SUGARCANE MAPPING AND HARVEST MONITORING USING ML AND GOOGLE EARTH ENGINE

By

Ala Ud Din Awan
(2020-NUST-MS-RS&GIS-330615)

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Remote Sensing & GIS.

Institute of Geographical Information Systems
School of Civil and Environmental Engineering
National University of Sciences and Technology
Islamabad, Pakistan

August, 2024

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS/MPhil thesis written by **Ala Ud Din Awan (Registration No. MSRSGIS 00000330615), of Session 2020 (Institute of Geographical Information Systems)** has been vetted by undersigned, found complete in all respects as per NUST Statutes/Regulation, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of MS/MPhil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis.

Signature: ______________

Name of ~~Supervisor: Dr~~ Muhammad Ali Tahir

Date: 9-7-24

Signature (HOD): ______________

Date: 9-7-24

DR. MUHAMMAD ALI TAHIR
Head of Department
Associate Professor
SCEE (IGIS)
NUST. H-12 Islamabad

Signature (Principal & Dean SCEE): ______________

Date: 24 JUL 2024

PROF DR MUHAMMAD IRFAN
Principal & Dean
SCEE, NUST

DEDICATION

I dedicate this thesis to the loving memory of my parents, whose guidance, optimism, and love instilled in me the values of hard work, resilience, and compassion. Their legacy inspires me to strive for excellence and make a positive impact. I am deeply grateful to my family, whose support, patience, and encouragement have driven my academic pursuits. Your presence in my life has been a constant source of comfort, motivation, and strength. I also extend my heartfelt appreciation to my dear friend, whose friendship has been a beacon of hope and joy throughout my academic journey. Your encouragement and belief in me have made a lasting impact.

ACADEMIC THESIS: DECLARATION OF AUTHORSHIP

I, **Ala Ud Din Awan,** declare that this thesis and the work presented in it are my own and have been generated by me as the result of my original research.

"Sugarcane Mapping and Harvest Monitoring using ML and Google Earth Engine"

I confirm that:

1. This work was done wholly by me in candidature for an M.S. research degree at the National University of Sciences and Technology, Islamabad.
2. Wherever I have consulted the published work of others, it has been attributed.
3. Wherever I have quoted from the work of others, the source has been always cited.
4. I have acknowledged all main sources of help.
5. Where the work of the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.
6. None of this work has been published before submission. This work is not plagiarized under the H.E.C. plagiarism policy.

Signed: 

Date: 2024-08-02

ACKNOWLEDGEMENTS

In the name of Allah, the Most Gracious, the Most Merciful. I am grateful for the blessings and mercy of Allah, which enabled me to complete this thesis. I would like to express my sincere gratitude to my thesis supervisor, Dr. Ali Tahir, and my thesis committee, Dr. Muhammad Azmat and Mr. Muhammad Hasan Mustafa, for their guidance, support, and valuable feedback. I am also thankful to my family, friends, and colleagues for their love, motivation, and encouragement throughout this journey. Their collective efforts and support have been instrumental in the completion of this thesis, and I am honored to acknowledge their contributions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Explanation	**Abbreviation**
Khyber Pakhtun Khuwa	KPK
Support Vector Machine	SVM
Random Forest	RF
Classification and Regression Tree	CART
k-Nearest Neighbor	KNN
Machine Learning	ML
Synthetic Aperture Radar	SAR
Modified Soil Adjusted Vegetation Index	MSAVI
Normalized Difference Vegetation Index	NDVI
Normalized Difference Senescent Vegetation Index	NDSVI
Normalized Difference Moisture Index	NDMI
Normalized Difference Turbidity Index	NDTI
Enhanced Vegetation Index	EVI
Green Chlorophyll Index	GCI
Near-infrared	NIR
Dera Ismail Khan	D.I Khan
Google Earth Engine	GEE
Near Real Time	NRT
Short-wave Infrared	SWIR
Gross Domestic Product	GDP

ABSTRACT

This study addresses the spreading disparity between global food demand and production, particularly focusing on optimizing agricultural processes in Dera Ismail Khan, Pakistan, with a specific emphasis on sugarcane. Leveraging ground survey data and Sentinel 2 satellite imagery, the research employs a two-step methodology, utilizing NDVI variations for sample selection and conducting a comprehensive field survey with the Kobo Toolbox. A total of 450 samples were collected, with 85% used for model training and 15% for validation. The methodology incorporates RF, Support Vector Machine (SVM), Classification and Regression Tree (CART), K-Nearest Neighbor (KNN) and K-Means clustering classification models, with SVM demonstrating the highest overall accuracy of 90% followed by Random Forest (RF) at 81.43%. The study identifies a total sugarcane area of 54,055 acres. A novel approach for sugarcane harvest monitoring is introduced, establishing a threshold NDVI value of 0.25 for identifying harvested areas, and providing real-time insights into cropping patterns. The study also addresses challenges posed by atmospheric conditions and advocates for integrating SAR data. The objectives include enhancing precision in sugarcane identification and mapping, improving growth and harvest tracking, and comparing classification models. The results showcase the potential of this SVM for sustainable crop monitoring, contributing to informed decision-making in agriculture and addressing the global food demand-production gap.

Chapter 1

INTRODUCTION

Agriculture plays a crucial role in the economies of least developing countries like Pakistan and is a fundamental part of people's livelihood. It is an important sector in which our modes of life and business innovativeness combine. It has manifold roles in the economy of any nation and these roles include food security, poverty reduction, industrial revolution, and economic growth (Gardner et al., 2001). Pakistan is a semi-industrialized economy with a well-integrated agricultural sector, contributing about 22.9% to Pakistan's Gross Domestic Product (GDP) and sustaining 37.4% of the labor force (*Economic Survey of Pakisatn*, 2022-23). Pakistan is cultivating about 23.9 million hectares out of its total 79.6 million hectares, constituting 30% of its land area. Pakistan has the world's largest adjoining irrigation system, and it has vast agricultural resources, with around 80% of its cultivated area being irrigated. Agriculture is negatively impacted by factors such as inadequate water (Afzal, 1996; Li et al., 2023), poor crop yields (Aslam, 2016), limited financing (Elahi et al., 2018), conventional farming practices (Ahmad et al., 2013; Talib et al., 2019), natural disasters (Farooq & Fatima, 2022; Rehman et al., 2016), and others (Badar et al., 2007; Chandio et al., 2016; Rehman et al., 2019).

1.1 Background

The gap between food demand and food production has become larger in recent years due to increasing global population (Godfray et al., 2010). In July-August 2022, Pakistan experienced a heavy monsoon spell, which damaged two major sub-sectors: agriculture and livestock. The entire damage in the agriculture industry is estimated to be Rs 800 billion (US$ 3.725 billion) (Azam & Shafique, 2017). This sector is producing in the stage of increasing return to scale, which means that the allocation of inputs in this sector is not optimal (Azam & Khan, 2010). In Khyber Pakhtun Khuwa (KPK) – Pakistan, majority of farmers own small size of landholding and having poor finical resources, 90% of farmers are illiterate and use traditional methods of cultivating sugarcane (Pervaiz et al., 2013).

The Economic Survey of Pakistan for the year 2022-23 provides valuable insights into the agricultural landscape of the country, highlighting the substantial contributions of essential crops to both agricultural value addition and the GDP. During this period, essential crops accounted for 18.23 percent of agricultural value addition and 4.18 percent of GDP, while other crops contributed 14.49 percent to agricultural value addition and 3.32 percent to GDP

(Economic Survey of Pakisatn, 2022-23). Among the pivotal crops in Pakistan are Cotton, Sugarcane, Rice, Maize, and Wheat, each making a distinct contribution to the nation's GDP and agricultural value addition. Specifically, Sugarcane, recognized as one of the world's most economically significant crops (Chen et al., 2021; Huang et al., 2018; Sumesh et al., 2021). In the fiscal year 2022-23, Sugarcane occupied 1,319 thousand hectares of land, marking a notable 4.7% increase from the previous year. The crop's output experienced a year-on-year growth of 2.8%, reaching an impressive 91.111 million tons (Wing, 2023).

1.2 Literature Review

Geographically, Sugarcane cultivation is concentrated in the provinces of Punjab, Sindh, and Khyber Pakhtunkhwa, underscoring its significance in these regions. However, despite its economic importance, most sugarcane farmers in Pakistan continue to rely on manual agronomic practices for crop monitoring and measurement. Various factors contribute to the persistence of manual practices among sugarcane farmers, including a lack of knowledge, limited availability of technological tools, high initial costs, poor awareness, and a general reluctance to adopt novel technologies. Researchers have actively engaged in enhancing sugarcane productivity, often employing human-based field monitoring and measurement techniques (Braithwaite et al., 2007; Cherry, 2001; He et al., 2021; Huang et al., 2019; Vennila et al., 2021; Verma et al., 2021; Wilson, 2021). These efforts highlight the gap between the potential benefits of technology and the current practices in the sugarcane farming sector.

In a pioneering study, Wang et al. (2019) harnessed the capabilities of Sentinel-2 imaging and crop phenology data to effectively map sugarcane cultivation in the challenging weather conditions of Longzhou county, China. Utilizing the Google Earth Engine (GEE) platform, the researchers employed a comprehensive approach that fused images captured at critical stages of sugarcane growth, including seedling, elongation, and harvest. This integration resulted in the generation of cloud-free remote sensing images, proving crucial for accurate land use mapping in subtropical areas prone to adverse weather conditions. The significance of fine-resolution Sentinel-2 data became evident in addressing the challenges posed by the region's specific weather conditions. The study underscored the pivotal role of advanced remote sensing technology in overcoming weather-related impediments, emphasizing the need for high-quality data for effective land use mapping. Moreover, the research proposed a streamlined method for sugarcane mapping, leveraging Machine Learning (ML) classifiers such as Polynomial-SVM, RBF-SVM, and RF. This innovative approach capitalizes on the capabilities of ML algorithms to discern and classify sugarcane fields, providing a more efficient and accurate alternative to traditional mapping methods. Notably, the application of various machine learning algorithms in remote sensing (RS) data analysis has gained widespread recognition. Decision trees (DT), SVM, Genetic Algorithm (GA), and Ensemble Learning (EL) have all demonstrated successful outcomes in the realm of sugarcane crop mapping, showcasing commendable accuracy levels (Canata et al., 2021; de Almeida et al., 2021; Jiang et al., 2019; Nihar et al., 2022; Singh et al., 2020).

In a recent study, Lonare et al. (2022) conducted a comprehensive evaluation of the efficacy of the cloud-based remote sensing platform, GEE, for the identification of sugarcane crops in the village of June Khed in India. Leveraging advanced technology, the research employed various vegetation indices, including NDVI, Enhanced Vegetation Index (EVI), and Land Surface Water Index (LSWI). These indices have been proven effective in prior studies for vegetation analysis (Chandrasekar et al., 2010; Chivasa et al., 2021; Imran et al., 2020; Júnior et al., 2020; Rahman et al., 2004; Rudorff et al., 2009; Xavier et al., 2006; Xin et al., 2020). In addition to vegetation indices, machine learning models, including SVM, RF, and Classification and Regression Trees (CART), were employed, utilizing both Sentinel-2 and Landsat-8 satellite data. Notably, SVM with Sentinel-2 data emerged as the most effective combination, achieving the highest accuracy at 78% and an F1-score of 0.80. The superior performance of Sentinel-2 data was attributed to its impressive spatial resolution of 10 meters and temporal resolution of 5 days, outperforming Landsat-8 in both aspects. The outcome of this research extends beyond the realm of academic exploration, as the sugarcane maps generated through these advanced technologies hold practical applications. These maps can serve as valuable Decision Support Tools, facilitating crucial functions such as acreage estimation, yield prediction, growth monitoring, and overall farm management. For local farmers in India and similar agricultural landscapes, these tools offer a means to enhance profitability by optimizing resource allocation and improving decision-making processes. This study contributes to the growing body of knowledge regarding the application of remote sensing and machine learning in agriculture, particularly in the context of sugarcane cultivation. The findings underscore the potential for advanced technologies to revolutionize farming practices, providing farmers with actionable insights and tools for sustainable and profitable agricultural management.

