Master Thesis

submitted within the UNIGIS MSc. programme

at the Department of Geoinformatics - Z_GIS

University of Salzburg, Austria

under the provisions of UNIGIS framework

Wheat Mapping Using Sentinel-2 Imagery

by

Bikesh Bade

u107051

A thesis submitted in partial fulfillment of the requirements of

the degree of

Master of Science (Geographical Information Science & Systems) – MSc (GISc)

Advisor (s):

Dr. Him Lal Shrestha

Kathmandu, 8th Nov 2023

Science Pledge

By my signature below, I certify that my project report is entirely the result of my own work. I have cited all sources of information and data I have used in my project report and indicated their origin.

Kathmandu, 8th Nov 2023

Place and Date	Signature

Acknowledgements

I would like to express my profound gratitude to all those who have played a pivotal role in the successful completion of my thesis on wheat mapping using Sentinel-2 images.

First and foremost, I extend my heartfelt appreciation to Dr. Him Lal Shrestha, my dedicated supervisor, for his unwavering guidance, continuous support, and unparalleled expertise throughout the entire thesis journey. His invaluable insights have been instrumental in shaping the direction and execution of this research.

I extend my sincere appreciation to the International Maize and Wheat Improvement Center (CIMMYT) and its dedicated team for their generous support and collaborative efforts, which have significantly elevated the quality and relevance of this study. The invaluable contributions from Urs Schulthess, Gerald Blasch Sonam Rinchen SHERPA, Bhavani Pinjarla, Dristy Bajimaya have been instrumental in enhancing various aspects of our research. Their expertise, access to resources, and overall commitment have played a vital role in the success of this project.

Furthermore, I extend my sincere acknowledgments to the data providers and researchers who have made Sentinel-2 imagery datasets publicly available. Their commitment to open science and data sharing has immensely facilitated the analysis and interpretation of the results, making this thesis possible.

In conclusion, I wish to convey my heartfelt thanks to all individuals and organizations who have directly or indirectly contributed to the successful completion of my thesis on wheat mapping with Sentinel-2 images. Your unwavering support and guidance have been indispensable, and I am truly grateful for the opportunity to undertake this research.

Abstract

Cropland classification is a crucial in agriculture and it enables the identification and monitoring of crop types over large geographical areas. Numerous uses for this data exist, such as urban development, environmental monitoring, and agricultural planning. Nepal, being faced with unique agricultural challenges with diversified terrain, can benefit significantly from crop mapping also with the adoption of remote sensing technology. This study investigated the performance of SVM and RF algorithms in the context of cropland classification. The primary objective was to assess their efficacy in accurately categorizing agricultural land into wheat and non-wheat regions.

The study utilized a dataset of high-resolution Sentinel 2 (L2A and L2B) satellite imagery covering a large agricultural region. In recent year, Sentinel-2 imagery has gained significant momentum, offering a highly effective and comprehensive approach to monitoring wheat fields. Sentinel-2, part of the European Space Agency's Copernicus program, provides high-resolution multispectral data, allowing for a detailed assessment of agricultural landscapes. Satellite data alone is not enough for accurate classification so for in situ data, we used Kobo toolbox to collect different crops spatially scattered in the study area. The imagery was preprocessed to extract spectral features, which were then fed into the SVM and RF classifiers. For farmland classification applications, machine learning methods like Random Forest (RF) and Support Vector Machine (SVM) have been used extensively. These algorithms can leverage large datasets of satellite imagery and other geospatial data to learn the complex relationships between spectral characteristics and crop types.

The classifiers were trained and evaluated on a subset of the dataset, and the best performing models were used to predict the crop types for the remaining pixels. The results revealed that the Random Forest algorithm consistently outperformed the Support Vector Machine, achieving a significantly higher level of accuracy of 99% training accuracy and

86% validation accuracy based on the in-situ data. Specifically, the Random Forest algorithm demonstrated the ability to predict that 56% of the total area under consideration was cultivated with wheat, while the remaining 44% was classified as non-wheat land. The superior performance of the Random Forest algorithm can be attributed to its robustness to image noise and its capacity to comprehend the intricate relationships between spectral characteristics and crop types. These advantages make it a well-suited tool for crop land classification in challenging agricultural landscapes.

The findings of this study highlight the potential of the Random Forest algorithm for generating reliable cropland maps. Such maps can be used for a variety of applications, including agricultural planning, environmental monitoring, and urban development. Future research could explore the use of Random Forest algorithm in conjunction with other machine learning techniques, such as deep learning, to further improve the accuracy of crop land classification. Additionally, the algorithm could be applied to classify other crop types, such as rice, maize, and soybeans, in diverse agricultural regions.

In conclusion, the results of this study emphasize the Random Forest algorithm as a promising tool for crop classification and land cover analysis. Its exceptional accuracy positions it as a valuable resource for generating dependable crop type maps with diverse applications across multiple domains.

Table of Contents

List of Tables

List of Abbreviations

RF	Random Forest
SVM	Support Vector Machine
CIMMYT	International Maize and Wheat Improvement Center
NDVI	Normalized Difference Vegetation Index
EVI	Enhanced Vegetation Index
SOS	Start of Season
EOS	End of Season
LOS	Length of Season
GDP	Gross Domestic Product
ESA	European Space Agency
SWIR	Shortwave Infrared
GEE	Google Earth Engine

List of Figures

Chapter-1: Introduction

1.1 Background

Wheat is one of the most important cereal grains in the world and a staple diet for a large proportion of the world's population. This versatile crop has been cultivated for thousands of years and plays a crucial role in human nutrition and agriculture. Due to population growth and economic expansion that raises per capita food consumption, it is anticipated that food demand will rise from 59% to 98% (Mishra et al., 2023). Wheat is prized for its high carbohydrate content, which provides a valuable source of energy, as well as its protein content, especially gluten, which gives wheat dough its elastic texture and makes it ideal for breadmaking. From bread and pasta to cereals and pastries, wheat products are a cornerstone of diets worldwide, making it a cornerstone of global agriculture and food security. Wheat plays a fundamental role in ensuring food security for a significant portion of the global population.

