Performance assessment of the Sen2-Agri automated system in various cropping systems around the world. *Remote Sensing of Environmen*t, 221, 551–568.

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, 245–261.

Deloitte (2017). Agricultural opportunities in Africa Crop farming in Ethiopia, Nigeria and Tanzania Introduction Africa remains the only region yet to experience a "Green Revolution". 1–36. Available at: https://www.deloitte.com/content/dam/Deloitte/za/Documents/africa/DeloitteZA_Agricultural_opportunites_in_Africa_Oct2017-updated.pdf. (accessed on 03 August 2023).

Devi, D., Biswas, S. K. and Purkayastha, B. (2019). Learning in presence of class imbalance and class overlapping by using one-class SVM and undersampling technique. *Connection Science*, 31, 105–142.

Dietterich, T.G. (1998). Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. *Neural Computation*, 10, 1895–1923.

Djamai, N. and Fernandes, R. (2018). Comparison of SNAP-Derived Sentinel-2A L2A Product to ESA Product over Europe. *Remote Sensing*, 10, 926.

Duveiller, G., López-Lozano, R. and Baruth, B. (2013). Enhanced processing of 1-km spatial resolution fAPAR time series for sugarcane yield forecasting and monitoring. *Remote Sensing*, 5 (3), 1091–1116.

Elders, A., Carroll, M. L., Neigh, C. S. R., D'Agostino, A. L., Ksoll, C., Wooten, M. R. and Brown, M. E. (2022). Estimating crop type and yield of smallholder fields in Burkina Faso using multi-day Sentinel-2. *Remote Sensing Applications: Society and Environment*, 27, 100820.

ESA. (2015). ESA's Optical High-Resolution Mission for GMES Operational Services; ESA: Paris, France. European Space Agency (ESA) 2023. Available at: https://www.researchgate.net/publication/272907479_Sentinel-2_ESA's_Optical_High-Resolution_Mission_for_GMES_Operational_Services. (accessed on 01 September 2023).

Estes, L.D, Ye, S., Song, L., Luo, B., Eastman, J., Meng, Z., Zhang, Q., McRitchie, D., Debats, S.R., Muhando, J., Amukoa, A.H., Kaloo, B.W., Makuru, J., Mbatia, B.K., Muasa, I.M., Mucha, J., Mugami, A.M., Mugami, J.M., Muinde, F.W., Mwawaza, F.M., Ochieng, J., Oduol, C.J., Oduor, P., Wanjiku, T., Wanyoike, J.G., Avery, R.B. and Caylor, K.K.,(2022). High Resolution, Annual Maps of Field Boundaries for SmallholderDominated Croplands at National Scales. *Frontiers of Artificial Intelligence*. 4,744863.

,

FAO. (2014). The Future of Food and Agriculture: Trends and Challenges; FAO: Rome, Italy, Volume 4.

Faye, G., Mbengue, F., Coulibaly, L., Sarr, M.A., Mbaye, M., Tall, A., Tine, D., Marigo, O. and Ndour, M.M.M. (2020). Complementarity of Sentinel-1 and Sentinel-2 data for mapping agricultural areas in Senegal. *Advances in Remote Sensing*, 9, 101–115.

Feng, F., Gao, M., Liu, R., Yao, S. and Yang, G. (2023). A deep learning framework for crop mapping with reconstructed sentinel-2 time series images. *Computer and Electronics in Agriculture*, 213,108227

Feng, S., Zhao, J., Liu, T., Zhang, H., Zhang, Z. and Guo, X. (2019). Crop Type Identification and Mapping Using Machine-Learning Algorithms and Sentinel-2 Time Series Data. *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing*, 12, 3295–3306.

Feyisa, G.L., Palao, L.K., Nelson, A., Gumma, M.K., Paliwal, A., Win, K.T., Nge, K.H. and Johnson, D.E. (2020). Characterizing and mapping cropping patterns in a complex agro-ecosystem: An iterative participatory mapping procedure using machine learning algorithms and MODIS vegetation indices. *Computer and Electronics in Agriculture,* 175, 105595.

Figueroa-Figueroa, D.K., Ramírez-Dávila, J.F., Antonio-Némiga, X. and González-Huerta, A. (2020). Mapping of avocado in the south of the state of Mexico by digital image-processing -2. *Revista Mexicana de ciencias agrícolas*, 4, 865–879.

Foley, J.A., Ramankutty, N., Brauman, K.A., Cassidy, E.S., Gerber, J.S., Johnston, M., Mueller, N.D., O'Connell, C., Ray, D.K. and West, P.C. (2011). Solutions for a cultivated planet. *Nature*, 478, 337–342.

Forkuor, G., Dimobe, K., Serme, I. and Tondoh, J.E. (2018). Landsat-8 vs. Sentinel-2: Examining the added value of sentinel-2's red-edge bands to land-use and land-cover mapping in Burkina Faso. *GIScience Remote Sensing*, 55, 331–354.

Fowler, J., Waldner, F. and Hochman, Z. (2020). All pixels are useful, but some are more useful: Efficient in situ data collection for crop-type mapping using sequential exploration methods. *International Journal of Applied Earth Observation and Geoinformation*, 91, 102114.

Fraser, A. (2008). White farmers' dealings' with land reform in South Africa: Evidence from Northern Limpopo Province. *Tijdschrift voor Economische en Sociale Geografie*, 99, 24–36.

Freund, Y. and Schapire, R.E. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, 55, 119–139.

Friedman, J.H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics,* 29, 1189–1232.

Gallo, I., Ranghetti, L, Landro, N., La Grassa, R. and Boschetti, M. (2023). In-season and dynamic crop mapping using 3D convolution neural networks and sentinel-2 time series. *ISPRS Journal of Photogrammetry and Remote Sensing*, 195, 335–52.

Gao, F. and Zhang, X. (2021). Mapping crop phenology in near real-time using satellite remote sensing: challenges and opportunities. *Journal of Remote Sensing*, 2021, 1-14.

Gao, F., Anderson, M., Daughtry, C., Karnieli, A. and Hively, D. (2020). A within-season approach for detecting early growth stages in corn and soybean using high temporal and spatial resolution imagery. *Remote Sensing of Environment*, 242, 111752.

Gao, F., Fang, W., Sun, X., Wu, Z., Zhao, G., Li, G. and Li, R. (2022). A novel apple fruit detection and counting methodology based on deep learning and trunk tracking in modern orchards. *Computers and Electronics in Agriculture*, 197, 107000.

Gao, S., Niu, Z., Huang, N. and Hou, X. (2013). Estimating the Leaf Area Index, height and biomass of maize using HJ-1 and RADARSAT-2. *International Journal of Applied Earth Observations and Geoinformation*, 24, 1–8.

Gašparović, M.; Dobrinić, D. (2020). Comparative assessment of machine learning methods for urban vegetation mapping using multitemporal Sentinel-1 imagery. *Remote Sensing*, 12, 1952.

Ge, Y., Bai, G., Stoerger, V. and Schnable, J.C. (2016). Temporal dynamics of maize plant growth, water use, and leaf water content using automated high throughput RGB and hyperspectral imaging. *Computers and Electronic in Agriculture*, 127, 625–632.

Geetha, M. N., Assistant, D., Gowda, A. and Associate, K. (2016). Crop Mapping using SAR Imagery: A Review. *International Journal of Advanced Research in Computer Science*, 7, 47–52.

Gelaye, Y. and Getahun, S. (2024). A review of the carbon sequestration potential of fruit trees and their implications for climate change mitigation: The case of Ethiopia. *Cogent Food and Agriculture*, 10, 1-16.