In a groundbreaking approach proposed by Zhu et al. (2019), a method for extracting sugarcane plantation areas was introduced, utilizing deep learning techniques and multi-temporal images from GF-2 and BJ-2 satellites over a span of four months. The methodology involved the extraction of non-vegetation areas followed by temporal processing of sugarcane areas using a Deep Convolution Neural Network (DCNN). The input to the DCNN included images from the sowing period, growing period, matured period, and additional relevant data. This innovative fusion of satellite imagery and deep learning technology demonstrated the potential for automating sugarcane mapping and health monitoring, providing a foundation for transformative advancements in precision agriculture. In alignment with this technological trajectory, the integration of advanced technologies such as satellite imagery, ML algorithms, and the GEE presents a promising avenue for revolutionizing the field. This research seeks to build upon these foundations by developing a comprehensive remote sensing methodology. The primary objectives include accurate identification, mapping, and growth tracking of sugarcane using ML algorithms on the GEE platform.

By harnessing the capabilities of these cutting-edge technologies, the research endeavors to deliver valuable insights for both farmers and policymakers. Furthermore, the proposed study aims to implement an innovative approach leveraging remote sensing to analyze temporal harvest patterns of sugarcane. This involves the application of machine learning algorithms in conjunction with remote sensing data to gain insights into the dynamics of sugarcane harvesting. These objectives are central to enhancing the efficiency and sustainability of sugarcane cultivation and management practices. In a related study, Mafuratidze et al. (2023) conducted a spectral analysis to assess the severity of hailstorm damage to sugarcane. The research used six spectral indices (GCI, NDVI, NDSVI, RECI, NDTI, and MSAVI2) to extract spectral differences for each index separately. This approach provides a quantitative means of evaluating the impact of environmental factors on sugarcane health, further emphasizing the potential for remote sensing techniques to inform decision-making in agricultural management. Together, these studies contribute to the overarching theme of leveraging advanced technologies and methodologies for enhancing the precision, efficiency, and sustainability of sugarcane cultivation. The integration of remote sensing, deep learning, and machine learning holds promise for transforming agricultural practices, offering practical tools for farmers, policymakers, and researchers alike.

This research endeavors to confront pivotal challenges in sugarcane cultivation through the development and implementation of an innovative remote sensing-based methodology. The primary objective is to significantly augment precision in sugarcane identification and mapping, aiming for higher accuracy and reliability by seamlessly integrating remote sensing and machine learning techniques. The study further seeks to automate and enhance the efficacy of growth tracking, leveraging advanced technologies to streamline and optimize the monitoring process.

1.3 Objectives

The objectives of the research are as follows:

a. **Enhance Precision in Sugarcane Identification and Mapping**: Develop a remote sensing-based method that leverages machine learning algorithms to improve the accuracy and reliability of sugarcane identification and mapping, providing a robust foundation for informed decision-making in agriculture.
b. **Improve Growth and Harvest Tracking**: Implement cutting-edge technologies to monitor the growth tracking process, utilizing remote sensing data and machine learning models in GEE. This objective aims to streamline monitoring procedures, providing real-time insights into the developmental and harvesting stages of sugarcane crops.

By achieving these objectives, the research endeavors to contribute to the broader goal of integrating cutting-edge technologies for the advancement of sustainable and informed agricultural practices in the sugarcane industry.

Chapter 2

MATERIALS AND METHODS

2.1 Study Area

The study area for this research is Dera Ismail Khan (D.I Khan), located in Khyber Pakhtunkhwa province of Pakistan is an agricultural district as shown in Figure 1 and has a total population of around 1.8 million (Statistics, 2023). It has a total geographical area of 0.73 million hectares, of which only 0.24 million hectares are cultivated (Khan et al., 2008). About one-third of the farmed area is irrigated, while the remaining two-thirds depends on rainfall and hill torrents for moisture. Major crops grown in the district include wheat, cotton, sugarcane, and rice. In the fiscal year 2020–21, the total cultivated area for sugarcane in D.I Khan encompassed 23,600 hectares, constituting 21.97% of the overall sugarcane cultivation area within the province. Concurrently, the production output reached 1.4 million tons, contributing significantly to the regional sugarcane production at 26.54% of the total (Bureau of Statistics 2022). DI Khan is an arid region with an estimated yearly rainfall of around 230–268 millimeters, which can change periodically and have different intensities (Iffat, 2012) while the average maximum and minimum temperatures are 45 °C (June) and 8 °C (January) respectively. River Indus is the main physical feature flowing southward in the eastern part of the study area and the major source of irrigation is the Chashma right bridge canal.

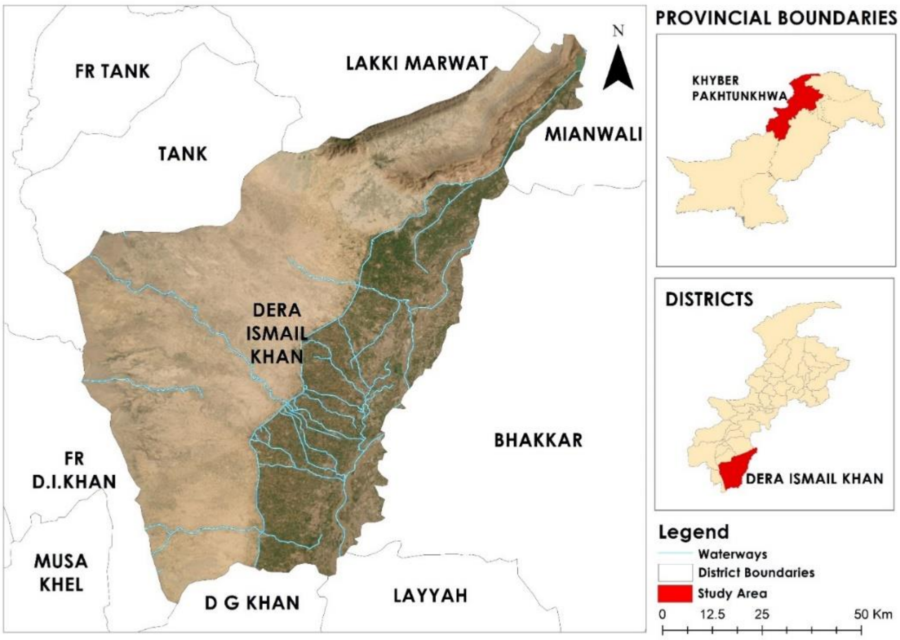

Figure 1: Regional context of the study area, outlining provincial and district boundaries, and highlighting the study area itself (red) and its water network (blue).

2.1.1 Datasets

A ground survey was conducted to gather comprehensive data of different crops in the area, resulting in the collection of over 450 ground samples of different crops. These samples encompass a wide range of vegetation classes, including Wheat, Sugarcane, Gram, Canola, and others. The distribution of the collected samples is as follows:

1. Sugarcane: 230 samples
2. Other Vegetation: 237 samples

Sugarcane crop is sown 2 times in a year, February and September (Khaliq et al., 2023). And harvesting starts by the end of October. To train the model for identification of all sugarcane, we utilized satellite data of Feb 2022 to October 2023. This is the duration in which sugarcane from 2021 September and 2022 February, both can be identified and are present. Sentinel 2 images, Level-2A Surface Reflectance was used for this purpose. GEE provides an efficient way to process remote sensing data and perform classification using Machine Learning Models. The collected satellite data is also processed on GEE where images from the same period are grouped together to create a mosaic. Each of these mosaics consists of 8 bands including Red, Green, Blue, NIR, NDVI, NDMI, SWIR, Red Edge 1 bands (Mafuratidze et al., 2023). The images are for the months of Feb, March, April, May, June and October. Each of the images was carefully selected to have less than 5% cloud cover over the area. Images for the months of July, August and September were not included due to the higher percentage of clouds. All these mosaics from each month are then stacked together to create one single image with 48 bands. After that various image processing techniques such as masking of cloud and shadows, extraction of a region of interest, resampling of bands from 20 m to 10 m spatial resolution, calculation of vegetation indices as well as extraction of images for further analysis were all performed in GEE.

Sugarcane crops are of two types i.e. planted and ratoon crop based on method of propagation. The planted sugarcane is sown as seed or seedling in April–May and harvested in Feb–March of the second year (12-month crop). The subsequent regenerated crop from the existing buds is known as ratoon sugarcane or stubble crop and it is harvested in November (9-month crop) (Singh et al., 2020). To cater this kind of sugarcane, satellite data for the months of October 2021, November 2021, December 2021 and January 2021 were later added in the mosaic as well. Making a total of 80 bands in 1 single image. This image was then subsequently used for classification. The primary objective of this data collection effort was to train a Machine Learning model specifically designed for the identification of sugarcane. By including samples from various vegetation classes, the model can learn to distinguish between different types of vegetation and accurately identify sugarcane instances.

Detailed list of Datasets is mentioned in Table 1.

Table 1: Compilation of Primary and Secondary data including satellite imagery, administrative boundaries, field survey samples, and global land cover data.

Sr.	Datasets	Data Source
1	Harmonized Sentinel-2 MSI: Multispectral Instrument, Level-2A	European Union/ESA (GEE)
2	Administrative boundaries	Global Administrative Areas, DivaGIS
3	Samples Collected	Field Survey
4	Dynamic World V1	GEE (World Resources Institute) (Brown et al., 2022)

2.2 Flow Diagram

The overall workflow of this study is divided into 2 steps. First is the sample selection and second is extraction of sugarcane crop mask in the study area and analyzing the harvest patterns. Sample Fields were pre-identified using NDVI patterns and only those fields were surveyed. First, NDVI stack from satellite imagery for the sugarcane crop cycle duration was calculated. Final output is difference raster that is the result of minimum – maximum NDVI raster data. Areas with higher NDVI differences were then surveyed. Detailed steps are shown in Figure 2. Second Step is the Machine Learning Classifier training, Execution and validation for sugarcane crop identification. After the sugarcane fields are identified harvest monitoring patterns are also studied for the balance and harvested sugarcane for the cropping season of 2022-2023.

