



Sugarcane Classification and Age Estimation using Sentinel-2

A Case Study Transmara West, Narok County, Kenya.

Master Thesis

for the attainment of the Master's degree "Master of Science",
abbreviated "MSc"

submitted within the University Master Program for Further Education
"Geographical Information Science & Systems – (UNIGIS MSc)"
at the Department of Geoinformatics - Z_GIS,
Faculty of Digital and Analytical Sciences,
University of Salzburg

submitted by

Gareth John Jones

Supervisor:
Dr. Karl Atzmanstorfer

Enoosaen, Kenya, August 2026

DECLARATION

I hereby declare that I have written this thesis on my own and that I have not used any sources or other tools and resources other than those indicated. All passages that have been taken literally or in substance from the cited sources are clearly identified as such

I confirm that I have not used any illustrations from textbooks or other sources unless they are subject to a license permitting free use.

I further confirm that I have used artificial intelligence (AI)-based applications exclusively for those purposes explicitly permitted in the [guidelines on the use of generative AI in the UNIGIS program](#), in its current version. Generative AI tools were used in a limited capacity for minor code debugging, grammar clarification and sentence refinement to improve clarity and readability. All analytical methods, spatial analysis, interpretations, and content were independently developed, verified, and written by the author. I bear full responsibility for all content of this thesis.

This thesis has not previously been submitted in the same or a similar form as a Bachelor's/Master's/Diploma thesis/Dissertation/Master Thesis."

A handwritten signature in black ink, consisting of a stylized 'J' followed by a horizontal line and a small loop.

ACKNOWLEDGEMENTS

This research would not have been possible without the support, guidance and encouragement of many individuals and organisations.

I would like to express my gratitude to my supervisor, Dr. Karl Atzmanstorfer, for his guidance, constructive feedback and encouragement throughout this research. I would also like to extend my appreciation to the UNIGIS Salzburg team for their continued support, assistance and encouragement throughout my master's studies. The knowledge and learning opportunities provided through the UNIGIS programme have contributed significantly to both my academic and professional development.

My sincere thanks go to Transmara Sugar Company Limited (TSCL) for providing access to operational data and resources that made this research possible. I would also like to thank the Executive Management of Transmara Sugar Company Limited for their support and encouragement in pursuing my postgraduate studies.

I would finally like to thank my wife and family for their support, patience and understanding throughout this process. Their belief in me provided the motivation and strength needed to complete both this research and the UNIGIS programme successfully.

To all those who contributed directly or indirectly to the completion of this study, I offer my sincere appreciation and gratitude.

ABSTRACT

Accurate information on sugarcane classification and crop age estimation is important for harvest planning, cane supply forecasting and mill planning. In Kenya the monitoring cane availability relies heavily on miller databases and field inspections, which may not effectively capture changes in sugarcane extent, particularly for Non-Contracted Cane (NCC). This study investigated the use of Sentinel-2 time-series imagery and machine learning techniques for sugarcane classification and crop age estimation within the Transmara Sugar Company Limited (TSCL) supply area in Narok County, Kenya. Sentinel-2 imagery captured between September 2023 and August 2025 for age estimation using Sentinel Scene Classification Layer, and Sentinel-2 imagery sourced between December 2024 and August 2025 was processed using Google Earth Engine for sugarcane classification. Spectral bands and vegetation indices were used as variables in a Random Forest classification model. Classification accuracy was assessed using a validation dataset and validated against TSCL Farmer Registry contracted area. The Random Forest classifier achieved an Overall Accuracy of 98.15% and a Kappa Coefficient of 0.963. Validation against TSCL Farmer Registry produced an area accuracy of 91.5%, also additionally confirming the reliability of the classification. The study identified a total sugarcane area of 18,899.1ha, of which 10,064.1ha was contracted sugarcane to TSCL and 8,835.0ha was classified as NCC. The NCC represented 46.8% of the classified sugarcane area. Crop age estimation achieved an accuracy of 87.2% when compared with TSCL Farmer Registry contracted fields age records. The age estimations therefore transferred to the NCC classified fields, assigning an estimated age to 90.4% of the classified NCC area. The results show that Sentinel-2 time-series imagery and machine learning techniques provide a reliable approach for sugarcane monitoring. The developed framework can act as a supporting tool for monitoring and verifying sugarcane supply within Kenya and monitoring of NCC sugarcane production within Kenya's sugar catchments.

Keywords: Agricultural Monitoring, Crop Age Estimation, Google Earth Engine, Kenya, Non-Contracted Cane, Random Forest, Remote Sensing, Sentinel-2, Sugarcane Classification, Time-Series.

LIST OF ABBREVIATIONS

Abbreviation	Meaning
BOA	Bottom of Atmosphere
COMESA	Common Market for Eastern and Southern Africa
DEM	Digital Elevation Model
EAC	East African Community
ESA	European Space Agency
GEE	Google Earth Engine
ITCZ	Inter-Tropical Convergence Zone
KNBS	Kenya National Bureau of Statistics
KAM	Kenya Association of Manufacturers
L2A	Level 2A
LTM	Long Term Mean
MGRS	Military Grid Reference System
NCC	Non-Contracted Cane
NDMI	Normalized Difference Moisture Index
NDRE	Normalized Difference Red Edge Index
NDVI	Normalized Difference Vegetation Index
NIR	Near Infrared
RF	Random Forest
SCL	Scene Classification Layer
SWIR	Shortwave Infrared
TCH	Tons Cane per Hectare
TCHM	Tons Cane per Hectare per Month
TOA	Top of Atmosphere
TSCL	Transmara Sugar Company Limited
USDA	United States Department of Agriculture

FIGURES

Figure 1.1: Kenya's main sugarcane growing regions.....	2
Figure 1.2: Conceptual Framework.....	13
Figure 2.1: Study area showing the regions of TSCL	15
Figure 2.2: Five years mean monthly rainfall for TSCL	17
Figure 2.3: Five years mean monthly minimum and maximum temperatures for TSCL	18
Figure 2.4: Elevation of TSCL	19
Figure 2.5: Soil suitability classification of TSCL.....	20
Figure 2.6: Southern sugar catchment area	21
Figure 2.7: Sentinel-2 tiles (36MXD and 36MYD) of TSCL	24
Figure 2.8: Percentage of TSCL field polygons affected by cloud cover (Classification).....	30
Figure 2.9: Percentage of TSCL field polygons affected by cloud cover (SCL Series).....	32
Figure 2.10: Transition Duration from Bare Soil to Vegetation	35
Figure 2.11: Illustration of sugarcane growth following planting and harvesting	36
Figure 2.12: Example NDVI profile for timber, bush, and tea.....	40
Figure 2.13: Example NDVI profile for open area and grassland.....	40
Figure 2.14: Example NDVI profile for maize	41
Figure 2.15: Example of centroid generation following the application of a 5m negative	45
Figure 2.16: Spatial distribution of training and validation samples across TSCL.....	46
Figure 3.1: Farmer Registry sugarcane area by TSCL regions	60
Figure 3.2: Area of registered sugarcane fields classified as sugarcane and NCC by TSCL regions	61
Figure 3.3: Percentage Agreement Between Classified Sugarcane Area and TSCL Farmer Registry by Region.....	62
Figure 3.4: RF Classification Results TSCL Field Polygons.....	63
Figure 3.5: TSCL Field Polygons Contracted & Non-contracted.....	64
Figure 3.6: Regional Summary of Contracted and Non-Contracted (NCC) Sugarcane Classification.....	65
Figure 3.7: NCC Classification Age Groups.....	69
Figure 3.8: NCC Age Estimated SCL Series	70

TABLES

Table 2.1: Sentinel-2 Spectral Bands Used in this Study.....	22
Table 2.2: Sentinel-2 Scene Classification Layer (SCL) Classes.....	23
Table 2.3: Sentinel-2 Scene Classification Layer (SCL) Images Used for Crop Age Estimation	25
Table 2.4: Sentinel-2 Image Acquisition Dates Used for Sugarcane Classification	26
Table 2.5: Summary of Data Sources Used in this Study	27
Table 2.6: Software Used in this Study.....	28
Table 2.7: Percentage of Cloud-Covered TSCL Fields in the Selected Sentinel-2 Images	30
Table 2.8: Percentage of Cloud-Covered TSCL Field Polygons in the Selected Sentinel-2 Images	31
Table 2.9: Distribution of Sugarcane Training and Validation Samples by TSCL Regions.....	37
Table 2.10: Distribution of Training and Validation Samples by Farmer Registry Crop Age....	38
Table 2.11: Distribution of Other Land-Cover Training Samples by TSCL Region	42
Table 2.12: Distribution of Other Land-Cover Training and Validation Samples by TSCL Region.	43
Table 2.13: Distribution of Training Samples by Land-Cover Class and TSCL Region.....	43
Table 2.14: Distribution of Validation Samples by Land-Cover Class and TSCL Region.....	44
Table 2.15: Sentinel-2 Scene Classification Layer (SCL) Classes used in this Study	50
Table 3.1: Confusion Matrix for the Random Forest Sugarcane Classification.....	57
Table 3.2: Class Specific Classification Accuracy Statistics	58
Table 3.3: Summary of Classification Accuracy Statistics	59
Table 3.4: Regional Comparison of Classified Sugarcane Area and TSCL Farmer Registry Data.	61
Table 3.5: Summary of Random Forest Classification Results.....	63
Table 3.6: Classified Contracted and Non-Contracted (NCC) Sugarcane Area by TSCL Regions	65
Table 3.7: Validation Results for Sugarcane Crop Age Estimation.....	67
Table 3.8: Summary of Crop Age Estimation Agreement	67

Contents

DECLARATION	i
ACKNOWLEDGMENTS.....	ii
ABSTRACT	iii
LIST OF ABBRIVIATIONS	iv
FIGURES	v
TABLES	vi
1 Introduction	1
1.1 Background	1
1.2 Problem Statement	4
1.3 Aims and Objectives	5
1.3.1 Research Question.....	5
1.3.2 Research Hypothesis	5
1.3.3 Research Objectives	6
1.3.4 Study Limitations	6
1.4 Literature Review	7
1.5 Conceptual Framework	13
2 Materials and Methods	15
2.1 Study Area	15
2.2 Data Sources	22
2.3 Data Preprocessing	28
2.4 Vegetation Indices	33
2.5 Training Data.....	34
2.5.1 Sugarcane Training Sample Development.....	34
2.5.2 Non-Sugarcane Training Sample Development.....	39

2.5.3 Training Sample Preparation.....	44
2.6 Random Forest Classification	46
2.7 Crop Age Estimation	49
2.7.1 Development of the Sentinel-2 SCL Time Series	49
2.7.2 Identification of Bare Soil Events	50
2.7.3 Determination of Date Last Bare Soil	50
2.7.4 Estimation of Cane Age	51
2.7.5 Crop Cycle Attribution.....	51
2.7.6 Validation Against Farmer Registry Records	52
2.8 Accuracy Assessment.....	52
2.8.1 Validation Dataset	52
2.8.2 Confusion Matrix Assessment	53
2.8.3 Overall Accuracy, Producers Accuracy and Users Accuracy	54
2.8.4 Kappa Coefficient	54
2.8.5 Farmer Registry Validation.....	55
2.8.6 Crop Age Estimation Validation	56
3 Results.....	57
3.1 Classification Accuracy Assessment.....	57
3.2 Validation Using Farmer Registry Database.....	59
3.3 Spatial Distribution of Contracted and Non-Contracted Sugarcane.....	62
3.4 Crop Age Estimation Validation	66
3.5 NCC Age Estimation Results	68
3.6 Summary	70
4 Discussion	72
5 Conclusion	77
REFERENCES	80
APPENDIX	87

1: Introduction

1.1: Background

Sugarcane has long been part of Kenya's agricultural environment, both for chewing and traditional brews, before colonial rule (Mati & Thomas, 2019). Industrial sugarcane was introduced in 1902 with trial areas in Kisumu, followed by the first factories at Miwani (1922) and Ramisi (1927) on the Coast (Mati & Thomas, 2019). During the post-independence period, policy reforms such as Sessional Paper No. 10 of 1965 allowed African farmers to grow sugarcane, while the government invested in large public sugar mills, including Muhoroni (1966), Chemelil (1968), Mumias (1973), Nzoia (1978), and South Nyanza (1979) (Kombo & Ndiema, 2023). By 1976 production of 296,000 tons briefly exceeded national consumption of 253,000 tons, and therefore briefly making Kenya making a net exporter, and the establishment of the Kenya Sugar Authority in 1973 sought to regulate and coordinate the sector (Mati & Thomas, 2019).

Since the 1980s, Kenya has shifted from exporting more sugar than it imported to importing more than it exported (Mati & Thomas, 2019; Kombo & Ndiema, 2023). Although the area under sugarcane increased to about 220,000 hectares, the average yield remained at around 55 tons per hectare, while the production costs rose to more than \$1,000 per ton (Mati & Thomas, 2019). These costs were much higher than global averages and therefore reduced profitability and the long-term sustainability of the sector (Mati & Thomas, 2019). Policy liberalization in the 1990s was intended to encourage private investment but instead increased inefficiencies through unregulated competition, cane poaching and delayed payments to farmers (Kombo & Ndiema, 2023). By the early 2000s, Kenya relied on Common Market for Eastern and Southern Africa (COMESA) safeguard measures to support the sugar industry, with repeated extensions highlighting the gap between local sugar production and national demand (Gitonga & Ferrese, 2025). These challenges grew during the 2010s as increasing mill debt, declining sugar production rates and the closure of major mills, including Mumias and Miwani, highlighted ongoing challenges in the industry (Kombo & Ndiema, 2023).

Percentage Distribution of Kenya's Sugarcane Growing Regions

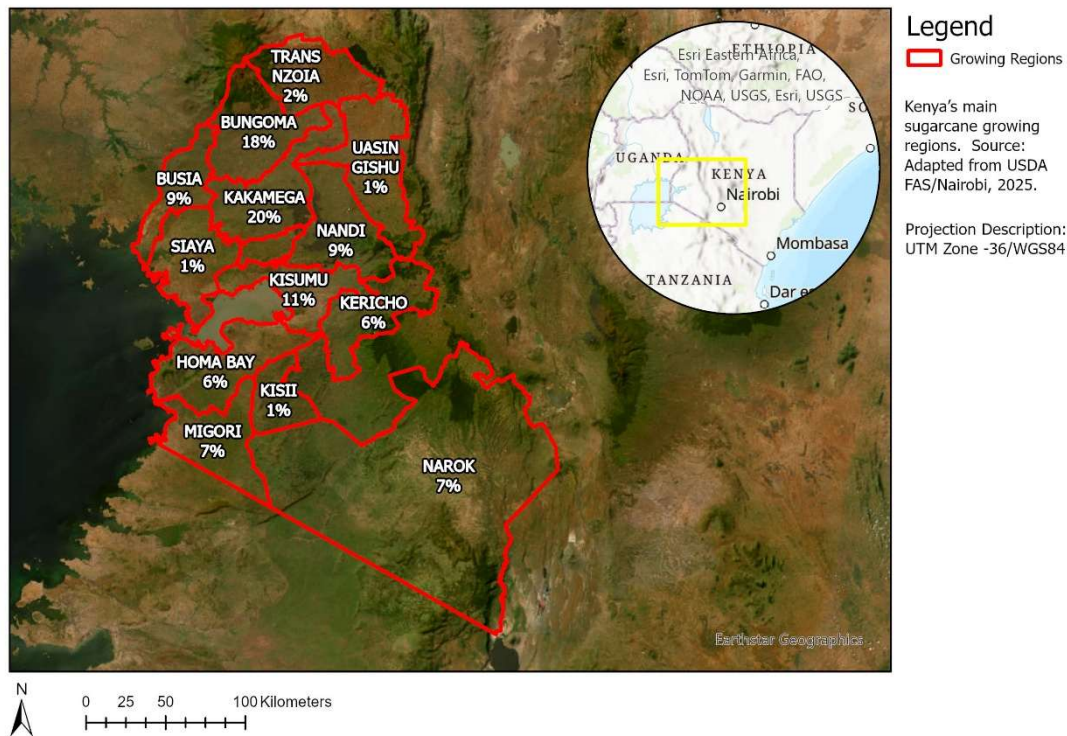


Figure 1.1 - Kenya's main sugarcane growing regions. Source: Created by the author using county boundary data provided by TSCL and sugarcane production statistics from Gitonga and Ferrese (2025).

Sugarcane plays a critical role in Kenya's socio-economic and agricultural infrastructure, (see Figure 1.1 that maps the sugarcane growing regions of Kenya) (Gitonga & Ferrese, 2025). The sugarcane production distribution supports over 8 million people, and contributes towards food security, regional development and employment (KAM, 2021). The industry also contributes 15% of Kenya's agricultural GDP and supports about 250,000 smallholder farmers directly (KAM, 2021). In Narok County, sugarcane production has become a key economic activity, driving rural development and improving household income (KNBS, 2024; Tiema et al., 2024). According to the National Agricultural Report (KNBS, 2024), sugarcane ranks among the top four cash crops in terms of area under cultivation and gross value to households. Sugarcane therefore supports rural livelihoods by helping sustain schools, health facilities and local markets in Kenya's sugar growing regions (Tiema et al., 2024). However, the closure of major sugar mills has reduced livelihood opportunities and forced many farmers to find other sources of income, weakening rural communities (Ambasa & Calistus, 2025). Despite

the important of sugarcane, Kenya's sugar industry has faced a long-term crisis due to poor management of public mills, ineffective policies, reliance on imports and declining productivity (Kombo & Ndiema, 2023). Between 2024-2025, production is forecast to fall by 19.8% to 650,000 metric ton against consumption of 1.25 million ton, creating a deficit that is filled by COMESA and East African Community (EAC) imports (Gitonga & Ferrese, 2025).

The Kenyan sugar sector therefore currently characterized by declining production, rising production costs and increasing reliance on imports (Kombo & Ndiema, 2023; Gitonga & Ferrese, 2025). Sugarcane remains a vital cash crop for many smallholder farmers, however inefficiencies in miller operations and weak enforcement of zoning policies have exacerbated the problems of cane poaching, as millers encroach into other mill areas to harvest from farmers already contracted (Kombo & Ndiema, 2023). Remote sensing technologies have been identified as potential tools for improving crop monitoring, yield estimation and agricultural decision making in developing agricultural systems (Belgiu & Csillik, 2018; Immitzer et al., 2016). The need for accurate spatial data on sugarcane fields is therefore essential for policy development, cane availability forecasts and miller planning (KAM, 2021). Traditional methods of monitoring sugarcane, including field surveys and miller records, are time-consuming and difficult to keep up to date (KAM, 2021). Similar challenges are found across Sub-Saharan Africa, where limited baseline data and technical capacity make agricultural monitoring difficult (Nakalembe et al., 2021). These challenges highlight the need for modern, data-driven approaches to sugarcane monitoring (KAM, 2021). High-resolution satellite data, such as Sentinel-2, and advances in machine learning have made it possible to improve sugarcane monitoring through crop classification (Belgiu & Csillik, 2018; Immitzer et al., 2016). Sentinel-2's frequent revisit time and spectral data allow crop growth to be monitored, while Random Forest (RF) has shown strong performance in distinguishing sugarcane from other land uses (Immitzer et al., 2016; Som-ard et al., 2021).

1.2: Problem Statement

Over the last years, different studies have highlighted the potential of remote sensing for sugarcane mapping and monitoring using optical satellite imagery, vegetation indices and machine learning techniques (Mulianga et al., 2015; Wang et al., 2020; Som-ard et al., 2021). In specific, Sentinel-2 imagery and machine learning have improved crop classification through higher spatial resolution and more frequent observations (Immitzer et al., 2016; Eisfelder et al., 2024). With sugarcane production being an important pillar for Kenya's economy, previous studies have shown that remote sensing can be used to map sugarcane in Kenya (Mulianga et al., 2015). In Kenya, strategic planning documents have also highlighted gaps in the use of coordinated, data-driven approaches to improve planning, efficiency and decision-making in the sugar industry (KAM, 2021). As a result, millers continue to rely mainly on field surveys, farmer records and other conventional methods to monitor cane supply and support operational planning. Existing research in specific focusses on mapping sugarcane, monitoring crop growth and estimating crop yield (Som-ard et al., 2021; Wang et al., 2020; Zheng et al., 2022). However, only a few studies have explored methods for estimating sugarcane crop age from satellite imagery or identifying non-contracted sugarcane (NCC) for operational planning (Som-ard et al., 2021). Hence in this research, the Sentinel-2 Scene Classification Layer (SCL) is used to identify bare soil associated with planting and harvesting events, making it possible to estimate crop age in the Transmara Sugar Company Limited (TSCL) region of Transmara West Sub-County, Narok County, Kenya. This reduces the need for field surveys and farmer records and provides a more efficient and consistent approach to crop monitoring for operational planning. In addition, the proposed framework integrates Sentinel-2 spectral bands, vegetation indices (NDVI, NDMI and NDRE) and RF classification to map contracted and NCC. The identification of NCC is important as it provides information on sugarcane occurring outside the TSCL Farmer Registry, which can support cane supply assessment, farmer contracting and operational planning within the TSCL region (Kombo & Ndiema, 2023).

1.3: Aims and Objectives

The aim of this study is to develop a remote sensing framework for sugarcane classification to map contracted and NCC, as well as for crop age estimation, using Sentinel-2 Level-2A surface reflectance imagery acquired between 13 September 2023 and 30 August 2025, together with TSCL Farmer Registry data, as of 1 September 2025. The framework combines Sentinel-2 spectral bands, vegetation indices (NDVI, NDMI and NDRE) with a RF classifier to map contracted and NCC, while the SCL is used to estimate crop age. Although the framework is developed using the TSCL region in Transmara West Sub-County, Narok County, Kenya, as a case study, it is intended to support sugarcane monitoring in other smallholder production systems with similar conditions. This approach aligns with the Kenya Association of Manufacturers' Sugar Sub-Sector Strategic Plan (KAM, 2021), which identifies the lack of reliable spatial information for planning and decision-making as a key challenge in the Kenyan sugar sector.

1.3.1: Research Question

How can a remote sensing framework using Sentinel-2 L2A surface reflectance imagery be developed to classify contracted and NCC fields and estimate crop age in smallholder sugarcane production systems?

1.3.2: Research Hypothesis

The proposed research framework uses Sentinel-2 L2A time-series imagery to generate spectral bands and vegetation indices (NDVI, NDMI and NDRE) with a RF classifier. The classifier is used to identify contracted and NCC fields. Sentinel-2's SCL is used to identify bare soil linked with planting and harvesting events to support crop age estimation.

The research hypothesis of this study is:

The proposed remote sensing framework can accurately classify contracted and NCC fields and estimate sugarcane crop age within 3-months of the crop age recorded in the TSCL Farmer Registry for the study period from September 2023 to August 2025.

1.3.3: Research Objectives

The study seeks to develop a remote sensing framework for mapping sugarcane, estimating crop age and identifying NCC within the TSCL supply area. To achieve this aim, the study is guided by the following specific objectives:

- To generate multi-temporal Sentinel-2 spectral bands and vegetation index (NDVI, NDRE and NDMI) time series to support sugarcane classification, and to apply a RF classifier to distinguish sugarcane from other land-cover classes within the TSCL supply area,
- To identify bare-soil events from the SCL time-series catalogue from 13 September 2023 to 30 August 2025 to estimate sugarcane crop age within the TSCL,
- To assess the accuracy of the sugarcane classification using Overall Accuracy, the Kappa Coefficient and area comparison with the TSCL Farmer Registry, and to evaluate crop age estimates as of 30 August 2025 by comparing satellite-estimated crop ages with the TSCL Farmer Registry dated 1 September 2025 using a three-month accuracy threshold within the TSCL supply area,
- To identify NCC within the TSCL supply area by comparing the classified sugarcane map as of 30 August 2025 with the TSCL Farmer Registry dated 1 September 2025.

1.3.4: Study Limitations

This study has several limitations. First, cloud cover reduced the number of cloud-free Sentinel-2 images available during the study period. To reduce this limitation, up to two Sentinel-2 images were selected each month, allowing the image with the least cloud

cover to be used for classification. In addition, cloud masking was applied in Google Earth Engine (GEE) to remove cloud and cloud-shadow pixels before classification. Although these steps reduced the impact of cloud cover, some cloud contamination may still have affected the image time series. Second, the Sentinel-2 SCL, which was used to identify bare-soil events for crop age estimation, may have misclassified cloud shadows as bare soil, introducing uncertainty into the crop age estimates (Main-Knorn et al., 2017). Third, the TSCL Farmer Registry used to validate the sugarcane classification and crop age estimates may not have fully reflected the latest field conditions. Although TSCL has procedures to update farmer records and correct duplicate or overlapping field polygons, these updates may be delayed because of administrative processes. Consequently, some records or field boundaries may not have reflected the latest field conditions when the data were shared for this study. Finally, the framework was developed and evaluated using the TSCL supply area as a case study. Although the methodology may be applicable to other sugarcane-growing regions, it should be validated using local datasets before wider application.