Gené-Mola, J., Gregorio, E., Cheein, F.A., Guevara, J., Llorens, J., Sanz-Cortiella, R., Escolà, A. and Rosell-Polo, J.R. (2020). Fruit detection, yield prediction and canopy geometric characterization using LiDAR with forced airflow. *Computers and Electronic in Agriculture*, 168, 105121.

Genuer, R., Poggi, J.M. and Tuleau-Malot, C. (2015). VSURF: An R package for variable selection using random forests. R J., 7, 19–33.

Ghaseminik, F., Aghamohammadi, H. and Azadbakht, M. (2021). Land cover mapping of urban environments using multispectral LiDAR data under data imbalance. *Remote Sensing Applications: Society and Environment*, 21, 100449.

Gilbertson, J.K. and Van Niekerk, A. (2017). Value of dimensionality reduction for crop differentiation with multi-temporal imagery and machine learning. *Computers and Electronic in Agriculture*, 142, 50–58.

Gómez, C., White, J. C. and Wulder, M. A. (2016). Optical remotely sensed time series data for land cover classification: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 116, 55–72.

Gourlay, S., Kilic, T. and Lobell, D. (2017). Could Debate Be Over? Errors Farmer-Reported Production and Their Implications for the Inverse Scale-Productivity Relationship in Uganda; SSRN: Rochester, NY, USA.

Grabska, E. and Frantz, D.; Ostapowicz, K. (2020). Evaluation of machine learning algorithms for forest stand species mapping using Sentinel-2 imagery and environmental data in the Polish Carpathians. *Remote Sensing of Environment*. 251, 112103.

Griffiths, P., Nendel, C. and Hostert, P. (2017). Intra-annual reflectance composites from Sentinel-2 and Landsat for national scale crop and land cover mapping. *Remote Sensing of Environment*, 220, 135–151.

Gutierrez-Coarite, R., Mollinedo, J., Cho, A. and Wright, M.G. (2018). Canopy management of macadamia trees and understory plant diversification to reduce macadamia felted coccid (*Eriococcus ironsidei*) populations. *Crop Protection*, 113, 75–83.

Haas, J. and Ban, Y. (2017). Sentinel-1A SAR and sentinel-2A MSI data fusion for urban ecosystem service mapping. *Remote Sensing Applications: Society and Environment*, 8, 41–53.

Hadria, R., Duchemin, B., Jarlan, L., Dedieu, G., Baup, F., Khabba, S., Olioso, A. and Le Toan, T. (2010). Potentiality of optical and radar satellite data at high spatio-temporal resolutions for the monitoring of irrigated wheat crops in Morocco. *International Journal of Applied Earth Observation and Geoinformation*, 12, 32–37.

Hansch, R., Ley, A. and Hellwich, O. Correct and still wrong: The relationship between sampling strategies and the estimation of the generalization error. (2017). *In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS*), Fort Worth, TX, USA, 23–28 July 2017, 3672–3675.

Hao, P., Wu, W., Niu, Z., Wang, L. and Zhan, Y. (2018). Estimation of different data compositions for early-season crop type classification. *PeerJ*, 1-21.

Hebbar, R. and Ravishankar, H.M.; Trivedi, S.; Subramoniam, S.R.; Raj, U.; Dadhwal, V.K. (2014). Object-oriented classification of high-resolution data for inventory of horticultural crops. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40, 745–749.

Heckel, K., Urban, M., Schratz, P., Mahecha, M.D. and Schmullius, C. (2020). Predicting forest cover in distinct ecosystems: The potential of multi-source Sentinel-1 and -2 data fusion. *Remote Sensing*, 12, 302.

Heydari, S.S. and Mountrakis, G. (2018). Effect of classifier selection, reference sample size, reference class distribution and scene heterogeneity in per-pixel classification accuracy using 26 Landsat sites. *Remote Sensing of Environment*, 204, 648–658.

Hong-xia, L.U.O., Sheng-pei, D., Mao-fen, L., En-ping, L., Qian, Z. and Ying-ying, H.U. (2020). Comparison of machine learning algorithms for mapping mango plantations based on Gaofen-1 imagery. *Journal of Integrative Agriculture 19*, 2815–2828.

Htitiou, A., Boudhar, A., Lebrini, Y., Hadria, R., Lionboui, H., Elmansouri, L., Tychon, B. and Benabdelouahab, T. (2019). The Performance of Random Forest Classification Based on Phenological Metrics Derived from Sentinel-2 and Landsat 8 to Map Crop Cover in an Irrigated Semi-arid Region. *Remote Sensing in Earth Systems Sciences*, 2, 208–224.

Hu, Q., Wu, W., Song, Q., Yu, Q., Lu, M., Yang, P., Tang, H., and Long, Y. (2016). Extending the pairwise separability index for multi crop identification using time-series MODIS images. *IEEE Transactions on Geoscience and Remote Sensing*, 54, 6349-6361.

Hu, Q., Wu, W.B., Song, Q., Lu, M., Chen, D., Yu, Q., and Tang, H.J. (2017). How do temporal and spectral features matter in crop classification in Heilongjiang Province, China? *Journal of Integrative Agriculture*, 16, 324–336.

Ibrahim, A., Saito, K., Bado, V. B. and Wopereis, M. C. S. (2021). Thirty years of agronomy research for development in irrigated rice-based cropping systems in the West African Sahel: Achievements and perspectives. *Field Crops Research*, 266, 108149.

Ismail, N., and Malik, O.A. (2018). Real-time visual inspection system for grading fruits using computer vision and deep learning techniques. *Information Processing in Agriculture*. 9, 24-37.

Jahel, C., Baron, C., Vall, E., Karambiri, M., Castets, M., Coulibaly, K., Bégué, A. and Lo Seen, D. (2017). Spatial modelling of agro-ecosystem dynamics across scales: A case in the cotton region of West-Burkina Faso. *Agricultural Systems*, 157, 303–315.

Jamnadass, R. H., Dawson, I. K., Franzel, S., Leakey, R. R. B., Mithfer, D., Akinnifesi, F. K. and Tchoundjeu, Z. (2011). Improving livelihoods and nutrition in Sub-Saharan Africa through the promotion of indigenous and exotic fruit production in smallholders' agroforestry systems: A review. *International Forestry Review*, 13, 338–354.

Jain, M., Mondal, P., Defries, R. S., Small, C. and Galford, G. L. (2013). Mapping cropping intensity of smallholder farms: A comparison of methods using multiple sensors. *Remote Sensing of Environment*. 134, 210–223.

Jin, Z., Azzari, G., You, C., Di Tommaso, S., Aston, S., Burke, M., and Lobell, D.B. (2019). Smallholder area and yield mapping at national scales with Google Earth Engine. *Remote Sensing of Environment*, 228, 115–128.

Johansen, K., Duan, Q., Tu, Y.H., Searle, C., Wu, D., Phinn, S., Robson, A., and McCabe, M.F. (2020). Mapping the condition of macadamia tree crops using multi-spectral UAV and WorldView-3 imagery. *ISPRS Journal of Photogrammetry and Remote Sensing, 165,* 28–40.

Karthikeyan, L., Chawla, I. and Mishra, A. K. (2020). A review of remote sensing applications in agriculture for food security: Crop growth and yield, irrigation, and crop losses. *Journal of Hydrology,* 586, 124905.

Kellengere Shankarnarayan, V. and Ramakrishna, H. (2020). Paradigm change in Indian agricultural practices using Big Data: Challenges and opportunities from field to plate. *Information Processing in Agriculture*, 7, 355–368.

Kestur, R., Meduri, A. and Narasipura, O. (2019). MangoNet: a deep semantic segmentation architecture for a method to detect and count mangoes in an open orchard. *Engineering Applications of Artificial Intelligence*, 77, 59–69.