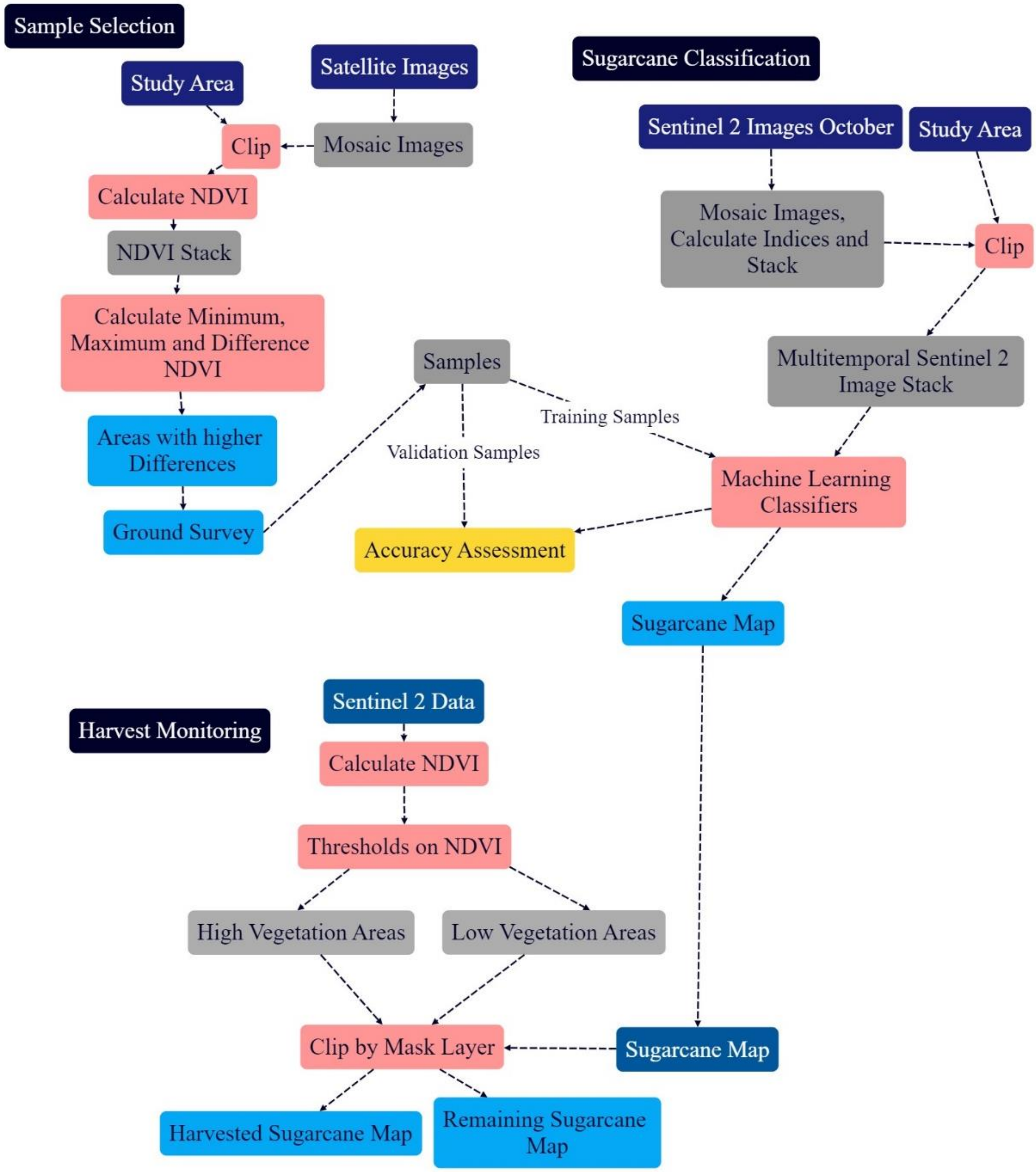

Figure 2: Flow Diagram for Sugarcane Crop Identification and Monitoring: Illustrates methodology using NDVI values for sample selection, ML model trained with temporal images and field data, and monthly NDVI for harvest monitoring, enhancing precision in agricultural management.

2.3 Ground Survey

NDVI stack of October 2021 – October 2022 was created on GEE using Harmonized Sentinel-2 MSI: Multispectral Instrument, Level-2A (Mehmood et al., 2023). These images were carefully selected to be cloud-free and cover the entire study area. Minimum and maximum rasters were generated from this NDVI stack using .max() and . min() filters in GEE. The maximum NDVI raster shows areas with higher NDVI values meaning higher vegetation and the minimum raster shows areas with lower NDVI values. A difference raster was generated using these minimum and maximum rasters as shown in Figure 3. This difference raster is a representation of NDVI changes during the complete period of the sugarcane crop cycle. Areas with higher difference values of NDVI are those where some sort of vegetation was present during this complete cycle, and it was sown and harvested. These are the areas that represent active vegetation. And are subsequently used for surveying. This helps us to pre-identify areas for the survey and make sure that they are well spread in the overall area. Only these vegetation areas were selected for ground survey to avoid data redundancy, time management, speed and efficiency.

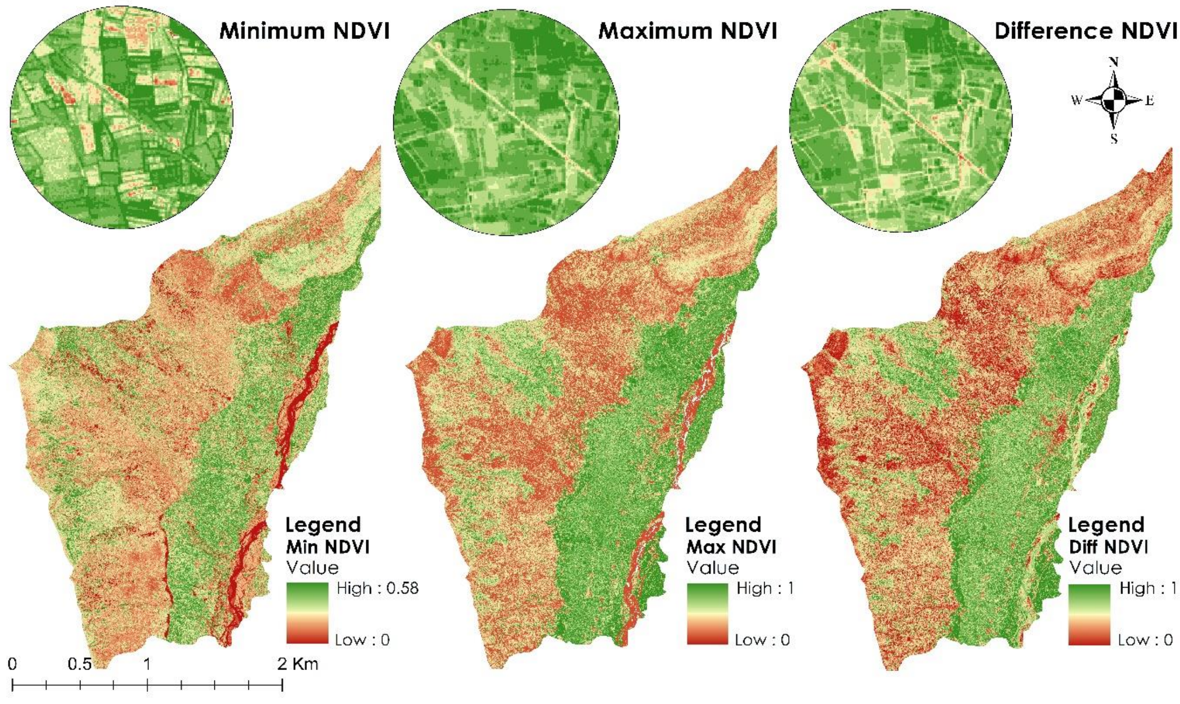

Figure 3: Exploring Dynamic Vegetation Patterns: Minimum, Maximum, and Differential NDVI Maps highlight active vegetation variations, guiding targeted field surveys for ML model training.

The study area does not have a stable internet connection, to tackle with this situation kobo Toolbox was used (Lakshminarasimhappa, 2022). Kobo Toolbox is an open-source tool developed by USAID, Harvard Humanitarian Initiative, and Brigham and Women's Hospital for mobile data collection. It offers accuracy, speed, data quality, analysis, and cost-effectiveness, allowing researchers to design customized questionnaires and distribute surveys via the Kobo Collect Android app (Poloju et al., 2022). Kobo Toolbox comes with a mobile application and a web-based dashboard. Data collection form was created on the dashboard with fields including Current crop in the field, Sugarcane type (in case of sugarcane), current stage, cropping history of last 2 years (if farmer is on the field), crop sowing date, Field location (Point), Field boundary (Polygon), image of the field/crop and any additional information available. The ground survey activity was done using the kobo Toolbox Android application, took 5 days in the field and collected more than 450 samples as shown in Figure 4. These samples are well distributed in the study area and are exactly on the fields pre identified. Major crops samples collected were Sugarcane, Wheat, Gram, Maize and Canola. Most of the sugarcane samples were of type Ratoon. Another phenomenon of mix cropping was also identified where people are sowing 2 crops together in a field. For example, sugarcane and wheat or canola and sugarcane are both sown in the same field once. And they keep on growing together. Once the wheat or canola is mature it is harvested, and the underlying Sugarcane grows up.

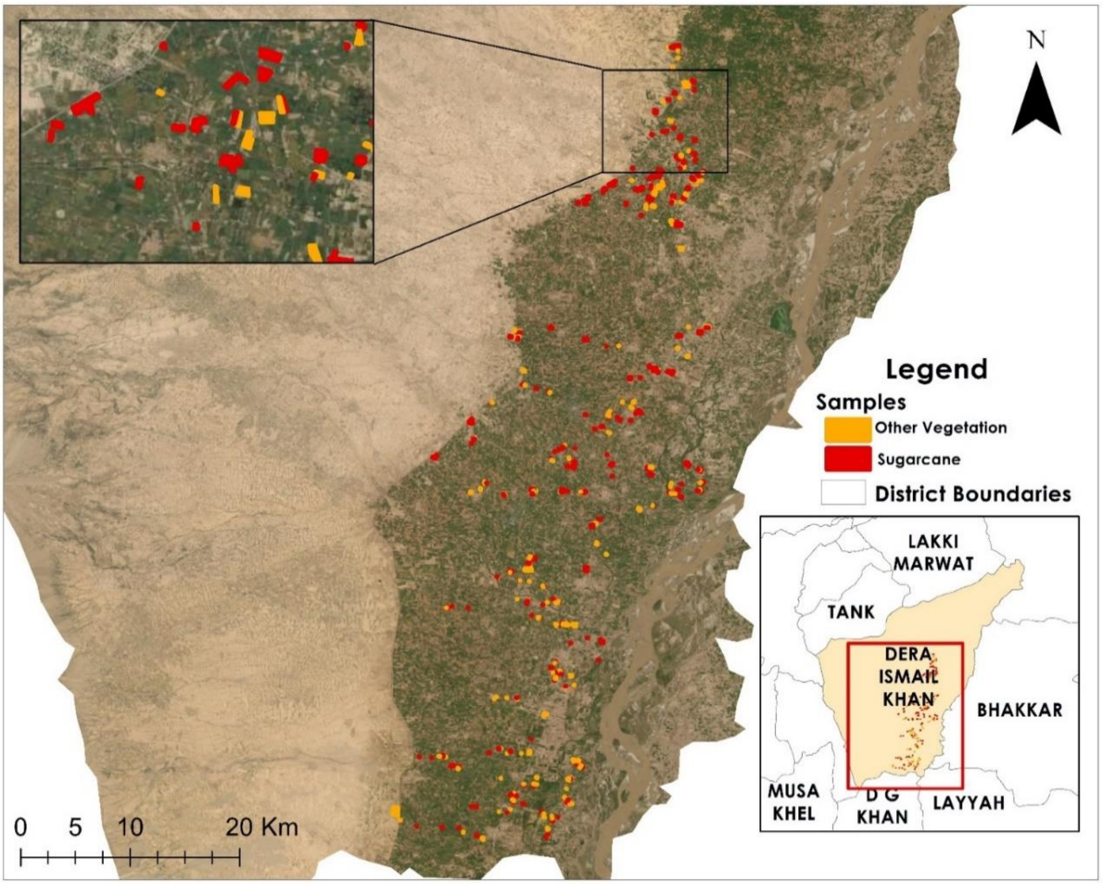

Figure 4: Map depicting the Extent and Locations of Samples Collected: A total of 467 samples, encompassing diverse vegetation types such as sugarcane, wheat, canola, gram, and others, were

systematically gathered through field surveys. The map illustrates the spatial distribution and coverage of the collected samples across the study area.