1.4: Literature Review

Agricultural monitoring in Sub-Saharan Africa remains challenging because it often depends on field surveys, farmer reports and limited agricultural records that are costly, time-consuming and difficult to keep up to date (Bégué et al., 2020; Nakalembe et al., 2021). Operational crop monitoring is also constrained by limited technical capacity, inconsistent data and restricted access to reliable information (Bégué et al., 2020; Nakalembe et al., 2021). Earth observation technology has therefore made it possible to monitor large agricultural areas quickly and consistently, reducing the need for field surveys (Segarra et al., 2020). The availability of free satellite imagery has further increased the use of remote sensing in precision agriculture, particularly in developing countries where conventional monitoring is limited by cost and logistical constraints (Segarra et al., 2020; Nakalembe et al., 2021). In Kenya, remote sensing has been applied to agricultural monitoring, crop mapping and land-use analysis (Mulianga et al., 2015). Mulianga et al. (2015) demonstrated the potential of satellite-derived vegetation indices for monitoring sugarcane production, achieving a classification accuracy of 83%. Sentinel-2 imagery has also improved agricultural monitoring by providing higher spatial

and temporal resolution, making it well suited to smallholder farming systems (Segarra et al., 2020). This is particularly important in Kenya, where smallholder farming and fragmented agricultural landscapes make conventional field-based monitoring challenging (Mulianga et al., 2015; Nakalembe et al., 2021). The integration of multi-temporal Sentinel-2 imagery, vegetation indices and machine learning has improved crop mapping by capturing seasonal crop development and distinguishing between different crop types (Immitzer et al., 2016; Vuolo et al., 2018). These developments highlight the potential of remote sensing to support more efficient sugarcane monitoring and operational decision-making within the TSCL regions.

Sentinel-2 is one of the most widely used satellite missions for agricultural monitoring. Its 10 m and 20 m spatial resolutions, five-day revisit time and red-edge bands make it effective for monitoring crop condition, growth and classification (Drusch et al., 2012; Immitzer et al., 2016). The combination of 10 m and 20 m spectral bands provides a balance between spatial detail and spectral information, making Sentinel-2 well suited to agricultural monitoring across a range of agricultural landscapes (Phiri et al., 2020). The combination of Sentinel-2 imagery and cloud-based platforms, such as Google Earth Engine (GEE), has improved agricultural monitoring by making it easier to process large amounts of satellite data at local, regional and continental scales (Gorelick et al., 2017). Using Sentinel-2 images collected over time improves crop classification by capturing changes in crop growth throughout the growing season, making it easier to distinguish between crops with similar spectral characteristics than single-date imagery (Immitzer et al., 2016; Vuolo et al., 2018). This has been highlighted in several studies. Vuolo et al. (2018) found that multi-temporal Sentinel-2 imagery combined with a RF classifier improved crop classification, making it a useful approach for agricultural monitoring. Wang et al. (2019) indicated that multi-temporal Sentinel-2 imagery and machine learning could accurately distinguish sugarcane from other crops in complex agricultural landscapes. Eisfelder et al. (2024) successfully used Sentinel-1 and Sentinel-2 time-series imagery in Google Earth Engine to map cropland and different crop types in Ethiopia. Together, these studies highlight that multi-temporal Sentinel-2 imagery provides a reliable approach for crop classification and large-scale agricultural monitoring. The spectral bands of Sentinel-2 can also be used to derive vegetation indices, which provide

additional information on vegetation condition, crop health and seasonal development (Xue & Su, 2017; Segarra et al., 2020).

Vegetation indices generated from Sentinel-2 imagery further improve crop classification by providing additional information on vegetation condition and crop health (Xue & Su, 2017; Segarra et al., 2020). The Normalized Difference Vegetation Index (NDVI) is widely used to monitor vegetation greenness and biomass (Xue & Su, 2017), while the Normalized Difference Moisture Index (NDMI) measures vegetation moisture (Gao, 1996). The Normalized Difference Red Edge Index (NDRE) uses the Sentinel-2 red-edge and near-infrared bands to estimate chlorophyll content and detect subtle changes in crop condition that may not be captured by NDVI alone (Delegido et al., 2011; Clevers & Gitelson, 2013). The Sentinel-2 red-edge bands provide valuable information on chlorophyll content and canopy structure, improving the classification of crop types and supporting more accurate crop classification (Delegido et al., 2011; Immitzer et al., 2016).

Sugarcane phenology describes the seasonal progression of crop growth and development from planting to harvest (Lisson et al., 2005; Som-ard et al., 2021). In remote sensing, vegetation indices derived from multi-temporal satellite imagery can be used to monitor these seasonal changes, identify crop growth stages and distinguish between different crop types (Belgiu & Csillik, 2018; Vuolo et al., 2018). Sugarcane is well suited to phenological analysis because of its long growing season, that produces distinct seasonal growth patterns that can be captured using multi-temporal Sentinel-2 imagery (Wang et al., 2019; Som-ard et al., 2021). Studies by Wang et al. (2019) and Som-ard et al. (2021) showed that by combining multi-temporal Sentinel-2 imagery, vegetation indices and machine learning, improves sugarcane classification by capturing crop phenology and distinguishing sugarcane from other vegetation with similar spectral characteristics. These studies highlight that integrating temporal information with vegetation indices provides a reliable approach for monitoring sugarcane growth, identifying crop growth stages and improving crop classification (Wang et al., 2019; Som-ard et al., 2021).

Land cover classification methods in remote sensing are generally divided into pixel-based and object-based approaches (Blaschke, 2010). Pixel-based classification assigns each image pixel to a land cover class based on its spectral and temporal characteristics,

while object-based classification groups neighbouring pixels into image objects before classification (Blaschke, 2010; Belgiu & Drăguț, 2016). Pixel-based classification is widely used with medium-resolution satellite imagery such as Sentinel-2 because multi-temporal data capture seasonal changes in crop growth, improving crop classification (Immitzer et al., 2016; Vuolo et al., 2018). Although object-based methods can improve classification when high-resolution imagery and clearly defined field boundaries are available, pixel-based machine learning approaches continue to provide accurate and reliable results with Sentinel-2 imagery (Belgiu & Drăguț, 2016; Immitzer et al., 2016; Belgiu & Csillik, 2018). For these reasons, this study used a pixel-based RF classifier to identify sugarcane fields.

Machine learning algorithms are widely used in remote sensing because they learn relationships between spectral, temporal and environmental variables from labelled training data, allowing them to classify land-cover types without making assumptions about the data (Belgiu & Drăguț, 2016). Unlike traditional classification methods that rely on manually defined decision rules, machine learning algorithms identify complex patterns within the data and use these patterns to classify unknown pixels (Belgiu & Drăguț, 2016). RF is one of the most widely used supervised machine learning algorithms for agricultural classification because it can process large datasets, handle many input variables and is less prone to overfitting than a single decision tree (Breiman, 2001; Belgiu & Drăguț, 2016). The algorithm constructs multiple decision trees using random subsets of the training samples and predictor variables (Breiman, 2001). At each decision point, only a random selection of variables is evaluated to determine the best split, increasing diversity among the trees and reducing the influence of individual variables (Breiman, 2001; Belgiu & Drăguț, 2016). Each tree independently predicts the land-cover class of a pixel, and the final classification is determined by majority voting across all trees, resulting in a robust and reliable classification (Breiman, 2001). These characteristics make RF particularly suitable for analyzing multi-temporal Sentinel-2 imagery containing numerous spectral bands, vegetation indices and temporal features (Belgiu & Drăguț, 2016; Vuolo et al., 2018). Studies have demonstrated that RF achieves high classification accuracy for agricultural applications using Sentinel-2 imagery (Immitzer et al., 2016; Vuolo et al., 2018). For these reasons, RF was selected in this

study to classify sugarcane and other land-cover classes using multi-temporal Sentinel-2 spectral bands and vegetation indices.

Most remote sensing approaches to crop age estimation use multi-temporal vegetation indices, such as NDVI, to monitor crop growth and estimate planting dates by tracking changes in vegetation throughout the growing season (Atzberger & Eilers, 2011; Wang et al., 2020). These approaches are effective for monitoring crop development and phenology but do not directly identify potential planting or harvesting events (Atzberger & Eilers, 2011; Wang et al., 2020). Sugarcane crop age estimation is particularly challenging because planting and harvesting dates vary according to management practices, ratooning cycles and environmental conditions (Inman-Bamber & Smith, 2005). Under normal growing conditions, sugarcane typically reaches maturity after approximately 12–18 months, depending on the variety, climate and management practices (Inman-Bamber & Smith, 2005; Som-ard et al., 2021). An alternative approach is to use the Sentinel-2 SCL, generated by the Sen2Cor atmospheric correction processor, which classifies each pixel as vegetation, bare soil, water, clouds, cloud shadows and other land-cover or atmospheric classes (Main-Knorn et al., 2017). Although SCL is primarily used for cloud masking and image quality improvement, it also provides useful information on land surface conditions, including the occurrence of bare soil (Main-Knorn et al., 2017). In this study, historical bare-soil events derived from the Sentinel-2 SCL are investigated as indicators of planting or harvesting events for estimating sugarcane crop age. This study investigates the potential of SCL observations to estimate sugarcane crop age by identifying the earliest bare-soil event associated with planting or harvesting.

Accuracy assessment is an important component of remote sensing because it measures how well a classification represents the actual land-cover classes (Congalton & Green, 2019). Classification accuracy is commonly evaluated using a confusion matrix, from which Overall Accuracy (OA), Producer's Accuracy (PA), User's Accuracy (UA) and the Kappa coefficient are calculated (Foody, 2002). Overall Accuracy measures the proportion of correctly classified samples, while Producer's Accuracy and User's Accuracy quantify omission and commission errors for each land-cover class (Foody, 2002; Congalton & Green, 2019). The Kappa coefficient provides an additional measure

of agreement by accounting for agreement that could occur by chance (Cohen, 1960; Foody, 2002). These metrics are widely used to evaluate the performance and reliability of supervised classification models (Stehman, 1997; Foody, 2002). In operational agricultural applications, classification results are commonly validated using independent reference data to assess their suitability for decision-making (Stehman, 1997; Congalton & Green, 2019). In this study, classification accuracy was assessed using a confusion matrix generated from independent validation samples. The classified sugarcane area was also compared with the TSCL Farmer Registry to assess the agreement between the satellite generated classification and operational records. In addition, the estimated crop ages were compared with the crop ages recorded in the TSCL Farmer Registry using a 3-month agreement threshold to evaluate the performance of the proposed SCL-based crop age estimation approach.

1.5: Conceptual Framework

The conceptual framework (Figure 1.2) provides an overview of the proposed methodology and illustrates how the research objectives are addressed through the integration of Sentinel-2 imagery, the TSCL Farmer Registry and the analytical processes used in this study.

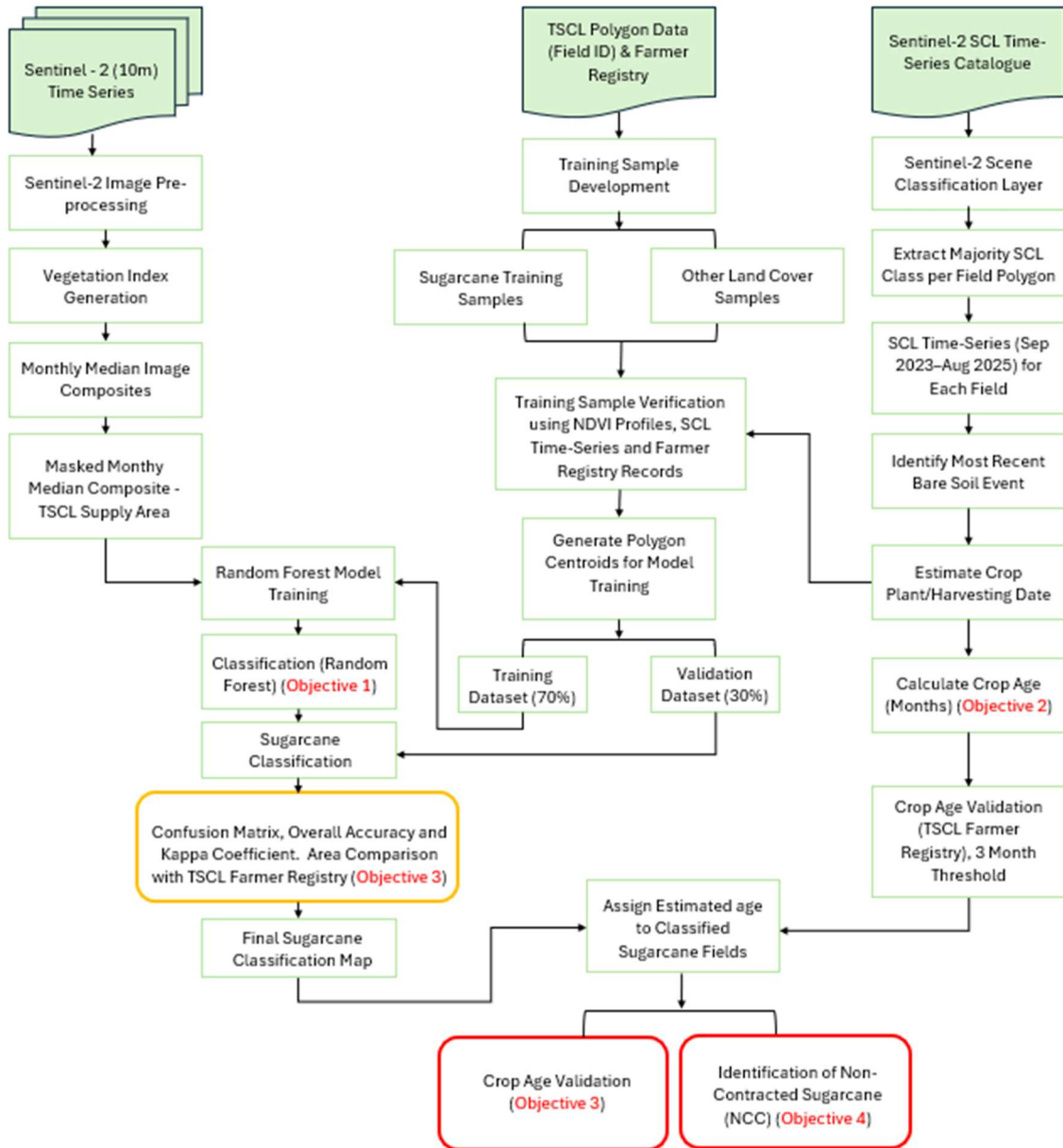


Figure 1.2 - Conceptual framework illustrating the relationship between the research objectives, methodological workflow and study outputs for sugarcane classification and crop age estimation

Figure 1.2 summarizes the methodological framework developed for sugarcane classification, crop age estimation and identification of NCC within the TSCL supply area. The framework consists of two interlinked components that address the four research objectives. Objective 1 involves generating multi-temporal Sentinel-2 spectral bands and the NDVI, NDMI and NDRE, which are used with a RF classifier to distinguish sugarcane from other land-cover classes. Training and validation samples are developed from TSCL polygon data (Field ID) and Farmer Registry records and verified using NDVI profiles, the Sentinel-2 SCL time-series catalogue and Farmer Registry records before polygon centroids are generated for model training. The RF model is trained using 70% of the samples and validated using the remaining 30%. Objective 2 uses the Sentinel-2 SCL time-series catalogue from 13 September 2023 to 30 August 2025 to identify the most recent bare-soil event within each TSCL field polygon. The identified bare-soil events are interpreted as planting or harvesting events, and crop age is estimated by calculating the time difference between the identified event and the Sentinel-2 reference date of 30 August 2025. Objective 3 assesses the accuracy of the sugarcane classification using Overall Accuracy, the Kappa Coefficient and area comparison with the TSCL Farmer Registry, while the satellite-derived crop age estimates are validated against the TSCL Farmer Registry dated 1 September 2025 using a three-month accuracy threshold. The validated crop age estimates are subsequently assigned to the classified sugarcane fields. Finally, Objective 4 identifies NCC by comparing the final sugarcane classification map as of 30 August 2025 with the TSCL Farmer Registry dated 1 September 2025. Thus, the framework directly links the Sentinel-2 data processing, RF classification, crop age estimation, accuracy assessment and NCC identification procedures to the four research objectives.

2: Materials and Methods

2.1 Study Area

The study area is based in Transmara West Sub-County, Narok County, Kenya which is located approximately 220km west of Nairobi and within the western Rift Valley region (TSCL, 2025). The focus area is the region of TSCL, their operational area covers approximately 93,408ha and is made up of six regions; Enosaen, Oloontare, Shankoe, Moita/Nkarar, Osinoni and Osupuko (see Figure 2.1) (TSCL, 2025). Situated around 1°03'S and 34°47'E, with elevations ranging from approximately 1,440 m to 2,000 m above sea level (Snyman, 2016). Differences in elevation influence the local climate, soil development, and agricultural productivity (Snyman, 2016).

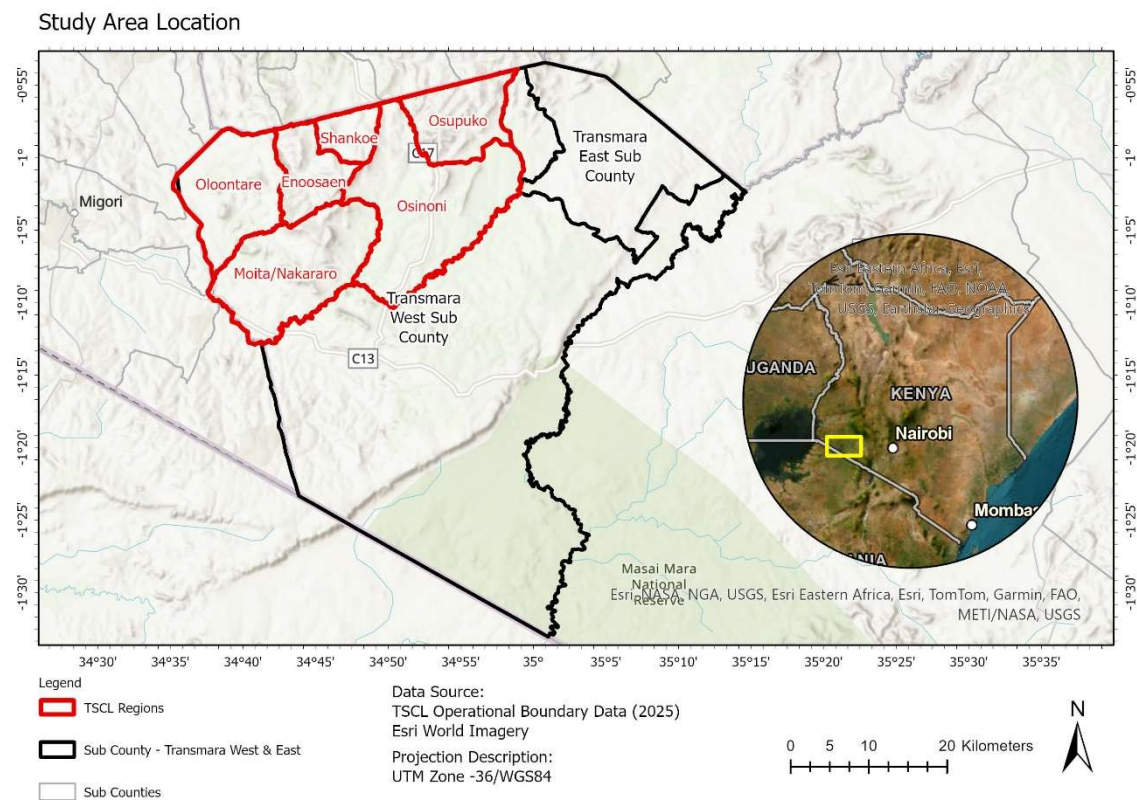


Figure 2.1 - Study area showing the regions of Transmara Sugar Company Limited (TSCL) in Transmara West Sub-County, Narok County, Kenya. Source: Created by the author using TSCL regional boundary data (2025) and Esri World Imagery basemap.

Agriculture is the dominant economic activity within the study area and Narok County is recognized as one of Kenya's major agricultural regions, supporting both crop and livestock production due to favorable climatic conditions and relatively fertile soils (KNBS, 2023). Unlike many sugar producing regions where large estates contribute significantly to cane supply, sugarcane production within the TSCL regions is entirely dependent on smallholder out-grower farmers (Snyman, 2016; TSCL, 2025). The upland areas are largely used for rainfed sugarcane farming, with the cane supplied to the TSCL sugar mill (Snyman, 2016; TSCL, 2025). Sugarcane is cultivated continuously throughout the year due to the climatic conditions and planting practices adopted by farmers (TSCL, 2025). These conditions create sugarcane fields of different ages and growth stages across the landscape, making the area well suited for remote sensing studies of crop classification and age estimation.

Maize is the second most important crop in the study area and is generally grown during two rainy seasons each year (TSCL, 2025). Planting and harvesting typically occurs during January to February and June to July, corresponding with seasonal rainfall patterns (TSCL, 2025). In addition to sugarcane and maize, smaller areas are occupied by tea plantations, eucalyptus woodlots and subsistence crops such as vegetables and cassava (TSCL, 2025). These land-cover types are difficult to classify because they have similar spectral responses during some stages of crop growth (Immitzer et al., 2016; Vuolo et al., 2018).

The climate of the study area is classified as temperate with a warm summer according to the Koppen-Geiger climate classification system (Snyman, 2016). Despite being located close to the equator, the relatively high elevation results in moderate temperatures throughout the year (Snyman, 2016). The climate is strongly influenced by the Inter-Tropical Convergence Zone (ITCZ), which contributes to the bimodal rainfall pattern characteristic of much of East Africa (Snyman, 2016). TSCL rainfall records for the five-year period from 2020/21 to 2024/25 indicate a mean annual rainfall of approximately 1,889 mm, while monthly minimum and maximum temperatures range from approximately 13–15°C and 26–28°C, respectively (TSCL, 2025).

According to recent data from the TSCL Ennoosaen Automatic Weather Station, the average annual rainfall between the 2020/21 and 2024/25 seasons was approximately 1,890 mm (TSCL, 2025). Rainfall is distributed throughout the year, although the highest levels occur during March to May and October to November. Average monthly rainfall during March, April and May was approximately 170mm, 243mm and 197mm, while drier conditions were observed during July and August (see Figure 2.2). Other than these high months, the area receives enough rainfall throughout the year (Snyman, 2016).

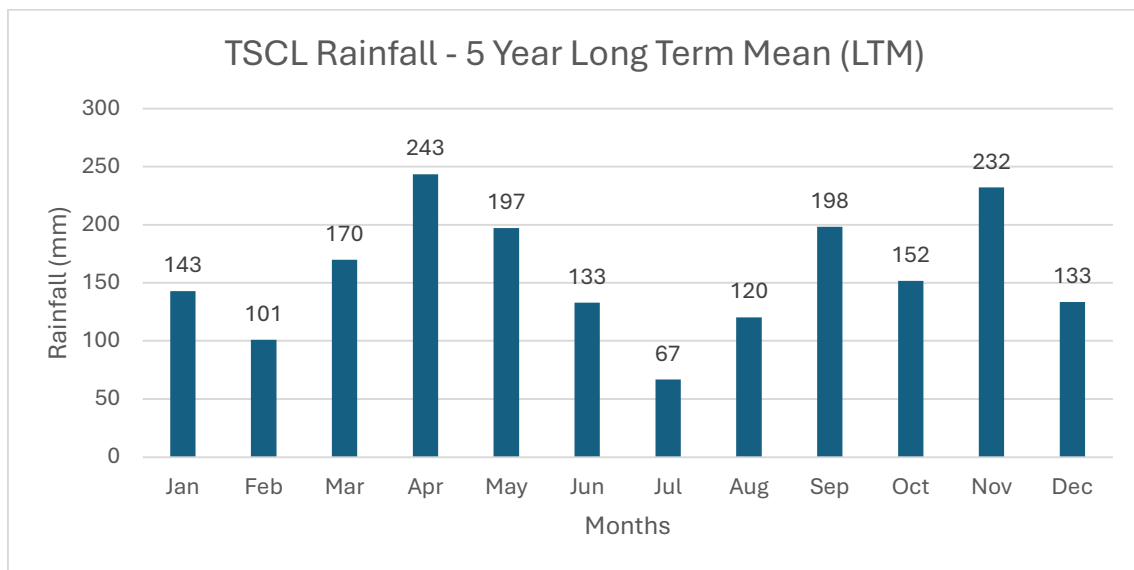


Figure 2.2 – Five-year mean monthly rainfall for the TSCL study area (2020/21–2024/25). Source: Created by the author using TSCL rainfall records (2025).