Kganyago, M., Mhangara, P. and Adjorlolo, C. (2021). Estimating crop biophysical parameters using machine learning algorithms and Sentinel-2 imagery. *Remote Sensing,* 13, 4314.

Khaldoon, A., Sujan, R., Ali, A., and Agrawal, D.P. (2020). Enhancing Imbalanced Dataset by Utilizing (K-NN Based SMOTE_3D Algorithm). *Annals of Robotics and Automation*, 4, 1–6.

Kiyohara, S.; Miyata, T.; Mizoguchi, T. (2015). Prediction of grain boundary structure and energy by machine learning. *Science Advances*, 18, 1–5.

Kolecka, N., Ginzler, C., Pazur, R., Price, B. and Verburg, P.H. (2018). Regional scale mapping of grassland mowing frequency with Sentinel-2 time series. *Remote Sensing*, 10, 1221.

Kordi F. and Yousefi H. (2022). Crop classification based on phenology information by using time series of optical and synthetic-aperture radar images. *Remote Sensing Applications: Society and Environment*, 27, 100812.

Kraaijvanger, R., Singh, V., Ajay, N.P., Apurva, D., Jaydipsinh, K. and Jignesh, S.K. (2019). Exploring the Potential of Mapping Cropping Patterns on Smallholder Scale Croplands Using Sentinel-1 SAR Data. *Remote Sensing*, 11, 1–18.

Kranjcˇic´, N., Medak, D., Župan, R. and Rezo, M. (2019). Support vector machine accuracy assessment for extracting green urban areas in towns. *Remote Sensing*, 11, 655.

Krawczyk, B. (2016). Learning from imbalanced data: Open challenges and future directions. *Progress in Artificial Intelligence*, *5*, 221–232.

Kroschel, J., Sporleder, M., Tonnang, H. E. Z., Juarez, H., Carhuapoma, P., Gonzales, J. C. and Simon, R. (2013). Predicting climate-change-caused changes in global temperature on potato tuber moth Phthorimaea operculella (Zeller) distribution and abundance using phenology modelling and GIS mapping. *Agricultural and Forest Meteorology*, 170, 228–241.

Kuhn, M. Futility Analysis in the Cross-Validation of Machine Learning Models. (2014), 1-22.

Kumar R, Krishna H, Rajan S, Alizadeh M. and Barman P. (2021). Understanding phenological stages of pomegranates vis-à-vis flowering and fruiting regulation. *Journal of Horticultural Research*, 29, 1–22.

Kuter, S. (2021). Completing the machine learning saga in fractional snow cover estimation from MODIS Terra reflectance data: Random forests versus support vector regression. *Remote Sensing of Environment*, 255, 112294.

Kutya, L. (2013). Small Scale Agriculture. Afesis-corplan, 1–2.

Kyere, I., Astor, T., Graß, R. and Wachendorf, M. (2020). Agricultural crop discrimination in a heterogeneous low-mountain range region based on multi-temporal and multi-sensor satellite data. *Computer and Electronics in Agriculture*, 179, 105864.

Lahlou, O., Benmansour, S., Zennayi, Y., Bourzeix, F. (2022). CerealNet: A Hybrid Deep Learning Architecture for Cereal Crop Mapping Using Sentinel-2 Time-Series. *Informatics*, 9, 96.

Lakshmanan, G. and Mathiyazhagan, K. (2019). Machine-learning classifiers on Sentinel-2 satellite image for the classification of banana (Musa Sp.) plantations of Theni district, *Tamil* Nadu, India. *International Journal of Chemical Studies*, 7, 1419–1425.

Lamour, J., Lechaudel, O.N.M., Le Moguédec, G. and Tisseyre, J.T.B. (2019). Spatial analysis and mapping of banana crop properties: issues of the asynchronicity of the banana production and proposition of a statistical method to take it into account. *Precision Agriculture*, 21, 897-921.

Lanaras, C., Bioucas-Dias, J., Galliani, S., Baltsavias, E. and Schindler. K. (2018). Super-resolution of sentinel-2 images: learning a globally applicable deep neural network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 146, 305-319.

Laris, P., Foltz, J. D. and Voorhees, B. (2015). Taking from cotton to grow maize: The shifting practices of small-holder farmers in the cotton belt of Mali. *Agricultural Systems*, 133, 1–13.

Lebourgeois, V., Dupuy, S., Vintrou, É., Ameline, M., Butler, S., and Bégué, A. (2017). A Combined random forest and OBIA classification scheme for mapping smallholder agriculture at different nomenclature levels using multisource data (simulated Sentinel-2 time series, VHRS and DEM). *Remote Sensing*, 9, 259.

Leroux, L., Castets, M., Baron, C., Escorihuela, M., Bégué, A. and Lo, D. (2019). Maize yield estimation in West Africa from crop process-induced combinations of multi-domain remote sensing indices. *European Journal of Agronomy*. 108, 11–26.

Li, B., Xu, X., Han, J., Zhang, L., Bian, C., Jin, L. and Liu, J. (2019). The estimation of crop emergence in potatoes by UAV RGB imagery. *Plant Methods*, 15, 1–13.

Li, G., Cui. J., Han, W., Zhang, H., Chen, H. and Ao, J. (2022). Crop type mapping using timeseries Sentinel-2 imagery and U-net in early growth periods in the Hetao irrigation district in China. *Computer and Electronics in Agriculture*, 203:107478.

Li, J., Du, Q.; Li, Y., and Li, W. (2018). Hyperspectral Image Classification with Imbalanced Data Based on Orthogonal Complement Subspace Projection. *IEEE Transactions on Geoscience and Remote Sensing*, 56, 3838–3851.

Li, Z., Yang, X., Cui, S., Yang, Q., Yang, X., Li, J. and Shen, Y. (2018). Developing sustainable cropping systems by integrating crop rotation with conservation tillage practices on the Loess Plateau, a long-term imperative. *Field Crops Research*, 222, 164–179.

Liu, C. (2016). Analysis of Sentinel-1 SAR data for mapping standing water in the Twente region. 37. Available at: www.itc.nl/library/papers_2016/msc/wrem/cliu.pdf. (accessed 13 December 2023)

Liu, M., Wang, T., Skidmore, A., and Liu, X. (2018). Heavy metal-induced stress in rice crops detected using multi-temporal Sentinel-2 satellite images. *Science of the Total Environment*, 637–638, 18–29.

Loggenberg, K., Strever, A., Greyling, B. and Poona, N. (2018). Modelling water stress in a shiraz vineyard using hyperspectral imaging and machine learning. Remote Sensing, 10, 202.

Louis, J., Debaecker, V., Pflug, B., Main-Knorn, M., Bieniarz, J., Mueller-Wilm, U., Cadau, E. and Gascon, F. (2016). Sentinel-2 SEN2COR: L2A processor for users. *In Proceedings of the Living Planet Symposium, Prague, Czech Republic*, 9–13.

Louw, D., and Flandorp, C. (2017). Horticultural Development Plan for the Thulamela Local Municipality: Agricultural Overview; OABS Development (Pty) Ltd.: Western Cape, South Africa, 27.

Luo, C., Liu, H., Lu, L., Liu, Z., Kong, F., and le Zhang, X. (2021). Monthly composites from Sentinel-1 and Sentinel-2 images for regional major crop mapping with Google Earth Engine. *Journal of Integrative Agriculture,* 20, 1944–1957.

Maimaitijiang, M., Sagan, V., Sidike, P., Hartling, S., Esposito, F. and Fritschi, F.B. (2020). Soybean yield prediction from UAV using multimodal data fusion and deep learning. *Remote Sensing of Environment*, 237, 111599.

Maldonado, S.; López, J. (2014). Imbalanced data classification using second-order cone programming support vector machines. *Pattern Recognition*, 47, 2070–2079.