In countries like Nepal and other Asian nations, wheat cultivation serves as a crucial staple food and contributes significantly to the agricultural economy. Wheat holds significant agricultural and dietary importance in Nepal. It is one of the major cereal crops cultivated in the country and contributes significantly to the nation's food security. Since the dawn of time, wheat has been grown across Nepal, especially in the hills to the far and mid-west. In Nepal, it ranks third in importance among crops, behind rice and maize (Joshi et al., 1970). Wheat cultivation in Nepal is predominantly practiced during the winter season, known as the Rabi season, in regions with suitable climatic conditions. The harvested wheat is not only consumed in its whole grain form but is also milled into flour for making traditional Nepali staples like roti and sel roti. In rural areas, wheat farming is often a vital source of income for many households, playing a crucial role in rural economies. Additionally, the Nepali government has been actively promoting wheat cultivation and providing support to

farmers through various agricultural programs to increase production and enhance food security in the country.

However, wheat cultivation and management practices in Nepal were predominantly characterized by traditional and subsistence farming methods (Qamer et al., 2014). Many smallholder farmers in Nepal continued to rely on age-old techniques, such as manual planting and limited use of modern inputs like fertilizers and mechanized equipment, leading to suboptimal yields. Challenges included the limited availability of high-quality seeds and modern wheat varieties, inadequate pest, and disease management, as well as inconsistent access to irrigation. In another context, achieving food security in these regions is challenging, and sustainable agricultural practices are essential for meeting the growing demands of a burgeoning population. Efficient management and monitoring of wheat fields become paramount in addressing food security challenges. Traditionally, ground-based methods have been employed for crop assessment. However, these methods are often inadequate to cover the vast and diverse agricultural landscapes in the region.

Relative to comprehensive on-field surveying, remote sensing techniques use spectral signatures to provide information on the status of crops at different growth levels at a lower cost. Using a satellite or an instrument positioned in the atmosphere, remote sensing technology gathers and analyzes data about the world and its objects without requiring direct human contact (Muruganantham et al., 2022). Understanding the location and timing of crop production sheds light on the region's ecology and ecological balance (Mishra et al., 2023). The sustainable evolution of human communities, food security, and agricultural productivity all greatly depend on crop mapping utilizing remote sensing data (TAO et al., 2022). The methods used to monitor the natural and human resources on the surface of the Earth have changed dramatically with the development of satellite remote sensing technology, which enables the monitoring of vast areas. (Phiri et al., 2020). Utilization of remote sensing technology has gained traction in recent years, offering a more comprehensive and efficient approach to crop field monitoring. The adoption of remote

sensing technology has surged, presenting a more comprehensive and efficient approach to monitoring wheat fields. This technology has revolutionized the way we gather data and insights about agricultural landscapes. With the aid of satellites and drones, remote sensing allows for real-time and large-scale monitoring of wheat crops, enabling farmers and agricultural experts to assess crop health, detect diseases or stress factors, and optimize resource allocation. This transformative shift from traditional field monitoring methods to remote sensing has the potential to significantly enhance wheat cultivation practices, ultimately contributing to improved yields and sustainable agriculture.

Another important factor that helps on crop mapping is phonology. Crop phenology information is crucial for assessing crop productivity and crop management (Sakamoto et al., 2005). Phenology is an essential area of ecological research with broad implications that examines the timing of natural events in plant life cycles. It helps us understand the intricate relationships between plant growth and their environment. The timing of events such as bud burst, flowering, migration, and hibernation in response to environmental cues like temperature and daylight plays a vital role in ecosystem functioning. It provides insights into climate change impacts, as shifts in phenological events can serve as early indicators of climate-related alterations. Furthermore, phenological data aids in agricultural planning, helping farmers optimize planting and harvesting times.

Phenology also plays vital importance in wheat mapping, where it serves as a guiding compass for effective crop management. By closely observing the phenological stages of wheat, from germination to maturation, farmers can make informed decisions about when to plant and harvest, optimizing crop yields while conserving resources. Furthermore, phenology aids in early detection and response to potential threats such as pests, diseases, and stress factors that may impact wheat crops. When integrated with remote sensing technologies like Sentinel-2 imagery, phenological data becomes a powerful tool for accurately assessing the health and growth of wheat fields, supporting sustainable agricultural practices, and ensuring food production security.

Also, in recent year, Sentinel-2 imagery has gained significant momentum, offering a highly effective and comprehensive approach to monitoring wheat fields. Sentinel-2, part of the European Space Agency's Copernicus program, provides high-resolution multispectral data, allowing for a detailed assessment of agricultural landscapes. This advanced satellite technology has revolutionized the way we collect data on wheat crops. With the frequent and free availability of Sentinel-2 imagery, farmers and agricultural experts can monitor wheat fields in near-real-time, enabling precise evaluations of crop health, early detection of diseases or stress factors, and informed decision-making for resource allocation. This remarkable shift towards incorporating Sentinel-2 imagery into agricultural practices has the potential to revolutionize wheat cultivation, fostering increased productivity and sustainability in agriculture.

Nepal, being faced with unique agricultural challenges with diversified terrain, can benefit significantly from the adoption of remote sensing technology. The use of Sentinel-2 imagery in Nepal covers a wide range of applications, providing valuable information for geographical mapping, seasonal monitoring, and the evaluation of development activities and agricultural landscapes. This technology holds promise in revolutionizing wheat mapping efforts and transforming agricultural practices in the region. By harnessing the capabilities of Sentinel-2 imagery and focusing on regional agricultural contexts, most of studies are making substantial contributions to improved food security, efficient resource management, and a more sustainable agricultural future for the entire region. Overall, Sentinel-2 imagery is a powerful tool that's helping Nepal deal with its unique challenges in geography and agriculture.