Temperatures remain stable throughout the year, with mean minimum temperatures ranging from 12.9°C to 14.8°C, mean maximum temperatures from 25.8°C to 28.0°C, and an average annual temperature of about 20.3°C (TSCL, 2025). Based on the data shared, the warmest months are generally around the September and April periods, while cooler months occurred during July and December periods (see Figure 2.3).

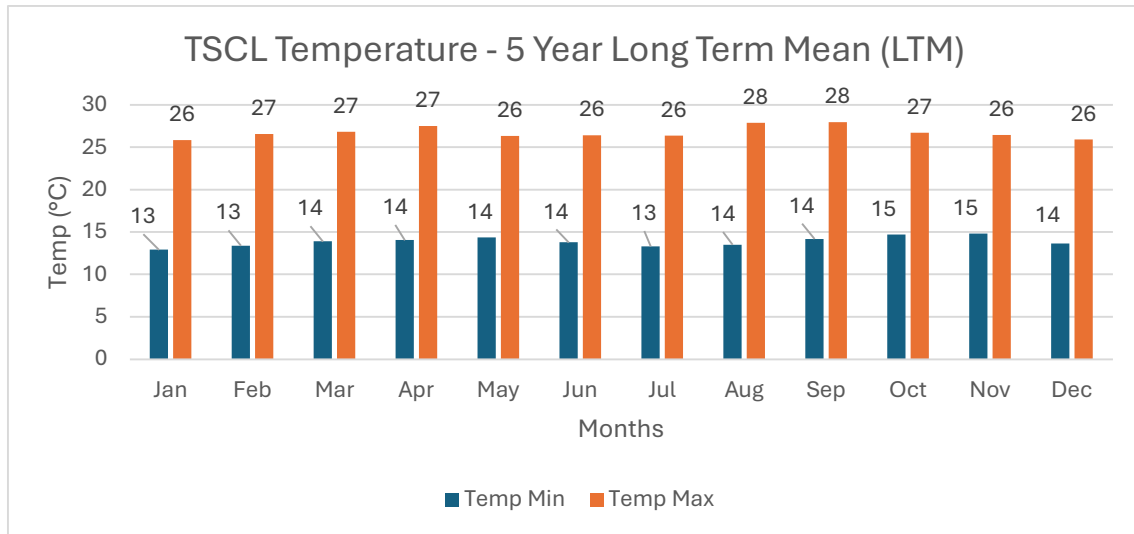


Figure 2.3 - Five-year mean monthly minimum and maximum temperatures for the TSCL study area (2020/21–2024/25). Source: Created the author using TSCL temperature data (2025).

Topographically, the study area can be divided into two major landscape units (see Figure 2.4). The northern and central portions consist of rolling uplands with elevations ranging between 1,700 m and 2,000 m above sea level, while lower-lying bottomland zone found in the west, south and east (Snyman, 2016). The upland regions generally receive higher rainfall and contain the most productive agricultural land, whereas the bottomland areas are characterized by poorer drainage and lower agricultural suitability (Snyman, 2016).

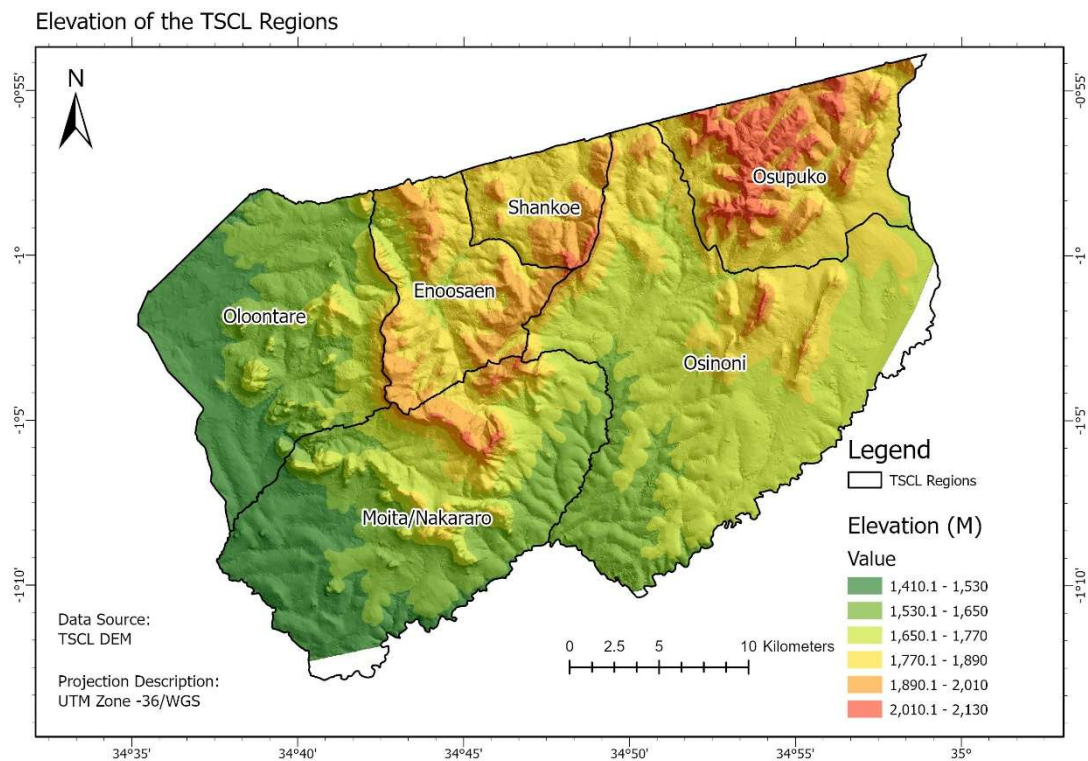


Figure 2.4 - Elevation of the TSCL operational regions. Source: Created by the author using the TSCL Digital Elevation Model (DEM) dataset (2025).

The soils of the study area are dominated by Humic Nitisols, which are deep, well-drained, highly weathered tropical soils characterized by red coloration, strong structural development and favorable water-holding capacity (Snyman, 2016; FAO, 2015). Nitisols are among the most productive agricultural soils in East Africa due to their deep profiles, stable structure and good internal drainage (FAO, 2015). These soils are considered highly suitable for rainfed sugarcane production due to their depth, rooting potential and moisture retention characteristics (Snyman, 2016). Humic Nitisols occupy most of the productive sugarcane-growing areas within the TSCL regions (Snyman, 2016). Humic Nitisols correspond predominantly with areas classified as S1 (Highly Suitable) and S2 (Moderately Suitable) in the TSCL soil suitability assessment (see Figure 2.5), reflecting their favorable physical and hydrological characteristics for sustained rainfed sugarcane production (Snyman, 2016). Less suitable soils, including Gleysols, Plinthosols and Planosols, occur predominantly within the lower-lying bottomland areas where drainage limitations restrict agricultural potential (Snyman, 2016).

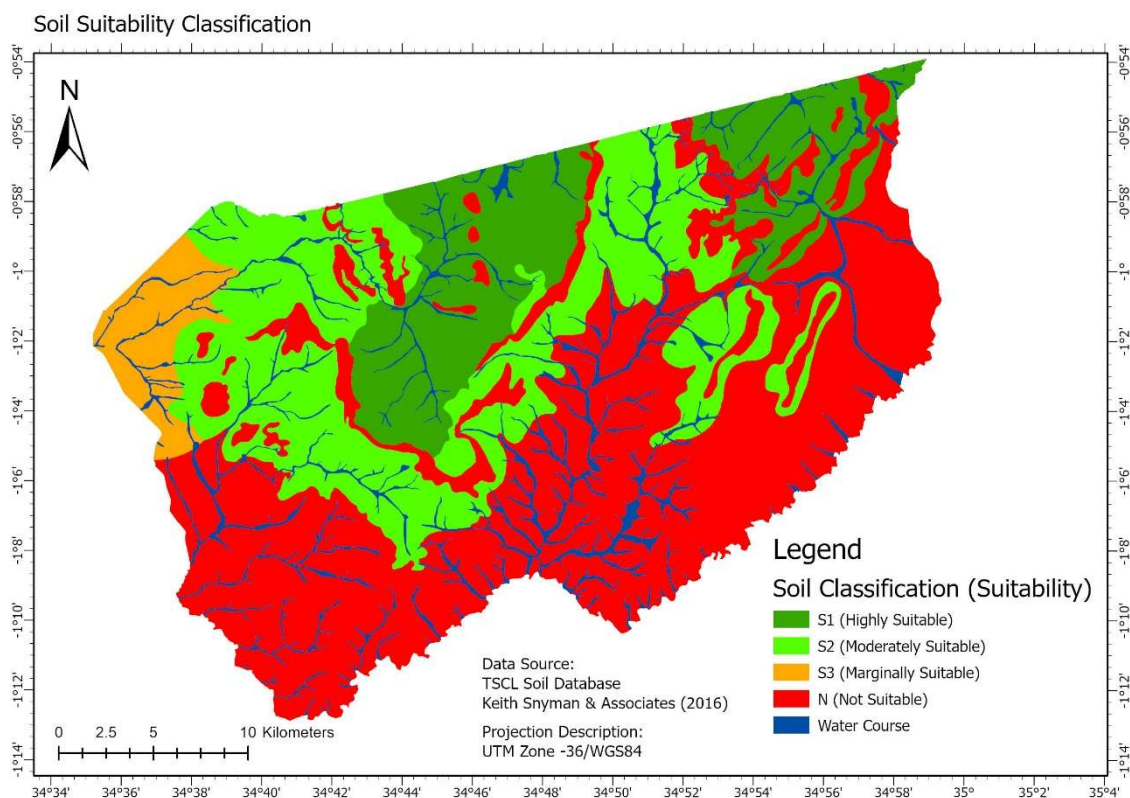


Figure 2.5 - Soil suitability classification of the TSCL operational area. Source: Created by the author using the TSCL soil suitability dataset, adapted from Snyman (2016).

TSCL records show that sugarcane plays an important role in the local agricultural economy. During Week 53 of the 2024/2025 crushing season, TSCL reported a cane tonnage of 767,369 tonnes harvested from 7,788 ha of contracted cane. The average cane yield was 98.5 TCH (tonnes cut per hectare) with a TCHM (tonnes cane per hectare per month) value of 5.3 (TSCL, 2025). The average harvest age of contracted sugarcane was reported as 18.7 months against a target harvest age of 18.0 months. These operational statistics indicate that sugarcane within the study area is generally harvested between 18 and 19 months after planting (TSCL, 2025).

The TSCL regions form part of the Southern Sugar Catchment Area established under the Sugar Act No. 11 of 2024. The Act formally divides Kenya's sugar industry into five geographical sugar catchment areas for cane management and governance, namely the Central, Upper Western, Lower Western, Southern and Coastal Catchments (Government of Kenya, 2024). The Southern Catchment comprises of Narok, Migori, Homa Bay and Kisii Counties and contains five sugar millers operating within proximity to one another

(see Figure 2.6) (Government of Kenya, 2024). Within this catchment, TSCL operates alongside neighboring millers including Sukari Industries Limited, South Nyanza Sugar Company (SONY), Soit Sugar Company and Angata Sugar Mills (TSCL, 2025) (see Figure 2.6).

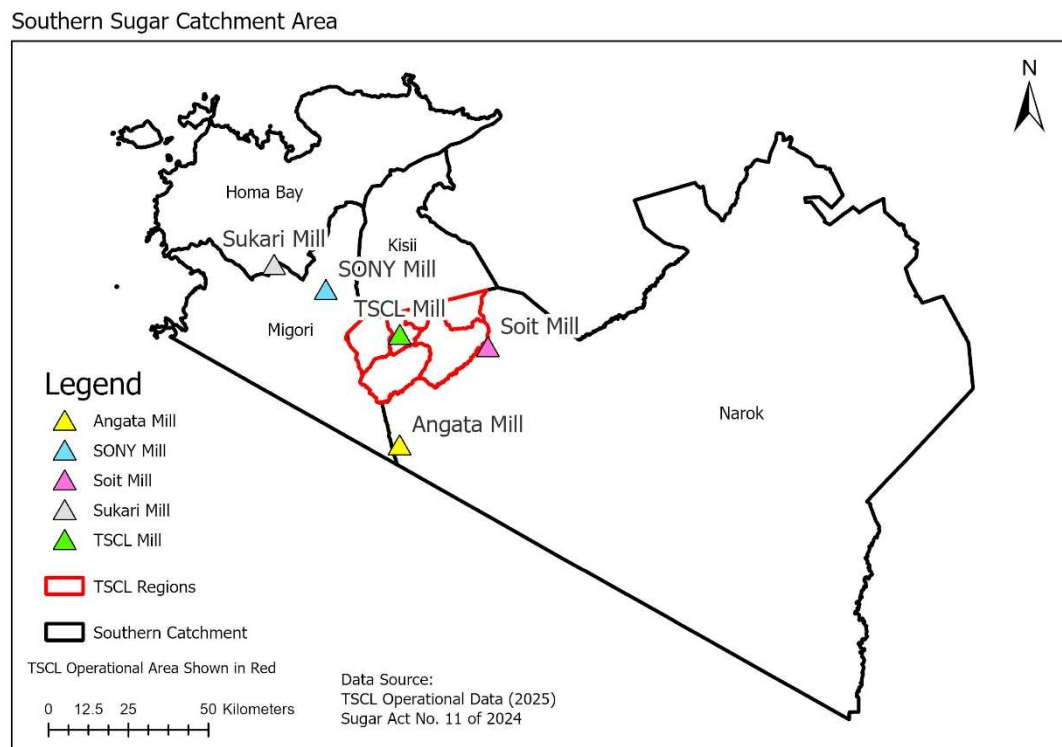


Figure 2.6 - Southern sugar catchment area showing the TSCL operational region and neighboring sugar mill catchments. Source: Created by author using TSCL operational boundary data (2025) and the sugar catchment boundaries defined under the Sugar Act No. 11 of 2024 (Government of Kenya, 2024).

The concentration of multiple millers within the same catchment creates significant competition for available sugarcane resources (TSCL, 2025). Although contractual agreements exist between farmers and millers, competition for mature cane can result in miller encroachment, whereby sugarcane contracted to one miller is harvested by another (TSCL, 2025). The requirement of accurate and up-to-date information regarding the spatial extent, distribution and maturity of sugarcane has become increasingly important for strategic planning and operational decision-making (KAM, 2021; TSCL, 2025).

2.2 Data Sources

The study uses a combination of satellite imagery, spatial datasets, and operational records to classify sugarcane and estimate crop age in the TSCL region. The primary remote sensing dataset consisted of Sentinel-2 L2A imagery sources through the European Space Agency's (ESA) Copernicus Programme. Sentinel-2 is a multispectral Earth observation mission that provides imagery in 13 spectral bands with spatial resolutions of 10m, 20m and 60m (Drusch et al., 2012; ESA, 2024). With a revisit frequency of approximately five days, provides sufficient temporal and spatial resolution for monitoring agricultural activities and tracking crop development throughout the growing period (Drusch et al., 2012). Sentinel bands used in this study included red, red-edge, near-infrared (NIR) and shortwave infrared (SWIR) spectral bands, including Band 4 (B4), Band 5 (B5), Band 8 (B8), Band 8A (B8A), Band 11 (B11) and Band 12 (B12) (ESA, 2024) (see Table 2.1). These bands were selected because they capture vegetation and moisture characteristics that are important for distinguishing sugarcane from surrounding land-cover classes within the TSCL study area (Gao, 1996; Delegido et al., 2011). They also provide the spectral information required to calculate the vegetation indices used as predictor variables in the RF classification model (Tucker, 1979; Gao, 1996; Delegido et al., 2011).

Table 2.1 - Sentinel-2 Spectral Bands Used in this Study.

Band	Description	Central Wavelength (nm)	Resolution (m)
B4	Red	665	10
B5	Red Edge	705	20
B8	Near Infrared (NIR)	842	10
B8A	Narrow NIR	865	20
B11	SWIR 1	1610	20
B12	SWIR 2	2190	20

Source: Adapted from ESA (2024) Sentinel-2 User Handbook and Drusch et al. (2012).

A Sentinel-2 SCL time series was developed to estimate crop age. The SCL product is generated through the Sen2Cor atmospheric correction processor and classifies each pixel into categories such as vegetation, bare soil, water, cloud shadow, clouds and cirrus (Main-Knorn et al., 2017). The SCL layer has a spatial resolution of 20 m and forms part

of the standard Sentinel-2 L2A product (ESA, 2024). The time-series covered the period from September 2023 to August 2025 and used imagery from Sentinel-2 tiles 36MXD and 36MYD (see Figure 2.7), which covered approximately 99.8% of the TSCL regional polygons, representing 56,773 of the 56,837 digitized field polygons. The vegetation (Class 4) and bare soil (Class 5) classes (ESA, 2024) were used to identify changes over time linked to crop planting and harvesting (see Table 2.2). These changes were analysed to estimate sugarcane crop age at field level and were compared against the TSCL Farmer Registry records for validation. An SCL time series (catalogue) was useful because planting and harvesting expose bare soil, creating a pattern that can be detected by satellite images and used to estimate crop age (Wang et al., 2020).

Table 2.2 - Sentinel-2 Scene Classification Layer (SCL) Classes

0	No Data
1	Saturated or Defective Pixels
2	Dark Area Pixels
3	Cloud Shadow
4	Vegetation
5	Bare Soil
6	Water
7	Clouds (Low Probability)
8	Clouds (Medium Probability)
9	Clouds (High Probability)
10	Thin Cirrus
11	Snow or Ice

Source: Adapted from European Space Agency (ESA, 2024).

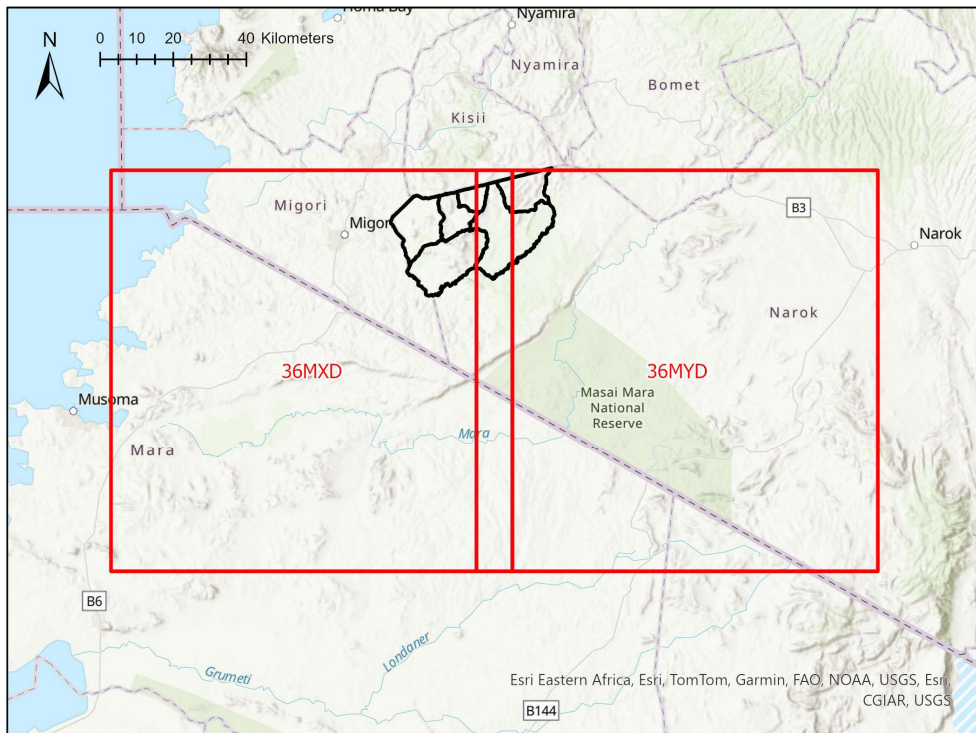


Figure 2.7 - Sentinel-2 tiles (36MXD and 36MYD) cover the TSCL study area. Source: Created by the author using the Copernicus Sentinel-2 Military Grid Reference System (MGRS) tile grid and the Esri World Topographic Basemap.

The SCL image catalogue consisted of twenty-one acquisition dates between September 2023 and August 2025. The aim was to use two images per month to reduce information loss and improve the continuity of observations. Where excessive cloud covers reduced image quality, alternative images from the same month were selected. The image dates used for crop age estimation are shown in Table 2.3.

Table 2.3 - Sentinel-2 Scene Classification Layer (SCL) Images Used for Crop Age Estimation

Period	Image Date 1	Image Date 2
Sept-23	13/09/2023	23/09/2023
Oct-23	13/10/2023	23/10/2023
Dec-23	07/12/2023	22/12/2023
Feb-24	05/02/2024	-
Mar-24	11/03/2024	16/03/2024
Apr-24	10/04/2024	-
May-24	-	30/05/2024
Jun-24	04/06/2024	24/06/2024
Jul-24	09/07/2024	19/07/2024
Aug-24	03/08/2024	28/08/2024
Sept-24	02/09/2024	22/09/2024
Oct-24	02/10/2024	12/10/2024
Dec-24	11/12/2024	26/12/2024
Jan-25	05/01/2025	20/01/2025
Feb-25	04/02/2025	24/02/2025
Mar-25	06/03/2025	21/03/2025
Apr-25	15/04/2025	25/04/2025
May-25	02/05/2025	-
Jun-25	11/06/2025	24/06/2025
Jul-25	04/07/2025	31/07/2025
Aug-25	03/08/2025	30/08/2025

Source: Prepared by the author using Sentinel-2 imagery obtained from the Copernicus Open Access Hub. *Note. Missing image dates indicate that suitable Sentinel-2 imagery was unavailable due to excessive cloud cover.*

For sugarcane classification, a separate Sentinel-2 image period was developed using imagery sourced between December 2024 and August 2025. Multiple image dates were used to capture seasonal changes in crops and improve the separation of sugarcane from

other land-cover types (Belgiu & Csillik, 2018; Vuolo et al., 2018). The imagery used for classification are shown in Table 2.4.

Table 2.4 - Sentinel-2 Image Acquisition Dates Used for Sugarcane Classification

Image Date
26/12/2024
05/01/2025
04/02/2025
06/03/2025
15/04/2025
02/05/2025
11/06/2025
31/07/2025
30/08/2025

Source: Prepared by the author using Sentinel-2 imagery obtained from the Copernicus Open Access Hub.

Spatial datasets were supplied by TSCL which were field boundary polygons containing unique Field IDs, the TSCL regional management boundaries and the locations of neighbouring sugar mills operating within the Southern Kenya Sugar Catchment. Also, spatial regional management areas comprised Enoosaen, Osinoni, Shankoe, Osupuko, Moita/Nkararo and Oloontare. Additional spatial datasets shared that included county and sub-county administrative boundaries used for map production, spatial referencing and the preparation of location maps. Operational datasets provided by TSCL included the Farmer Registry, Harvest Bulletin (Harvest Performance Data) and Cane Development Data. The Farmer Registry contained records of contracted sugarcane fields and associated farmer information, while the Harvest Bulletin provided historical harvesting records. The Cane Development dataset contained information relating to seed cane supply. All operational datasets were linked to the spatial field boundaries through a unique Field ID, enabling direct comparison between satellite derived observations and company production records. Summary of data sources used in the study are shown in Table 2.5.