Maldonado, W. and Barbosa, J.C. (2016). Automatic green fruit counting in orange trees using digital images. *Computer and Electronics in Agriculture*, 127, 572–581.

Mandanici, E., and Bitelli, G. (2016). Preliminary comparison of Sentinel-2 and Landsat 8 imagery for a combined use. *Remote Sensing*, 8, 1014.

Mandal, D., Kumar, V., Ratha, D., Lopez-sanchez, J. M., Bhattacharya, A., Mcnairn, H., Rao, Y. S. and Ramana, K. V (2020). Assessment of rice growth conditions in a semi-arid region of India using the Generalized Radar Vegetation Index derived from RADARSAT-2 polarimetric SAR data. *Remote Sensing of Environment*, 237, 111561.

Maponya, P. and Mpandeli, S. (2012). Climate Change and Agricultural Production in South Africa: Impacts and Adaptation Options. *Journal of Agricultural Science,* 4, 49–60.

Mashaba-Munghemezulu, Z., Chirima, G. and Munghemezulu, C. (2021). Mapping Smallholder Maize Farms Using Multi-Temporal Sentinel-1 Data in Support of the Sustainable Development Goals. *Remote Sensing*, 13, 1666.

Mashonganyika F, Mugiyo H, Svotwa E, Kutywayo D. (2021). Mapping of winter wheat using sentinel-2 NDVI data. a case of Mashonaland Central Province in Zimbabwe. *Frontiers in Climate*, 3, 1–10.

Maxwell, A.E., Warner, T.A. and Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*, 39, 2784–2817.

Mccarty, J. L., Neigh, C. S. R., Carroll, M. L. and Wooten, M. R. (2017). Remote Sensing of Environment Extracting smallholder cropped area in Tigray, Ethiopia with wall-to-wall sub-meter WorldView and moderate resolution Landsat 8 imagery. *Remote Sensing of Environment.* 202, 142–151.

McNairn, H., Champagne, C., Shang, J., Holmstrom, D. and Reichert, G. (2009). Integration of optical and synthetic aperture radar (SAR) imagery for delivering operational *annual crop inventories. ISPRS Journal of Photogrammetry and Remote Sensing,* 64, 434–49.

Mercier, A., Betdeer, J., Baudry. J., Le Roux, V., Spicher, F., Lacoux, J., Roger, D. and Hubert-Moy, L (2020). Evaluation of Sentinel-1 & 2 time series for predicting wheat and rapeseed phenological stages. *ISPRS Journal of Photogrammetry and Remote Sensing*, 163, 231–256.

Meroni, M., D'Andrimont, R., Vrieling, A., Fasbender, D., Lemoine, G., Rembold, F., Seguini, L. and Verhegghen, A. (2021). Comparing land surface phenology of major European crops as derived from SAR and multispectral data of Sentinel-1 and -2. *Remote Sensing of Environment*, 253, 112232.

Minaei, S., Soltanikazemi, M., Shafizadeh-Moghadam, H. and Mahdavian, A. (2022). Field-scale estimation of sugarcane leaf nitrogen content using vegetation indices and spectral bands of Sentinel-2: Application of random forest and support vector regression. *Computers and Electronics in Agriculture*. 200,107130.

Mittapalli, B., Desai, M. D. and Antonio, S. (2014). Estimation of Canopy Nitrogen Concentration Across C3 and C4 Grasslands Using WorldView-2 Multispectral Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7, 2322–2325.

Morin, D.; Planells, M.; Guyon, D.; Villard, L.; Mermoz, S.; Bouvet, A.; Thevenon, H.; Dejoux, J.-F.; Le Toan, T.; Dedieu, G. (2019). Estimation and mapping of forest structure parameters from open access satellite images: Development of a generic method with a study case on coniferous plantation. *Remote Sensing*, 11, 1275.

Moumni, A. and Lahrouni, A. (2021). Machine learning-based classification for crop-type mapping using the fusion of high-resolution satellite imagery in a semiarid area. *Scientifica*, 8810279.

Mukwada, G., Mazibuko, S., Moeletsi, M. and Robinson, G. (2021). Can famine be averted? A spatiotemporal assessment of the impact of climate change on food security in the Luvuvhu River Catchment of South Africa. *Land*, 10, 527.

Muro, J., Canty, M., Conradsen, K., Hüttich, C., Nielsen, A.A., Skriver, H., Remy, F., Strauch, A., Thonfeld, F. and Menz, G. (2016). Short-term change detection in wetlands using Sentinel-1 time series. *Remote Sensing*, 8, 795.

Nabil, M., Farg, E., Arafat, S.M., Aboelghar, M., Afify, N.M. and Elsharkawy, M.M. (2022). Tree-fruits crop type mapping from Sentinel-1 and Sentinel-2 data integration in Egypt's New Delta project. *Remote Sensing Applications: Society and Environment*, 27, 100776.

Nabil, M., Zhang, M., Bofana, J., Wu, B., Stein, A. and Dong, T. (2020). Assessing factors impacting the spatial discrepancy of remote sensing-based cropland products: A case study in Africa. *International Journal of Applied Earth Observation and Geoinformation*, 85, 102010.

Naboureh, A., Li, A., Bian, J., Lei, G., and Amani, M. (2020). A hybrid data balancing method for classification of imbalanced training data within Google Earth Engine: Case studies from mountainous regions. *Remote Sensing*, 12, 3301.

Nakalembe, C. (2020). Field-Level Crop Type Classification with k Nearest Neighbors: A Baseline for a Nearest Neighbors. *arXiv*.1-6.

Navarro, A., Rolim, J., Miguel, I., Catalão, J., Silva, J., Painho, M. and Vekerdy, Z. (2016). Crop monitoring based on SPOT-5 Take-5 and sentinel-1A data for the estimation of crop water requirements. *Remote Sensing*, 8, 1-20.

Ndegwa C, B. M. (2014). Using Hyperspectral Data to Identify Crops in a Cultivated Agricultural Landscape - A Case Study of Taita Hills, Kenya. *Journal of Earth Science & Climatic Change*, 5, 1-4.

Neigh, C.S., Carroll, M.L., Wooten, M.R., McCarty, J.L., Powell, B.F., Husak, G.J., Enenkel, M. and Hain, C.R. (2018). Smallholder crop area mapped with wall-to-wall WorldView sub-meter panchromatic image texture: A test case for Tigray, Ethiopia. *Remote Sensing of Environment*, 212, 8–20.

Ni, R., Tian, J., Xiaojuan, L., Yin, D., Li, J., Gong, H., Zhang, J., Zhu, L. and Wu, D. (2021). An enhanced pixel-based phenological feature for accurate paddy rice mapping with Sentinel-2 imagery in Google Earth Engine. *ISPRS Journal of Photogrammetry and Remote Sensing*, 178, 282–296.

Ni, R., Tian, J., Xiaojuan, L., Yin, D., Li, J., Gong, H., Zhang, J., Zhu, L. and Wu, D. (2020). Mapping sugarcane plantation dynamics in Guangxi, China, by time series Sentinel-1, Sentinel-2 and Landsat images. *Remote Sensing of Environment*, 247, 111951.

Noi, P.T. and Kappas, M. (2017). Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. *Sensors*, 18, 18.

Oeba, V. O. and Illiassou, S. A. (2020). Role of Tree-Based Systems in Enhancing Food Security and Nutrition, 712–723.

Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E. and Wulder, M.A. (2014). Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, 148, 42–57.

Orynbaikyzy, A., Gessner, U. and Conrad, C. (2019). Crop type classification using a combination of optical and radar remote sensing data: A review. *International Journal of Remote Sensing*, 40, 6553–6595.