In Nepal, the combination of phenology and Sentinel-2 imagery has emerged as a game-changing approach for wheat mapping and agricultural management. Phenology, by tracking the life cycle stages of wheat, provides critical insights into the optimal timing for planting and harvesting, allowing Nepali farmers to adapt to changing climate conditions and maximize crop yields. Sentinel-2 imagery, with its high-resolution and frequent data

capture, complements phenological observations by offering detailed visualizations of wheat fields. This technology enables precise monitoring of crop health, detection of stress factors, and early identification of potential issues like pest outbreaks or disease infestations. In Nepal's challenging agricultural landscape, the fusion of phenological data and Sentinel-2 imagery not only enhances wheat mapping accuracy but also empowers farmers and agricultural experts to make informed decisions, ultimately contributing to improved food security and sustainable farming practices.

The prolonged conflict between Russia and Ukraine has resulted in a reduction of wheat supply on the international market. This has had a substantial negative influence on the food security status of those countries whose local wheat supply is entirely imported. Crucially, many nations throughout the world rely on wheat imports from Russia and Ukraine to meet local food demand, as these two countries together account for almost 26% of the world's wheat export market. While Nepal is a producer of wheat, the country's domestic wheat production not always sufficient to meet its growing population's needs. Therefore, Nepal imports wheat from various sources, including neighboring countries like India. Wheat imports help supplement the domestic supply and ensure that there is an adequate amount of wheat available to meet the country's food requirements. This reduction of wheat supply in international market put lot of pressure on the wheat production in Nepal.

This study focuses on how Sentinel-2 imagery is making wheat mapping better in Nepal. We're looking at past studies to see what's been achieved and where we can do even better. Our goal is to give practical advice on how to use Sentinel-2 technology for growing wheat in Nepal. This research aims to make farming in Nepal stronger and more resilient for the future. This study also intends to demonstrate the advancements and contributions made by Sentinel-2 images in the mapping of wheat, which offers priceless insights for sustainable agriculture and the world's food supply. By evaluating existing studies in the region, we aim to identify potential areas of improvement and offer recommendations for the effective utilization of Sentinel-2 remote sensing technology in Nepal's wheat cultivation practices.

Ultimately, this research will contribute to promoting food security, efficient resource management, and sustainable agricultural practices, paving the way for a more resilient agricultural future in the region.

Sentinel-2 imagery has the potential to revolutionize wheat mapping in the western region of Nepal. However, there is a significant research gap in this area, with existing studies having only marginally explored the application of Sentinel-2 data for wheat mapping in the western region. This lack of research is particularly concerning given the unique agricultural landscape of the western region, which is characterized by a diversity of landforms, soil types, and climatic conditions. Existing studies have focused on developing generic wheat mapping methodologies using Sentinel-2 imagery. However, these methodologies may not be well-suited to the specific challenges of wheat mapping in the western region. For example, the unique landforms and soil types of the western region can lead to variations in spectral reflectance, which can make it difficult to distinguish wheat crops from other land cover types using generic Sentinel-2-based methodologies. Additionally, the diverse climatic conditions of the western region can lead to variations in the phenology of wheat crops, which can further complicate wheat mapping using Sentinel-2 imagery.

To address this research gap, this thesis proposes to develop a region-specific methodology for wheat mapping using Sentinel-2 imagery in the western region of Nepal. This methodology will be tailored to the unique agricultural landscape and climatic conditions of the western region and will consider the challenges of distinguishing wheat crops from other land cover types and tracking the phenology of wheat crops. The development of a region-specific methodology for wheat mapping using Sentinel-2 imagery will have several benefits for the western region of Nepal. Farmers will receive more precise and timely information regarding the conditions of the wheat crop in the first place. Decisions about crop management techniques, such as irrigation and fertilizer application, can be made with this information in hand. Second, it will help agricultural planners to better understand the spatial distribution of wheat production in the western region. This

information can be used to develop more effective agricultural policies and programs. Third, it will contribute to food security efforts in the western region by helping to ensure that farmers have access to the resources they need to produce sufficient wheat to meet the needs of the local population.

By 2030, it is anticipated that Nepal's wheat demand will increase by about 890 thousand metric tons (Khadka et al., 2020). Previous research on wheat mapping in Nepal has primarily relied on traditional ground-based methods and low-resolution satellite data, which are less effective in providing timely and accurate information. The conventional approaches, such as conducting field surveys, reviewing literature, interpreting maps, and analyzing collateral and ancillary data, are time-consuming, date-lagging, and frequently prohibitively expensive when it comes to acquiring vegetation covers. Remote sensing technology provides a useful and affordable way to research (Xie et al., 2008). The integration of Sentinel-2 imagery into wheat mapping offers several advantages, including frequent revisit times, higher spatial resolution, and multi-spectral capabilities that are well-suited for assessing wheat fields' health and growth.

In summary, wheat production in Nepal, especially in the western part, is a vital component of the country's agricultural landscape. The integration of Sentinel-2 imagery, with its advanced remote sensing capabilities, holds great promise for improving wheat mapping and monitoring practices. This thesis endeavors to bridge the existing research gap by developing a tailored methodology that leverages Sentinel-2 data to support agricultural sustainability and food security in the western part of Nepal.

1.2 Objectives

Main Objective

The main objective of the study is to create a workflow for wheat mapping using temporal Sentinel-2 satellite imagery.

Sub-objectives

Since the main objective of the study is to map the wheat land, there are several sub-objectives that need to be addressed. The sub-objectives of the study are as follows:

- To plot phenological chart with NDVI to address the phenological matrix such as SOS, EOS, and LOS.
- Compare and analyze the performance of RF and SVM machine learning algorithms for wheat mapping.
- Generate spatial distribution maps of wheat land with temporal sentinel 2 imagery with the help of the phenological metrics derived from NDVI plotted with survey data.
- Address different scenarios in the model calibration and validation based on the accuracy.

1.3 Scope of project

Wheat is a staple crop for billions of people around the world, and its cultivation is essential for food security. However, wheat production is facing increasing challenges due to climate change, pests, and diseases. To address these challenges, it is important to have accurate and up-to-date information on wheat cultivation.

This research study aims to map wheat cultivation in the western regions of the Terai using Sentinel-2 satellite data collected during the winter season of 2022-2023. The study will take a comprehensive approach, including phenological monitoring. In the study will extract a detailed phenological chart of the wheat crop, which will be used to create a precise and dynamic representation of its life cycle. This will be done by analyzing Sentinel-2 data for key phenological stages, such as germination, emergence, tillering, flowering, and grain filling.