Table 2.5 - Summary of Data Sources Used in this Study

Data Source	Description	Purpose
Sentinel-2 Level-2A Imagery	Multispectral imagery (10m and 20m resolution) from tiles 36MXD and 36MYD	Sugarcane classification
Sentinel-2 Scene Classification Layer (SCL)	20m resolution classification layer generated by Sen2Cor	Crop age estimation through identification of vegetation and bare soil transitions
SPOT 6 Imagery	High-resolution satellite imagery acquired in July 2025 and provided by TSCL	Visual interpretation and additional verification of training data used for classification
TSCL Field Polygons	Field boundaries containing unique Field IDs.	Spatial analysis
TSCL Milling Area Boundary	Operational boundary of TSCL	Definition of study area
TSCL Regional Boundaries	Enoosaen, Osinoni, Shankoe, Osupuko, Moita/Nkararo and Oloontare	Regional analysis and reporting
Administrative Boundaries	County and sub-county boundaries	Map production and spatial referencing
Neighbouring Miller Locations	Locations of surrounding sugar mills	Assessment of competitive operating area
Farmer Registry	Database of contracted sugarcane fields	Validation of classification and crop age estimation
Harvest Bulletin	Historical harvest performance records	Verification of harvesting events
Cane Development Data	Records of seed cane supply and crop establishment	Verification of planting events

Source: Compiled by the author from TSCL operational datasets and Copernicus Sentinel-2 imagery.

Data processing and analysis were conducted using Google Earth Engine (GEE) (Gorelick et al., 2017), ArcGIS Pro (ESRI, 2024) and Microsoft Excel. Google Earth Engine was used for Sentinel-2 image acquisition, preprocessing and classification. ArcGIS Pro was used for spatial data management, raster processing, field-level analysis and cartographic production. Microsoft Excel was used for data preparation, image catalogue management, random sample generation, accuracy assessment and comparison of estimated crop ages against Farmer Registry records. The software used in this study are shown in Table 2.6.

Table 2.6 - Software Used in this Study

Software	Purpose
Google Earth Engine	Satellite image acquisition, preprocessing and classification
ArcGIS Pro	Spatial analysis, raster processing and map production (age estimation)
Microsoft Excel	Data preparation, integration and statistical analysis (age estimation)

Source: Author.

2.3 Data Pre-processing

The following sections describe the main steps used to develop the remote sensing framework shown in Figure 1.2. These include Sentinel-2 image pre-processing, vegetation index generation, training sample development, RF classification, classification validation, SCL time-series analysis and crop age estimation. Where applicable, the results from these steps are presented in Chapter 3.

This study utilised Sentinel-2 L2A imagery obtained through Google Earth Engine (GEE) and Sentinel Hub. Sentinel-2 L2A products are generated from Level-1C Top of Atmosphere (TOA) reflectance data through atmospheric correction using the Sen2Cor processor developed by the ESA. The Sen2Cor algorithm corrects atmospheric effects including aerosol scattering, water vapour absorption and terrain influences to produce Bottom of Atmosphere (BOA) surface reflectance data suitable for quantitative remote

sensing applications (Main-Knorn et al., 2017). The Sentinel-2 imagery was already available as L2A surface reflectance data, so no additional atmospheric correction was needed (Main-Knorn et al., 2017; ESA, 2024). Atmospherically corrected imagery improves consistency between image dates and makes temporal analysis for sugarcane classification and crop age estimation more reliable (Main-Knorn et al., 2017; Drusch et al., 2012).

Pre-processing and image analysis were undertaken within Google Earth Engine (GEE), a cloud-based geospatial processing platform that provides access to large Earth observation data archives and enables efficient processing of multi-temporal satellite imagery (Gorelick et al., 2017). Sentinel-2 images were selected based on their temporal coverage of the sugarcane growing cycle and the availability of cloud-free observations within the study area (Belgiu & Csillik, 2018; Vuolo et al., 2018). Images with little cloud cover were selected to improve the quality of the data (see Table 2.7). Clouds can affect spectral reflectance values and vegetation indices (Main-Knorn et al., 2017). To reduce these effects, cloud and cloud-shadow masking was carried out using the Sentinel-2 SCL, which is produced as part of the Sen2Cor processing process (Main-Knorn et al., 2017). The SCL product was selected for crop age estimation because it provides a consistent classification of land surface conditions, including bare soil and vegetation classes (ESA, 2024). Changes between these classes can be used to identify field establishment, planting, and harvesting, making it possible to monitor crop development over time (Wang et al., 2020).

The study used nine Sentinel-2 L2A images, sourced between December 2024 and August 2025, for sugarcane classification through GEE (see Table 2.7). Images were selected to cover different stages of the sugarcane growing cycle and to ensure they had little cloud cover over the study area. As shown in Figure 2.8, Cloud cover over the field boundaries ranged from 1% to 11%, with eight of the nine selected images having about 1% cloud cover.

Table 2.7 - Percentage of Cloud-Covered TSCL Fields in the Selected Sentinel-2 Images.

Date	Cloud Percentage to Fields
26/12/2024	11%
05/01/2025	1%
04/02/2025	1%
06/03/2025	1%
15/04/2025	1%
02/05/2025	1%
11/06/2025	1%
31/07/2025	1%
30/08/2025	1%

Source: Created by the author using Sentinel-2 imagery processed in ArcGIS Pro and Microsoft Excel.

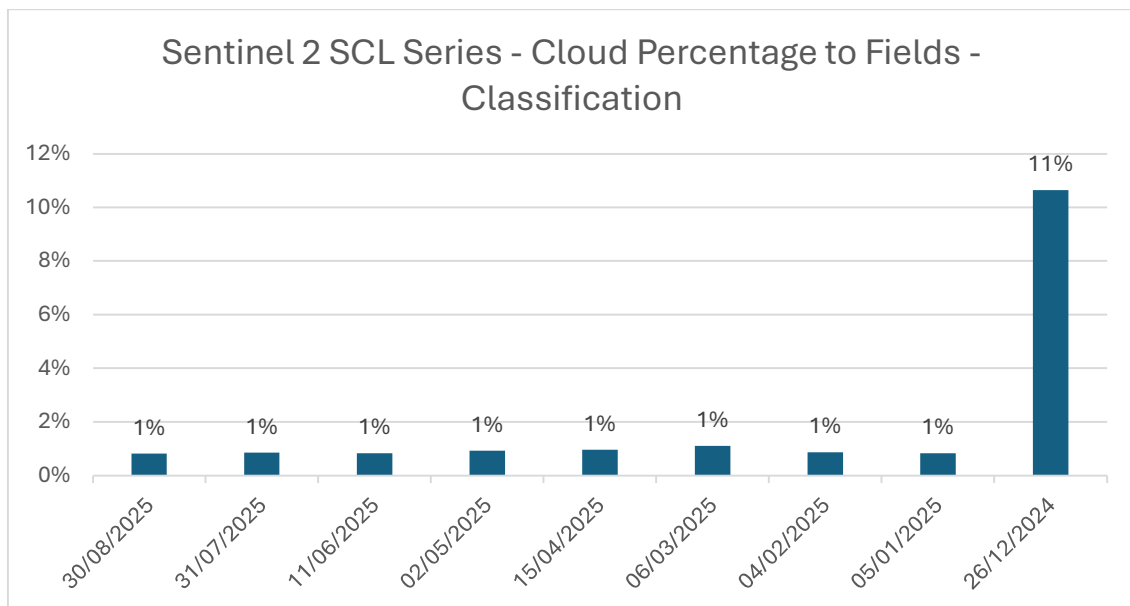


Figure 2.8 - Percentage of TSCL field polygons affected by cloud cover in the selected Sentinel-2 images used for sugarcane classification. Source: Created by the author using Sentinel-2 imagery processed in ArcGIS Pro, and Microsoft Excel.

Following cloud masking, the selected Sentinel-2 spectral bands were combined into a multi-temporal dataset for analysis. The imagery was processed on a common 10m analysis grid in GEE, allowing the spectral bands and derived vegetation indices to be

used together within a single classification model. The combined images showed changes in crop growth, planting, harvesting, and seasonal conditions across the study area. Using images from different dates improves crop classification by capturing changes in crop growth over time (Immitzer et al., 2016; Vuolo et al., 2018).

To estimate crop age, a Sentinel-2 SCL time series catalogue covering September 2023 to August 2025 was obtained from Sentinel Hub (see Table 2.8). The time-series included images from different dates throughout the study period to provide continuous coverage while keeping cloud cover low. Most images had less than 5% cloud cover within the field boundaries, providing reliable data for temporal monitoring (see Figure 2.9).

Table 2.8 - Percentage of Cloud-Covered TSCL Field Polygons in the Selected Sentinel-2 Images.

Date	Cloud Percentage to Fields	Date	Cloud Percentage to Fields
30/08/2025	1%	22/09/2024	1%
03/08/2025	1%	02/09/2024	1%
31/07/2025	1%	28/08/2024	1%
04/07/2025	1%	03/08/2024	8%
24/06/2025	1%	19/07/2024	1%
11/06/2025	1%	09/07/2024	5%
02/05/2025	1%	24/06/2024	1%
25/04/2025	2%	04/06/2024	1%
15/04/2025	1%	30/05/2024	9%
21/03/2025	2%	10/04/2024	17%
06/03/2025	1%	16/03/2024	5%
24/02/2025	1%	11/03/2024	1%
04/02/2025	1%	05/02/2024	16%
20/01/2025	10%	22/12/2023	1%
05/01/2025	1%	07/12/2023	1%
26/12/2024	11%	23/10/2023	1%
11/12/2024	17%	13/10/2023	1%
12/10/2024	1%	23/09/2023	1%
02/10/2024	1%	13/09/2023	1%

Source: Created by the author using Sentinel-2 imagery processed in Google Earth Engine (GEE), ArcGIS Pro, and Microsoft Excel.

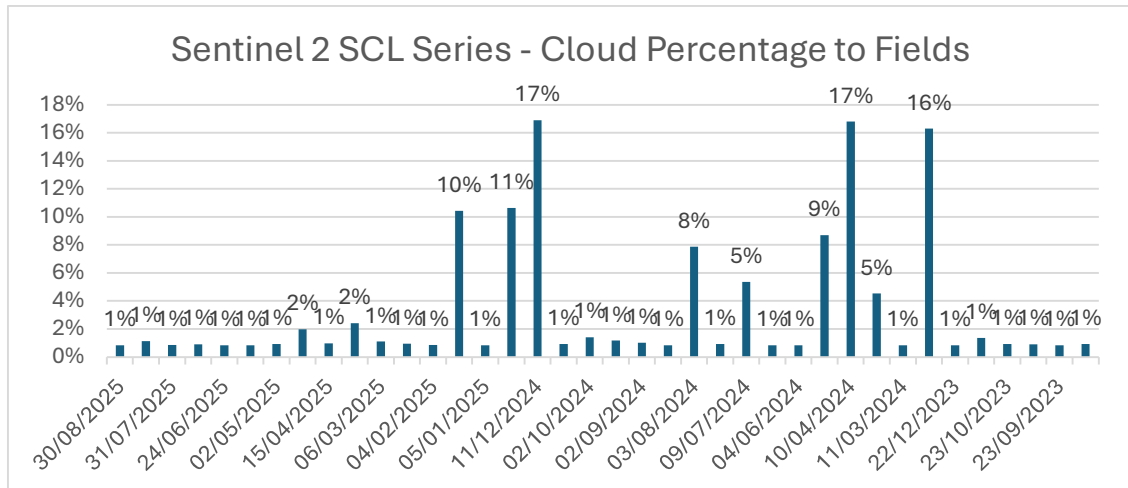


Figure 2.9 - Percentage of TSCL field polygons affected by cloud cover in the Sentinel-2 SCL image series used for crop age estimation. Source: Created by the author using Sentinel-2 imagery processed in ArcGIS Pro, and Microsoft Excel.

The TSCL regions cover two Sentinel-2 Military Grid Reference System (MGRS) tiles (36MXD and 36MYD). The imagery from both tiles was sourced from Sentinel Hub and mosaicked to ensure complete spatial coverage of the study area. The resulting rasters were masked to the TSCL field boundaries and converted to point features within ArcGIS Pro. SCL values extracted from the raster-derived point features were summarised at field level. To reduce the effect of isolated misclassified pixels, the most common SCL class in each field was identified using a majority-rule approach. The class occurring most frequently within a field was assigned as the representative land surface condition for that observation date. The resulting catalogue provided a record of land surface conditions within each field throughout the study period.

NDVI images for selected Sentinel-2 dates were also obtained from Sentinel Hub to monitor crop growth and changes over time (Tucker, 1979). NDVI values were extracted at field centroid locations and linked to individual Field IDs. The dataset provided field-level information on vegetation health and crop growth and was used to interpret crop growth patterns and estimate crop age.

2.4 Vegetation Indices

Vegetation indices are combinations of spectral reflectance bands used to highlight vegetation characteristics that are not easily identified from individual spectral bands (Huete et al., 2002). They are widely used in agricultural remote sensing to monitor vegetation condition, chlorophyll content and plant moisture, improving crop classification and monitoring (Xue & Su, 2017).

Three vegetation indices were derived from Sentinel-2 imagery in GEE following image pre-processing and cloud masking. These were the NDVI, NDMI and NDRE. These indices were selected because they provide complementary information on vegetation greenness, moisture content and chlorophyll concentration, which are important indicators of sugarcane growth and condition (Tucker, 1979; Gao, 1996; Delegido et al., 2011). The vegetation indices, together with the selected Sentinel-2 spectral bands, were used as the data inputs for the RF classification model.

NDVI was selected because it provides a measure of vegetation greenness and is widely used to monitor crop growth and biomass (Tucker, 1979; Xue & Su, 2017).

$$NDVI = (B8 - B4) / (B8 + B4)$$

Where B8 is the Near Infrared (NIR) band and B4 is the red band (Tucker, 1979).

NDMI was selected because the combination of near-infrared and short-wave infrared reflectance provides information on vegetation moisture and water status (Gao, 1996).

$$NDMI = (B8 - B11) / (B8 + B11)$$

Where B8 is the Near Infrared (NIR) band and B11 is the Short-Wave Infrared (SWIR) band (Gao, 1996).

NDRE was selected because it provides information related to chlorophyll content (Gitelson & Merzlyak, 1994; Delegido et al., 2011).

$$NDRE = (B8A - B5) / (B8A + B5)$$

Where B8A is the narrow near-infrared band and B5 is the red-edge band (Delegido et al., 2011; Clevers & Gitelson, 2013).

2.5 Training Data

Training data was developed for the RF classification to differentiate sugarcane from other land-cover classes. The following sections describe the development and verification of the sugarcane and other land-cover training samples used in the classification.

2.5.1 Sugarcane Training Sample Development

Training samples for sugarcane were taken from the TSCL Farmer Registry, which contained 12,992 contracted sugarcane field polygons distributed across the Enoosaen, Oloontare, Moita/Nkararo, Osupuko, Shankoe and Osinoni regions. The Farmer Registry provided field boundary information together with farmer information, contract details, crop ages, and was used as the main reference for developing sugarcane samples. The Farmer Registry provided crop age information, but the records were independently verified before the samples were selected. Each contracted field was matched to the Sentinel-2 SCL time-series catalogue described in 2.3. The catalogue contained monthly observations from September 2023 to August 2025 and was used to identify when sugarcane was established in each field. Crop age was estimated by examining the SCL time series for each field, working backwards from 30 August 2025 to identify the most recent change from bare soil to continuous vegetation cover. This transition was interpreted as the field establishment stage after planting or harvesting. This transition was taken as the crop establishment date after planting or harvesting, and the time between this date and the reference date was used to estimate crop age. Using this approach, crop-age estimates were successfully derived for 12,021 contracted fields. The satellite-derived crop age estimates were then compared with the crop ages recorded in the Farmer Registry. Fields exceeding the 3-month threshold, between the two age estimates were removed and therefore leaving 10,488 fields within the acceptable age range variance.

Historical cane development and harvest bulletins reports were compared with the dates identified from the SCL time series to measure the difference between the recorded planting or harvesting dates and the dates detected from the satellite data. A total of 100 fields were randomly selected for verification against TSCL cane development and harvest bulletin records (see Figure 2.10). The difference between the recorded planting and harvesting dates and the dates identified from the SCL time series was calculated to assess the accuracy of the satellite-derived crop age estimates.

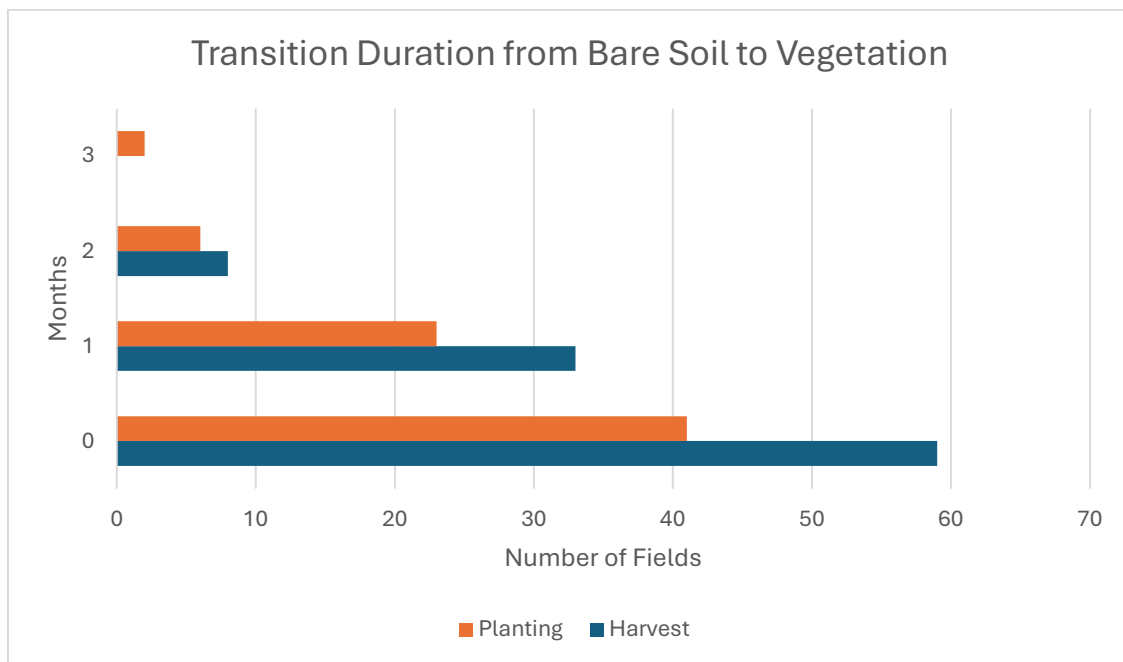


Figure 2.10 - Difference between planting and harvesting dates identified from the Sentinel-2 SCL time series and the corresponding TSCL report records. Source: Created by the author using Sentinel-2 SCL data and TSCL operational records.

Figure 2.10 indicates that most of the planting and harvesting transitions identified from the SCL time series occurred within 3-months of the corresponding TSCL reports. The difference for most observations was between 0 and 2 months, with only a small number exceeding this range. These differences were considered acceptable because of the monthly Sentinel-2 imagery, cloud cover, the time between images, delays between planting and visible crop growth and possible errors in the operational crop records (Foody, 2002). A 3-month age difference was used as the acceptance threshold for selecting training samples. Fields with a difference of 3-months or less were included in

the sugarcane reference dataset, while fields with larger differences were excluded from further analysis.

The 3-month threshold was also used to determine the minimum crop age for selecting sugarcane training samples. Analysis of the SCL time series indicated that recently planted or harvested fields typically remained classified as bare soil or exhibited sparse vegetation during the first three months following establishment. Sugarcane younger than four months was excluded from the training dataset to reduce confusion with bare soil. Training samples were therefore selected from sugarcane that was at least four months old, when a consistent vegetation signal could be detected in the SCL time series.

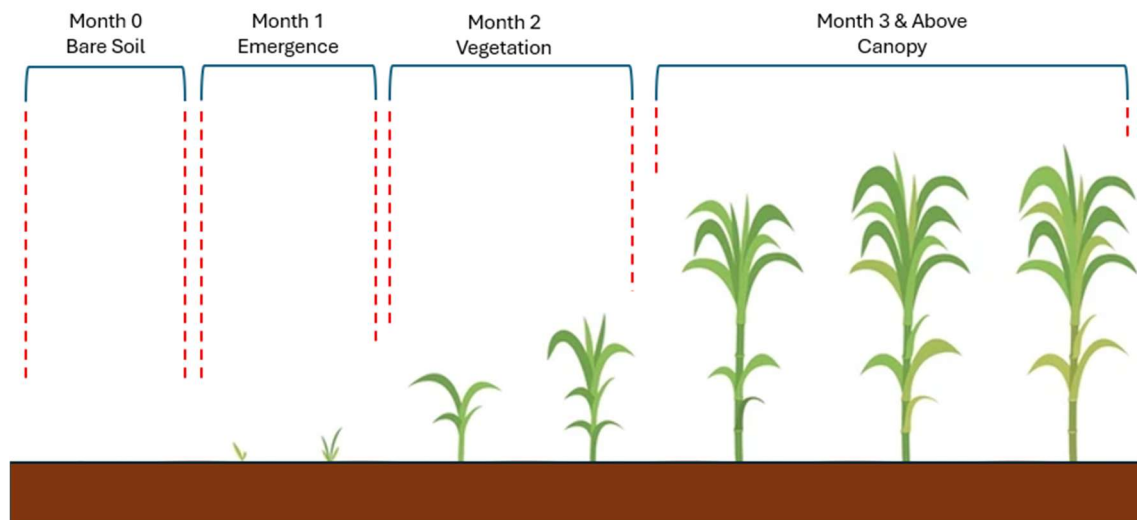


Figure 2.11 - Illustration of sugarcane growth following planting or harvesting, showing the transition from bare soil to vegetation used to support interpretation of the Sentinel-2 Scene Classification Layer (SCL) time series. Source: Author (2026).

Figure 2.11 provides an illustration of the relationship between the 3-month transition period and the minimum sugarcane training age used for sample selection and shows the growth stages of sugarcane after planting and harvesting. During the first three months after planting or harvesting, fields may remain classified as bare soil or sparse vegetation due to delayed crop emergence and incomplete canopy development (Som-ard et al., 2021). From about four months after planting, sugarcane can usually be detected consistently in the Sentinel-2 SCL time series, making it easier to distinguish from bare

soil and other non-vegetated land cover. For this reason, four months were adopted as the minimum crop age for sugarcane training sample selection.

After validation, a stratified random sampling method was used to select sugarcane training samples. Samples were selected across all TSCL regions and crop age groups to ensure they represented the different locations and crop ages found in the sugarcane production system (Belgiu & Drăguț, 2016; Congalton & Green, 2019). Samples were selected from crop-age categories ranging from 4 to 18 months, with 18 months representing the average harvest age reported by TSCL (TSCL, 2025) (see Table 2.9 and 2.10). This ensured that the classification model captured the full range of sugarcane growth stages and their spectral characteristics.

Table 2.9 - Distribution of Sugarcane Training and Validation Samples by TSCL Regions.

Regions	Training	Validation	Total
Enoosaen	211	100	311
Oloontare	174	82	256
Moita/Nkararo	111	53	164
Osupuko	128	48	176
Shankoe	133	49	182
Osinoni	133	49	182
Grand Total	890	381	1271

Source: Created by the author.

Table 2.10 - Distribution of Training and Validation Samples by Farmer Registry Crop Age.

Farmer Registry Age	Training	Validation	Total
4	70	30	100
5	68	32	100
6	71	29	100
7	57	28	85
8	41	24	65
9	56	27	83
10	60	25	85
11	93	43	136
12	88	36	124
13	68	36	104
14	83	27	110
15	41	13	54
16	32	7	39
17	37	16	53
18	25	8	33
Grand Total	890	381	1271

Source: Created by the author.

A total of 1,271 sugarcane samples were selected to train the classifier. These samples were divided into training and validation datasets using a 70:30 split, resulting in 890 training samples and 381 validation samples. The use of independent training and validation datasets is widely recommended in remote sensing classification studies to provide an unbiased assessment of classifier performance and classification accuracy (Belgiu & Drăguț, 2016).

2.5.2 Non-Sugarcane Training Sample Development

Training samples for the Other (non-sugarcane) class were developed using a combination of Sentinel-2 SCL time-series analysis, NDVI temporal profiles and visual interpretation of high-resolution imagery. Although the final classification framework grouped all non-sugarcane land-cover types into a single “Other” category, individual landcover classes were initially identified and evaluated separately to ensure that the training dataset adequately represented the range of non-sugarcane conditions present within the study area.

The Other class comprised maize, open area/grassland and timber/bush/tea landcover types. These classes were selected because they were the main non-sugarcane land cover types in the TSCL regions and had spectral characteristics that could be confused with sugarcane at different growth stages. Timber, bush, tea, and open grassland areas were first identified using the Sentinel-2 SCL time series. The SCL catalogue described in 2.3 was analyzed to identify areas with continuous vegetation cover throughout the study period that did not change to bare soil. Areas that remained covered by vegetation from September 2023 to August 2025 were selected as potential timber, bush, tea, and open grassland areas.