Ouzemou, J., El, Abderrazak, Lhissou, R., El, Ali, Bouch, N., El, R., Mostafa, E. and El, Abderrahmene (2018). Crop type mapping from pansharpened Landsat 8 NDVI data: A case of a highly fragmented and intensive agricultural system. *Remote Sensing Applications: Society and Environment*, 11, 94–103.

Ozdogan, M., Yang, Y., Allez, G. and Cervantes, C. (2010). Remote sensing of irrigated agriculture: Opportunities and challenges. *Remote Sensing*, 2, 2274–2304.

Pan L, Xia H, Zhao X, Guo Y, Qin Y. (2021). Mapping winter crops using a phenology algorithm time-series Sentinel l-2 and Landsat-7/8 images and Google Earth engine. *Remote Sensing*, 8, 1-21.

Paris, C., Weikmann, G. and Bruzzone, L. (2020). Monitoring of agricultural areas by using Sentinel 2 image time series and deep learning techniques. *Image and Signal Processing for Remote Sensing XXVI*; 115, 136-153.

Patil, B.M. and Burkpalli, V. (2021). A Perspective View of Cotton Leaf Image Classification Using Machine Learning Algorithms Using WEKA. *Advances in Human-Computer Interaction*, 2021, 9367778.

Peña, M.A. and Brenning, A. (2015). Assessing fruit-tree crop classification from Landsat-8 time series for the Maipo Valley, Chile. *Remote Sensing of Environment,* 171, 234–244.

Pena, M.A., Brenning, A. and Liao, R. (2017). Classifying fruit-tree crops by Landsat-8-time series. *IEEE 1st International Symposium on Geoscience and Remote Sensing, GRSS-CHILE 2017*, 19-22.

Peña-barragán, J.M; Ngugi, M.K; Plant, R.E; Six, J. (2011). Object-based crop identification using multiple vegetation indices, textural features, and crop phenology. *Remote Sensing of Environment*, 115, 1301–1316.

Peng, T., Zhou, J., Zhang, C. and Zheng, Y. (2017). Multi-step ahead wind speed forecasting using a hybrid model based on two-stage decomposition technique and AdaBoost-extreme learning machine. *Energy Conversion and Management*, *153*, 589–602.

Pérez-Hoyos, A., Rembold, F., Kerdiles, H. and Gallego, J. (2017). Comparison of global land cover datasets for cropland monitoring. *Remote Sensing*, 9, 1-24.

Persello, C., Tolpekin, V. A., Bergado, J. R. and de By, R. A. (2019). Delineation of agricultural fields in smallholder farms from satellite images using fully convolutional networks and combinatorial grouping. *Remote Sensing of Environment*, 231, 1-18.

Potgieter, A.B., Zhao, Y., Zarco-Tejada, P.J., Chen, K., Zhang, Y., Porker, K., Biddulph, B., Dang, Y.P., Neale, T., Roosta, F. and Chapman, S. (2021). Evolution and application of digital technologies to predict crop type and crop phenology in agriculture. *In Silico Plants*, 3, 1–23.

Preidl, S., Lange, M. and Doktor, D. (2020). Introducing APiC for regionalised land cover mapping on the national scale using Sentinel-2A imagery. *Remote Sensing of Environment*, 240, 111673.

Prins, A.J. and Van Niekerk, A. (2020). Crop type mapping using LiDAR, Sentinel-2 and aerial imagery with machine learning algorithms. *Geo-Spatial Information Science*, 24, 1–13.

Punalekar, S., Verhoef, A., Quaife, T.; Humphries, D., Bermingham, L. and Reynolds, C. (2018). Application of Sentinel-2A data for pasture biomass monitoring using a physically based radiative transfer model. *Remote Sensing of Environment*, 218, 207–220.

Qader, S. H., Dash, J., Alegana, V. A., Khwarahm, N. R., Tatem, A. J. and Atkinson, P. M. (2021). The role of earth observation in achieving sustainable agricultural production in arid and semi-arid regions of the world. *Remote Sensing*, 13, 1–27.

Qadir, A. and Mondal, P. (2020). Synergistic Use of Radar and Optical Satellite Data for Improved Monsoon Cropland Mapping in India. *Remote Sensing*, 12, 522.

Quan, Y., Zhong, X., Feng, W., Chan, J., Li, Q. and Xing, M. (2021). Smote-based weighted deep rotation forest for the imbalanced hyperspectral data classification. *Remote Sensing*, 13, 464.

Rajah, P., Odindi, J., Abdel-Rahman, E. and Mutanga, O. (2017). Determining the optimal phenological stage for predicting common dry bean (Phaseolus vulgaris) yield using field spectroscopy. *South African Journal of Plant and Soil*, 34, 379–388.

Rao, P., Zhou, W., Bhattarai, N., Srivastava, A., Singh, B., Poonia, S., Lobell, D. and Jain, M. (2021). Using Sentinel-1, Sentinel-2, and planet imagery to map crop type of smallholder farms. *Remote Sensing*, 13, 1870.

Remelgado, R., Zaitov, S. and Kenjabaev, S. (2020). A crop type dataset for consistent land cover classification in Central Asia. *Scientific Data*, 7, 1–6.

Ren, T., Xu, H., Cai, X., Yu, S. and Qi, J. (2022). Smallholder crop type mapping and rotation monitoring in mountainous areas with Sentinel-1/2 imagery. *Remote Sensing*, 14, 566.

Richard, K., Abdel-Rahman, E. M., Subramanian, S., Nyasani, J. O., Thiel, M., Jozani, H., Borgemeister, C. and Landmann, T. (2017). Maize cropping systems mapping using rapideye observations in agro-ecological landscapes in Kenya. *Sensors*, 17, 1–16.

Robert, P.C. (2002). Precision agriculture: A challenge for crop nutrition management. *Plant Soil.*, 247, 143–149.

Roth, S. I. B., Leiterer, R., Volpi, M., Celio, E., Schaepman, M. E. and Joerg, P. C. (2019). Automated detection of individual clove trees for yield quantification in northeastern Madagascar based on multi-spectral satellite data. *Remote Sensing of Environment*, 221,144–156.

Rumora, L., Miler, M. and Medak, D. (2020). Impact of various atmospheric corrections on sentinel-2 land cover classification accuracy using machine learning classifiers. *ISPRS International Journal of Geo-Information*, 9, 277.

Saedi, S. and Khosravi, I.H. (2020). A deep neural network approach towards real-time on-branch fruit recognition for precision horticulture. *Expert Systems with Applications*, 159:113594.

Saini, R. and Ghosh, S.K. (2018). Crop Classification on Single Date Sentinel-2 Imagery Using Random Forest and Support Vector Machine. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 5, 683–688.

Saini, R.; Ghosh, S.K. (2021). Crop classification in a heterogeneous agricultural environment using ensemble classifiers and single-date Sentinel-2A imagery. *Geocarto International*, 36, 2141–2159.

Santos, C.L.M., de, O. Lamparelli, R.A.C., Figueiredo, G.K.D.A., Dupuy, S., Bouty, J., Luciano, A.C. and dos, S. (2019). Classification of crops, pastures, and tree plantations along the season with multi-sensor image time series in a subtropical agricultural region. *Remote Sensing*, 11, 1–27.

Sarzynski, T., Giam, X., Carrasco, L., and Lee, J.S.H. (2020). Combining radar and optical imagery to map oil palm plantations in Sumatra, Indonesia, using the Google Earth Engine. *Remote Sensing*, 12, 1220.

Schirrmann, M., Giebel, A., Gleiniger, F. and Pflanz, L. (2016). Monitoring agronomic parameters of winter wheat crops with low-cost UAV imagery. *Remote Sensing*, 8, 1-19.