The outcomes of this research will be of paramount significance to regional agriculture. The project's findings will contribute substantially to an enhanced understanding of wheat crop mapping, yield estimation, and crop stress detection in the western Terai. This newfound knowledge and the resulting data products will, in turn, serve as indispensable tools for local authorities, agricultural practitioners, and decision-makers in their quest for more effective crop management and informed agricultural decisions.

In addition to the immediate benefits for regional agriculture, this research also has the potential to make a significant contribution to the broader field of wheat mapping and crop monitoring. The project's findings and methodologies can be adapted to other regions and crops, and the project's data products can be used to validate and improve existing crop mapping and monitoring algorithms.

1.4 Literature Review

1.4.1 Wheat Production in Nepal

The agriculture sector in Nepal accounts for almost one-third of the country's GDP, and two-thirds of the country's population make their living from it. However, a number of elements, including a high degree of temporal and geographical climate variability, irrigated and rain-fed agriculture systems, the fragile social and economic fabric of farmers, and distinctive mountain habits, make sustainable agriculture production throughout the nation a major task (Qamer et al., 2014). Wheat is a fundamental staple crop in Nepal, playing a pivotal role in the country's food security and livelihoods. After rice and maize, it is the third-most important cereal crop in Nepal, making up around 7% of the nation's total planted area (Joshi et al., 1970). In all three of Nepal's agroclimatic zones, which range in elevation from sea level to 4,000 meters, wheat is grown. The mid- and high-hill regions of Nepal are the next most important wheat-producing regions, after the Terai region. The rural economy is based primarily on traditional agriculture, with wheat being one of the most important cereal crops grown in the area. For the purpose of guaranteeing national food security and raising the standard of national land management, timely and accurate monitoring of cropping intensity is crucial (Guo et al., 2021).

Wheat production in Nepal has increased significantly in recent years, due to several factors, including:

- ✓ Increased irrigation facilities
- ✓ Improved access to input supplies
- ✓ Release of new high-yielding varieties
- ✓ Government support programs

However, wheat productivity in Nepal is still low compared to other countries. In terms of area and production, wheat is Nepal's third-most important cereal crop after rice and maize, although its productivity is extremely low when compared to other affluent nations like China

and Switzerland (Bhatta et al., 2020). The main cause of Nepal's low wheat yield is the improper and inadequate application of fertilizer without the application of site-specific nutrient management strategies. Despite its importance, wheat production in Nepal faces several challenges, including:

- ✓ **Low productivity**: The average wheat yield in Nepal is significantly lower than the yields in developed countries. This is due to several factors, including climate change, limited access to inputs, and poor agricultural practices.
- ✓ **Climate change**: Climate change is posing a growing threat to wheat production in Nepal. Changes in temperature, precipitation patterns, and pests and diseases are all negatively impacting wheat yields.
- ✓ **Limited access to inputs**: Many farmers in Nepal lacks access to adequate irrigation, fertilizers, and other inputs. This limits their ability to maximize wheat yields.
- ✓ **Poor agricultural practices**: Some farmers in Nepal still use traditional agricultural practices that are inefficient and unsustainable. This can lead to low yields and environmental degradation.

1.4.2 Remote Sensing for Wheat Mapping

The science of gathering data about an object or phenomenon without making physical touch with it is known as remote sensing. Remote sensing technologies, such as satellite imagery, can be used to map wheat production at a large scale.

There are several advantages to using remote sensing for wheat mapping:

- ✓ Remote sensing can provide timely and accurate information about wheat production.
- ✓ Remote sensing can be used to map wheat production in remote and inaccessible areas.
- ✓ Remote sensing can be used to map wheat production over large areas.

- ✓ Several different remote sensing techniques have been used for wheat mapping, including:

  Optical imagery: Optical imagery captures reflected sunlight from the Earth's surface. Wheat can be distinguished from other crops and vegetation using spectral information from optical imagery.

  Radar imagery: Radar imagery can penetrate clouds and vegetation, making it useful for wheat mapping in adverse weather conditions.

  LiDAR imagery: LiDAR imagery uses laser pulses to measure the distance to the Earth's surface. LiDAR imagery can be used to map wheat production by measuring the height of wheat plants.

Conventional methods of assessing crop conditions and yields are often labor-intensive, time-consuming, and limited in scope. Field crop monitoring and mapping can be done using remote sensing techniques (remotesensing_workshop_2014). In contrast, remote sensing technology has emerged as a transformative tool for crop monitoring. It allows for the collection of data about crop health, growth, and distribution over extensive geographical areas without the need for direct field observations. Technological advancements in agriculture have made it possible for farmers to monitor, quantify, and react to both temporal and spatial variations in their crops. These "precision farming" techniques seek to minimize waste and negative effects while precisely directing agricultural inputs (Hunt et al., 2019).

1.4.3 Phenology

The term “phenology” describes the periodic stage of the plant life cycle that is associated with environmental conditions and climatic change. Phenology often depicts the physiological stages of crop growth for cereal crops, including the dates of sowing, green-up, jointing, heading, and maturity (Zhao et al., 2022). It is a critical factor for wheat production, as it determines when plants are most susceptible to pests and diseases, and

when they are most productive. Time series of high temporal resolution remote sensing data have been frequently used in recent decades to identify the phenology of crops (Ashourloo et al., 2022).

Wheat phenology can be divided into four main stages:

- ✓ **Seedling stage**: This stage begins when the seed germinates and ends when the plant has developed its first true leaves.
- ✓ **Vegetative stage**: This stage begins when the plant has developed its first true leaves and ends when the plant begins to flower.
- ✓ **Flowering stage**: This stage begins when the plant begins to flower and ends when the plant has set grain.
- ✓ **Grain filling stage**: This stage begins when the grain has set and ends when the grain is mature and ready to harvest.

The timing of these stages can vary depending on the variety of wheat, the climate, and the growing conditions. However, there is a general phenological sequence that most wheat varieties follow. Understanding wheat phenology is important for farmers for a number of reasons. First, it can help farmers to identify periods when their crops are most susceptible to pests and diseases. For example, wheat plants are most susceptible to fungal diseases during the flowering stage. Farmers can use this information to apply pesticides at the appropriate time to protect their crops.