2.4 Pre-Processing of Datasets

Collected samples were analyzed and matched with cropping patterns and NDVI trends to verify collected data. A mosaic of sentinel-2 L2A Surface Reflectance was created for the months of October 2021 to October 2022 including the spectral bands and indices that are sensitive to different sort of vegetation greenness and water status (Di Vittorio & Georgakakos, 2018) and characterize the growth curve for individual crop types (Jackson et al., 2004; Zhang et al., 2010). NDVI (Tucker, 1979) and EVI (Huete et al., 2002) are highly related to leaf area index and chlorophyl in the canopy and are widely used to indicate vegetation greenness. It is also observed from literature review that Normalized Difference Moisture Index (NDMI) is the most used moisture index for determination of moisture in plants and bare soil (Davidson & Finlayson, 2007; Rahman & Mesev, 2019; Strashok et al., 2022) hence it has been used here due to the higher water content in sugarcane then other crops (Inman-Bamber & Smith, 2005). Hence for sugarcane crop identification, spectral bands as well as spectral indices including Red, Green, Blue, NIR, Red Edge 1, NDMI, NDVI and EVI from sentinel 2 for each month of the complete period were stacked together. As Sugarcane crop is at highest vegetation in the months of September and October, so a careful comparison of all the spectral bands and indices for these 2 months was carried out and results showed that blue band, EVI, NDMI and NDVI show significant spectral differences as shown in Figure 7. Hence only these bands and indices were used for further classification and model training to extract sugarcane. Dynamic World is a 10m Near Real Time (NRT) Land Use/Land Cover (LULC) dataset including class probabilities and label information for nine different (Venter et al., 2022). Dynamic Word Data contains crop class as well and this class was used to remove all those areas which do not fall in the category of vegetation from the stacked image collection. This helps eliminating all the extra areas including water, barren and builtup, only vegetation area is left to be classified further.

2.5 Machine Learning Classifiers

Supervised and unsupervised classification methods play crucial roles in crop classification. While supervised methods require labeled training data for classification, unsupervised methods like self-supervised deep learning and unsupervised domain adaptation can handle variations in environmental conditions without the need for extensive labeled data (Moumni & Lahrouni, 2021; Senaras et al., 2023). For instance, self-supervised methods have shown significant improvements in F1-score performance compared to traditional models like RF (Moumni & Lahrouni, 2021; Senaras et al., 2023). On the other hand, unsupervised domain adaptation frameworks like STDAN combine adversarial training with self-training to generate new training data in the target domain, proving effective in cross-domain crop type mapping without the need for extensive analyst intervention (Moumni & Lahrouni, 2021; Senaras et al., 2023). These approaches demonstrate the potential of leveraging both supervised and unsupervised techniques for accurate and efficient crop classification, each offering unique advantages in handling different challenges in agricultural management and monitoring.

RF, SVM, KNN, and Decision Trees (CART) have been extensively studied for crop classification (Abdelmalek & Assia, 2022; Alzhanov et al., 2023; Asgari & Hasanlou, 2023; Kolhe et al., 2022). RF (Breiman, 2001) has shown promising results in achieving high accuracy levels while reducing training time significantly (Alzhanov et al., 2023). SVM has been utilized for crop mapping with satisfactory results, especially when combined with Vegetation Indices (VIs) like ARVI (Asgari & Hasanlou, 2023). KNN has been employed for crop recommendation systems, showcasing high-accuracy rates (Mishra et al., 2021). Decision Trees, specifically Smile CART, have been compared with other models, demonstrating competitive performance in crop state identification (Abdelmalek & Assia, 2022). Additionally, K-means clustering has been used in agriculture for tasks like yield prediction, showcasing its versatility in agricultural applications (Mishra et al., 2021). Each algorithm brings unique strengths to crop classification, with RF and SVM standing out for their accuracy and efficiency in different agricultural contexts.

Recent advancements in machine learning models have significantly enhanced crop classification tasks. In this study, algorithms such as RF, SVM, CART, and KNN were employed for supervised classification, while K-means clustering was used for unsupervised classification to map sugarcane crops. To develop the supervised classification models, a field survey was conducted to collect training and validation samples, which were split into an 85:15 ratio for training and validation, respectively. These samples were categorized into two classes: sugarcane and other crops. The accuracy and error matrices were generated using validation samples, resulting in overall accuracies of 90% for SVM, 81% for RF, 78.57% for CART, and 75.71% for KNN.

SVM achieved the highest accuracy due to its effectiveness in handling high-dimensional spaces and its robustness to overfitting, particularly in binary classification problems (Braun et al., 2012). RF, with its ensemble learning approach, provided slightly lower accuracy due to its sensitivity to the quality of individual decision trees but was still effective in capturing complex patterns (Htitiou et al., 2019). CART, while simpler and more interpretable, achieved moderate accuracy as it is prone to overfitting on noisy datasets (Shao & Lunetta, 2012). K-NN yielded the lowest accuracy, likely due to its reliance on local neighborhood information, which can be less effective when the class boundaries are not well-defined (Jasim & Al-Taei, 2018).

SVM is often compared to other machine learning algorithms in agricultural studies due to its distinct advantages and disadvantages:

- **SVM vs. RF**: SVM tends to perform better in high-dimensional spaces and with a clear margin of separation between classes. However, RF is more robust to noise and can handle large datasets with many features, often requiring less preprocessing (Nitze et al., 2012).
- **SVM vs. CART**: SVM generally provides higher accuracy but at the cost of increased computational resources and complexity. CART is simpler and more interpretable but can overfit easily, especially with noisy data (Shao & Lunetta, 2012).
- **SVM vs. K-NN**: SVM is more robust to overfitting and handles high-dimensional data better. K-NN is easier to implement and understand but can be less accurate, especially with unevenly distributed classes or unclear boundaries (Ghosh et al., 2022).
- **SVM vs. K-means Clustering**: SVM, a supervised learning algorithm, typically outperforms K-means clustering, an unsupervised method, in terms of accuracy because SVM leverages labeled data to learn the decision boundary between classes. K-means, while useful for initial exploratory data analysis, often results in lower accuracy due to its reliance on distance metrics and the assumption that clusters are spherical and evenly sized, which is not always the case in agricultural data (Al-Tamimi & Al-Bakri, 2005).

Chapter 3

RESULTS AND DISCUSSION

3.1 Accuracy Assessment

The GPS location (polygons) of more than 450 samples of sugarcane and other crops in the study area were collected during ground survey that was used for model training and validation. These samples are error-free, contain pure vegetation of that same class and do not have any impurities. About 85% (375) of the data was used for model training and the rest 15% (70) field data was used as validation samples. The total area of Sugarcane in the study area was 54055 acres as shown in Figure 7. The objective of this study was to identify the sugarcane and differentiate it from other crops, so the confusion matrix was generated for both the classes as shown in Table 2. A total of 4 supervised ML classification models and 1 unsupervised classification (K-mean) were used in the study to correctly classify sugarcane in the study area and draw a comparison of these models. The classification results were compared based on the user's accuracy, producer's accuracy and overall accuracy in Table 3. To evaluate the performance of the classification models (RF, SVM, CART, Smile Gradient Tree Boost, Smile KNN, and Naive Bayes) in differentiating between other vegetation and sugarcane, a confusion matrix was employed (Table 2). This analysis revealed variations in model effectiveness for both the user's accuracy and the producer's accuracy.

User's accuracy reflects the proportion of pixels a model classified correctly within a specific class. For example, the RF model achieved a user's accuracy of 90% for sugarcane, indicating that 90% of pixels classified as sugarcane by the model were sugarcane based on reference data. However, the user's accuracy for other vegetation was only 60%, suggesting the model struggled to distinguish other vegetation from sugarcane in some cases. This pattern is repeated across other models, with SVM and Naive Bayes exhibiting similar user's accuracy discrepancies between the two classes.

Producer's accuracy, on the other hand, assesses the proportion of actual class pixels that the model correctly identified. The RF model, for instance, had a producer's accuracy of 70.59% for other vegetation, signifying that out of all the actual "other vegetation" pixels, the model correctly classified slightly more than 70%. However, the producer's accuracy for sugarcane was higher at 84.80%, indicating a better ability to identify true sugarcane pixels. This trend is again observed in other models, with some demonstrating a stronger ability to identify true positives for one class compared to the other.

Table 2: Confusion Matrix for Classification in GEE: Shows actual vs. predicted classes for Sugarcane and Other Vegetation, highlighting model performance.

Classifier	Class	Other Vegetation	Sugarcane	Total
RF	Other Vegetation	12	8	**20**
	Sugarcane	5	45	**50**
	Total	**17**	**53**	**70**
SVM	Other Vegetation	17	3	**20**
	Sugarcane	4	46	**50**
	Total	**21**	**49**	**70**
CART	Other Vegetation	10	10	**20**
	Sugarcane	5	45	**50**
	Total	**15**	**55**	**70**
KNN	Other Vegetation	11	9	**20**
	Sugarcane	8	42	**50**
	Total	**19**	**51**	**70**
k-Means	Other Vegetation	4	16	**20**
	Sugarcane	15	35	**50**
	Total	**19**	**51**	**70**

Finally, the overall accuracy summarizes the total percentage of pixels correctly classified across both classes. The Support Vector Mchine achieved the highest overall accuracy (90%), followed by the RF (81.43%). While these models performed well overall, k-Nearest Neighbor (75.71%) had a lower overall accuracy, highlighting the need for careful model selection based on the specific classification task.

Table 3: Performance Comparison of Machine Learning Models for Crop Classification.

Model	RF	SVM	CART	KNN	k-Means
User's Accuracy Other Vegetation	60	85	50	55	20
User's Accuracy Sugarcane	90	92	90	84.00	70
Producer's Accuracy Other Vegetation	70.59	80.95	66.67	57.89	21
Producer's Accuracy Sugarcane	84.80	93.87	81.81	82.35	69
Overall Accuracy	81.43	90	78.57	75.71	78
Kappa Coefficient	0.5237	0.758	0.432	0.395	-0.107

Figure 5 compares the area of sugarcane and other crops classified by all 6 classification models. The unsupervised classification model (k-Mean) gives the highest while KNN gives the lowest area for sugarcane. According to (FAO, 2024) total area for sugarcane cultivation for Dera Ismail Khan in 2022 was 53258 acres and in comparison with SVM (54055) and RF

(55163), we have achieved 98% and 96% accuracy respectively. CART and K-mean show over-estimation of sugarcane while KNN shows under-estimation.

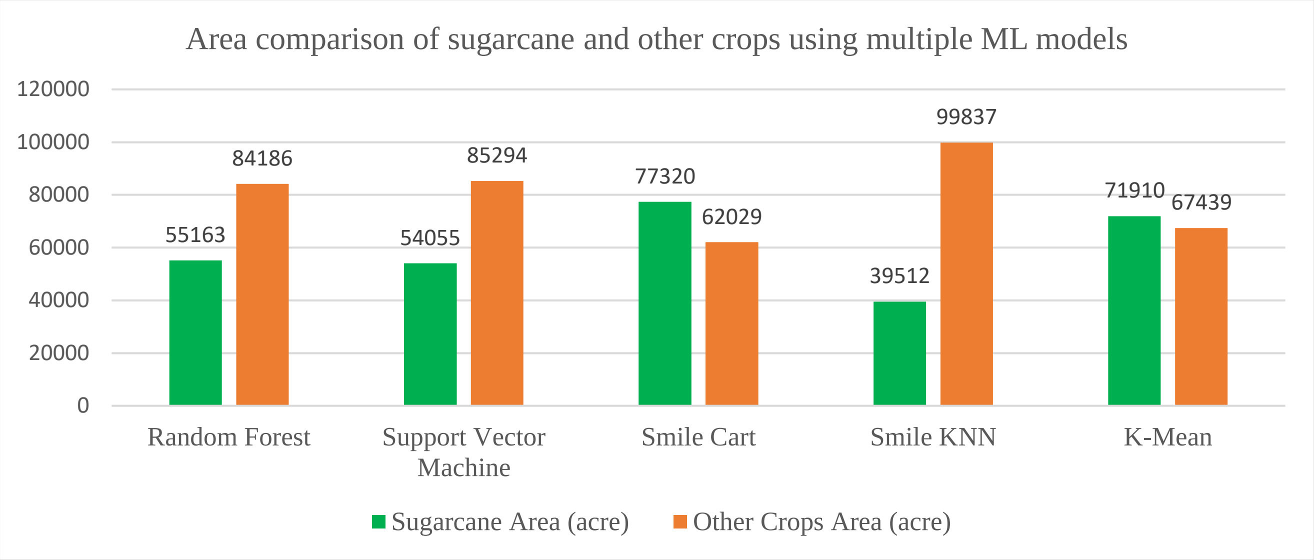

Figure 5: Area Comparison of Sugarcane and Other Crops by Machine Learning Models

A comparative analysis of the performance of all five classification models, including K-Nearest Neighbors (KNN), SVM, RF, CART, and K-Means, in distinguishing sugarcane from non-sugarcane fields across four different locations is presented in Figure 6. The results indicate that SVM outperforms the other models, achieving the highest accuracy in all four locations. RF follows closely, demonstrating a strong performance in sugarcane classification. In contrast, KNN, Smile Cart, and K-Means exhibit varying degrees of accuracy. Green color represents sugarcane while yellow represents other vegetation and samples are shown in red outline. SVM has correctly classified all the validation sample locations followed by RF which shows more stray pixels within the field. K-mean clustering unsupervised classification model shows more smooth results but that is highly overestimated. These findings suggest that SVM and RF are the most suitable models for sugarcane classification, and their performance is robust across different locations.