Schlemmera, M., Gitelson, A., Schepersa, J., Fergusona, R., Peng, Y., Shanahana, 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, 47–54.

Schmidt, M., Pringle, M., Devadas, R., Denham, R. and Tindall, D. (2016). A framework for large-area mapping of past and present cropping activity using seasonal Landsat images and time series metrics. *Remote Sensing*, 8, 312.

Schreier, J., Ghazaryan, G. and Dubovyk, O. (2021). Crop-specific phenomapping by fusing Landsat and sentinel data with MODIS time series. *European Journal of Remote Sensing*, 54:47–58.

Segarra, J., Buchaillot, M.L., Araus, J.L. and Kefauver, S.C. (2020). Remote sensing for precision agriculture: Sentinel-2 improved features and applications. *Agronomy*, 10, 641.

Shang, J., Jiao, X., Mcnairn, H., Kovacs, J., Walters, D., Ma, B. and Geng, X. (2013). Tracking crop phenological development of spring wheat using synthetic aperture radar (SAR) in northern Ontario, Canada. *2013 Second International Conference on Agro-Geoinformatics*. 517–521.

Shelestov, A., Lavreniuk, M., Kussul, N., Novikov, A. and Skakun, S. (2017). Exploring Google Earth Engine platform for big data processing: Classification of multi-temporal satellite imagery for crop mapping. *Frontiers of Earth Science*, 5, 1–10.

Shi, W., Wang, M. and Liu, Y. Crop yield and production responses to climate disasters in China. *Science of Total Enviroment*, 750, 141147.

Singh, R. K., Rizvi, J., Behera, M. D. and Biradar, C. (2021). Automated crop type mapping using time-weighted dynamic time warping-A basis to derive inputs for enhanced food and Nutritional Security. *Current Research in Environmental Sustainability*, 3, 1-9.

Singh, R., Patel, N.R. and Danodia, A. (2022). Deriving phenological metrics from landsat OLI for sugarcane crop type mapping: a case study in North India. *Journal of the Indian Society of Remote Sensing*, 50:1021–1030.

Singh, V. Patel, A.N., Dalwadi, A., Kathota, J. and Suthar, J. (2017). Horticultural Fruit Crop Plantations Mapping using Geoinformatics Technology in Gujarat State, India. *International Journal of Advanced Remote Sensing and GIS*, 6, 2033–2049.

Sinha, P., Robson, A., Schneider, D., Kiic, T., Mugera, H.K., Ilukor, J. and Tindamanyire, J.M. (2020). The potential of in-situ hyperspectral remote sensing for differentiating 12 banana genotypes grown in Uganda. *ISPRS Journal of Photogrammetry and Remote Sensing*, 167, 85–103.

Sinha, R., Quirós, J.J., Sankaran, S. and Khot, L.R. (2021). High resolution aerial photogrammetry-based 3D mapping of fruit crop canopies for precision inputs management. *Information Processing in Agriculture*, 9, 1–13.

Sishodia, R. P., Ray, R. L. and Singh, S. K. (2020). Applications of remote sensing in precision agriculture: A review. *Remote Sensing*, 12, 1–31.

Sivasankar, T., Kumar, D., Srivastava, H. and Patel, P. (2018). Advances in radar remote sensing of agricultural crops: A review. *International Journal of Advanced Science Engineering and Information Technology*, 8, 1126–1137.

Song, Q., Xiang, M., Hovis, C., Zhou, Q., Lu, M., Tang, H. and Wu, W. (2019). Object-based feature selection for crop classification using multi-temporal high-resolution imagery. *International Journal of Remote Sensing*, 40, 2053–2068.

Southworth, J. and Muir, C. (2021). Speciality grand challenge: remote sensing time series analysis. *Frontiers in Remote Sensing*, 2, 1-5.

Steinhausen, M.J., Wagner, P., Narasimhan, B. and Waske, B. (2018). Combining Sentinel-1 and Sentinel-2 data for improved land use and land cover mapping of monsoon regions. *International Journal of Applied Earth Observation and Geoinformation*, 73, 595–604.

Sun, L., Chen, J. and Han, Y. (2019). Joint use of time series Sentinel-1 and Sentinel-2 imagery for cotton field mapping in heterogeneous cultivated areas of Xinjiang, China. *In Proceedings of the 2019 8th International Conference on Agro-Geoinformatics 2019,* 2019-2022.

Sun, Z., Di, L. and Fang, H. (2020). Machine Learning on Greenest Pixels for Crop Mapping. *EarthArXiv*.

Taghizadeh-Mehrjardi, R., Schmidt, K., Eftekhari, K., Behrens, T., Jamshidi, M., Davatgar, N., Toomanian, N., and Scholten, T. (2020). Synthetic resampling strategies and machine learning for digital soil mapping in Iran. *European Journal of Soil Science,* 71, 352–368.

Talema, T. and Hailu, B.T. (2020). Mapping rice crop using sentinels (1 SAR and 2 MSI) images in tropical area: A case study in Fogera wereda, Ethiopia. *Remote Sensing Applications: Society and Environment*, 18, 100290.

Tariq, A., Yan, J., Gagnon, A. S., Riaz Khan, M. and Mumtaz, F. (2022). Mapping of cropland, cropping patterns and crop types by combining optical remote sensing images with decision tree classifier and random forest. *Geo-Spatial Information Science*. 302-320.

Tatsumi, K., Yamashiki, Y., Torres, M.A.C. and Taipe, C.L.R. (2015). Crop classification of upland fields using Random Forest of time-series Landsat 7 ETM+ data. *Computer Electronics in Agriculture*, 115, 171–179.

Tavares, P.A., Beltrão, N.E.S., Guimarães, U.S. and Teodoro, A.C. (2019). Integration of Sentinel-1 and Sentinel-2 for classification and LULC mapping in the urban area of Belém, eastern Brazilian Amazon. *Sensors,* 19, 1140.

Tian, H., Huang, N., Niu, Z., Qin, Y., Pei, J. and Wang, J. (2019). Mapping winter crops in China with multi-source satellite imagery and phenology-based algorithm. *Remote Sensing*, 11, 1–23.

Tian, Y., Yang, G., Wang, Z., Wang, H., Li, E.and Liang, Z. Apple detection during different growth stages in orchards using the improved YOLO-V3 model. *Remote Sensing Applications: Society and Environment.* 18, 100290.

Torbick, N., Chowdhury, D., Salas, W. and Qi, J. (2017). Monitoring rice agriculture across myanmar using time series Sentinel-1 assisted by Landsat-8 and PALSAR-2. *Remote Sensing*, 9, 119.

Tu, Y. H., Johansen, K., Phinn, S. and Robson, A. (2019). Measuring canopy structure and condition using multispectral UAS imagery in a horticultural environment. *Remote Sensing*, 11, 15–17.

Tu, Y.H., Phinn, S., Johansen, K., Robson, A. and Wu, D. (2020). Optimizing drone flight planning for measuring horticultural tree crop structure. *ISPRS Journal of Photogrammetry and Remote Sensing*, 160, 83–96.

Ukwuoma CC, Zhiguang Q, Bin Heyat MB, Ali L, Almaspoor Z, Monday HN. (2022). Recent advancements in fruit detection and classification using deep learning techniques. *Mathematical Problems in Engineering*, 2022, 1-29.

United Nations. (2015). *World Population Prospects: The 2015 Revision; Key Findings and Advance Tables*; United Nations: New York, NY, USA, 4, 88–100.

Useya, J. and Chen, S. (2019). Exploring the potential of mapping cropping patterns on smallholder scale croplands using Sentinel-1 SAR data. *Chinese Geographical Science*, 29, 626–639.

Usha, K. and Singh, B. (2013). Potential applications of remote sensing in horticulture: A review. *Scientia Horticulture*, 153, 71–83.