After identifying candidate locations using the SCL time series, NDVI time series were generated for each location. The resulting profiles were examined to assess seasonal variability and vegetation stability. Timber, bush and tea areas showed higher NDVI values with limited seasonal variation (see Figure 2.12). Grassland and open areas displayed lower NDVI values and greater temporal variability but did not exhibit the pronounced growth and harvest cycles associated with cultivated crops (see Figure 2.13). The use of temporal vegetation signatures to distinguish land-cover classes is well established in agricultural remote sensing and has been shown to improve classification performance when compared with single-date imagery (Belgiu & Csillik, 2018; Vuolo et al., 2018).

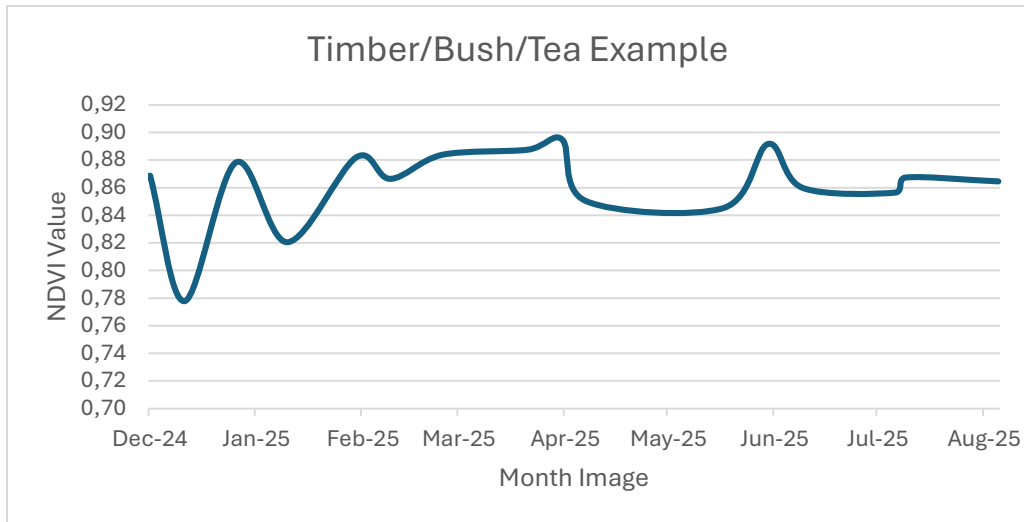


Figure 2.12 - Example NDVI temporal profile for timber, bush, and tea training samples.
Source: Created by the author using Sentinel-2 imagery.

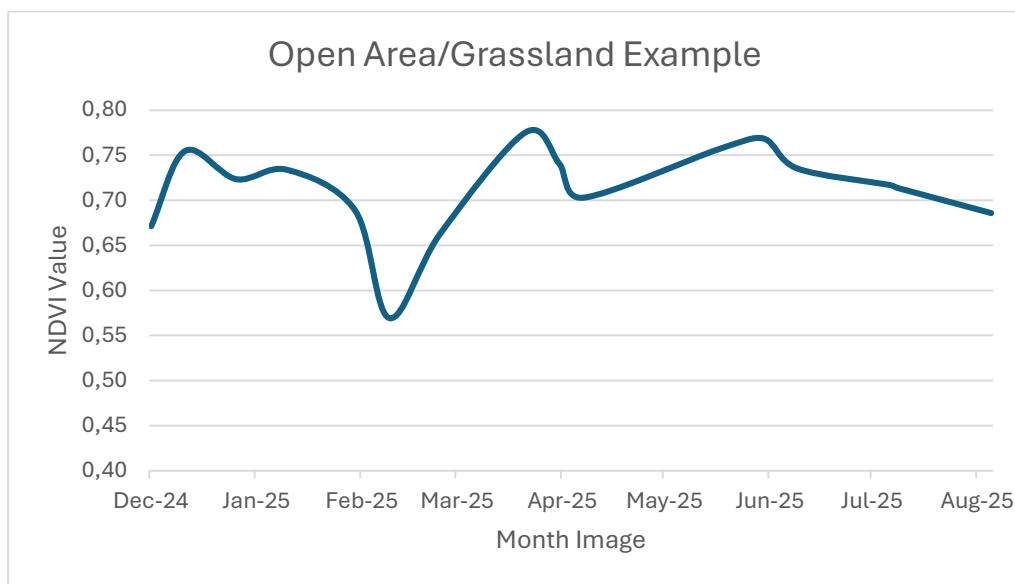


Figure 2.13 - NDVI temporal profile for open area and grassland training samples.
Source: Created by the author using Sentinel-2 imagery.

The selected samples were checked using high-resolution SPOT imagery provided by TSCL. The imagery was visually examined to confirm the land cover type and remove samples with mixed land cover or uncertain classification. Using additional imagery to verify training samples is a common practice in supervised classification and helps improve the reliability of the training data (Eisfelder et al., 2024).

Maize samples were identified using a combination of SCL temporal analysis, NDVI temporal profiles and known seasonal cropping patterns within the study area. Information supplied by TSCL indicated that maize planting and harvesting activities were concentrated during the January to February and July to August periods (TSCL, 2025). Potential maize fields were therefore identified by examining the SCL time series for changes between bare soil and vegetation during these periods. The identified sample fields were further assessed using NDVI temporal profiles (see Figure 2.14). Unlike timber, bush and tea classes, maize exhibited a characteristic phenological pattern consisting of low NDVI values following planting, a rapid increase in NDVI during vegetative growth, peak NDVI values during canopy development and a decline following harvest. Fields displaying these characteristics were retained as maize training samples. Phenological information derived from multi-temporal satellite observations has been widely used for crop discrimination and agricultural land-cover mapping (Wang et al., 2020).

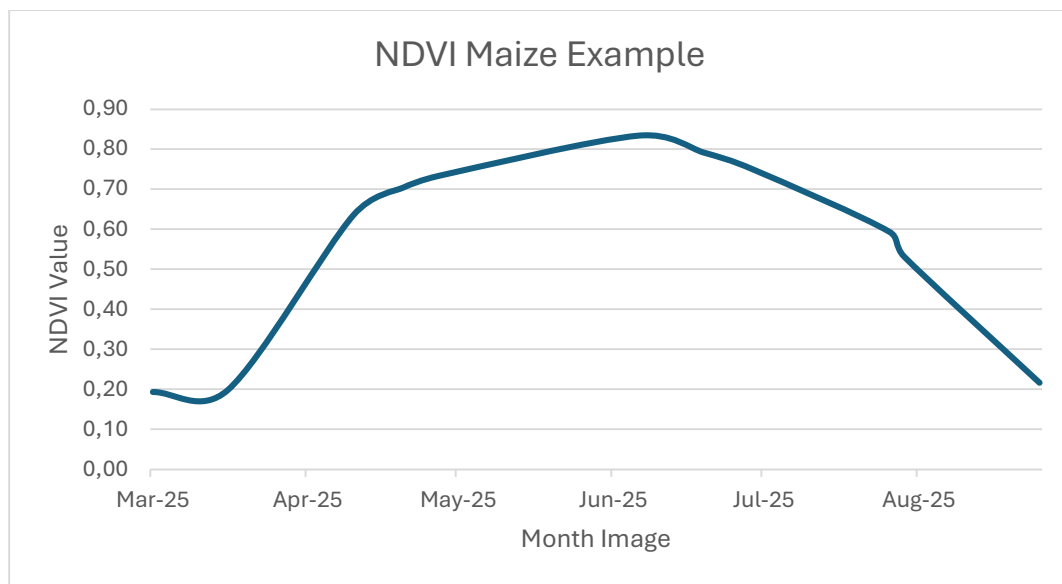


Figure 2.14 - NDVI temporal profile for maize training samples. Source: Created by the author using Sentinel-2 imagery.

Using SCL observations, NDVI time series, and high-resolution SPOT imagery provided several sources of information for selecting training samples and helped reduce classification errors. This approach improved the reliability of the Other training dataset and ensured that the selected samples represented the different non-sugarcane land cover

types found in the study area. A total of 1,257 samples were selected for the Other category, comprising 535 maize samples, 629 open area/grassland samples and 93 timber/bush/tea samples (see Table 2.11).

Table 2.11 - Distribution of Other Land-Cover Training Samples by TSCL Region.

Region	Maize	OpenArea/Grassland	Timber/Bush/Tea	Total
Enoosaen	60	31	15	106
Oloontare	62	76	26	164
Moita/Nkararo	163	133	11	307
Osupuko	69	127	16	212
Shankoe	21	31	15	67
Osinoni	160	231	10	401
Grand Total	535	629	93	1257

Source: Created by author.

After the Other samples were identified and verified, 1,257 samples were selected to train the classifier. These comprised 535 maize samples, 629 open area/grassland samples and 93 timber/bush/tea samples. To ensure that the training dataset adequately represented the spatial variability present within the study area, samples were distributed across all TSCL regions.

Table 2.12 - Distribution of Other Land-Cover Training and Validation Samples by TSCL Region.

Region	Training	Validation	Total
Enoosaen	77	29	106
Oloontare	113	51	164
Moita/Nkararo	223	84	307
Osupuko	142	70	212
Shankoe	43	24	67
Osinoni	282	119	401
Grand Total	880	377	1257

Source: Created by author.

The samples were divided into training and validation datasets using a 70:30 split. This resulted in 880 training samples and 377 validation samples (see Tables 2.13 and 2.14). The training dataset was used to develop the RF classification model, while the validation dataset was reserved for the accuracy assessment.

Table 2.13 - Distribution of Training Samples by Land-Cover Class and TSCL Region

Region	Maize	Open Area/Grassland	Timber/Bush/Tea	Total
Enoosaen	43	24	10	77
Oloontare	44	49	20	113
Moita/Nkararo	117	101	5	223
Osupuko	47	85	10	142
Shankoe	17	16	10	43
Osinoni	116	158	8	282
Grand Total	384	433	63	880

Source: Created by author.

The distribution of training and validation samples by TSCL regions is presented in Table 2.12. Additional distributions by landcover class and regions are presented in Tables 2.13 and 2.14.

Table 2.14: Distribution of Validation Samples by Land-Cover Class and TSCL Region

Region	Maize	OpenArea/Grassland	Timber/Bush/Tea	Total
Enoosaen	17	7	5	29
Oloontare	18	27	6	51
Moita/Nkararo	46	32	6	84
Osupuko	22	42	6	70
Shankoe	4	15	5	24
Osinoni	44	73	2	119
Grand Total	151	196	30	377

Source: Created by author.

The final sample distribution ensured that all major non-sugarcane landcover types were represented within both the training and validation datasets.

2.5.3 Training Sample Preparation

All land cover areas were represented as field polygons using the TSCL Farmer Registry and additional land cover datasets created during the training sample selection process. Representative sample locations were created by generating the center point (centroid) of each training polygon in ArcGIS Pro. Using centroids ensured that each training sample was represented by a single point and made it easier to extract spectral information from the Sentinel-2 time series.

To minimise the influence of boundary effects and mixed pixels, a negative buffer of 5m was applied to all training polygons prior to sample selection (see Figure 2.15). Polygons smaller than the applied buffer distance were excluded from the training dataset as they could not support a representative centroid location following buffering. This process

reduced the likelihood of training samples being influenced by adjacent land-cover types, field boundaries, roads or other edge features that could introduce classification uncertainty (Foody, 2002; Eisfelder et al., 2024).



Figure 2.15 - Example of centroid generation following the application of a 5m negative buffer to sugarcane and non-sugarcane training polygons.

The resulting centroid points were visually inspected and cross-referenced against high-resolution SPOT imagery to verify positional accuracy and landcover consistency (see Figure 2.15). Samples with uncertain classification, mixed land cover, or location errors were removed from the training dataset.

The final training dataset comprised 1,271 Sugarcane samples and 1,257 Other samples (see Figure 2.16). Together, these samples formed the reference dataset used to train and validate the RF classifier.

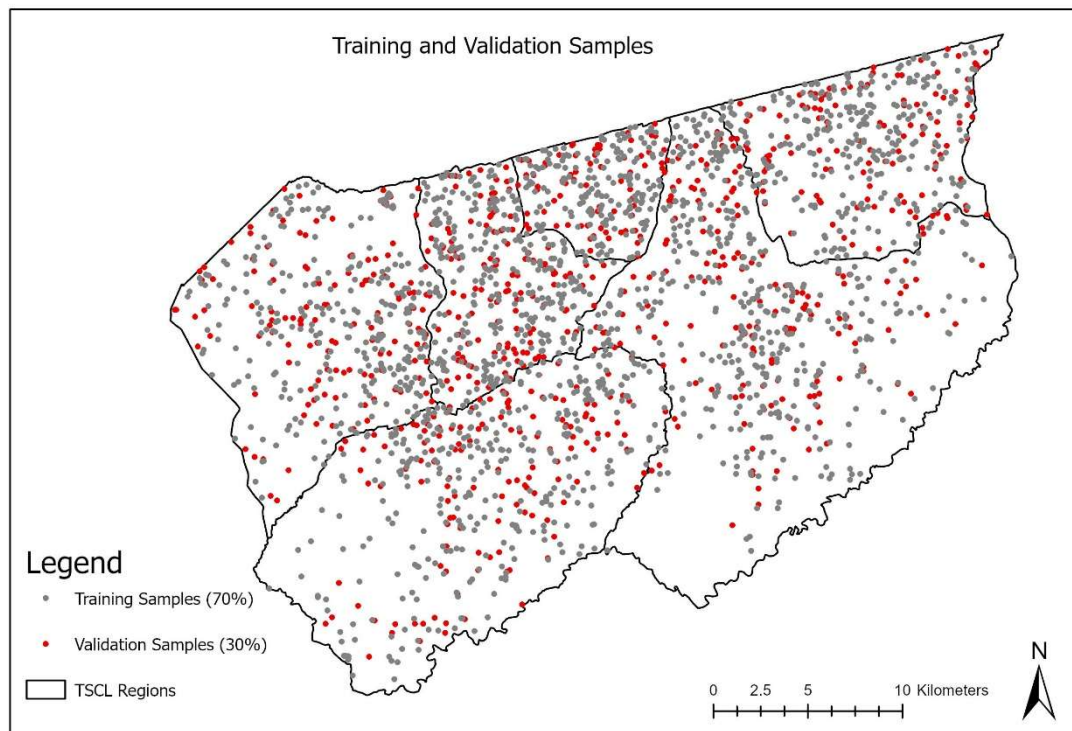


Figure 2.16 - Spatial distribution of the training (70%) and validation (30%) samples across the TSCL operational area. Source: Created by author.

2.6 Random Forest Classification

Sugarcane was classified using the RF algorithm in GEE. RF is a machine learning algorithm developed by Breiman (2001) that uses multiple decision trees to produce robust classification results and reduce overfitting. RF is widely used in remote sensing because it is reliable, can handle many input variables and performs well in land cover classification (Breiman, 2001; Belgiu & Drăguț, 2016).

The classifier was developed using the training dataset described in 2.5. Training samples were grouped into two classes; Sugarcane (Class 1) and Other (Class 2). The Sugarcane class represented sugarcane fields across multiple crop-age categories, while the Other class comprised maize, open area/grassland and timber/bush/tea landcover types. Including several non-sugarcane land-cover types helped provide representative training data for distinguishing sugarcane from vegetation with similar spectral characteristics (Belgiu & Drăguț, 2016). A multi-temporal Sentinel-2 dataset was created using nine

Sentinel-2 L2A image composites collected between December 2024 and August 2025. These image dates were selected to capture changes in vegetation and crop growth throughout the sugarcane growing cycle. Using Sentinel-2 images from different dates improves crop classification because it captures changes in crop growth that cannot be seen in a single image (Immitzer et al., 2016). Each image date, six Sentinel-2 spectral bands were retained, namely Band 4 (Red), Band 5 (Red Edge), Band 8 (Near Infrared), Band 8A (Narrow Near Infrared), Band 11 (Short-Wave Infrared 1) and Band 12 (Short-Wave Infrared 2). Sentinel-2 red-edge and short-wave infrared bands are valuable for agricultural monitoring because they help detect differences in plant structure, chlorophyll levels, and moisture within the crop canopy (Delegido et al., 2011; Immitzer et al., 2016). In addition to the spectral bands, three vegetation indices were calculated for each image date, the NDVI, the NDRE and the NDMI. These indices were selected because they provide complementary information on vegetation greenness, moisture content and chlorophyll concentration, which are important indicators of sugarcane growth and condition (Tucker, 1979; Gao, 1996; Delegido et al., 2011; Clevers & Gitelson, 2013). Cloud masking was undertaken using the Sentinel-2 SCL, as described in 2.3. For each spectral band and vegetation index, the median value from the available image dates was used to fill missing values in the image composites.

The final dataset contained 81 variables from the multi-temporal Sentinel-2 images, which were created from the spectral bands and vegetation indices for each of the nine image dates and used by the RF model to classify crops. Previous studies have found that using Sentinel-2 images collected at different times during the growing period improves crop classification (Belgiu & Csillik, 2018; Vuolo et al., 2018). This is because seasonal changes and crop growth help distinguish between different crop types and reduce classification errors (Belgiu & Csillik, 2018; Vuolo et al., 2018). Image segmentation using the Simple Non-Iterative Clustering (SNIC) algorithm (Achanta & Süsstrunk, 2017) was incorporated to support visual quality control during the development of the classification. The segmented imagery provided a way of checking field boundaries and assessing whether the pixel-based classification corresponded with the spatial structure of agricultural fields. However, the final classification used pixel-based spectral bands and vegetation indices from the multi-temporal Sentinel-2 image stack.

Training samples were extracted from the multi-temporal image stack using the “sampleRegions” function in GEE. Carefully selected training samples help improve the accuracy and reliability of supervised classification. (Belgiu & Drăguț, 2016). A total of 1,770 labelled training samples were used for the classification, 890 Sugarcane samples and 880 Other samples.

The RF classifier was configured to generate 50 decision trees. Each tree was built using a random sample of the training data and a random selection of the input variables (Breiman, 2001). Using different training samples and input variables for each tree helped improve classification accuracy and reduce overfitting (Breiman, 2001).

After training, the classifier was applied to the full dataset to produce the final sugarcane classification. Each decision tree predicted a class for every pixel, and the final class was based on the majority vote, with the class receiving the most votes assigned to the pixel (Breiman, 2001). The resulting classification was then clipped to the TSCL regional boundary. The classified raster was subsequently masked to the field boundaries and the majority class for each Field ID was determined from the classified pixels. Each field was then classified as either Sugarcane or Other based on most points within the Field boundary. The final field-based classification was used to estimate sugarcane area, assess classification accuracy, and analyze the distribution of sugarcane across the TSCL regions.

2.7 Crop Age Estimation

This section describes the method used to estimate sugarcane crop age using the Sentinel-2 SCL time series. The main steps include the identification of bare-soil events and the calculation of crop age from the most recent planting or harvesting event.

2.7.1 Development of the Sentinel-2 SCL Time Series

Crop age estimation was undertaken using a temporal series of Sentinel-2 L2A SCL imagery. The SCL product is generated as part of the Sen2Cor atmospheric correction process and provides a pixel-level classification of land surface and atmospheric conditions. The product contains twelve predefined classes representing vegetation, bare soil, water, cloud shadow, clouds, cirrus and other surface conditions (Main-Knorn et al., 2017; ESA, 2024).

To reduce the effects of cloud cover while maintaining good coverage throughout the study period, approximately two Sentinel-2 images with the lowest cloud cover were selected each month between September 2023 and August 2025. The majority TSCL regions covers two Sentinel-2 MGRS tiles (36MXD and 36MYD); therefore, corresponding SCL rasters for each acquisition date were mosaicked to create a single image covering the entire study area.

The merged SCL rasters were masked to the TSCL field boundaries using the Extract by Mask tool in ArcGIS Pro. The masked rasters were subsequently converted to point features while retaining the original SCL classification code for each pixel. A spatial join was then performed between the raster derived points and the registered field boundary polygons.

The TSCL field database contained 56,837 digitised field polygons. For each image acquisition date, the dominant SCL class within each field polygon was determined using a majority-count approach, whereby the most frequently occurring SCL code was assigned as the representative classification for that field (see Table 2.15). This process

produced a time-series catalogue containing a single SCL classification for each field polygon for every image date throughout the observation period.

Table 2.15: Sentinel-2 Scene Classification Layer (SCL) Classes used in this Study

0	No Data
1	Saturated or Defective Pixels
2	Dark Area Pixels
3	Cloud Shadow
4	Vegetation
5	Bare Soil
6	Water
7	Clouds (Low Probability)
8	Clouds (Medium Probability)
9	Clouds (High Probability)
10	Thin Cirrus
11	Snow or Ice

Source: Adapted from European Space Agency (ESA, 2024).

2.7.2 Identification of Bare Soil Events

The identification of the most recent bare-soil event provides an indication of the timing (date) of the latest field harvest or plant event. Each field polygon in the catalogue was analysed in reverse, chronological order, beginning with the most recent Sentinel-2 image sourced on 30 August 2025. The catalogue was searched backwards to identify the first occurrence of SCL Class 5 (Bare Soil). The date corresponding to this event was recorded as the date last Bare Soil.

2.7.3 Determination of Date Last Bare Soil

The date, last Bare Soil (Class 5), was extracted from the SCL catalogue using Microsoft Excel. The first event of the text value "Bare Soil" was identified, and the corresponding image date was returned as the date of last Bare Soil. Fields that did not contain a bare soil event within the available catalogue period were assumed to have remained under vegetation cover throughout the period. Such fields were interpreted as representing

relatively older crop stands or other land types for which no planting or harvesting event could be detected within the available Sentinel-2 records.

2.7.4 Estimation of Cane Age

Estimated cane age was calculated as the elapsed time (months) between the most recent image date in the catalogue, 30 August 2025, and the date of last Bare Soil. The calculation was performed in Microsoft Excel using the DATEDIF function and expressed in complete months.

The crop age estimation can be represented as “Estimated Cane Age” = Reference Date - Date Last Bare Soil where:

Reference Date = 30 August 2025 - Date Last Bare Soil = Most recent occurrence of SCL Class 5 (Bare Soil)

For example, a field with a date of last Bare Soil of 4 February 2025 would have an estimated crop age of approximately six months on 30 August 2025, whereas a field with a date of last Bare Soil of 21 March 2025 would have an estimated crop age of approximately five months.

2.7.5 Crop Cycle Attribution

Operational crop cycle information was obtained from the TSCL Farmer Registry and linked to the estimated crop ages using the unique Field ID associated with each registered field polygon. Crop cycle categories included Plant Cane (PC), First Ratoon, Second Ratoon and Third Ratoon crops. The crop cycle information was not used directly in the age estimation process but provided additional operational context for interpretation of the estimated ages and subsequent validation against company records. The integration of crop cycle information enabled assessment of whether the estimated ages were consistent with the expected growth stages associated with each production cycle.

2.7.6 Validation Against Farmer Registry Records

The satellite-derived sugarcane age estimates were compared with the crop ages recorded in the TSCL Farmer Registry on 1 September 2025. The Farmer Registry contains field information including crop age, crop cycle, farmer details and contract information. For each field, the age difference was calculated by comparing the crop age in the Farmer Registry with the crop age estimated from the SCL catalogue. A tolerance threshold of 3-months was adopted. Fields where the difference between the registry age and the estimated cane age exceeded 3-months were flagged for further investigation and assigned a “Check” status. The percentage of fields with an age difference of 3 months or less was then calculated to assess how well the Sentinel-2 SCL time series estimated sugarcane crop age.

2.8 Accuracy Assessment

This section describes the methods used to assess the accuracy of the sugarcane classification and crop age estimates. The assessment included classification accuracy metrics, comparison with the TSCL Farmer Registry and validation of the estimated crop ages.

2.8.1 Validation Dataset

The accuracy of the sugarcane classification was evaluated using a validation dataset created from the reference samples prepared during the training data stage. A stratified random sampling method was used to ensure that all major sugarcane-growing regions were well represented within the TSCL regions, namely Enosaaen, Oloontare, Moita/Nkararo, Osupuko, Shankoe and Osinoni (Congalton, 1991; Congalton & Green, 2019).