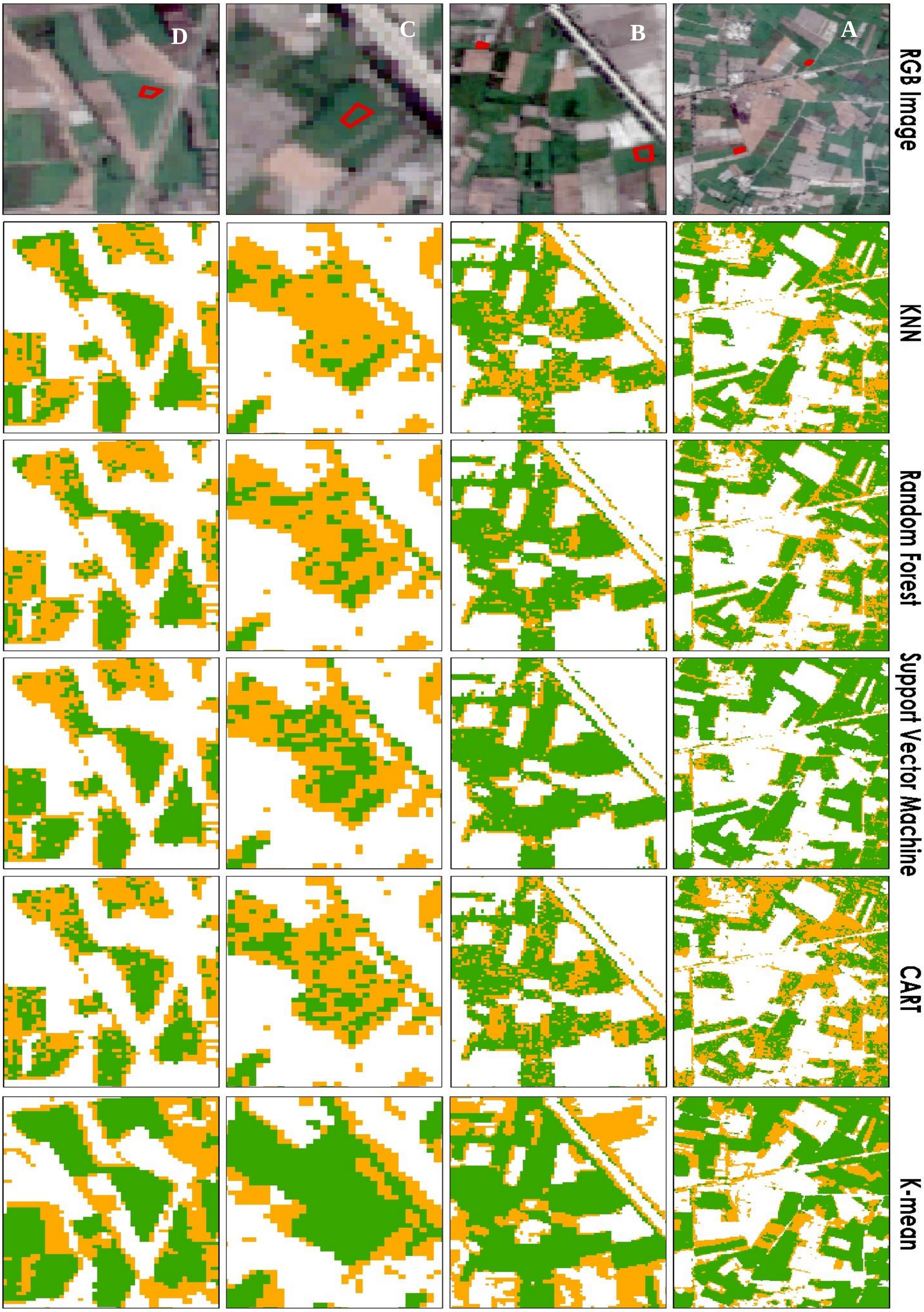

Figure 6: Comparison of Classification Models for Sugarcane and Non-Sugarcane Classification

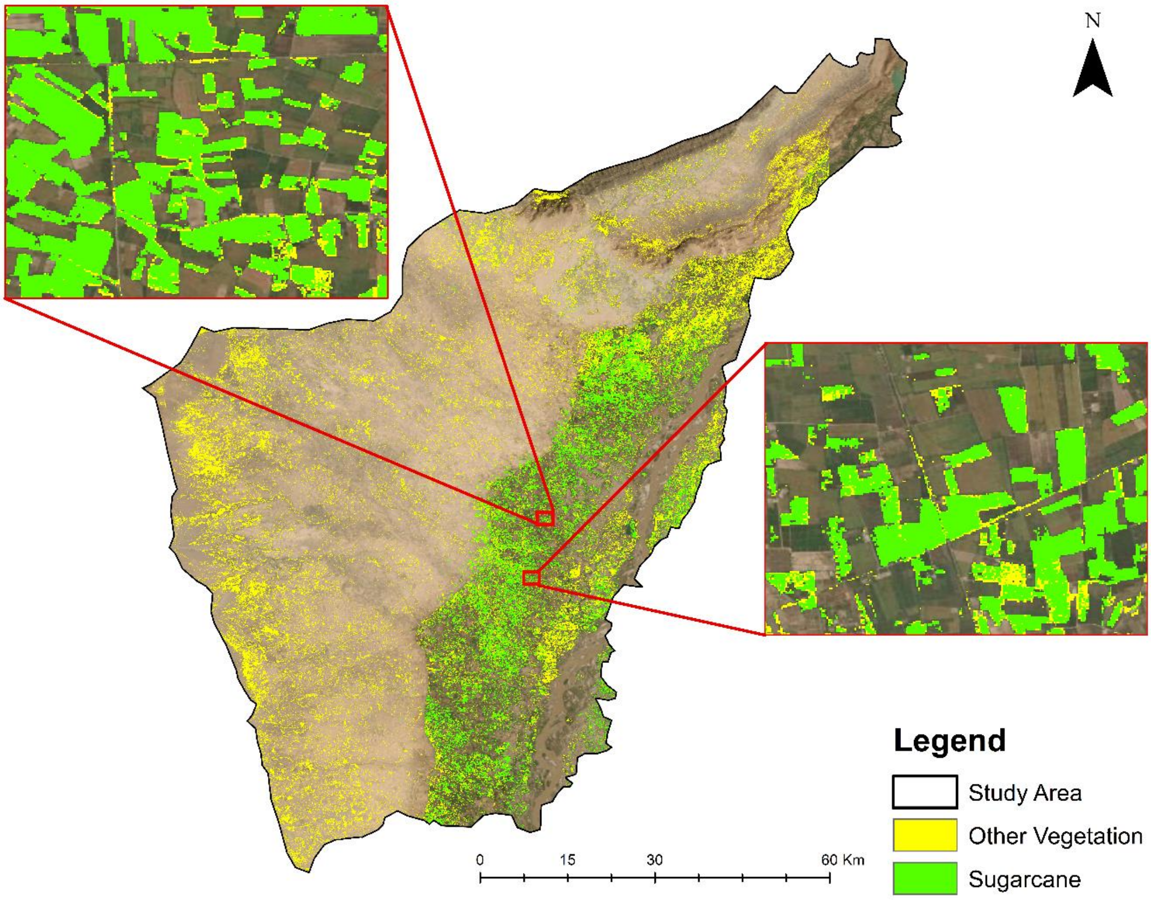

Figure 7: Sugarcane Fields map with a total area of 54055 acres for the year 2022-2023 identified through the Application of SVM in GEE. The classification result shows 2 sugarcane and other crops.

A detailed comparison of high-performing classification models RF and SVM is presented here. Among the five models tested, SVM and RF emerged as the top performers with overall accuracies of 90% and 81%, respectively as presented in Table 3.

The SVM classifier achieved the highest accuracy, indicating its effectiveness in distinguishing between the two classes (Figure 7). SVM works by finding the optimal hyperplane that maximally separates the classes in the feature space. This method is particularly effective in high-dimensional spaces and performs well even when the number of dimensions exceeds the number of samples. However, SVMs require careful parameter tuning, such as selecting the appropriate kernel and regularization term, which can be computationally intensive and time-consuming. Additionally, SVMs are less effective on noisy datasets with overlapping classes, making them less robust in certain scenarios.

On the other hand, the RF classifier, while slightly less accurate, offers several practical advantages. RF is an ensemble learning method that constructs multiple decision trees and merges their predictions to improve accuracy and control overfitting. This model handles large datasets and high-dimensional data more efficiently than SVM, making it scalable for larger applications. RF also provides insights into feature importance, enhancing interpretability and allowing researchers to understand which features contribute most significantly to the classification task. However, RF can be biased towards classes with more instances and is generally less interpretable compared to single decision trees. RF classification results show a total area of 55163 acres for Sugarcane as shown in Figure 8.

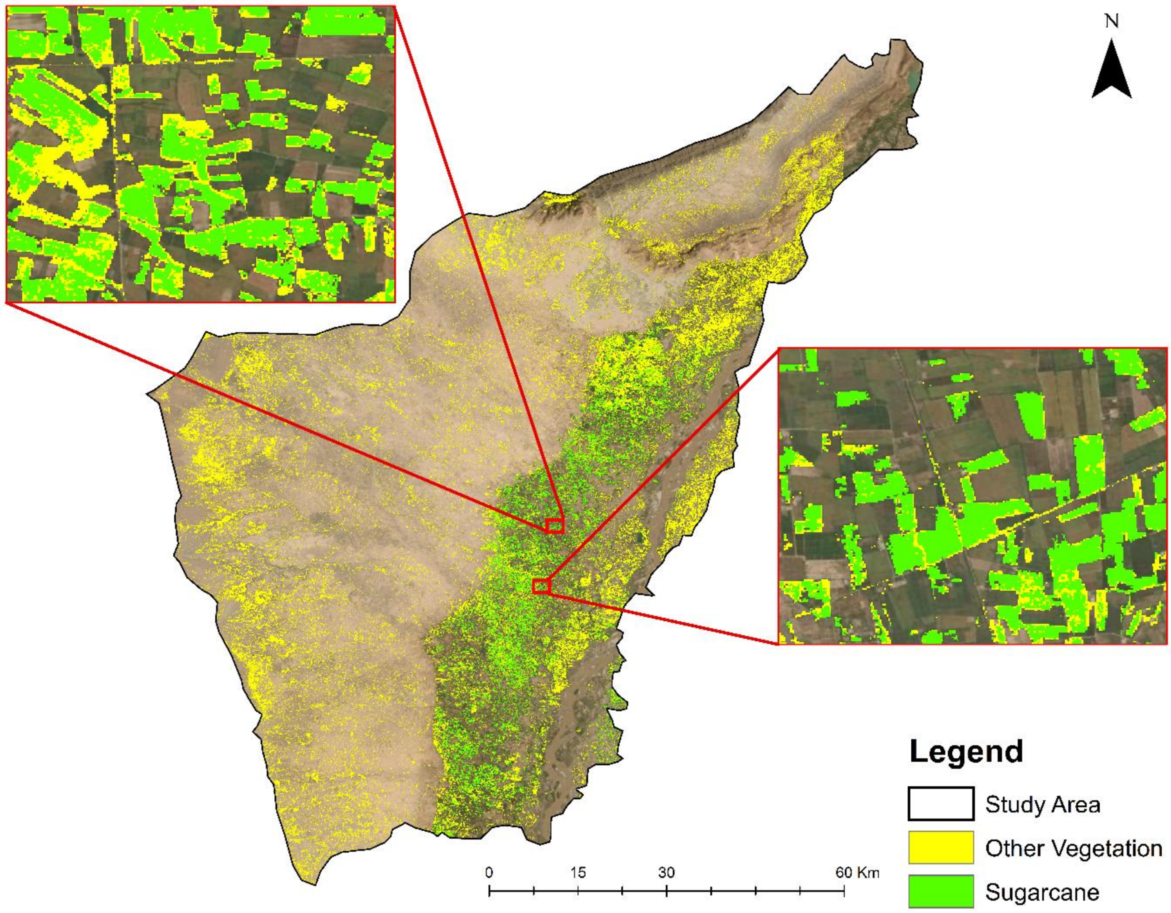

Figure 8: Sugarcane Fields map with a total area of 55163 acres for the year 2022-2023 identified through the Application of RF in GEE. The classification result shows 2 sugarcane and other crops.