Ustuner, M., Sanli, F. and Abdikan, S. (2016). Balanced vs imbalanced training data: Classifying Rapideye data with support vector machines. *International Archives of*

the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 41, 379–384.
Valcarce-Diñeiro, R., Arias-Pérez, B., Lopez-Sanchez, J. M. and Sánchez, N. (2019). Multi-temporal dual- and quad-polarimetric synthetic aperture radar data for crop-type mapping. *Remote Sensing*, 11, 1-20.
Valero, S., Arnaud, L., Planells, M. and Ceschia, E. (2011). Synergy of Sentinel-1 and Sentinel-2 imagery for early seasonal agricultural crop mapping. *Remote Sensing*, 13, 1–32.
Van Tricht, K., Gobin, A., Gilliams, S. and Piccard, I. (2018). Synergistic use of radar Sentinel-1 and optical Sentinel-2 imagery for crop mapping: A Case study for Belgium. *Remote Sensing*, 10, 1642.
Vanino, S., Nino, P., De Michele, C., Bolognesi, S.F., D'Urso, G., Di Bene, C., Pennelli, B., Vuolo, F., Farina, R. and Pulighe, G. (2018). Capability of Sentinel-2 data for estimating maximum evapotranspiration and irrigation requirements for tomato crop in Central Italy. *Remote Sensing of Environment*, 215, 452–470.
Vapnik, V.N. The Nature of Statistical Learning Theory, 2nd ed.; Springer: New York, NY, USA, 2000.
Vapnik, V.N. *The Nature of Statistical Learning Theory*; Springer: Berlin/Heidelberg, Germany, 1995, 334. Available at: https://ci.nii.ac.jp/naid/10020951890 (accessed on 6 October 2022).
Vasconez JP, Delpiano J, Vougioukas S, Cheein FA. (2020), Comparison of convolutional neural networks in fruit detection and counting: a comprehensive evaluation. *Computer and Electronics in Agriculture*, 173,105348.
VDM, "Vhembe District Municipality 2020/21 Idp Review. 2020. Available at: https://www.cogta.gov.za/cgta_2016/wp-content/uploads/2020/12/VHEMBE-DISTRICT-MUNICIPALITY-2020-21-IDP-REVIEW.pdf. (accessed on 16 December 2021).
Veloso, A., Mermoz, S., Bouvet, A., Le Toan, T., Planells, M., Dejoux, J.-F. and Ceschia, E. (2017). Understanding the temporal behaviour of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications. *Remote Sensing of Environment*, 199, 415–426.
Venkatesan, M., Pazhanivelan, S. and Sudarmanian, N. S. (2019). Multi-Temporal feature extraction for precise maize area mapping using time-series Sentinel SAR data. *XLII*, 18–20.
Villa, P., Stroppiana, D., Fontanelli, G., Azar, R. and Brivio, P.A. (2015). In-season mapping of crop type with optical and X-band SAR data: a classification tree approach using synoptic seasonal features. *Remote Sensing*,10, 12859.
Vira, B., Agarwal, B., Jamnadass, R., Kleinschmit, D., McMullin, S., Mansourian, S., Neufeldt, H., Parrotta, J. A., Sunderland, T. and Wildburger, C. (2015). Forests, trees and landscapes for food security and nutrition. *Forests and food: Addressing hunger and nutrition across sustainable landscapes*, 33, 9–28.
Vluymans, S. Learning from imbalanced data. (2019). *IEEE Studies in Computational Intelligence*, 807, 81–110.
Vogels, M. F. A., de Jong, S. M., Sterk, G., Douma, H. and Addink, E. A. (2019). Spatio-temporal patterns of smallholder irrigated agriculture in the horn of Africa using GEOBIA and Sentinel-2 imagery, *Remote Sensing*, 11.
Vogels, M. F. A., Jong, S. M. De, Sterk, G., Wanders, N., Bierkens, M. F. P. and Addink, E. A. (2020). An object-based image analysis approach to assess irrigation-water consumption from MODIS products in Ethiopia. *International Journal of Applied Earth Observation and Geoinformation*, 88, 102067.

Vreugdenhil, M.; Wagner, W.; Bauer-Marschallinger, B.; Pfeil, I.; Teubner, I.; Rüdiger, C.; Strauss, P. (2018). Sensitivity of Sentinel-1 backscatter to vegetation dynamics: An Austrian case study. *Remote Sensing*, 10, 1396.

Vuolo, F., Neuwirth, M., Immitzer, M., Atzberger, C. and Ng, W. (2018). How much does multi-temporal Sentinel-2 data improve crop-type classification? *International Journal of Applied Earth Observation & Geoinformation*,72, 122–130.

Waldner, F., Bellemans, N., Hochman, Z., Newby, T., de Abelleyra, D., Verón, S. R., Bartalev, S., Lavreniuk, M., Kussul, N., Maire, G. Le, Simoes, M., Skakun, S. and Defourny, P. (2019) Roadside collection of training data for cropland mapping is viable when environmental and management gradients are surveyed. *International Journal of Applied Earth Observation and Geoinformation*, 80, 82–93.

Waldner, F., Chen, Y., Lawes, R. and Hochman, Z. (2019). Needle in a haystack: Mapping rare and infrequent crops using satellite imagery and data balancing methods. *Remote Sensing of Environment*, 233, 111375.

Waldner, F., Fritz, S., Di Gregorio, A. and Defourny, P. (2015). Mapping priorities to focus cropland mapping activities: Fitness assessment of existing global, regional and national cropland maps. *Remote Sensing*, 7, 7959–7986.

Wang, M., Liu, Z., Ali Baig, M.H., Wang, Y., Li, Y. and Chen, Y. (2019). Mapping sugarcane in complex landscapes by integrating multitemporal Sentinel-2 images and machine learning algorithms. *Land Use Policy*. 88, 104190.

Wang, Q. and Zhang, J. (2016). A data transfer fusion method for discriminating similar spectral classes. *Sensors*, 16, 1895.

Wang, S., Azzari, G., and Lobell, D.B. (2019). Crop type mapping without field-level labels: Random Forest transfer and unsupervised clustering techniques. *Remote Sensing of Environment*, 222, 303–317.

Wang, W., Liu, X., and Chan, W.K.V. (2020). Imbalanced classification problem using data-driven and random forest method. *ACM International Conference Proceeding Series*, 26–30.

Wardlow, B.D., Egbert, S.L. and Kastens, J.H. (2007). Analysis of time-series MODIS 250 m vegetation index data for crop classification in the U.S. Central Great Plains. *Remote Sensing of Environment*, 108, 290–310.

Weier, S.M., Grass, I., Linden, V., Tscharntke, T. and Taylor, P.J. (2018). Natural vegetation and bug abundance promote insectivorous bat activity in macadamia orchards, South Africa. *Biological Conservation*, 226, 16–23.

Wolfert, S., Ge, L., Verdouw, C. and Bogaardt, M.-J. (2017). Big data in smart farming—A review. *Agricultural Systems*, 153, 69–80.

Woodruff, K. (2017). Introduction to boosted decision trees. *In Proceedings of the Machine Learning Group Meeting, New Mexico State University, Las Cruces, NM, USA, 29 September*, 1–11.

Wu, J., Yang, G., Yang, H., Zhu, Y., Li, Z. and Lei, L. (2020). Extracting apple tree crown information from remote imagery using deep learning. *Computers and Electronics in Agriculture*, 174, 105504.

Xie, G. and Niculescu, S. (2022). Mapping crop types using sentinel-2 data machine learning and monitoring crop phenology with sentinel-1 backscatter time series in Pays de Brest, Brittany France. *Remote Sensing*, 14, 1-27.