The sugarcane reference dataset consisted of 1,271 verified sugarcane polygons distributed across a range of crop ages, while the Other (non-sugarcane) reference dataset comprised 1,257 polygons representing maize, open grassland areas and tea/bush/timber vegetation. To minimise spatial uncertainty associated with mixed pixels along field

boundaries (Foody, 2002), a negative buffer of 5m was applied to all reference polygons prior to sample extraction. This ensured that the training and validation samples came from the middle of the fields, where the land cover was more uniform.

Following buffering and verification, the sugarcane dataset was divided into training (70%) and validation (30%) datasets, resulting in 890 sugarcane training samples and 381 sugarcane validation samples. The Other (non-sugarcane) dataset was divided into 880 training samples and 377 validation samples. The validation dataset was not used during model training and therefore provided an independent assessment of classification performance (Belgiu & Drăguț, 2016; Congalton & Green, 2019).

The Other (non-sugarcane) classes included maize, open/grassland areas and tea/bush/timber vegetation. Although the main aim of the study was to identify sugarcane rather than create a detailed land cover map, these classes were included because they are the main land cover types that can have similar spectral characteristics to sugarcane at different growth stages. For the final accuracy assessment, all Other (non-sugarcane) classes were combined into one category, allowing the results to be evaluated using a simple sugarcane and non-sugarcane confusion matrix.

2.8.2 Confusion Matrix Assessment

Classification performance was assessed using a confusion matrix. The confusion matrix compares the class predicted by the RF classifier with the actual class for each validation sample (Congalton, 1991). The matrix shows the number of correctly and incorrectly classified samples and is used to calculate standard accuracy measures (Congalton, 1991).

For this study, the confusion matrix consisted of two classes:

- Sugarcane (Class 1)
- Other (Class 2)

The diagonal values show the correctly classified samples, while the off-diagonal values show classification errors (Congalton & Green, 2019). The confusion matrix was

generated using the validation dataset and was used to calculate Overall Accuracy, Producer's Accuracy, User's Accuracy and the Kappa Coefficient. These metrics are widely used to assess the reliability of land cover classifications from remotely sensed imagery (Foody, 2002; Congalton & Green, 2019).

2.8.3 Overall Accuracy, Producers Accuracy and Users Accuracy

Overall Accuracy (OA) represents the proportion of validation samples that were correctly classified and provides a general measure of classification performance. It is calculated as the ratio of correctly classified samples to the total number of validation samples.

$$OA = (\text{Correctly Classified Samples} / \text{Total Validation Samples}) * 100$$

where:

- Correctly classified samples are shown along the main diagonal of the confusion matrix,
- Total Validation Samples refers to the total number of validation samples used in the assessment.

Along with Overall Accuracy, Producer's Accuracy and User's Accuracy were calculated for each class. Producer's Accuracy shows how often reference samples were classified correctly and indicate omission error (Congalton & Green, 2019). User's Accuracy shows how often samples classified into a category belong to that category and indicates commission error (Congalton & Green, 2019).

2.8.4 Kappa Coefficient

The Kappa Coefficient (K) was calculated to measure the agreement between the classified results and the reference data after accounting for chance agreement (Congalton & Green, 2019). The Kappa statistics are widely used in remote sensing accuracy assessment because it provides an additional measure of classification agreement that accounts for chance agreement (Foody, 2002).

$$K = (P_o - P_e) / (1 - P_e) \text{ (Congalton \& Green, 2019)}$$

where:

- P_o represents the observed agreement derived from the confusion matrix.
- P_e represents the expected agreement occurring by chance.

Kappa values range from -1 to 1, where values approaching 1 indicate strong agreement between the classification and the reference data, values near 0 indicate random agreement and negative values indicate disagreement (Congalton & Green, 2019). According to Landis and Koch (1977), Kappa values greater than 0.80 indicate almost perfect agreement and are generally considered indicative of highly reliable classification results.

2.8.5 Farmer Registry Validation

In addition to the statistical accuracy assessment based on validation samples, the final sugarcane classification will be compared against operational records contained within the TSCL Farmer Registry. The Farmer Registry contains field-level information for contracted sugarcane production (TSCL, 2025) and therefore provides an independent operational dataset for assessing classification performance. After classification, each registered field was classified as either Sugarcane or Other based on most classified pixels within the field. The resulting classification was compared against Farmer Registry records using the unique Field ID associated with each contracted field. Validation was carried out at the field level by calculating the percentage of registered sugarcane area that was correctly classified in each TSCL region. Regional summaries were produced for Enosaaen, Oloontare, Moita/Nkararo, Osinoni, Osupuko and Shankoe. These results would provide an additional measure of classification performance and support the confusion matrix results from the independent validation dataset. This comparison was valuable because it assessed the classification using actual commercial production records rather than only the validation samples.

2.8.6 Crop Age Estimation Validation

The reliability of the crop age estimation methodology was assessed by comparing satellite derived sugarcane ages with corresponding Farmer Registry crop age records. The age estimation process uses the Sentinel-2 SCL time series to identify the most recent bare-soil event associated with planting or harvesting. Crop age was calculated as the time difference (months) between the reference image date and the identified bare-soil event. The satellite estimated crop ages was compared against Farmer Registry ages for all fields, where both operational records and satellite-derived age estimates were available. A tolerance threshold of 3 months was adopted to account for uncertainties associated with image acquisition frequency, cloud contamination, mixed pixels, harvesting schedules, field management practices and differences between actual field operations and the nearest available satellite observation date. Fields where the difference between the estimated cane age and Farmer Registry age was less than or equal to three months were then seem as agreeing with the TSCL data, while fields exceeding this threshold were flagged (“Check”). The proportion of fields falling within the tolerance threshold was calculated as an operational agreement metric used to evaluate the reliability of the Sentinel-2 SCL time-series approach for sugarcane crop age estimation. The resulting agreement statistics provide an indication of the likely reliability of crop age estimation across the wider study area, including NCC fields for which age validation records were not available.

3: Results

This chapter presents the results obtained from the remote sensing framework developed to address the four research objectives of the study. The results include the generation and application of multi-temporal Sentinel-2 spectral bands and vegetation indices within the RF classification, assessment of classification accuracy and agreement with the TSCL Farmer Registry, estimation and validation of sugarcane crop age using the Sentinel-2 SCL time series, and the identification and spatial assessment of NCC within the TSCL supply area.

3.1 Classification Accuracy Assessment

The accuracy of the RF sugarcane classification was evaluated using an independent validation dataset comprising 757 validation samples. A confusion matrix was used to compare the classification results with the reference data and to calculate the Overall Accuracy, Producer's Accuracy, User's Accuracy and the Kappa Coefficient (Congalton, 1991; Congalton & Green, 2019).

Table 3.1 - Confusion Matrix for the Random Forest Sugarcane Classification.

Reference Class	Classified as Sugarcane	Classified as Other	Total
Sugarcane	375	6	381
Other	8	368	376
Total	383	374	757

Source: Created by the author using the classification validation dataset. Note: A total of 377 "Other" validation samples were initially generated through random selection in Microsoft Excel. During validation, one duplicate sample (Osupuko Region) was identified and excluded from the analysis, resulting in 376 unique "Other" samples included in the confusion matrix.

Of the 757 validation samples, 743 were correctly classified while only 14 samples were misclassified (see Table 3.1). Of the 381 sugarcane reference samples, 375 were correctly

classified. For the Other class, 368 of the 376 reference samples were correctly classified. These results show a high level of agreement between the classified map and the validation dataset.

The Overall Accuracy (OA) was calculated as:

$$OA = (743 / 757) * 100$$

$$OA = 98.15\%$$

Producer Accuracy and User Accuracy were also calculated for each class. Sugarcane achieved a Producer Accuracy of 98.4% and a User Accuracy of 97.9%, while the Other class achieved a Producer Accuracy of 97.9% and a User Accuracy of 98.4% (see Table 3.2).

Table 3.2 - Class Specific Classification Accuracy Statistics.

Metric	Sugarcane (%)	Other (%)
Producer Accuracy	98.4	97.9
User Accuracy	97.9	98.4

Source: Created by the author.

The Kappa Coefficient (K) was also calculated to measure agreement beyond chance and had a value of 0.963.

According to the interpretation framework proposed by Landis and Koch (1977), Kappa values between 0.81 and 1.00 indicate almost perfect agreement between classified and reference datasets. The Kappa value of 0.963 indicates that the classification was highly reliable (see Table 3.3).

Table 3.3 - Summary of Classification Accuracy Statistics.

Metric	Value
OA	98.15%
Kappa	0.963
Sugarcane PA	98.4%
Sugarcane UA	97.9%
Other PA	97.9%
Other UA	98.4%

Source: Created by author.

The high Overall Accuracy (OA) and Kappa Coefficient (Kappa) show that the multi-temporal Sentinel-2 imagery, together with the spectral bands and vegetation indices, accurately distinguished sugarcane from other vegetation classes. Similar studies have shown that multi-temporal Sentinel-2 imagery improves crop classification by capturing differences in crop growth over time (Belgiu & Csillik, 2018; Vuolo et al., 2018). The low number of classification errors shows that the Sentinel-2 time series effectively distinguished sugarcane from maize, grassland, tea, bush, and timber.

Overall, the results confirm that the RF classification approach provides a reliable methodology for operational sugarcane mapping within the TSCL regions, consistent with findings reported by Belgiu & Drăguț (2016) and Som-ard et al. (2021).

3.2 Validation Using Farmer Registry Database

The sugarcane classification was compared against TSCL Farmer Registry records. The Farmer Registry contained a total contracted sugarcane area of 10,997.8ha distributed across six operational management regions (see Figure 3.1). Comparison between the classified sugarcane map and the Farmer Registry revealed that 10,064.1ha was successfully identified as sugarcane by the RF classification.

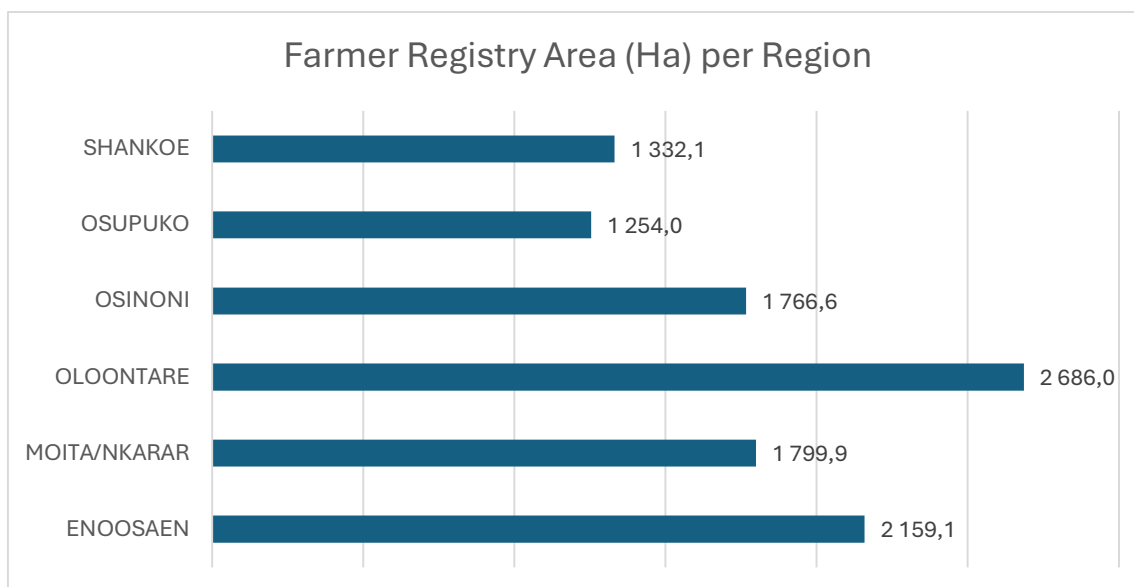


Figure 3.1 - Farmer Registry sugarcane area by TSCL supply region. Source: Created by the author using TSCL Farmer Registry data.

The overall area agreement was calculated as:

$$\text{Area Agreement} = (10,064.1 / 10,997.8) * 100$$

$$\text{Area Agreement} = 91.5\%$$

The results indicate that more than nine out of every ten hectares of registered sugarcane are successfully identified by the classification process. The remaining 933.7 ha (8.5%) were classified as non-cane areas (see Figure 3.2).

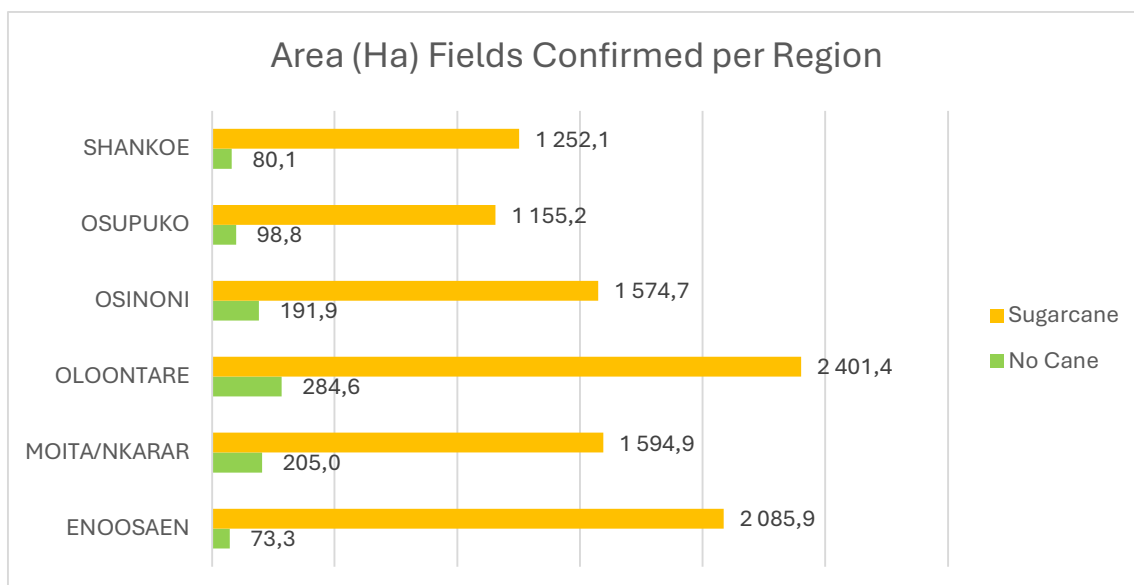


Figure 3.2 - Area of registered sugarcane fields classified as sugarcane and non-sugarcane by TSCL regions. Source: Created by the author using TSCL Farmer Registry data and RF classification results.

TSCL regional classification results, Enooaen achieved the highest agreement (96.6%), followed by Shankoe (94.0%) and Osupuko (92.1%). Slightly lower agreement levels were observed in Moita/Nkararo (88.6%), Osinoni (89.1%) and Oloontare (89.4%) (see Table 3.4 and Figure 3.3).

Table 3.4 - Regional Comparison of Classified Sugarcane Area and TSCL Farmer Registry Data.

Region	Registered Area (ha)	Classified Sugarcane (ha)	Agreement (%)
Enooaen	2,159.1	2,085.9	96.6
Moita/Nkararo	1,799.9	1,594.9	88.6
Oloontare	2,686.0	2,401.4	89.4
Osinoni	1,766.6	1,574.7	89.1
Osupuko	1,254.0	1,155.2	92.1
Shankoe	1,332.1	1,252.1	94.0
Total	10,997.8	10,064.1	91.5

Source: Created by author using TSCL farmer Registry.

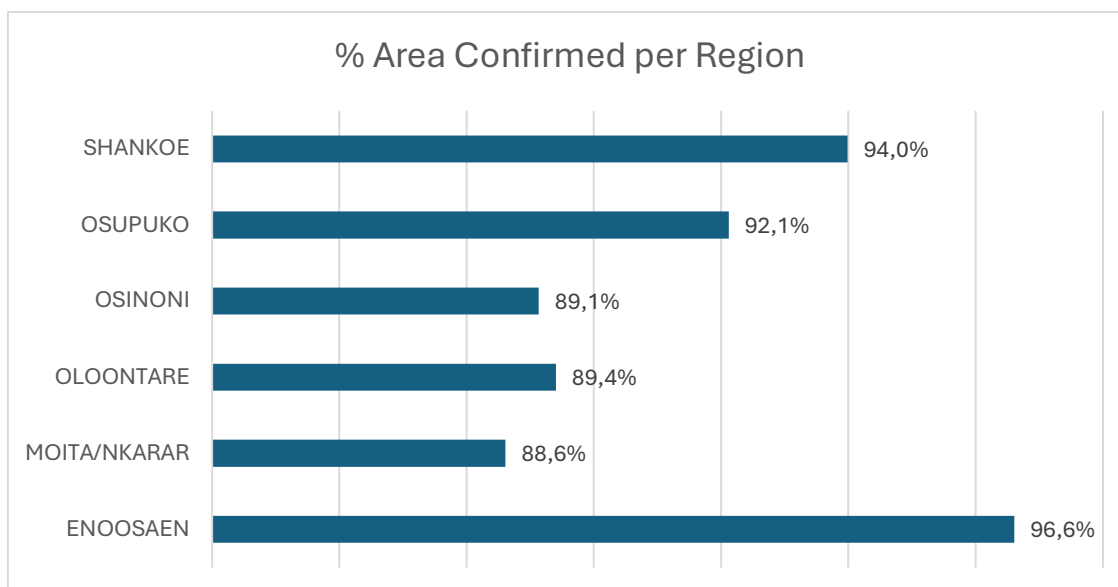


Figure 3.3 - Percentage Agreement Between Classified Sugarcane Area and TSCL Farmer Registry by Region. Source: Created by the author using TSCL Farmer Registry data and RF classification results

Overall, the Farmer Registry validation supports the findings of the confusion matrix assessment and shows that the classification methodology is suitable for sugarcane monitoring within the TSCL regions.

3.3 Spatial Distribution of Contracted and Non-Contracted Sugarcane

The classification identified a total sugarcane area of 18,899.1ha within the study area (see Figure 3.4 and Table 3.5). This comprised 10,064.1ha, successfully identified against the Farmer Registry, contracted sugarcane fields and 8,835.0ha classified as NCC (see Figure 3.5 and Table 3.6).

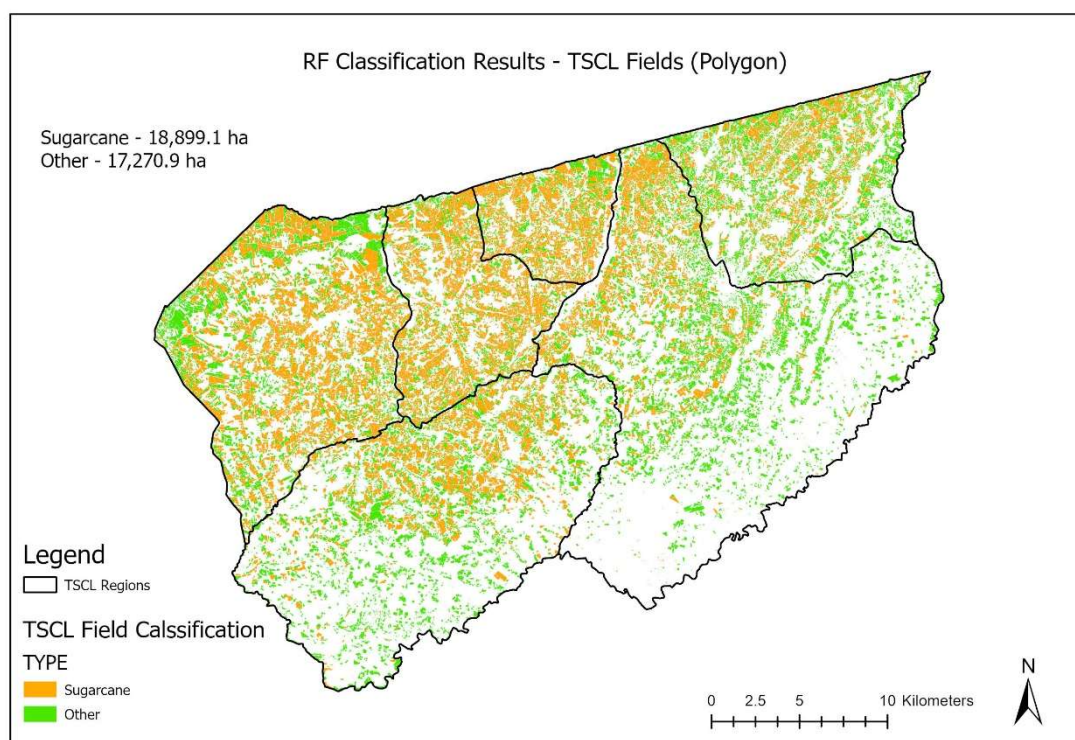


Figure 3.4 - Field-Level Sugarcane Classification Produced Using the Random Forest (RF) Classification. Source: Created by author, based on Sentinel-2 imagery, TSCL field boundaries and Random Forest (RF) classification results.

Table 3.5 summarizes the field-level classification results produced by the RF classifier, showing the total area classified as Sugarcane and Other within the TSCL study area.

Table 3.5 - Summary of Random Forest Classification Results

Classification	Sum of Area Ha
Sugarcane	18 899,1
Other	17 270,9
Grand Total	36 170,0

Source: Created by author, based on Random Forest (RF) classification results.

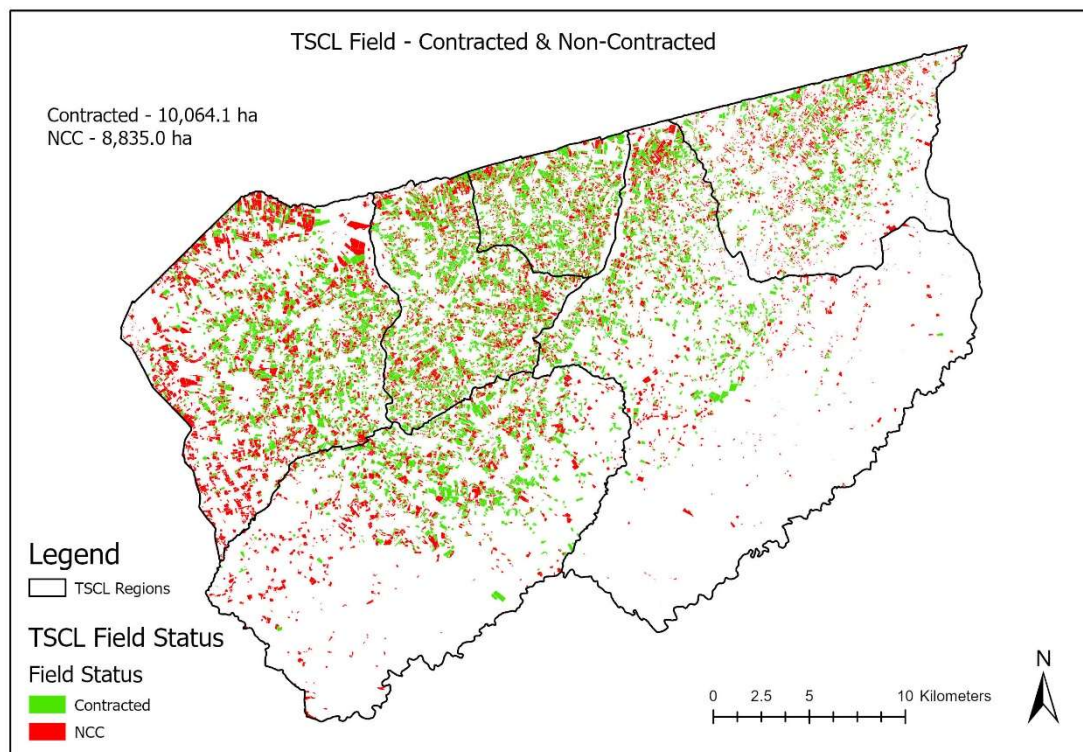


Figure 3.5 - TSCL Field Status: Distribution of Contracted and Non-Contracted Sugarcane Fields. Source: Created by author, based on TSCL field boundary and contract status data.

Figure 3.6 compares the classified sugarcane area for contracted and NCC across the six TSCL regions. The information highlights regional differences in contract status and shows an overview of the distribution of classified sugarcane within the study area.