Second, understanding wheat phenology can help farmers to optimize their crop management practices. For example, farmers can adjust their irrigation schedules based on the stage of crop growth. During the vegetative stage, wheat plants require more water than during the flowering stage. Third, understanding wheat phenology can help farmers to predict crop yields. By tracking the progress of their crops through the phenological stages, farmers can get a better idea of when to expect to harvest their crops.

1.4.4 Importance of Phenology for Wheat Mapping

Phenology is also important for wheat mapping. By monitoring the timing of wheat phenological stages, remote sensing can be used to map wheat areas and yields. A practical method for defining the spatiotemporal patterns of vegetation phenology at the pixel level is remote sensing (Pan et al., 2015). This information can be used to support sustainable agriculture and food security efforts. For example, remote sensing can be used to identify areas where wheat crops are stressed. This information can be used to target interventions to improve crop health and productivity. Remote sensing can also be used to monitor the effects of climate change on wheat phenology. Numerous studies have demonstrated that phenology data and temporal satellite image stacking may be used to greatly increase wheat mapping accuracy (Y. Li et al., 2020). In the context of wheat mapping in Nepal, the integration of phenology and temporal satellite image stacking has the potential to provide significant benefits. For example, this approach could be used to:

- ✓ Improve the accuracy of wheat yield estimates.
- ✓ Support sustainable wheat cultivation practices.
- ✓ Identify areas of wheat cultivation that are most vulnerable to climate change.
- ✓ Develop strategies to mitigate the impact of climate change on wheat production.
- ✓ Overall, phenology and temporal satellite image stacking are valuable tools for wheat mapping that can be used to support sustainable agriculture and food security in Nepal.

1.4.5 Sentinel-2 Imagery

Sentinel-2 is a pair of Earth observation satellites that work together to produce optical imagery of the surface of the Earth with high resolution. Wheat mapping is only one of the many uses for the publicly available Sentinel-2 imagery. Within the realm of remote sensing, the Sentinel-2 satellite imagery system, developed, and operated by the European Space Agency (ESA), has garnered considerable attention for its utility in agricultural applications. Sentinel-2 is made up of two satellites, Sentinel-2A and Sentinel-2B, that offer high-

resolution, multispectral imaging with a special blend of spectral and spatial capabilities. The spectral bands, including visible, near-infrared, and shortwave-infrared, offer critical information for characterizing vegetation and crop health. The spatial resolution, down to 10 meters, enables detailed monitoring of agricultural fields. Sentinel-2 images have given researchers a promising way to examine crop phenology (Ashourloo et al., 2022). Sentinel-2 satellite imagery provides precise tracking of the regional distribution patterns of various crop kinds, with a spatial precision of up to 10 m. During the crop-growing season, the five-day return interval aids in gathering phenology data. (G. Li et al., 2022).

Sentinel-2 imagery has several advantages for wheat mapping:

- ✓ Sentinel-2 imagery is available at a high spatial resolution (10-20 meters). This allows for the identification of individual wheat fields.
- ✓ Sentinel-2 imagery is available at a high temporal resolution (5 days). This allows for the monitoring of wheat growth and development.
- ✓ Sentinel-2 imagery captures spectral information in the visible, near-infrared, and shortwave infrared bands. This spectral information can be used to distinguish wheat from other crops and vegetation.

1.4.6 Machine Learning Accuracy general practice

Feature Screening

Feature screening is the process of identifying and removing features that are not informative or that are correlated with other features. A number of techniques, including recursive feature reduction, information gain, and correlation analysis, can be used for this. Feature screening is important for improving the performance of machine learning models because it reduces the noise in the data and simplifies the model (Feng et al., 2019). This can lead to improved accuracy, reduced overfitting, and increased interpretability.

- ✓ Importance: Identify and remove features that are not informative or that are correlated with other features.
- ✓ Benefits: Improve the performance of machine learning models by reducing the noise in the data and simplifying the model.
- ✓ Methods: Correlation analysis, information gain, recursive feature elimination.

Back Feature Engineering

The technique of generating new features from preexisting features to increase their prediction ability or better fit them for a certain machine learning algorithm is known as back feature engineering. This can be done using a variety of methods, such as principal component analysis, spectral index calculation, and spatial filtering. Back feature engineering is important for improving the accuracy of machine learning models because it can create new features that are more informative than the original features. For machine learning algorithms that are sensitive to the distribution of the features, this can be quite helpful.

1. Importance: Create new features from existing features to improve their predictive power or to make them more suitable for a particular machine learning algorithm.
2. Benefits: Improve the accuracy of machine learning models by creating more informative features.
3. Methods: Principal component analysis, spectral index calculation, spatial filtering.

Some specific feature that helps in crop mapping

i. **Spectral reflectance**: The spectral reflectance of wheat crops varies over the growing season. This variation can be used to distinguish wheat crops from other land cover types.

ii. **NDVI (Normalized Difference Vegetation Index)**: An indicator of spectral sensitivity to plant greenness is the NDVI. Wheat crops are typically greener than other land cover types, so the NDVI can be used to identify wheat crops.

iii. **EVI (Enhanced Vegetation Index)**: The EVI is a modified version of the NDVI that is more sensitive to wheat crops.

iv. **SWIR (Shortwave Infrared)**: The amount of water in plants may be detected using the SWIR bands. Wheat crops have a higher water content than other land cover types, so the SWIR bands can be used to identify wheat crops.

v. **Temporal features**: Temporal features can be used to track the changes in wheat crops over time. For example, the NDVI can be used to track the growth and development of wheat crops.

Specific Benefits for Wheat Mapping in Nepal

- ✓ Improve the accuracy of wheat yield estimates.
- ✓ Support sustainable wheat cultivation practices.
- ✓ Identify areas of wheat cultivation that are most vulnerable to climate change.
- ✓ Develop strategies to mitigate the impact of climate change on wheat production.