Xiong, Y., Zeng, X., Liao, J., Lai, W., Chen, Y. and Zhu, M. (2022). An approach to detecting and mapping individual fruit trees integrated YOLOv5 with UAV remote sensing.

Yandún Narváez, F.J., Salvo del Pedregal, J.P., Prieto, A., Torres-Torriti, M. and Auat Cheein, F.A. (2016). LiDAR and thermal images fusion for ground-based 3D characterisation of fruit trees. *Biosystems Engineering*, 151, 479–494.

Yang, H., Pan, B., Li, N., Wang, W., Zhang, J. and Zhang, X. (2021). A systematic method for spatio-temporal phenology estimation of paddy rice using time series Sentinel-1 images. *International Journal of Remote Sensing*, 259, 112394.

Yedage, A.S., Gavali, R.S. and Jarag, A.P. (2013). Land assessment for horticulture (Pomegranate) crop using GIS and Fuzzy decision analysis in the Sangola taluka of Solapur District. *International Journal of Remote Sensing*, 2,104–13.

Yi, Z., Jia, L. and Chen, Q. (2020). Crop classification using multi-temporal Sentinel-2 data in the Shiyang river basin of China. *Remote Sensing*, 12, 1–21.

Yin, L. (2023). Mapping smallholder cashew plantations to inform sustainable tree crop expansion in Benin. *Remote Sensing of Environment*, 295, 113695.

You, L., Wood, S. and Wood-sichra, U. (2009). Generating plausible crop distribution maps for Sub-Saharan Africa using a spatially disaggregated data fusion and optimization approach. *Agricultural Systems*, 99,126–140.

You, N. and Dong, J. (2019). Examining earliest identifiable timing of crops using all available Sentinel 1/2 imagery and Google Earth Engine. *ISPRS Journal of Photogrammetric Remote Sensing*,161, 109–123.

Yousefi, D.M., Rafie, A.M., Aziz, S.A., Azrad, S., Masri, M.M.M., Shahi, A. and Marzuki, O. (2021). Classification of oil palm female inflorescences anthesis stages using machine learning approaches. *Information Processing in Agriculture*, 8, 537–549.

Yu, Y., Li, M., and Fu, Y. (2018). Forest type identification by random forest classification combined with SPOT and multitemporal SAR data. *Journal of Forestry Research*, 29, 1407–1414.

Zeyada, H.H., Ezz, M., Nasr, A., Shokr, M. and Harb, H.M. (2016). Evaluation of the discrimination capability of full polarimetric SAR data for crop classification. *International Journal of Remote Sensing*, 37, 2585–2603.

Zhang, D., Pan, Y., Zhang, J., Hu, T., Zhao, J., Li, N. and Chen, Q. (2020). A generalized approach based on convolutional neural networks for large area cropland mapping at very high resolution. *Remote Sensing of Environment*, 247,111912.

Zhang, H. (2018). Object-based crop classification using multi-temporal SPOT-5 imagery and textural features with a Random Forest classifier. *Geocarto International*, 33, 1017–1035.

Zhang, H., Kang, J., Xu, X. and Zhang, L. (2020). Accessing the temporal and spectral features in crop type mapping using multi-temporal Sentinel-2 imagery: A case study of Yi'an County, Heilongjiang province, China. *Computer and Electronics in Agriculture*, 176, 105618.

Zhang, H.Y., Zhu Q.B., Huang M., and Guo Y. (2018). Automatic determination of optimal spectral peaks for classification of Chinese tea varieties using laser induced breakdown spectroscopy. *International Journal of Agricultural and Biological Engineering*,11, 154–8.

Zhang, M., Qin, Z., Liu, X. and Ustin, S. L. (2003). Detection of stress in tomatoes induced by late blight disease in California, USA, using hyperspectral remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 4, 295–310.

Zhang, M., Wu, B., Yu, M., Zou, W. and Zheng, Y. (2014). Crop condition assessment with adjusted NDVI using the uncropped arable land ratio. *Remote Sensing*, 6, 5774–5794.

Zhang, P., Hu, S., Li, W. and Zhang, C. (2020). Parcel-level mapping of crops in a smallholder agricultural area: A case of central China using single-temporal VHSR imagery. *Computer and Electronics in Agricultur*e, 175, 105581.

Zhang, S., Guo, J., Luo, N., Wang, L., Wang, W. and Wen, K. (2019). Improving Wi-Fi fingerprint positioning with a pose recognition assisted SVM algorithm. *Remote Sensing*, 11, 652.

Zhang, W., Wang J., Liu Y., Chen K., Li H., Duan Y., Wu W., Shi Y., and Guo W. (2021). Deep-learning-based in-field citrus fruit detection and tracking. *Horticulture Research*, 9, 1-14.

Zhao, G., Wang, L., Zheng, J., Tuerxun, N., Han, W. and Liu, L. (2023). Optimized Extraction Method of Fruit Planting Distribution Based on Spectral and Radar Data Fusion of Key Time Phase. *Remote Sensing and Multidisciplinary Digital Publishing Institute*, 15.

Zhao, K., Valle, D., Popescu, S., Zhang, X. and Mallick, B. (2013). Hyperspectral remote sensing of plant biochemistry using Bayesian model averaging with variable and band selection. *Remote Sensing of Environment*, 132, 102–119.

Zhao, W., Qu, Y., Chen, J. and Yuan, Z. (2020). Deeply synergistic optical and SAR time series for crop dynamic monitoring. *Remote Sensing of Environment*, 247, 111952.

Zheng, B., Myint, S.W., Thenkabail, P.S. and Aggarwal, R.M. (2015). A support vector machine to identify irrigated crop types using time-series Landsat NDVI data. *International Journal of Applied Earth Observation and Geoinformation*, 34, 03–112.

Zhi, F., Dong, Z., Guga, S., Bao, Y., Han, A., Zhang, J. and Bao, Y. (2022). Rapid and Automated Mapping of Crop Type in Jilin Province Using Historical Crop Labels and the Google Earth Engine. *Remote Sensing*, 14, 4028.

Zhong, L., Hu, L. and Zhou, H. *(*2019). Deep learning based multi-temporal crop classification ☆. *Remote Sensing of Environment*, 221, 430–443.

Zhong, Y., Hu, X., Wei, L. and Zhang, L. (2020). A robust spectral-spatial approach to identifying heterogeneous crops using remote sensing imagery with high spectral and spatial resolutions. *Remote Sensing of Environment*, 239, 111605.

Zhou, T., Pan, J., Zhang, P., Wei, S. and Han, T. (2017). Mapping winter wheat with multi-temporal SAR and optical images in an urban agricultural region. *Sensors*, 17, 1210.

Zhou, X.X., Li, Y.Y., Luo. Y.K., Sun, Y.W., Su, Y.J., Tan, C.W. and Liu, Y.J. (2022). Research on remote sensing classification of fruit trees based on sentinel-2 multi-temporal imageries. *Scientific Reports*, 12:1–15.

Zhou, Z., Huang, J., Wang, J., Zhang, K., Kuang, Z., Zhong, S. and Song, X. (2015). Object-oriented classification of sugarcane using time series middle-resolution remote sensing data based on AdaBoost. *PLoS ONE*, 10, e0142069.

Zhu, L., Radeloff, V. and Ives, A.R. (2017). Improving the mapping of crop types in the Midwestern U.S. by fusing Landsat and MODIS satellite data. *International Journal of Applied. Earth Observation and. Geoinformation*, 58, 1–11.

Zurita-Milla, R., Izquierdo-Verdiguier, E. and De By, R.A. (2017). Identifying crops in smallholder farms using time series of WorldView-2 images. *9th International Workshop Analysis of Multitemporal Remote Sensing Images*, 21–23.