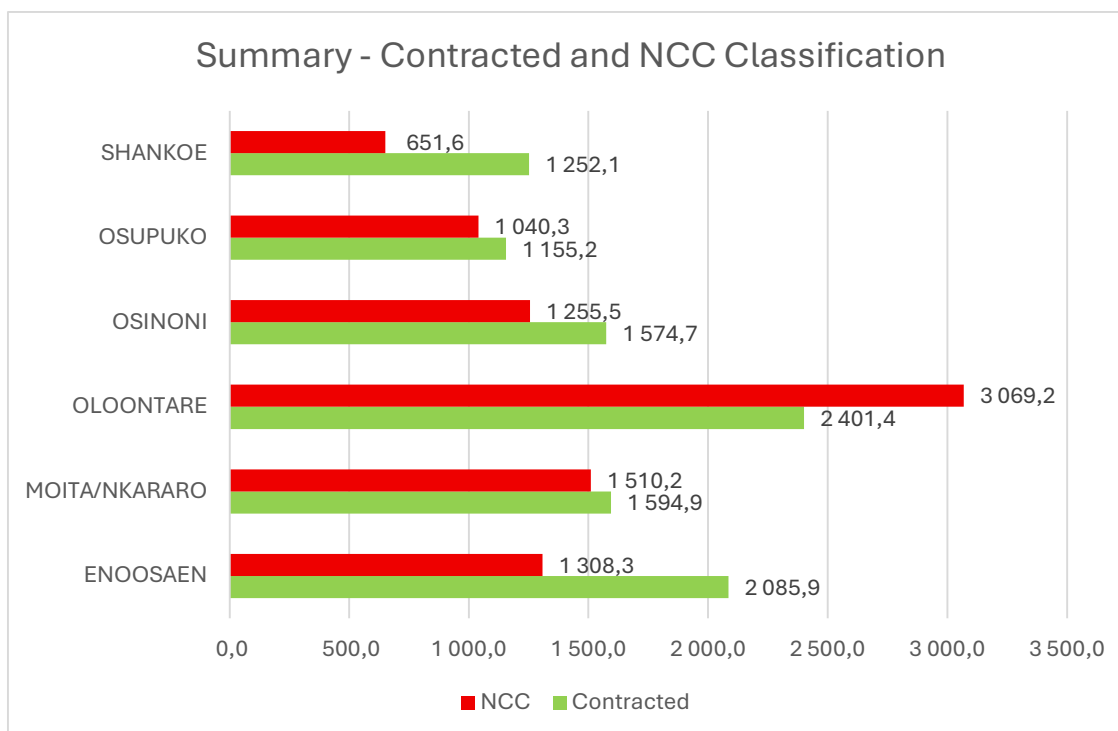


Figure 3.6 - Regional Summary of Contracted and Non-Contracted (NCC) Sugarcane Classification. Source: Created by author, based on TSCL contract status data and Random Forest (RF) classification results.

Table 3.6 summarizes the classified sugarcane area for contracted and NCC across the six TSCL regions.

Table 3.6 - Classified Contracted and Non-Contracted (NCC) Sugarcane Area by TSCL Regions.

Region	Contracted	NCC
ENOOSAEN	2 085,9	1 308,3
MOITA/NKARARO	1 594,9	1 510,2
OLOONTARE	2 401,4	3 069,2
OSINONI	1 574,7	1 255,5
OSUPUKO	1 155,2	1 040,3
SHANKOE	1 252,1	651,6
Grand Total	10 064,1	8 835,0

Source: Created by the author, based on TSCL Farmer Registry contract status data and Random Forest (RF) classification results.

The largest NCC extent was observed in Oloontare, where 3,069.2ha was classified as NCC. Enoosaen contained the largest contracted sugarcane area (2,085.9ha) while indicating the lower NCC extent (1,308.3ha).

NCC accounted for 8,835.0ha of the total mapped sugarcane area:

$\text{NCC Proportion} = (8,835.0 / 18,899.1) * 100$

NCC Proportion = 46.8%

The results showed the spatial distribution of contracted and NCC production across the study area. Regions such as Oloontare and Osinoni indicated higher NCC areas, suggesting significant levels of sugarcane occurring outside the contractual agreements. The ability to identify these areas highlights an important advantage of satellite-based monitoring, as it provides information across the entire study area rather than only where administrative records are available.

3.4 Crop Age Estimation Validation

The Sentinel-2 SCL method for estimating crop age was evaluated by comparing it with the Farmer Registry records. The Farmer Registry contained 12,992 contracted sugarcane fields. Of these, 12,987 fields were successfully linked to the SCL catalogue, while five Field IDs were not present in the Field ID database and were excluded from the crop age analysis. Of the 12,987 linked fields, 12,021 fields (92.6%) contained bare-soil changes from which a crop age estimate could be generated using the SCL time-series methodology (see Table 3.7). For each field, the crop age estimated from the SCL catalogue was compared with the crop age recorded in the Farmer Registry. A 3-month difference was accepted to allow for image frequency, cloud cover, harvesting schedules, and possible differences in field operations. A total of 10,488 fields fell within the 3-month tolerance threshold, while 1,533 fields exceeded the threshold and were flagged.

The operational agreement was calculated as:

$$\text{Agreement} = (10,488 / 12,021) * 100$$

$$\text{Agreement} = 87.2\%$$

Table 3.7 summarizes the validation results of the sugarcane crop age estimation using TSCL Farmer Registry data. The results include the number of contracted fields assessed, the proportion of fields with crop age estimates and the agreement between the estimated and registered crop ages.

Table 3.7 - Validation Results for Sugarcane Crop Age Estimation.

Metric	Value
Contracted Fields	12,987
Fields with Age Estimate	12,021
Coverage (%)	92.6
Fields within 3 Months	10,488
Fields Not within 3 Months ("Check")	1,533
Agreement (%)	87.2

Source: Created by the author, based on TSCL Farmer Registry data and Random Forest (RF) crop age estimation results.

Table 3.8 summarizes the agreement between the estimated crop ages and the TSCL SCL catalogue. Fields with estimated crop ages that matched or were within the 3 months threshold of the registered crop age were classified as "Agreement", while the remaining fields were flagged for further review.

Table 3.8 - Summary of Crop Age Estimation Agreement.

Status	Count	%
Agreement	10,488	87.2
Check	1,533	12.8

Source: Created by the author, based on TSCL Farmer Registry (SCL catalogue) data and Random Forest (RF) crop age estimation results.

The strong agreement between satellite-derived ages and Farmer Registry records illustrates that the Sentinel-2 SCL time-series methodology (see Table 3.8) can effectively identify planting and harvesting events and estimate crop age at a field level. Overall, the results demonstrate that the proposed crop age estimation methodology provides a reliable operational tool for estimating sugarcane age across both contracted and NCC production systems.

3.5 NCC Age Estimation Results

After validating the crop age estimation method, it was applied to the NCC fields identified by the RF classification. Of the total 8,835.0 ha classified as NCC, 7,987.9 ha were successfully assigned to an age category using the Sentinel-2 SCL time-series. This represents 90.4% of the total NCC area. The remaining 847.1 ha (9.6%) were categorised as “No Data” because insufficient temporal changes (Bare Soil) were available to reliably determine crop age (see Figure 3.8).

Figure 3.7 illustrates the spatial distribution of NCC fields by estimated crop age group, generated from the Sentinel-2 SCL time series. The map highlights the distribution of NCC fields across the TSCL study area according to four crop age classes.

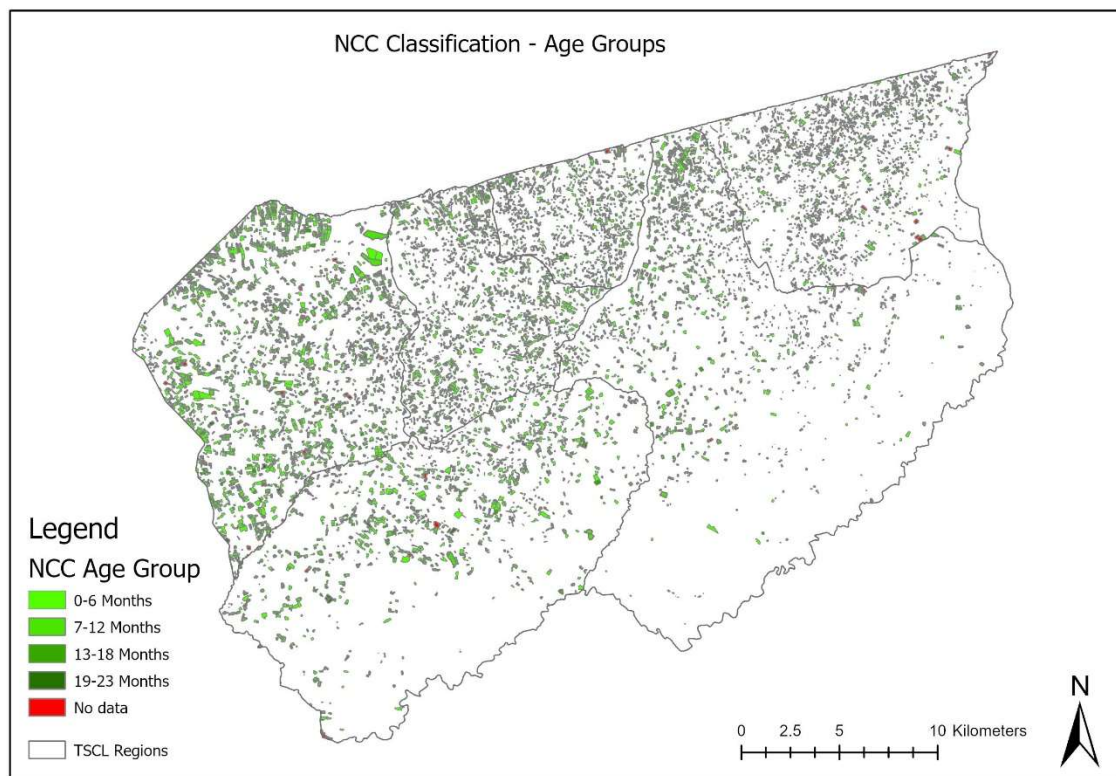


Figure 3.7 - Spatial Distribution of Non-Contracted Cane (NCC) Age Groups Derived from the Sentinel-2 SCL Time Series. Source: Created by the author, based on Sentinel-2 SCL time series, TSCL field boundary data and crop age estimation results.

Figure 3.8 illustrates the distribution of estimated crop ages for NCC, generated from the Sentinel-2 SCL time series. The chart summarizes the estimated sugarcane area within each monthly crop age category.

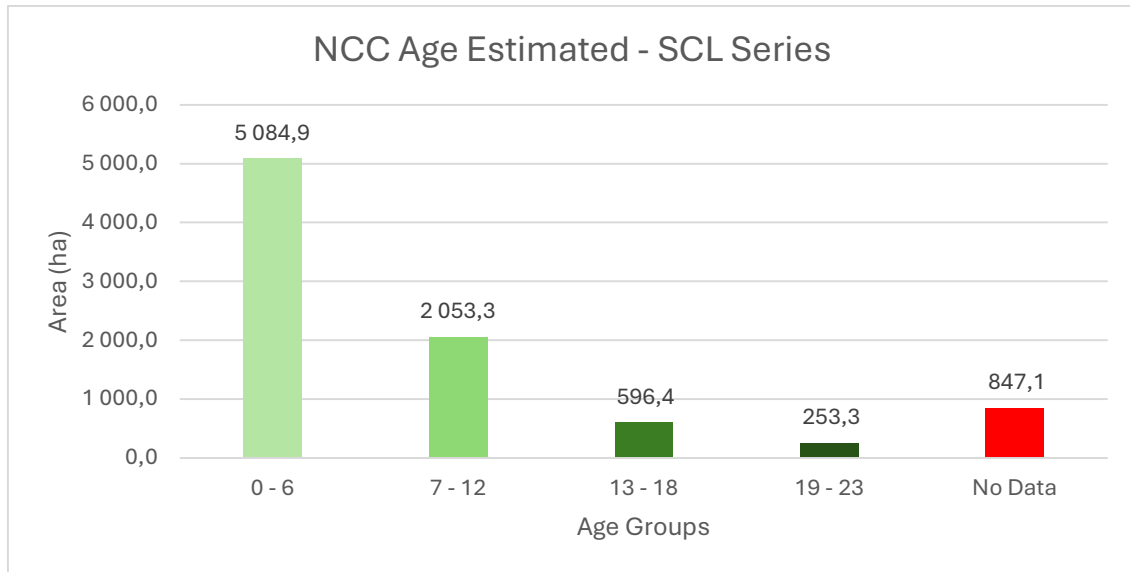


Figure 3.8 - Distribution of Estimated Non-Contracted Cane (NCC) Crop Ages Derived from the Sentinel-2 SCL Time Series. Source: Created by the author, based on Sentinel-2 SCL time series, TSCL field boundary data and crop age estimation results.

The largest age category corresponded to age class 5, accounting for 1,846.9ha of NCC. Additional large age groups were observed in age class 0 (1,525.0ha), age class 10 (789.3ha) and age class 11 (697.3ha). Smaller areas were distributed across the remaining age categories. The large number of fields assigned to an age class shows that the Sentinel-2 SCL time series can effectively estimate the age of sugarcane fields that are not included in the Farmer Registry.

3.6 Summary

The results show that the RF classification approach successfully identified sugarcane fields within the TSCL regions with an Overall Accuracy of 98.15% and a Kappa Coefficient of 0.963. Validation against Farmer Registry records confirmed an area agreement of 91.5%, while the Sentinel-2 SCL time-series methodology achieved an operational crop age agreement of 87.2%. The classification and assessment against Contracted Fields from TSCL Farmer Registry, enabled the identification of 8,835.0ha of NCC sugarcane and the estimation of crop age for 90.4% of the NCC area. These findings show the potential of integrating Sentinel-2 imagery, RF classification and SCL time-

series analysis for sugarcane monitoring and crop age estimation within the TSCL regions.

4: Discussion

The results demonstrate the potential of the developed remote sensing framework for sugarcane classification, crop age estimation and identification of NCC within the TSCL supply area. The framework combined multi-temporal Sentinel-2 imagery, vegetation indices, RF classification and Sentinel-2 SCL time-series information. The findings address the four research objectives and help answer the research question and hypothesis.

The first research objective focused on generating multi-temporal Sentinel-2 spectral bands and vegetation indices (NDVI, NDRE and NDMI) to assist with sugarcane classification and applying a RF classifier to differentiate sugarcane from other land-cover classes. The RF classification achieved an Overall Accuracy of 98.15% and a Kappa Coefficient of 0.963, indicating a high level of agreement between the classified results and the validation data. The use of multi-temporal Sentinel-2 imagery captured changes in sugarcane growth over time, which helped improve the classification. These changes over time helped distinguish sugarcane from other land-cover classes, including maize, grassland/open areas, and timber/tea/bush. The importance of multi-temporal information for differentiating crops and land-cover classes is consistent with the findings of Belgiu and Csillik (2018) and Vuolo et al. (2018), while the strong classification results also show that RF is suitable for analyzing complex remote sensing data (Belgiu & Drăguț, 2016). The inclusion of NDVI, NDMI and NDRE provided additional information on vegetation condition, moisture and red-edge characteristics that supported the differentiation of crop and land-cover types (Tucker, 1979; Gao, 1996; Delegido et al., 2011; Xue & Su, 2017). The classification results were then used together with the SCL-derived crop age estimates and for the identification of NCC.

The second research objective was to identify bare-soil events from the Sentinel-2 SCL time-series catalogue between 13 September 2023 and 30 August 2025 and use these events to estimate sugarcane crop age. The SCL time-series approach achieved an 87.2% agreement with the TSCL Farmer Registry crop-age records using the specified three-month threshold. This result indicates that historical bare-soil events identified from Sentinel-2 imagery can provide a useful basis for estimating sugarcane crop age where complete or current field records may not be available. The methodology identified the

most recent occurrence of SCL Class 5 (Bare Soil) within each field polygon as an indication of a planting or harvesting event. Crop age was then calculated as the time between the identified event and the reference date of 30 August 2025. The ability of multi-temporal satellite observations to capture temporal changes associated with crop development is consistent with previous research demonstrating the value of Sentinel-2 time-series information for crop and phenological monitoring (Belgiu & Csillik, 2018; Wang et al., 2019). However, the present study extends this temporal approach by using changes in the Sentinel-2 SCL, particularly bare-soil observations, as indicators of potential planting or harvesting events for crop age estimation.

The 12.8% of crop-age estimates that did not agree with the Farmer Registry within the three-month threshold demonstrate that uncertainty remains in the SCL-based approach. Several factors may explain these differences. Cloud contamination and cloud masking can reduce the number of usable Sentinel-2 observations and therefore resulting in a planting or harvesting event occurring between available image dates (Phiri et al., 2020). In addition, the SCL was not specifically designed for determining crop age and partially exposed soil or other surface conditions may affect the identification of Class 5 bare-soil events (Main-Knorn et al., 2017; ESA, 2024). Differences may also arise from inaccuracies or differences in the timing of information recorded in the Farmer Registry (Foody, 2002). Saying this, the 87.2% agreement within the three-month threshold indicates that the SCL time-series approach was able to reproduce the Farmer Registry crop-age information for most of the evaluated fields.

The third research objective assessed the accuracy of the sugarcane classification using Overall Accuracy, the Kappa Coefficient and area comparison with the TSCL Farmer Registry, while also checking whether the satellite-estimated crop ages agreed with the Farmer Registry within three months. The Overall Accuracy of 98.15% and Kappa Coefficient of 0.963 demonstrate strong classification performance based on the validation samples. In addition, comparison of the classified sugarcane area with the TSCL Farmer Registry resulted in an area agreement of 91.5%. Although this value is lower than the sample-based Overall Accuracy, the two measures evaluate different aspects of the classification and should therefore not be interpreted as equivalent accuracy measures (Stehman, 1997; Foody, 2002). The confusion matrix assesses agreement

between classified and validation samples, whereas the area comparison evaluates the correspondence between the satellite-derived sugarcane area and the area represented by the Farmer Registry.

The difference between the classification accuracy and the area agreement may be associated with several sources of uncertainty. These include differences between the date of satellite observation and the Farmer Registry, outdated field information, changes in land use, field-boundary differences, harvesting or planting events occurring around the period of image acquisition, mixed pixels along field boundaries and possible field digitisation errors (Foody, 2002; Congalton & Green, 2019). Such factors illustrate the difference between thematic classification accuracy and agreement with an operational agricultural database. Despite these potential sources of error or uncertainty, the combination of the 98.15% Overall Accuracy, Kappa Coefficient of 0.963 and 91.5% area agreement indicates a strong correspondence between the remotely sensed classification and the available reference information. The high classification performance is also consistent with previous studies demonstrating the effectiveness of Sentinel-2 imagery and RF classification for agricultural and land-cover mapping (Immitzer et al., 2016; Belgiu & Drăguț, 2016).

Crop-age validation was the second part of the third objective. The results showed that 87.2% of the satellite-estimated crop ages agreed with the TSCL Farmer Registry within the three-month threshold. This shows that the framework was able to classify sugarcane and estimate crop age with a high level of agreement with the TSCL reference data. These results also support the research hypothesis that the remote sensing framework can classify sugarcane and estimate crop age within three months of the Farmer Registry records. However, the identified limitations and uncertainties should still be considered when interpreting the results.

The fourth research objective was to identify NCC by comparing the classified sugarcane map from 30 August 2025 with the TSCL Farmer Registry dated 1 September 2025. The comparison identified 8,835.0ha of NCC, representing about 46.8% of the total classified sugarcane area. This is an important finding because it shows that a large amount of the sugarcane identified from Sentinel-2 imagery was not recorded as contracted sugarcane

in the Farmer Registry. This shows the value of remote sensing for identifying differences between sugarcane observed from satellite imagery and existing farmer records. Satellite monitoring can provide consistent information across large agricultural areas (Drusch et al., 2012). The identification of NCC could also be useful for TSCL by supporting field verification, grower recruitment, extension planning and cane supply forecasting. Regional differences were found, with Oloontare having the largest area of NCC. This suggests that field verification and grower engagement could be focused on areas where the largest differences occur between the classified sugarcane map and the Farmer Registry. The NCC results also help answer the research question. The RF classifier did not directly classify sugarcane as contracted or non-contracted. Instead, it first classified sugarcane from other land-cover classes. The classified sugarcane map was then compared with the TSCL Farmer Registry to identify NCC. Therefore, the high RF classification accuracy provides confidence in the identification of sugarcane, while field verification would still be needed to confirm whether the identified NCC fields are non-contracted.

Several limitations should be considered when interpreting the results. Although cloud masking was applied in GEE, cloud cover reduced the number of usable Sentinel-2 images and may have affected the detection of bare-soil events (Phiri et al., 2020). Sentinel-2 SCL was designed for scene classification and not specifically for estimating sugarcane crop age (Main-Knorn et al., 2017; ESA, 2024). Therefore, exposed soil or incorrect SCL classification may have affected the identification of planting and harvesting events. The study also used the TSCL Farmer Registry for area comparison and crop-age validation. Outdated records, inaccurate field boundaries, or differences in recorded planting and harvesting dates may therefore explain some of the differences between the satellite results and the Farmer Registry (Foody, 2002).

Another limitation is the identification of NCC. Ground validation data were not available for these fields, so the NCC results were based on the accuracy of the sugarcane classification and comparison with the Farmer Registry. Although the high classification accuracy provides confidence that these areas are sugarcane, it does not confirm that they are non-contracted. Future studies should include field verification to confirm the estimated 8,835.0ha of NCC. The framework should also be tested in other sugarcane-

growing regions of Kenya to determine whether similar classification and crop-age results can be achieved under different environmental and farming conditions.

Overall, the results answer the research question by showing that the remote sensing framework can be used to classify sugarcane, estimate crop age and identifying of NCC within the TSCL supply area. The high classification accuracy, 91.5% area agreement and 87.2% crop-age agreement within the three-month threshold support the research hypothesis. However, the lack of ground validation of NCC, possible inaccuracies in the Farmer Registry and limitations of the SCL bare-soil detection should be considered. Therefore, field verification is still important to confirm the results.

5: Conclusion

This study developed and assessed a remote sensing framework for sugarcane classification, crop age estimation and identification of NCC using Sentinel-2 time-series imagery within the TSCL regions in Transmara West Sub-County, Narok County, Kenya. The study shows the potential of Earth Observation data to support sugarcane monitoring and operational decision-making.

The study achieved the four research objectives. Multi-temporal Sentinel-2 spectral bands and vegetation indices (NDVI, NDRE and NDMI) were generated and used with a RF classifier to distinguish sugarcane from other land-cover classes. The classification achieved an Overall Accuracy of 98.15% and a Kappa Coefficient of 0.963. Comparison of the classified sugarcane area with the TSCL Farmer Registry resulted in an area agreement of 91.5%.

The Sentinel-2 SCL time-series catalogue was used to identify bare-soil events associated with potential planting or harvesting events and to estimate sugarcane crop age. The crop-age estimates achieved an 87.25% agreement with the TSCL Farmer Registry within the three-month threshold. Crop age was also estimated for 90.4% of the classified NCC area.

A total of 18,899.1 ha of sugarcane was classified, of which 10,064.1 ha corresponded with contracted sugarcane in the TSCL Farmer Registry and 8,835.0 ha was identified as NCC. The NCC represented approximately 46.8% of the total classified sugarcane area. This shows the value of remote sensing for identifying sugarcane areas that may not be represented as contracted sugarcane within the Farmer Registry.

Overall, the results answer the research question by showing that the developed remote sensing framework can be used to classify sugarcane, estimate crop age and identify NCC within the TSCL regions. The Overall Accuracy of 98.15%, Kappa Coefficient of 0.963, 91.5% area agreement and 87.2% crop-age agreement within the three-month threshold provide support for the research hypothesis. The results therefore show that combining Sentinel-2 imagery, RF classification, SCL time-series analysis and GEE provides a useful approach for sugarcane monitoring and Farmer Registry comparison.

The main conclusions of the study are:

- Multi-temporal Sentinel-2 spectral bands and vegetation indices (NDVI, NDRE and NDMI) were successfully generated following cloud masking in GEE, providing the dataset required for sugarcane classification.
- The RF classifier successfully distinguished sugarcane from other land-cover classes, achieving an Overall Accuracy of 98.15% and a Kappa Coefficient of 0.963.
- The Sentinel-2 SCL time-series catalogue identified bare-soil events associated with potential planting or harvesting events and provided a method for estimating sugarcane crop age.
- The classified sugarcane area achieved a 91.5% area agreement with the TSCL Farmer Registry, while crop-age estimation achieved an 87.25% agreement within the three-month threshold.
- Comparison of the classified sugarcane map with the TSCL Farmer Registry identified 8,835.0 ha of NCC, representing approximately 46.8% of the total classified sugarcane area. Crop age was estimated for 90.4% of the NCC area.