Accurate and reliable wheat mapping is critical for agricultural management, food security, and efficient resource allocation. With the increasing availability of high-dimensional data like satellite imagery, machine learning algorithms play a crucial role in this process. However, choosing the optimal algorithm requires a nuanced understanding of the specific challenges associated with wheat mapping and the distinct characteristics of different machine learning approaches.

Two prominent contenders for wheat mapping are Support Vector Machines (SVM) and Random Forest (RF). Both offer unique strengths and weaknesses, making a comprehensive comparison essential.

SVM excels in handling complex decision boundaries, a key requirement for wheat mapping. This is particularly advantageous for analyzing high-dimensional satellite imagery, where intricate patterns and subtle spectral variations differentiate wheat from other land cover types and crops. SVM's ability to capture these subtle variations translates to higher classification accuracy.

RF, on the other hand, leverages the power of ensemble learning, combining multiple decision trees for enhanced robustness and generalization. This approach proves valuable in dealing with the dynamic nature of agricultural landscapes and varying environmental conditions that can impact wheat growth. RF's adaptability and ability to generalize well to unseen data are crucial for achieving reliable results in real-world scenarios.

To identify the optimal algorithm, a comparative analysis delving into various factors is necessary. This includes:

Classification Accuracy: quantifying the ability to correctly distinguish wheat from other land cover types and different growth stages.

Computational Efficiency: analyzing the time and resources required for training and deploying each algorithm, especially important for large-scale applications.

Scalability: evaluating the ability of each algorithm to handle vast spatial extents and large datasets.

Model Complexity: understanding the interpretability of each algorithm and the ease of tuning its parameters for practical implementation.

1.5 Data and Software

1.5.1 In-Situ Data

For gathering in-situ data in the field, we teamed up with surveyor from CIMMYT and visited the farmland in the study area. To address the crop land accurate details, we interviewed farmer with series of questionnaire related to the cropland. We worked together to collect specific information about crops in a study area. We used a survey form on the Kobo Toolbox app to gather crop specific data. We made sure to address different questions and consider the layout of crop land during our data collection. To make sure our data was accurate and efficient, we chose Kobo Toolbox as our data collection platform. This decision, along with the support of our partners, played a crucial role in ensuring the quality of our dataset. After collecting the data, we carefully cleaned it using ArcGIS Pro. We also compared our survey data with satellite imagery to validate our results. This collaboration and the tools we used helped us create a precise and complete dataset, providing a strong foundation for our research. The ground data collection flow can be explained in detail in following chart figure 1.

Figure 1: In Situ data collection steps

There was total 359 farmland polygon is collected for the study. The farmland is chosen based on the preliminary farm study done by CIMMYT surveyors also considering the yearly survey samples. The crop distribution in the in-situ data revealed wheat as the dominant crop, constituting approximately 60.3% of the total dataset. Lentils followed with a significant 17.5%, while maize, mustard, and vegetables contributed 5.6%, 4.7%, and 4.1%, respectively. Banana, potato, and shrub tree comprised smaller proportions at 5%, 1.9%, and 0.9%, respectively. to Further details on the ground data collection can be found in figure below.

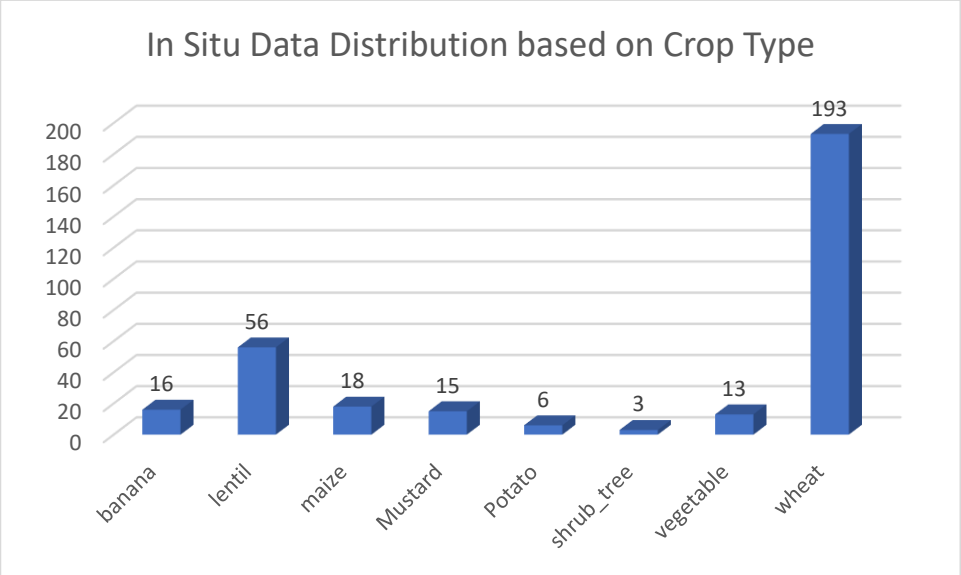


Figure 2: In Situ Data Distribution based on Crop Type

1.5.2 Satellite data

This study leverages GEE's advanced capabilities to access and retrieve Sentinel-2 satellite imagery. It was found that sentinel 2 tiles; T44RNS and T44RMS fall in the study area. Sentinel-2 (2A and 2B) imagery, encompassing the seasonal period from November 2022 to April 2023, has been selected for its alignment with the study's objectives and based the phenology chart that is plotted from the survey data, offering a valuable temporal perspective. Table 1 represents the details about the satellite and relative information on

the study area. Furthermore, Also the satellite orbit is taken in consideration. Study area falls on orbit 19 and orientation of orbit descending. Consideration of satellite orbit ensures optimal data capture conditions. This holistic approach, marked by meticulous consideration and strategic technology utilization, underpins the study's goal of enhancing efficiency wheat area mapping. Following are the satellite imagery data summary:

Table 1: Satellite data attribute on study area

Satellite	Sentinel 2 (2A and 2B)
Season date	November 2022 to April 2023
Orbit number	19
Satellite Orbit	Descending
Tile	T44RNS and T44RMS
Spectral bands	13 bands (4 visible and near-infrared bands and 6 shortwave infrared)
Spatial Resolutions	10 to 60 meters
Temporal resolutions	5 days

1.5.3 Google Earth Engine (GEE)

Google Earth Engine (GEE) is a cloud-based platform for planetary-scale geospatial analysis. Being an integrated platform that not only empowers conventional remote sensing experts but also a much larger audience that lacks the technical competence to use traditional supercomputers or large-scale commodity cloud computing resources, it is unique in the sector (Gorelick et al., 2017). Google provides customers with access to a vast collection of satellite images and geospatial information, including priceless sources like Sentinel-2 data. This scalable platform supports complex geospatial analyses, offers a

diverse set of analysis tools, and includes an integrated code editor for custom scripting in JavaScript or Python.