3.2 Harvest Patterns

Sugarcane Sown areas were identified because of Machine Learning classification. NDVI, NDMI, EVI, and Red Edge 1 showed larger differences in spectral response for sugarcane and were considered as the differentiating features to distinguish sugarcane from other crops. Hence these were used for model training. The total area of sugarcane identified for Dera Ismail Khan was 60808 Acres. Sugarcane Harvesting starts in late October and ends in late February to March (Farooq & Gheewala, 2020; Hassan et al., 2017). Sugarcane shows high values of NDVI in October when it is at full maturity while other crops in the same area are already harvested, or new crops are being sown including gram and wheat. Based on this fact we identified the harvesting patterns for sugarcane in the study area using NDVI values drop for fields where sugarcane is already identified. By the end of December 19934 acres of sugarcane was harvested leaving behind 40874 acres. Subsequently, these areas show a drop in NDVI where sugarcane was harvested, and it is completed by the end of February to the first week of March. The detailed harvesting trends are shown in Table 3 and Harvesting Maps/trends are shown in Figure 9.

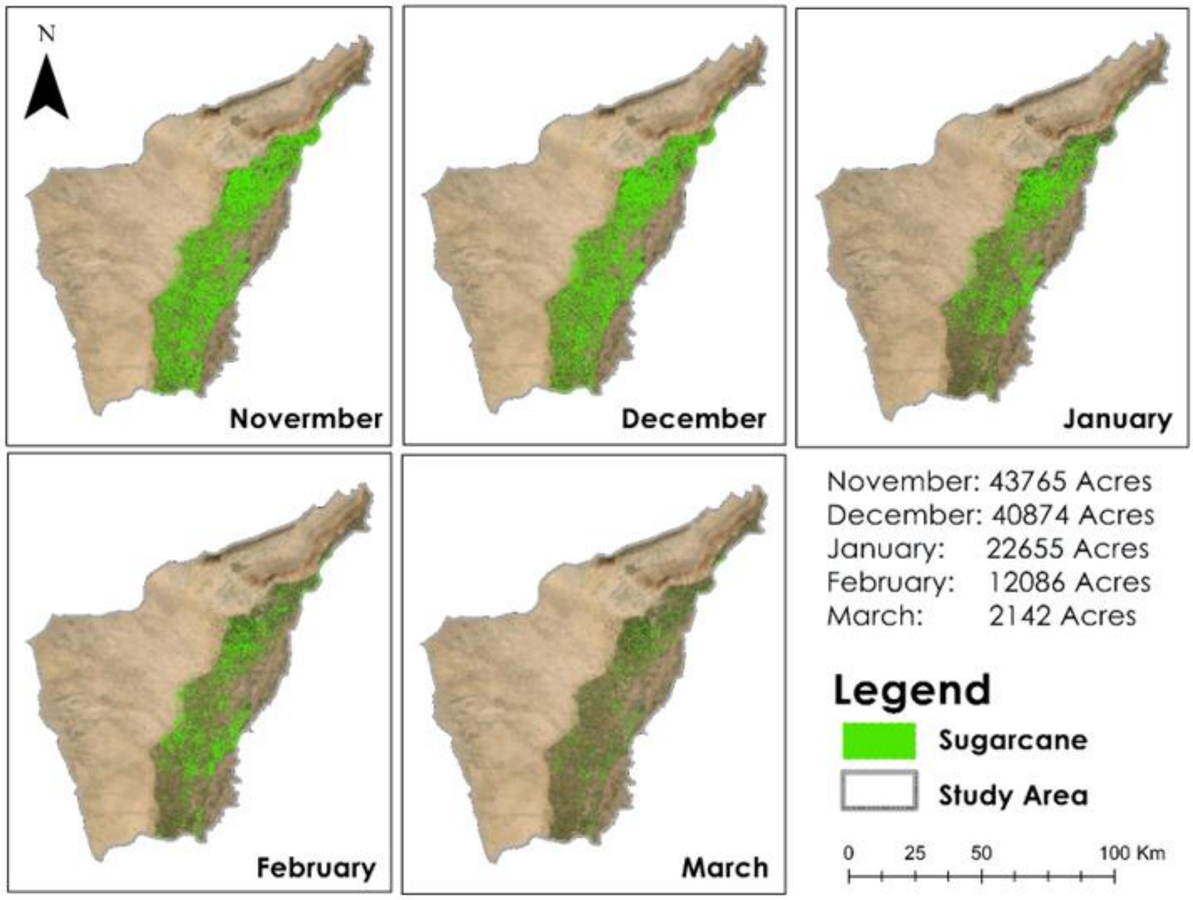

Figure 9: Temporal Dynamics of Sugarcane Harvest: Visualizes acreage trends, starting at 43,765 acres in November and declining to 2,142 acres in March, highlighting seasonal harvest impacts.

3.3 Spectral Features

In our endeavor to develop a robust machine learning model for sugarcane identification, meticulously selecting the most informative spectral features was paramount. A comprehensive analysis of all Sentinel-2 images features, including bands and derived indices, was conducted. Recognizing that sugarcane exhibits heightened vegetation and becomes readily distinguishable from other land covers during its peak growth period, we focused on the months of September and October for feature comparison. This in-depth comparison revealed noteworthy spectral differences in the blue band, EVI, NDMI, and NDVI. Notably, previous research by (Wang et al., 2022) and (Rauf et al., 2022) identified the Blue band as particularly effective in discerning sugarcane due to its sensitivity to chlorophyll absorption. Additionally, studies by (Jones & Vaughan, 2010) and (Rossini et al., 2012) highlighted the discriminatory power of EVI and NDVI in capturing vegetation health and biomass, while NDMI's sensitivity to moisture content (Gao, 1996) proved valuable for differentiating sugarcane from drier land covers. Therefore, recognizing the enhanced discriminative power of these specific features, we strategically selected them for model training and validation Figure 10. This data-driven approach ensures our model leverages the most informative spectral characteristics, ultimately aiming to yield robust and reliable sugarcane identification, similar to the success achieved by previous studies utilizing machine learning for crop classification (Suresh Kumar & Mohan, 2023).

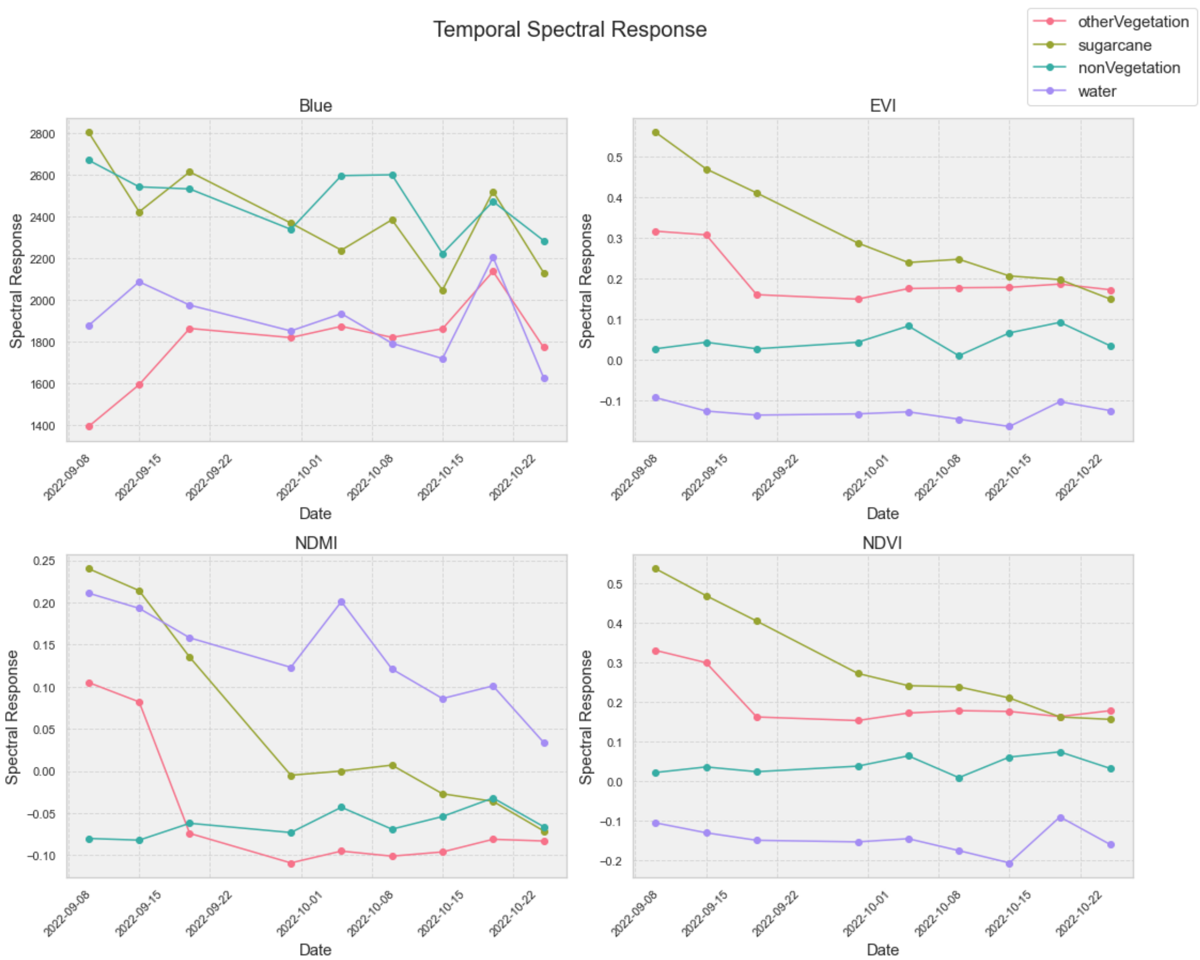

Figure 10: Spectral Features for ML Model Training: Shows distinctive responses of Blue Band, EVI, NDMI, and NDVI for sugarcane and other vegetation, aiding effective classification and extraction.