Although the results were strong, several limitations should be considered. Cloud cover reduced the availability of usable Sentinel-2 imagery and may have affected the identification of bare-soil events. The Sentinel-2 SCL was not specifically developed for crop-age estimation, which may introduce uncertainty when identifying planting or harvesting events. The study also relied on the TSCL Farmer Registry for area comparison and crop-age validation, and any outdated or inaccurate records could affect the comparison.

Independent ground-validation data was not available for the classified NCC fields. Therefore, the identified 8,835.0ha should require field verification to confirm the contractual status of these fields.

Future research should include field validation of the classified NCC and further testing of the SCL-based crop-age estimation method. The framework should also be tested in other sugarcane producing regions of Kenya to determine whether similar classification and crop-age results can be achieved under different conditions. Further development of the framework

could support sugarcane monitoring, Farmer Registry verification, grower recruitment, harvest planning and cane supply forecasting.

REFERENCES:

- Achanta, R. and Süsstrunk, S. (2017) ‘Superpixels and polygons using Simple Non-Iterative Clustering’, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4651–4660. <https://doi.org/10.1109/CVPR.2017.520>
- Ambasa, D.M. and Calistus, L.A. (2025) ‘Economic impacts of the declining sugar industry on livelihoods at the household level in the context of Western Kenya’, East African Journal of Arts and Social Sciences, 8(2), pp. 78–97. <https://doi.org/10.37284/eajass.8.2.2986>
- Atzberger, C. and Eilers, P.H.C. (2011) ‘Evaluating the effectiveness of smoothing algorithms in the absence of ground reference measurements’, International Journal of Remote Sensing, 32(13), pp. 3689–3709. <https://doi.org/10.1080/01431161003762405>
- 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, pp. 509–524. <https://doi.org/10.1016/j.rse.2017.10.005>
- Belgiu, M. and Drăguț, L. (2016) ‘Random Forest in remote sensing: A review of applications and future directions’, ISPRS Journal of Photogrammetry and Remote Sensing, 114, pp. 24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>
- Blaschke, T. (2010) ‘Object based image analysis for remote sensing’, ISPRS Journal of Photogrammetry and Remote Sensing, 65(1), pp. 2–16. <https://doi.org/10.1016/j.isprsjprs.2009.06.004>
- Breiman, L. (2001) ‘Random forests’, Machine Learning, 45(1), pp. 5–32. <https://doi.org/10.1023/A:1010933404324>
- 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 Sentinel-3’, International Journal of Applied Earth Observation and Geoinformation, 23, pp. 344–351. <https://doi.org/10.1016/j.jag.2012.10.008>

Cohen, J. (1960) ‘A coefficient of agreement for nominal scales’, *Educational and Psychological Measurement*, 20(1), pp. 37–46. <https://doi.org/10.1177/001316446002000104>

Congalton, R.G. (1991) ‘A review of assessing the accuracy of classifications of remotely sensed data’, *Remote Sensing of Environment*, 37(1), pp. 35–46. [https://doi.org/10.1016/0034-4257\(91\)90048-B](https://doi.org/10.1016/0034-4257(91)90048-B)

Congalton, R.G. and Green, K. (2019) *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. 3rd edn. Boca Raton, FL: CRC Press. <https://doi.org/10.1201/9780429052729>

Delegido, J., Verrelst, J., Alonso, L. and Moreno, J. (2011) ‘Evaluation of Sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content’, *Sensors*, 11(7), pp. 7063–7081. <https://doi.org/10.3390/s110707063>

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F. and Bargellini, P. (2012) ‘Sentinel-2: ESA’s optical high-resolution mission for GMES operational services’, *Remote Sensing of Environment*, 120, pp. 25–36. <https://doi.org/10.1016/j.rse.2011.11.026>

Eisfelder, C., Klein, I., Bekkum, E., Debela, H. and Kuenzer, C. (2024) ‘Cropland and crop type classification with Sentinel-1 and Sentinel-2 time series using Google Earth Engine for agricultural monitoring in Ethiopia’, *Remote Sensing*, 16(5), Article 866. <https://doi.org/10.3390/rs16050866>

Esri (2024) *ArcGIS Pro*. Redlands, CA: Environmental Systems Research Institute.

European Space Agency (ESA), 2024. *Sentinel-2 User Handbook*. Paris: European Space Agency. https://sentinels.copernicus.eu/documents/247904/685211/Sentinel-2_User_Handbook (Accessed: 17 April 2026)

Food and Agriculture Organization of the United Nations (FAO) (2015) *World Reference Base for Soil Resources 2014, Update 2015*. Rome: FAO. Available at:

<https://openknowledge.fao.org/server/api/core/bitstreams/bcdecec7-f45f-4dc5-beb1-97022d29fab4/content> (Accessed: 12 May 2026)

Foody, G.M. (2002) ‘Status of land cover classification accuracy assessment’, *Remote Sensing of Environment*, 80(1), pp. 185–201. [https://doi.org/10.1016/S0034-4257\(01\)00295-4](https://doi.org/10.1016/S0034-4257(01)00295-4)

Gao, B.-C. (1996) ‘NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space’, *Remote Sensing of Environment*, 58(3), pp. 257–266. [https://doi.org/10.1016/s0034-4257\(96\)00067-3](https://doi.org/10.1016/s0034-4257(96)00067-3)

Gitelson, A.A. and Merzlyak, M.N. (1994) ‘Spectral reflectance changes associated with autumn senescence of *Aesculus hippocastanum* leaves’, *Journal of Plant Physiology*, 143(3), pp. 286–292. [https://doi.org/10.1016/S0176-1617\(11\)81633-0](https://doi.org/10.1016/S0176-1617(11)81633-0)

Gitonga, P. and Ferrese, J. (2025) *Kenya Sugar Annual Report 2025*. Nairobi: USDA Foreign Agricultural Service. Available at: <https://www.fas.usda.gov/data/gain/2026/04/kenya-sugar-annual> (Accessed: 10 January 2026)

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D. and Moore, R. (2017) ‘Google Earth Engine: Planetary-scale geospatial analysis for everyone’, *Remote Sensing of Environment*, 202, pp. 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>

Government of Kenya (2024) *Sugar Act No. 11 of 2024*. Nairobi: Government Printer. Available at: <https://faolex.fao.org/docs/pdf/ken231795.pdf> (Accessed: 10 January 2026)

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X. and Ferreira, L.G. (2002) ‘Overview of the radiometric and biophysical performance of the MODIS vegetation indices’, *Remote Sensing of Environment*, 83(1–2), pp. 195–213. [https://doi.org/10.1016/S0034-4257\(02\)00096-2](https://doi.org/10.1016/S0034-4257(02)00096-2)

Immitzer, M., Vuolo, F. and Atzberger, C. (2016) ‘First experience with Sentinel-2 data for crop and tree species classifications in central Europe’, *Remote Sensing*, 8(3), Article 166. <https://doi.org/10.3390/rs8030166>

Inman-Bamber, N.G. and Smith, D.M. (2005) 'Water relations in sugarcane and response to water deficits', *Field Crops Research*, 92(2–3), pp. 185–202. <https://doi.org/10.1016/j.fcr.2005.01.023>

Kenya Association of Manufacturers (KAM) (2021) *Revitalising Kenya's Sugar Industry*. Nairobi: Kenya Association of Manufacturers. Available at: <https://www.kam.co.ke/> (Accessed: 15 January 2026)

Kenya National Bureau of Statistics (KNBS) (2023) *Economic Survey 2023*. Nairobi: KNBS. Available at: <https://www.knbs.or.ke/wp-content/uploads/2023/09/2023-Economic-Survey.pdf> (Accessed: 15 January 2026).

Kenya National Bureau of Statistics (KNBS) (2024) *Economic Survey 2024*. Nairobi: KNBS. Available at: <https://www.knbs.or.ke/wp-content/uploads/2024/05/2024-Economic-Survey.pdf> (Accessed: 15 January 2026).

Kombo, J.B. and Ndiema, A.C. (2023) 'A review of the state of sugar cane crisis in Kenya', *International Journal of Agricultural Policy and Research*, 11(1), pp. 27–34. [doi:10.15739/IJAPR.23.003](https://doi.org/10.15739/IJAPR.23.003)

Landis, J.R. and Koch, G.G. (1977) 'The measurement of observer agreement for categorical data', *Biometrics*, 33(1), pp. 159–174. <https://doi.org/10.2307/2529310>

Lisson, S.N., Inman-Bamber, N.G., Robertson, M.J. and Keating, B.A. (2005) 'The historical and future contribution of crop physiology and modelling research to sugarcane production systems', *Field Crops Research*, 92(2–3), pp. 321–335. <https://doi.org/10.1016/j.fcr.2005.01.010>

Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U. and Gascon, F. (2017) 'Sen2Cor for Sentinel-2', *Proceedings of the Living Planet Symposium*, Prague, Czech Republic. <https://doi.org/10.1117/12.2278218>

Mati, B.M. and Thomas, M.K. (2019) 'Overview of sugar industry in Kenya and prospects for production at the coast', *Agricultural Sciences*, 10(11), pp. 1477–1485. <https://doi.org/10.4236/as.2019.1011108>

Mulianga, B., Bégué, A., Clouvel, P. and Todoroff, P. (2015) 'Mapping cropping practices of a sugarcane-based cropping system in Kenya using remote sensing', *Remote Sensing*, 7(11), pp. 14428–14444. <https://doi.org/10.3390/rs71114428>

Nakalembe, C., Becker-Reshef, I., Bonifacio, R., Hu, G., Justice, C., Mwangi, K., Rembold, F., Shukla, S., Urbano, F., Whitcraft, A.K., Li, Y., Zappacosta, M., Jarvis, I. and Sanchez, A. (2021) 'A review of satellite-based global agricultural monitoring systems available for Africa', *Global Food Security*, 29, Article 100543. <https://doi.org/10.1016/j.gfs.2021.100543>

Phiri, D., Simwanda, M., Salekin, S., Nyirenda, V.R., Murayama, Y. and Ranagalage, M. (2020) 'Sentinel-2 data for land cover/use mapping: A review', *Remote Sensing*, 12(14), Article 2291. <https://doi.org/10.3390/rs12142291>

Segarra, J., Buchaillet, M.L., Araus, J.L. and Kefauver, S.C. (2020) 'Remote sensing for precision agriculture: Sentinel-2 improved features and applications', *Agronomy*, 10(5), Article 641. <https://doi.org/10.3390/agronomy10050641>

Snyman, S.J. (2016) *Agronomic Assessment of Sugarcane Production Systems in Transmara, Kenya*. Pretoria: South African Sugarcane Research Institute.

Som-ard, J., Atzberger, C., Izquierdo-Verdiguier, E., Vuolo, F. and Immitzer, M. (2021) 'Remote sensing applications in sugarcane cultivation: A review', *Remote Sensing*, 13(20), Article 4040. <https://doi.org/10.3390/rs13204040>

Stehman, S.V. (1997) 'Selecting and interpreting measures of thematic classification accuracy', *Remote Sensing of Environment*, 62(1), pp. 77–89. [https://doi.org/10.1016/S0034-4257\(97\)00083-7](https://doi.org/10.1016/S0034-4257(97)00083-7)

Tiema, C., Ochung, A. and Omwoma, S. (2024) ‘Sugarcane production in Kenya’, *Asian Journal of Advanced Research and Reports*, 18(9), pp. 30–47.
<https://doi.org/10.9734/ajarr/2024/v18i9732>

Transmara Sugar Company Limited (TSCL) (2025) Internal operational reports and Farmer Registry records. Unpublished internal company records. Narok County, Kenya: Transmara Sugar Company Limited.

Tucker, C.J. (1979) ‘Red and photographic infrared linear combinations for monitoring vegetation’, *Remote Sensing of Environment*, 8(2), pp. 127–150.
[https://doi.org/10.1016/0034-4257\(79\)90013-0](https://doi.org/10.1016/0034-4257(79)90013-0)

Vuolo, F., Neuwirth, M., Immitzer, M., Atzberger, C. and Ng, W.-T. (2018) ‘How much does multi-temporal Sentinel-2 data improve crop type classification?’, *International Journal of Applied Earth Observation and Geoinformation*, 72, pp. 122–130.
<https://doi.org/10.1016/j.jag.2018.06.007>

Wang, M., Liu, Z., Baig, M.H.A., Wang, Y., Li, Y. and Chen, Y. (2019) ‘Mapping sugarcane in complex landscapes by integrating multi-temporal Sentinel-2 images and machine learning algorithms’, *Land Use Policy*, 88, Article 104190.
<https://doi.org/10.1016/j.landusepol.2019.104190>

Wang, J., Xiao, X., Liu, L., Wu, X., Qin, Y., Steiner, J.L. and Dong, J. (2020) ‘Mapping sugarcane plantation dynamics in Guangxi, China, by time series Sentinel-1, Sentinel-2 and Landsat images’, *Remote Sensing of Environment*, 247, Article 111951.
<https://doi.org/10.1016/j.rse.2020.111951>

Xue, J. and Su, B. (2017) ‘Significant remote sensing vegetation indices: A review of developments and applications’, *Journal of Sensors*, 2017, Article 1353691.
<https://doi.org/10.1155/2017/1353691>

Zheng, Y., Li, Z., Pan, B., Lin, S., Dong, J., Li, X. and Yuan, W. (2022) ‘Development of a phenology-based method for identifying sugarcane plantation areas in China using high-

resolution satellite datasets', Remote Sensing, 14(5), Article 1274.
<https://doi.org/10.3390/rs14051274>

APPENDIX

Appendix A: ArcGIS Pro Code

Appendix B: Google Earth Engine (GEE) Code – Classification

Appendix C: SCL Catalogue & Training Data

Appendix A-1:

Code ArcGIS Pro – Merge SCL Titles

Python (ArcPy)

The code presented in this appendix was originally generated with the assistance of Claude Code and was subsequently adapted, tested and refined throughout the research until the final operational version was achieved.

```
import os
import arcpy

arcpy.env.overwriteOutput = True

aprx = arcpy.mp.ArcGISProject("CURRENT")
m = aprx.activeMap

scl_mxd = None
scl_myd = None

for lyr in m.listLayers():
    if lyr.name == "T36MXD_20240410T074609_SCL_20m.jp2":
        scl_mxd = lyr.dataSource
    elif lyr.name == "T36MYD_20240410T074609_SCL_20m.jp2":
        scl_myd = lyr.dataSource

if not scl_mxd or not scl_myd:
    raise Exception("One or both raster layers were not found in the active map.")

output_dir = r"C:\Users\gareth.jones\Documents\ARCGIS"
os.makedirs(output_dir, exist_ok=True)

output_name = "merged_temp.tif"

arcpy.management.MosaicToNewRaster(
    input_rasters=[scl_mxd, scl_myd],
    output_location=output_dir,
    raster_dataset_name_with_extension=output_name,
    coordinate_system_for_the_raster="",
    pixel_type="8_BIT_UNSIGNED",
    cellsize="20",
    number_of_bands=1,
    mosaic_method="LAST",
    mosaic_colormap_mode="FIRST"
)
```

Appendix A-2:

Code ArcGIS Pro – Majority Count Point Class within Polygon

Python (ArcPy)

The code presented in this appendix was originally generated with the assistance of Claude Code and was subsequently adapted, tested and refined throughout the research until the final operational version was achieved.

```
import arcpy
import os
from collections import Counter, defaultdict

arcpy.env.overwriteOutput = True

aprx = arcpy.mp.ArcGISProject("CURRENT")
m = aprx.activeMap

polygon_layer = None
point_layer = None

for lyr in m.listLayers():
    if lyr.name == "Narok_Fields":
        polygon_layer = lyr
    elif lyr.name == "SCL_Points":
        point_layer = lyr

if polygon_layer is None:
    raise Exception("Layer 'Narok_Fields' not found in the active map.")

if point_layer is None:
    raise Exception("Layer 'SCL_Points' not found in the active map.")

polygon_fc = polygon_layer.dataSource
point_fc = point_layer.dataSource

out_folder = r"D:\Sentinel_TSCL\Thesis\SCL\Fields"
out_fc = os.path.join(out_folder, "Narok_Fields_SCL.shp")

if arcpy.Exists(out_fc):
    arcpy.management.Delete(out_fc)

arcpy.management.CopyFeatures(polygon_fc, out_fc)

existing_fields = [f.name for f in arcpy.ListFields(out_fc)]

if "Majority_V" not in existing_fields:
    arcpy.management.AddField(out_fc, "Majority_V", "LONG")

sj_fc = os.path.join(out_folder, "temp_poly_point_join.shp")

if arcpy.Exists(sj_fc):
    arcpy.management.Delete(sj_fc)

arcpy.analysis.SpatialJoin(
    target_features=out_fc,
    join_features=point_fc,
```

```

    out_feature_class=sj_fc,
    join_operation="JOIN_ONE_TO_MANY",
    join_type="KEEP_ALL",
    match_option="INTERSECT"
)

sj_fields = [f.name for f in arcpy.ListFields(sj_fc)]

if "TARGET_FID" not in sj_fields:
    raise Exception("Field 'TARGET_FID' not found in spatial join output.")

if "grid_code" not in sj_fields:
    raise Exception("Field 'grid_code' not found in spatial join output.")

oid_field = arcpy.Describe(out_fc).OIDFieldName

poly_value_counts = defaultdict(list)

with arcpy.da.SearchCursor(sj_fc, ["TARGET_FID", "grid_code"]) as cursor:
    for target_fid, grid_code in cursor:
        if target_fid is not None and grid_code is not None:
            poly_value_counts[target_fid].append(grid_code)

majority_dict = {}

for poly_id, values in poly_value_counts.items():
    if values:
        majority_dict[poly_id] = Counter(values).most_common(1)[0][0]

with arcpy.da.UpdateCursor(out_fc, [oid_field, "Majority_V"]) as cursor:
    for row in cursor:
        row[1] = majority_dict.get(row[0], None)
        cursor.updateRow(row)

m.addDataFromPath(out_fc)

```

Appendix B:

Code Google Earth Engine (GEE) – Classification

JavaScript (Google Earth Engine JavaScript API)

The code presented in this appendix was originally generated with the assistance of Claude Code and was subsequently adapted, tested and refined throughout the research until the final operational version was achieved.

```
var training = ee.FeatureCollection("projects/ee-gjohnjones/assets/Training_Data");
var aoi = ee.FeatureCollection("projects/ee-gjohnjones/assets/TSCL_Narok");
```

```
Map.centerObject(aoi, 10);
Map.addLayer(training, {color: 'red'}, "Training Data");
Map.addLayer(aoi, {color: 'black'}, "TSCL Narok");
```

```
var targetDates = [
  '2024-12-26',
  '2025-01-05',
  '2025-02-04',
  '2025-03-06',
  '2025-04-15',
  '2025-05-02',
  '2025-06-11',
  '2025-07-31',
  '2025-08-30'
];
```

```
function maskS2Clouds(image) {
  var scl = image.select('SCL');

  var mask = scl.neq(3)
    .and(scl.neq(8))
    .and(scl.neq(9))
    .and(scl.neq(10))
    .and(scl.neq(11));

  return image.updateMask(mask);
}
```

```
function getDateComposite(targetDate) {
  var date = ee.Date(targetDate);
  var start = date.advance(-3, 'day');
  var end = date.advance(3, 'day');

  var s2 = ee.ImageCollection("COPERNICUS/S2_SR_HARMONIZED")
    .filterBounds(aoi)
    .filterDate(start, end)
    .map(maskS2Clouds)
    .median()
    .clip(aoi);

  var bands = ['B4', 'B5', 'B8', 'B8A', 'B11', 'B12'];
  var s2Image = s2.select(bands).divide(10000);

  var ndvi = s2Image.normalizedDifference(['B8', 'B4']).rename('NDVI');
  var ndre = s2Image.normalizedDifference(['B8A', 'B5']).rename('NDRE');
```

```

var ndmi = s2Image.normalizedDifference(['B8', 'B11']).rename('NDMI');

var composite = s2Image.addBands([ndvi, ndre, ndmi]);

return composite.rename([
  'B4_' + targetDate,
  'B5_' + targetDate,
  'B8_' + targetDate,
  'B8A_' + targetDate,
  'B11_' + targetDate,
  'B12_' + targetDate,
  'NDVI_' + targetDate,
  'NDRE_' + targetDate,
  'NDMI_' + targetDate
]);
}

var stackedImage = ee.ImageCollection(
  targetDates.map(function(date) {
    return getDateComposite(date);
  })
).toBands().clip(aoi);

var cleanBandNames = stackedImage.bandNames().map(function(name) {
  name = ee.String(name);
  return name.replace('[0-9]+_', '');
});

stackedImage = stackedImage.rename(cleanBandNames);

var baseBandNames = [
  'B4', 'B5', 'B8', 'B8A', 'B11', 'B12',
  'NDVI', 'NDRE', 'NDMI'
];

function createMedianBand(baseName) {
  baseName = ee.String(baseName);

  var matchingBands = stackedImage.bandNames().filter(
    ee.Filter.stringContains('item', baseName.cat('_'))
  );

  var medianBand = stackedImage
    .select(matchingBands)
    .reduce(ee.Reducer.median())
    .rename(baseName.cat('_median'));

  return medianBand;
}

var medianFillImage = ee.ImageCollection(
  baseBandNames.map(function(baseName) {
    return createMedianBand(baseName);
  })
).toBands();

var cleanMedianNames = medianFillImage.bandNames().map(function(name) {
  name = ee.String(name);
  return name.replace('[0-9]+_', '');
});

```

```

medianFillImage = medianFillImage.rename(cleanMedianNames);

var filledBands = stackedImage.bandNames().map(function(bandName) {
  bandName = ee.String(bandName);

  var baseName = bandName.split('_').get(0);
  var medianName = ee.String(baseName).cat('_median');

  var originalBand = stackedImage.select([bandName]);
  var fillBand = medianFillImage.select([medianName]).rename([bandName]);

  return originalBand.unmask(fillBand);
});

var stackedImageFilled = ee.ImageCollection(filledBands).toBands();

var finalBandNames = stackedImageFilled.bandNames().map(function(name) {
  name = ee.String(name);
  return name.replace("[0-9]+_", "");
});

stackedImageFilled = stackedImageFilled
  .rename(finalBandNames)
  .clip(aoi);

var seeds = ee.Algorithms.Image.Segmentation.seedGrid(30);

var segmented = ee.Algorithms.Image.Segmentation.SNIC({
  image: stackedImageFilled,
  size: 60,
  compactness: 2,
  connectivity: 8,
  seeds: seeds
}).select(['clusters']);

Map.addLayer(segmented.randomVisualizer(), {}, 'Segmented Image', false);

var trainingSamples = stackedImageFilled.sampleRegions({
  collection: training,
  properties: ['TYPE'],
  scale: 10,
  geometries: true
});

var classifier = ee.Classifier.smileRandomForest(50).train({
  features: trainingSamples,
  classProperty: 'TYPE',
  inputProperties: stackedImageFilled.bandNames()
});

var classified = stackedImageFilled.classify(classifier).clip(aoi);

var visParams = {
  min: 1,
  max: 2,
  palette: ['green', 'brown']
};

Map.addLayer(

```

```

    classified,
    visParams,
    'Classified: Sugarcane = 1, Other = 2',
    false
  );

  var sugarcaneOnly = classified.eq(1).selfMask();

  Map.addLayer(
    sugarcaneOnly,
    {palette: ['green']},
    'Sugarcane Only',
    true
  );

  var sugarcaneAreaHa = ee.Image.pixelArea()
    .divide(10000)
    .updateMask(sugarcaneOnly);

  var totalSugarcaneArea = sugarcaneAreaHa.reduceRegion({
    reducer: ee.Reducer.sum(),
    geometry: aoi.geometry(),
    scale: 10,
    maxPixels: 1e13
  });

  Export.image.toDrive({
    image: classified,
    description: 'Sugarcane_Classification_TSCL_Narok_MedianGapFill',
    folder: 'GEE_Exports',
    fileFormat: 'GeoTIFF',
    scale: 10,
    region: aoi.geometry(),
    maxPixels: 1e13
  });

  Export.image.toDrive({
    image: sugarcaneOnly,
    description: 'Sugarcane_Only_TSCL_Narok_MedianGapFill',
    folder: 'GEE_Exports',
    fileFormat: 'GeoTIFF',
    scale: 10,
    region: aoi.geometry(),
    maxPixels: 1e13
  });

```

Appendix C:

Scene Classification Layer Time-Series Catalogue TSCL Fields & Training Data

<https://drive.google.com/drive/folders/1lhcd0gt14EdFXOioToLSmbctoFRaNap5?usp=sharing>

Contact – gjohnjones@gmail.com