1.5.4 ArcGIS Pro

ArcGIS Pro, developed by Esri, is a robust and comprehensive geospatial analysis software renowned for its versatility in handling geospatial data. It plays a pivotal role in this study by serving as a crucial tool for the cleaning and pre-processing of ground truth data collected within the study area. ArcGIS Pro's capabilities enable meticulous data refinement, ensuring the accuracy and consistency of the collected information. This software offers a suite of geoprocessing tools, data management functionalities, and spatial analysis capabilities that empower researchers to address complex geospatial challenges. Its integration with various data formats and compatibility with external data sources make it an essential component of the research workflow, facilitating the transformation of raw field data into reliable, analysis-ready datasets. Thus, ArcGIS Pro stands as a cornerstone in the research process, contributing to the quality and reliability of the data used for subsequent analysis and interpretation.

1.6 Study Area

The study area is in the western Terai region of Nepal, encompassing 6 municipalities spanning two districts: Kailali and Bardiya. Table 2 list out all the municipalities and respective district. The geographical extent of the study area is defined by coordinates, with latitude range from 28° 21' 49'' to 28° 37' 26'' the longitude ranges from 81° 15' 44'' to 80° 53' 33". This region is characterized by the fertile plains of the Terai, featuring a mix of flatlands and low-lying areas. Geographically, it constitutes a vital part of the Terai landscape, renowned for its agricultural significance. The study area's vast extent encompasses diverse topographical features, and it experiences distinct wet and dry seasons, influencing land use patterns, cropping cycles, and agricultural practices. Understanding the geographical and seasonal intricacies of this region is pivotal for

interpreting the outcomes of the research, which primarily focuses on land use and crop analysis.

The study area experiences a subtropical climate with distinct wet and dry seasons. Average temperatures vary, with warm to hot conditions during the dry season and cooler temperatures in the wet season. Agricultural area is mostly fertile, and the fertile plains are characterized by alluvial soils, primarily conducive to agriculture. The landscape includes flatlands, valleys, and low-lying areas, forming an integral part of the Terai's diverse topography. Agriculture, a key focus, involves both irrigated and rainfed practices, with smallholder farming prevailing. Fields vary in size, and crops like rice, wheat, sugarcane, and assorted fruits and vegetables are commonly grown. Cereal crop is most dominant crop in the area. Rice in the summer and wheat in the winter

Table 2: List of local municipalities in study area

SN	District	Local Unit	Municipality Type
1	BARDIYA	Geruwa	Gaunpalika
2	BARDIYA	Rajapur	Nagarpalika
3	KAILALI	Bhajani	Nagarpalika
4	KAILALI	Janaki	Gaunpalika
5	KAILALI	Joshipur	Gaunpalika
6	KAILALI	Tikapur	Nagarpalika