Chapter 4

CONCLUSION AND RECOMMENDATIONS

The importance of Machine Learning models and RS technology for sugarcane crop classification and harvest patterns was investigated in this study. NDVI, NDMI, EVI, Blue band and EVI from sentinel 2 show higher spectral differences between sugarcane and other vegetation and hence these are features, helpful to distinguish sugarcane from other crops in the area and show great potential for detecting crop type, crop conditions (harvested or growing) and mapping sugarcane cropped areas for small sized farms over 1 acre in Dera Ismail Khan. Temporal and spatial data was utilized for the Machine Learning Models, SVM, RF, CART, KNN and k-Means classifiers and SVM gave highest 90% overall accuracy. The sugarcane maps prepared in this study will be used as a basis for precise acreages for increased accuracy in yield forecasting. Given that sugarcane is a highly dynamic crop with essential phenological characteristics, it is determined that the ML methods using RS data is certified to be usable, with good accuracy, in sugarcane crop classification. However, the optical RS data can perform very little in case of cloudy weather. Sugarcane and other crops were affected by haze during the month of February and due to that a significant drop in NDVI values is observed while the on-ground sugarcane field was not harvested. The monsoon season in the region also adds in the situation when Optical Images fail due to the cloudy weather. This creates a data gap that represents a certain growth stage of the sugarcane crop. However, while assimilation of optical and SAR data aids in covering that phenological stage data, it does not give the best accuracy. As a result, there is room to examine the technique for dealing with foggy optical satellite images. Future work will focus on sugarcane variety identification, distinguishing between plant cane and ratoon cane, sugarcane phenology discrimination and identification and mapping of other crops in various regions of Pakistan.

REFERENCES

Abdelmalek, D., & Assia, K. (2022). Machine Learning Classification Models Comparison for Crop Damage Identification. 2022 2nd International Conference on Advanced Electrical Engineering (ICAEE)

Afzal, M. (1996). Managing water resources for environmentally sustainable irrigated agriculture in Pakistan. The Pakistan Development Review, 977-988. https://www.jstor.org/stable/41260010

Ahmad, K. F. Z., Muhammad, S., Ul, H. M., Tahira, G. H., Feehan, H., Amir, M. S., & Atif, W. (2013). Agricultural dynamics in Pakistan: current issues and solutions. Russian Journal of Agricultural and Socio-Economic Sciences, 20(8), 20-26. https://doi.org/http://dx.doi.org/10.18551/rjoas.2013-08.03

Al-Tamimi, S., & Al-Bakri, J. (2005). Comparison between supervised and unsupervised classifications for mapping land use/cover in Ajloun area. Jordan Journal of Agricultural Sciences, 1(1).

Alzhanov, A., Nugumanova, A., & Sutula, M. (2023). Research on crop classification methods based on machine learning using wavelet transformations. Eurasian Journal of Applied Biotechnology(2).

Asgari, S., & Hasanlou, M. (2023). A Comparative Study of Machine Learning Classifiers for Crop Type Mapping Using Vegetation Indices. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 10, 79-85.

Aslam, M. (2016). Agricultural productivity current scenario, constraints and future prospects in Pakistan. Sarhad Journal of Agriculture, 32(4), 289-303. https://doi.org/http://dx.doi.org/10.17582/journal.sja/2016.32.4.289.303

Azam, A., & Shafique, M. (2017). Agriculture in Pakistan and its Impact on Economy. A Review. Inter. J. Adv. Sci. Technol, 103, 47-60. https://doi.org/http://dx.doi.org/10.14257/ijast.2017.103.05

Azam, M., & Khan, M. (2010). Significance of the sugarcane crops with special and ference to NWFP. Sarhad J. agric, 26(2), 289-295.

Badar, H., Ghafoor, A., & Adil, S. A. (2007). Factors affecting agricultural production of Punjab (Pakistan). Pak. J. Agri. Sci, 44(3).

Braithwaite, K., Croft, B., & Magarey, R. (2007). Progress in identifying the cause of Ramu stunt disease of sugarcane. Proc. Aust. Soc. Sugar Cane Technology,

Braun, A. C., Weidner, U., & Hinz, S. (2012). Classification in high-dimensional feature spaces—Assessment using SVM, IVM and RVM with focus on simulated EnMAP data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 5(2), 436-443.

Breiman, L. (2001). Random forests Machine Learning 45 (1): 5–32. In.

Brown, C. F., Brumby, S. P., Guzder-Williams, B., Birch, T., Hyde, S. B., Mazzariello, J., Czerwinski, W., Pasquarella, V. J., Haertel, R., & Ilyushchenko, S. (2022). Dynamic

World, Near real-time global 10 m land use land cover mapping. Scientific Data, 9(1), 251. https://doi.org/http://dx.doi.org/10.1038/s41597-022-01307-4

Bureau of Statistics , K. P. (2022). Important District-Wise Socio-Economic Indicators of Khyber Pakhtunkhwa 2022.

Canata, T. F., Wei, M. C. F., Maldaner, L. F., & Molin, J. P. (2021). Sugarcane yield mapping using high-resolution imagery data and machine learning technique. Remote Sensing, 13(2), 232. https://doi.org/https://doi.org/10.3390/rs13020232

Chandio, A. A., Jiang, Y., Koondhar, M. A., & Guangshun, X. (2016). Factors affecting agricultural production: an evidence from Sindh (Pakistan). Advances in Environmental Biology, 10(9), 164-172.

Chandrasekar, K., Sesha Sai, M., Roy, P., & Dwevedi, R. (2010). Land Surface Water Index (LSWI) response to rainfall and NDVI using the MODIS Vegetation Index product. International Journal of Remote Sensing, 31(15), 3987-4005. https://doi.org/https://doi.org/10.1080/01431160802575653

Chen, J., Wu, J., Qiang, H., Zhou, B., Xu, G., & Wang, Z. (2021). Sugarcane nodes identification algorithm based on sum of local pixel of minimum points of vertical projection function. Computers and Electronics in Agriculture, 182, 105994. https://doi.org/https://doi.org/10.1016/j.compag.2021.105994

Cherry, R. H. (2001). Insect management in sugarcane. University of Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences, EDIS.

Chivasa, W., Mutanga, O., & Burgueno, J. (2021). UAV-based high-throughput phenotyping to increase prediction and selection accuracy in maize varieties under artificial MSV inoculation. Computers and Electronics in Agriculture, 184, 106128. https://doi.org/https://doi.org/10.1016/j.compag.2021.106128

Davidson, N., & Finlayson, C. (2007). Earth observation for wetland inventory, assessment and monitoring. Aquatic Conservation: Marine and Freshwater Ecosystems, 17(3), 219-228. https://doi.org/https://doi.org/10.1002/aqc.846

de Almeida, G. M., Pereira, G. T., de Souza Bahia, A. S. R., Fernandes, K., & Júnior, J. M. (2021). Machine learning in the prediction of sugarcane production environments. Computers and Electronics in Agriculture, 190, 106452. https://doi.org/https://doi.org/10.1016/j.compag.2021.106452

Di Vittorio, C. A., & Georgakakos, A. P. (2018). Land cover classification and wetland inundation mapping using MODIS. Remote Sensing of Environment, 204, 1-17. https://doi.org/https://doi.org/10.1016/j.rse.2017.11.001

Economic Survey of Pakisatn. (2022-23). Accessed at: https://www.finance.gov.pk/survey/chapters_23/Economic_Survey_2022_23.pdf

Elahi, E., Abid, M., Zhang, L., Ul Haq, S., & Sahito, J. G. M. (2018). Agricultural advisory and financial services; farm level access, outreach and impact in a mixed cropping district of Punjab, Pakistan. Land use policy, 71, 249-260. https://doi.org/https://doi.org/10.1016/j.landusepol.2017.12.006

FAO, M. (2024). Ministry of National Food Security and Research, Pakistan.

Farooq, M. S., & Fatima, H. (2022). GLOBAL CLIMATE CHANGE AND NATURAL DISASTERS: A THREAT TO SUSTAINABLE FOOD PRODUCTION AND FOOD SECURITY OF PAKISTAN. BRAZILIAN JOURNAL OF AGRICULTURE-Revista de Agricultura, 97(2), 186-214. https://doi.org/http://dx.doi.org/10.37856/bja.v97i2.4303

Farooq, N., & Gheewala, S. H. (2020). Assessing the impact of climate change on sugarcane and adaptation actions in Pakistan. Acta Geophysica, 68, 1489-1503. https://doi.org/http://dx.doi.org/10.1007/s11600-020-00463-8

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), 257-266.

Gardner, B. L., Rausser, G. C., Evenson, R. E., & Pingali, P. (2001). Handbook of Agricultural Economics: Agricultural Development: Farmers, Farm Production and Farm Markets (Vol. 3). Elsevier. https://doi.org/http://dx.doi.org/10.1016/S1574-0072(06)03051-9

Ghosh, S., Singh, A., Jhanjhi, N., Masud, M., & Aljahdali, S. (2022). SVM and KNN Based CNN Architectures for Plant Classification. Computers, Materials & Continua, 71(3).

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J., Robinson, S., Thomas, S. M., & Toulmin, C. (2010). Food security: the challenge of feeding 9 billion people. science, 327(5967), 812-818. https://doi.org/https://doi.org/10.1126/science.1185383

Hassan, M., Fiaz, N., Mudassir, M. A., & Yasin, M. (2017). Exploring the ratooning potential of sugarcane (Saccharum officinarum L.) genotypes under varying harvesting times of plant crop. Pak. J. Agric. Res, 30, 303-309. https://doi.org/http://dx.doi.org/10.17582/journal.pjar/2017.30.3.303.309

He, S.-S., Zeng, Y., Liang, Z.-X., Jing, Y., Tang, S., Zhang, B., Yan, H., Li, S., Xie, T., & Tan, F. (2021). Economic evaluation of water-saving irrigation practices for sustainable sugarcane production in Guangxi Province, China. Sugar Tech, 1-7. https://doi.org/http://dx.doi.org/10.1007/s12355-021-00965-9

Htitiou, A., Boudhar, A., Lebrini, Y., Hadria, R., Lionboui, H., Elmansouri, L., Tychon, B., & Benabdelouahab, T. (2019). The performance of random forest classification based on phenological metrics derived from Sentinel-2 and Landsat 8 to map crop cover in an irrigated semi-arid region. Remote Sensing in Earth Systems Sciences, 2(4), 208-224.

Huang, H., Deng, J., Lan, Y., Yang, A., Zhang, L., Wen, S., Zhang, H., Zhang, Y., & Deng, Y. (2019). Detection of helminthosporium leaf blotch disease based on UAV imagery. Applied Sciences, 9(3), 558. https://doi.org/http://dx.doi.org/10.3390/app9030558

Huang, Y.-K., Li, W.-F., Zhang, R.-Y., & Wang, X.-Y. (2018). Color illustration of diagnosis and control for modern sugarcane diseases, pests, and weeds. Springer. https://doi.org/http://dx.doi.org/10.1007/978-981-13-1319-6

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sensing of Environment, 83(1-2), 195-213. https://doi.org/http://dx.doi.org/10.1016/S0034-4257(02)00096-2

Iffat, T. (2012). Desertification Dynamics and its Control Mechanisms in semi arid region of Pakistan: A case study of Karak district Ph. D dissertation to the Institute of Geography, University of Peshawar Pakistan].

Imran, A., Khan, K., Ali, N., Ahmad, N., Ali, A., & Shah, K. (2020). Narrow band based and broadband derived vegetation indices using Sentinel-2 Imagery to estimate vegetation biomass. Global Journal of Environmental Science and Management, 6(1), 97-108. https://doi.org/https://doi.org/10.22034/GJESM.2020.01.08

Inman-Bamber, N., & Smith, D. M. (2005). Water relations in sugarcane and response to water deficits. Field crops research, 92(2-3), 185-202. https://doi.org/https://doi.org/10.1016/j.fcr.2005.01.023

Jackson, T. J., Chen, D., Cosh, M., Li, F., Anderson, M., Walthall, C., Doriaswamy, P., & Hunt, E. R. (2004). Vegetation water content mapping using Landsat data derived normalized difference water index for corn and soybeans. Remote Sensing of Environment, 92(4), 475-482. https://doi.org/https://doi.org/10.1016/j.rse.2003.10.021

Jasim, S. S., & Al-Taei, A. A. M. (2018). A Comparison Between SVM and K-NN for classification of Plant Diseases.

Jiang, H., Li, D., Jing, W., Xu, J., Huang, J., Yang, J., & Chen, S. (2019). Early season mapping of sugarcane by applying machine learning algorithms to Sentinel-1A/2 time series data: a case study in Zhanjiang City, China. Remote Sensing, 11(7), 861. https://doi.org/https://doi.org/10.3390/rs11070861

Jones, H. G., & Vaughan, R. A. (2010). Remote sensing of vegetation: principles, techniques, and applications. Oxford University Press, USA.

Júnior, C. C., Johann, J. A., Antunes, J. F. G., & Deppe, F. D. (2020). Sugarcane mapping in Paraná State Brazil using MODIS EVI images. International Journal of Advanced Remote Sensing and GIS, 9(1), 3205-3221. https://doi.org/https://doi.org/10.23953/cloud.ijarsg.451

Khaliq, A., Ahmad, N., Yasin, M., Bashir, H. M., Afzal, M. S., & Mahmood, A. (2023). NOVEL HIGH CANE, SUGAR RESILIENT AND EARLY MATURING SUGARCANE VARIETY CPF 250 FOR PUNJAB PROVINCE. Life Science Journal, 20(8). https://doi.org/http://www.dx.doi.org/10.7537/marslsj200823.04

Khan, M. A., Zafar, J., & Bakhash, A. (2008). Energy requirement and economic analysis of sugarcane production in Dera Ismail Khan district of Pakistan. Gomal University Journal of Research, 24(1), 1-7.

Kolhe, J., Deshpande, G., Patel, G., & Rajani, P. (2022). Crop Decision Using Various Machine Learning Classification Algorithms. In IOT with Smart Systems: Proceedings of ICTIS 2022, Volume 2 (pp. 495-502). Springer.

Lakshminarasimhappa, M. (2022). Web-based and smart mobile app for data collection: Kobo Toolbox/Kobo collect. Journal of Indian Library Association, 57(2), 72-79.

Li, H., Chen, Q., Liu, G., Lombardi, G. V., Su, M., & Yang, Z. (2023). Uncovering the risk spillover of agricultural water scarcity by simultaneously considering water quality and quantity. Journal of Environmental Management, 343, 118209. https://doi.org/https://doi.org/10.1016/j.jenvman.2023.118209

Lonare, A., Maheshwari, B., & Chinnasamy, P. (2022). Village level identification of sugarcane in Sangali, Maharashtra using open source data. Journal of Agrometeorology, 24(3), 249-254. https://doi.org/http://dx.doi.org/10.54386/jam.v24i3.1688

Mafuratidze, P., Chibarabada, T. P., Shekede, M. D., & Masocha, M. (2023). A new four-stage approach based on normalized vegetation indices for detecting and mapping sugarcane hail damage using multispectral remotely sensed data. Geocarto International, 38(1), 2245788. https://doi.org/http://dx.doi.org/10.1080/10106049.2023.2245788

Mehmood, V., Murtaza, R., Zafar, Z., Shahzad, M., Berns, K., & Fraz, M. (2023). Time series-based active labeling framework for curating a multispectral sentinel 2 imagery dataset for crop type mapping. IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium

Mishra, T. K., Mishra, S. K., Sai, K. J., Alekhya, B. S., & Nishith, A. R. (2021). Crop Recommendation System using KNN and Random Forest considering Indian Data set. 2021 19th OITS International Conference on Information Technology (OCIT)

Moumni, A., & Lahrouni, A. (2021). Machine Learning-Based Classification for Crop-Type Mapping Using the Fusion of High-Resolution Satellite Imagery in a Semiarid Area. Scientifica, 2021(1), 8810279.

Nihar, A., Patel, N., & Danodia, A. (2022). Machine-Learning-Based Regional Yield Forecasting for Sugarcane Crop in Uttar Pradesh, India. Journal of the Indian Society of Remote Sensing, 50(8), 1519-1530. https://doi.org/http://dx.doi.org/10.1007/s12524-022-01549-0

Nitze, I., Schulthess, U., & Asche, H. (2012). Comparison of machine learning algorithms random forest, artificial neural network and support vector machine to maximum likelihood for supervised crop type classification. Proceedings of the 4th GEOBIA, Rio de Janeiro, Brazil, 79, 3540.

Pervaiz, U., Khan, F., Jan, D., & Zafarullah, M. (2013). An analysis of sugarcane production with reference to extension services in Union Council Malakandher-Peshawar. Sarhad Journal of Agriculture, 29(1), 145-150.

Poloju, K. K., Naidu, V. R., Rollakanti, C. R., Manchiryal, R. K., & Joe, A. (2022). New method of data collection using the kobo toolbox. Journal of Positive School Psychology, 1527-1535.

Rahman, M. R., Islam, A., & Rahman, M. A. (2004). NDVI derived sugarcane area identification and crop condition assessment. Plan Plus, 1(2), 1-12.

Rahman, S., & Mesev, V. (2019). Change vector analysis, tasseled cap, and NDVI-NDMI for measuring land use/cover changes caused by a sudden short-term severe drought: 2011 Texas event. Remote Sensing, 11(19), 2217. https://doi.org/https://doi.org/10.3390/rs11192217

Rauf, U., Qureshi, W. S., Jabbar, H., Zeb, A., Mirza, A., Alanazi, E., Khan, U. S., & Rashid, N. (2022). A new method for pixel classification for rice variety identification using spectral and time series data from Sentinel-2 satellite imagery. Computers and Electronics in Agriculture, 193, 106731.

Rehman, A., Chandio, A. A., Hussain, I., & Jingdong, L. (2019). Fertilizer consumption, water availability and credit distribution: Major factors affecting agricultural productivity in Pakistan. Journal of the Saudi Society of Agricultural Sciences, 18(3), 269-274. https://doi.org/http://dx.doi.org/10.1016/j.jssas.2017.08.002

Rehman, A., Jingdong, L., Du, Y., Khatoon, R., Wagan, S. A., & Nisar, S. K. (2016). Flood disaster in Pakistan and its impact on agriculture growth (a review). Environ Dev Econ, 6(23), 39-42.

Rossini, M., Cogliati, S., Meroni, M., Migliavacca, M., Galvagno, M., Busetto, L., Cremonese, E., Julitta, T., Siniscalco, C., & Morra di Cella, U. (2012). Remote sensing-based estimation of gross primary production in a subalpine grassland. Biogeosciences, 9(7), 2565-2584.

Rudorff, B. F. T., Adami, M., De Aguiar, D. A., Gusso, A., Da Silva, W. F., & De Freitas, R. M. (2009). Temporal series of EVI/MODIS to identify land converted to sugarcane. 2009 IEEE International Geoscience and Remote Sensing Symposium

Senaras, C., Holden, P., Davis, T., Rana, A. S., Grady, M., Wania, A., & de Jeu, R. (2023). Self-supervised learning for crop classification using planet fusion. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 48, 309-315.

Shao, Y., & Lunetta, R. S. (2012). Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points. ISPRS Journal of Photogrammetry and Remote Sensing, 70, 78-87.

Singh, R., Patel, N., & Danodia, A. (2020). Mapping of sugarcane crop types from multi-date IRS-Resourcesat satellite data by various classification methods and field-level GPS survey. Remote Sensing Applications: Society and Environment, 19, 100340. https://doi.org/http://dx.doi.org/10.1016/j.rsase.2020.100340

Statistics, P. B. o. (2023). HOUSEHOLDS, POPULATION, HOUSEHOLD SIZE AND ANNUAL GROWTH RATE (7th Population and Housing Census-2023 'The Digital Census', Issue.

Strashok, O., Ziemiańska, M., & Strashok, V. (2022). Evaluation and Correlation of Sentinel-2 NDVI and NDMI in Kyiv (2017–2021). Journal of Ecological Engineering, 23(9). https://doi.org/http://dx.doi.org/10.12911/22998993/151884

Sumesh, K., Ninsawat, S., & Som-ard, J. (2021). Integration of RGB-based vegetation index, crop surface model and object-based image analysis approach for sugarcane yield estimation using unmanned aerial vehicle. Computers and Electronics in Agriculture, 180, 105903. https://doi.org/https://doi.org/10.1016/j.compag.2020.105903

Suresh Kumar, M., & Mohan, S. (2023). Selective fruit harvesting: Research, trends and developments towards fruit detection and localization–A review. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 237(6), 1405-1444.

Talib, U., Ashraf, I., Agunga, R., & Ashraf, S. (2019). Resource-poor farmers and environmental degradation in Pakistan: How extension can help. Pak. J. Agric. Res, 32, 110-114. https://doi.org/http://dx.doi.org/10.17582/journal.pjar/2019/32.1.110.114

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

Vennila, A., Palaniswami, C., Durai, A. A., Shanthi, R., & Radhika, K. (2021). Partitioning of major nutrients and nutrient use efficiency of sugarcane genotypes. Sugar Tech, 23, 741-746. https://doi.org/http://dx.doi.org/10.1007/s12355-020-00948-2

Venter, Z. S., Barton, D. N., Chakraborty, T., Simensen, T., & Singh, G. (2022). Global 10 m Land Use Land Cover Datasets: A Comparison of Dynamic World, World Cover and Esri Land Cover. Remote Sensing, 14(16), 4101. https://doi.org/https://doi.org/10.3390/rs14164101

Verma, A. K., Garg, P. K., Hari Prasad, K., Dadhwal, V. K., Dubey, S. K., & Kumar, A. (2021). Sugarcane yield forecasting model based on weather parameters. Sugar Tech, 23, 158-166. https://doi.org/https://doi.org/10.1007/s12355-020-00900-4

Wang, M., Liu, Z., Baig, M. H. A., Wang, Y., Li, Y., & Chen, Y. (2019). Mapping sugarcane in complex landscapes by integrating multi-temporal Sentinel-2 images and machine learning algorithms. Land use policy, 88, 104190. https://doi.org/http://dx.doi.org/10.1016/j.landusepol.2019.104190

Wang, Y., Fan, L., Tao, R., Zhang, L., & Zhao, W. (2022). Research on cropping intensity mapping of the Huai River Basin (China) based on multi-source remote sensing data fusion. Environmental Science and Pollution Research, 1-19.

Wilson, B. E. (2021). Successful integrated pest management minimizes the economic impact of Diatraea saccharalis (Lepidoptera: Crambidae) on the Louisiana sugarcane industry. Journal of Economic Entomology, 114(1), 468-471. https://doi.org/https://doi.org/10.1093/jee/toaa246

Wing, P. E. A. s. (2023). Pakistan Economic Survey. Government of Pakistan, Finance Division, Economic Adviser's Wing.

Xavier, A. C., Rudorff, B. F., Shimabukuro, Y. E., Berka, L. M. S., & Moreira, M. A. (2006). Multi-temporal analysis of MODIS data to classify sugarcane crop. International Journal of Remote Sensing, 27(4), 755-768. https://doi.org/http://dx.doi.org/10.1080/01431160500296735

Xin, F., Xiao, X., Cabral, O. M., White Jr, P. M., Guo, H., Ma, J., Li, B., & Zhao, B. (2020). Understanding the land surface phenology and gross primary production of sugarcane plantations by eddy flux measurements, MODIS images, and data-driven models. Remote Sensing, 12(14), 2186. https://doi.org/http://dx.doi.org/10.3390/rs12142186

Zhang, X., Nan, Z., Sheng, Y., Zhao, L., Zhou, G., Yue, G., & Wu, J. (2010). Analysis of time-series modis 250M vegetation index data for vegetation classifiation in the wenquan area over the qinghai-tibet plateau. 2010 IEEE International Geoscience and Remote Sensing Symposium

Zhu, M., Yao, M., He, Y., He, Y., & Wu, B. (2019). Studies on high-resolution remote sensing sugarcane field extraction based on deep learning. IOP conference series: earth and environmental science