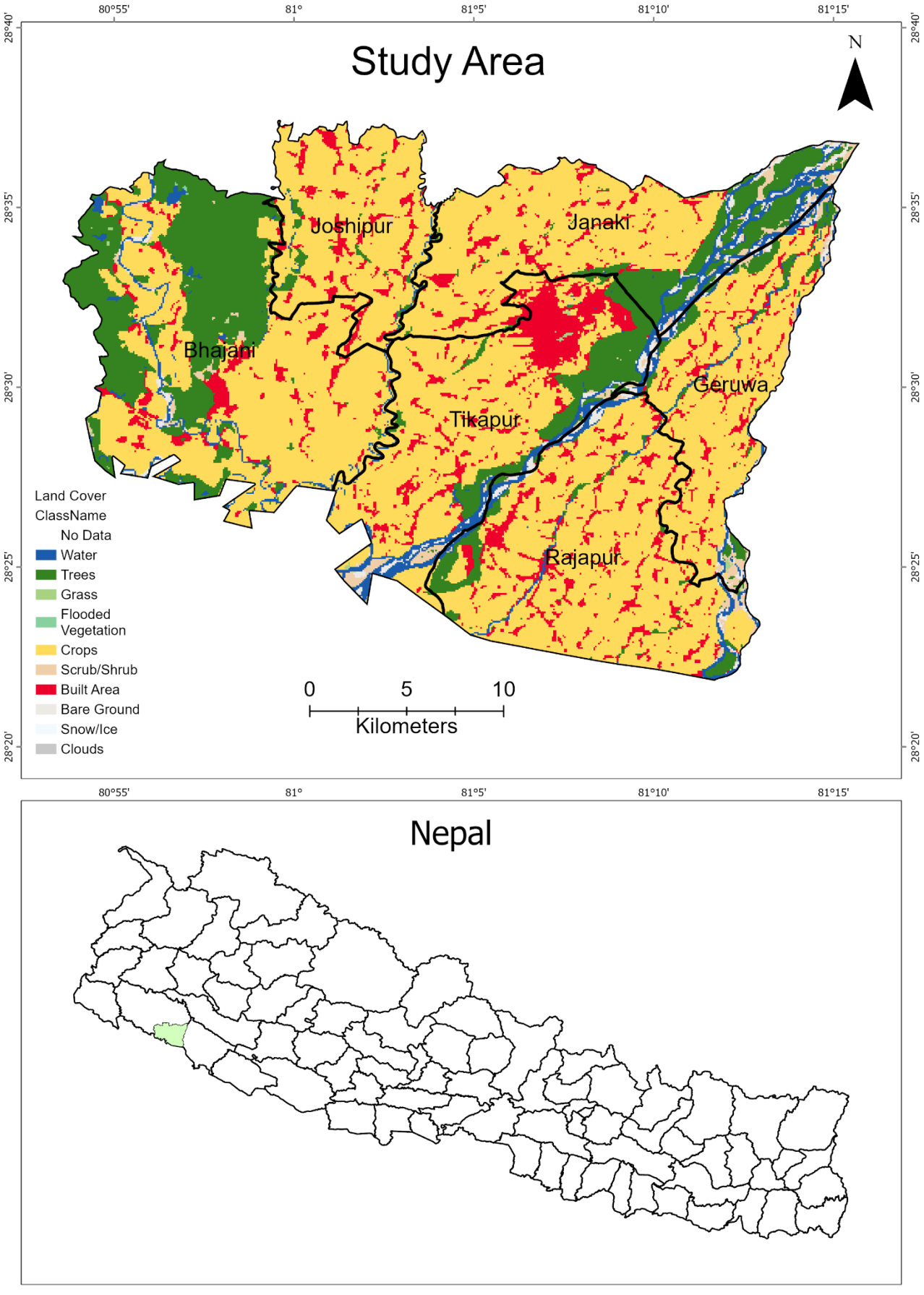


Figure 3: Map of study area

Chapter-2: Methodology

The methodology for this research involves a comprehensive approach to investigate agricultural dynamics in the study area to map cropland followed by the step in the figure 3.

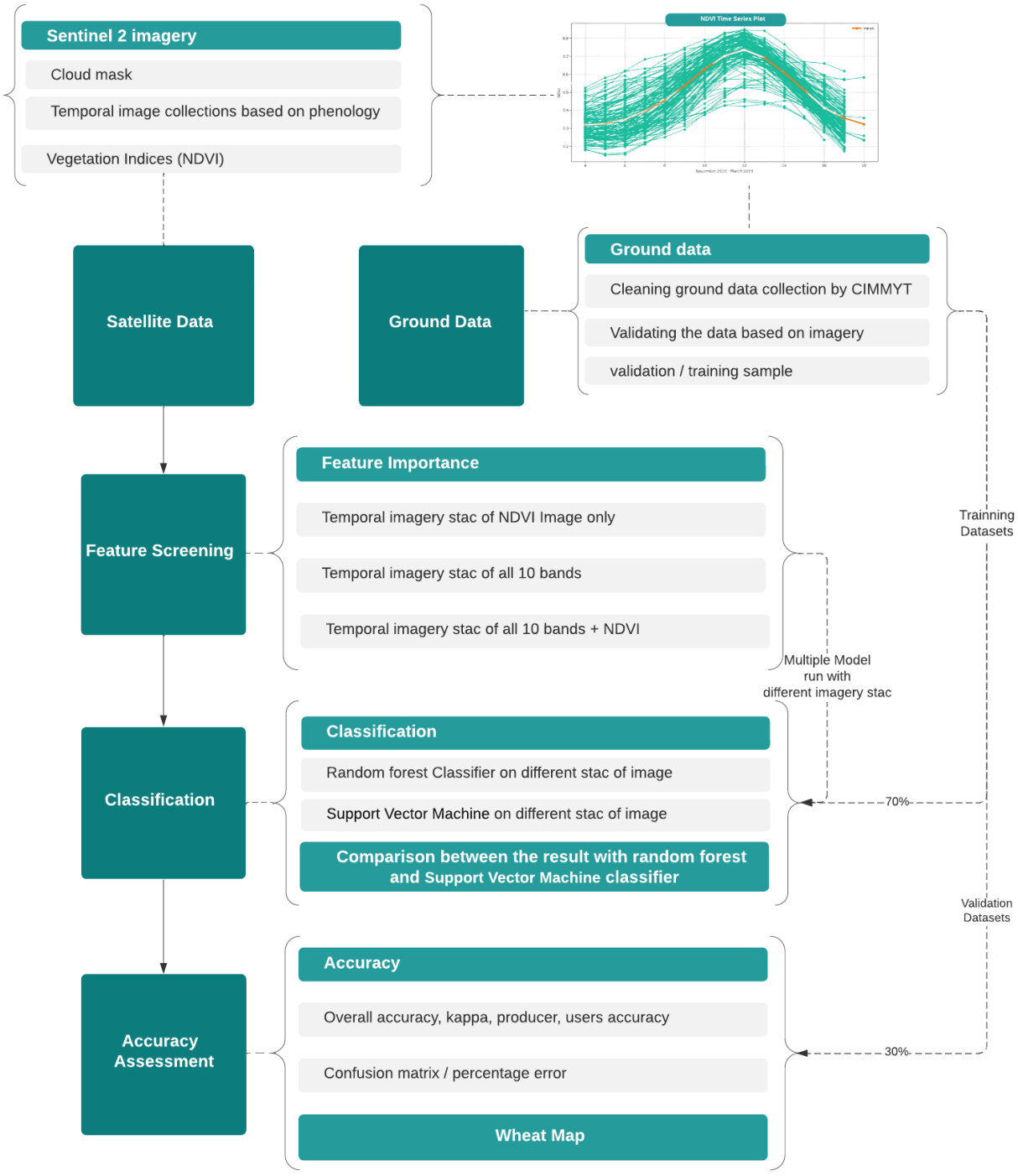


Figure 4: Methodology flow diagram

2.1 Data Acquisition

In this study, two types of data sources are utilized: satellite data and in-situ data. For satellite data, multi-temporal Sentinel-2 imagery is obtained, providing a comprehensive view of the study area's land cover and vegetation dynamics over time. The imagery's orbital properties for Level-2A (L2A) and Level-2B (L2B) data are carefully checked to ensure proper alignment and suitability for the analysis. Also taken in the consideration of tile and satellite orbit and its orientation. Tile T44RNS and T44RMS falls in the study area and satellite orbit is 19 and orientation is descending as represented in figure 4 and table 4. Satellite Imagery with cloud cover below 10% is chosen for more reliable results. Cloud masks and gap filling techniques are applied to address cloud cover and data gaps, ensuring a continuous and reliable dataset.

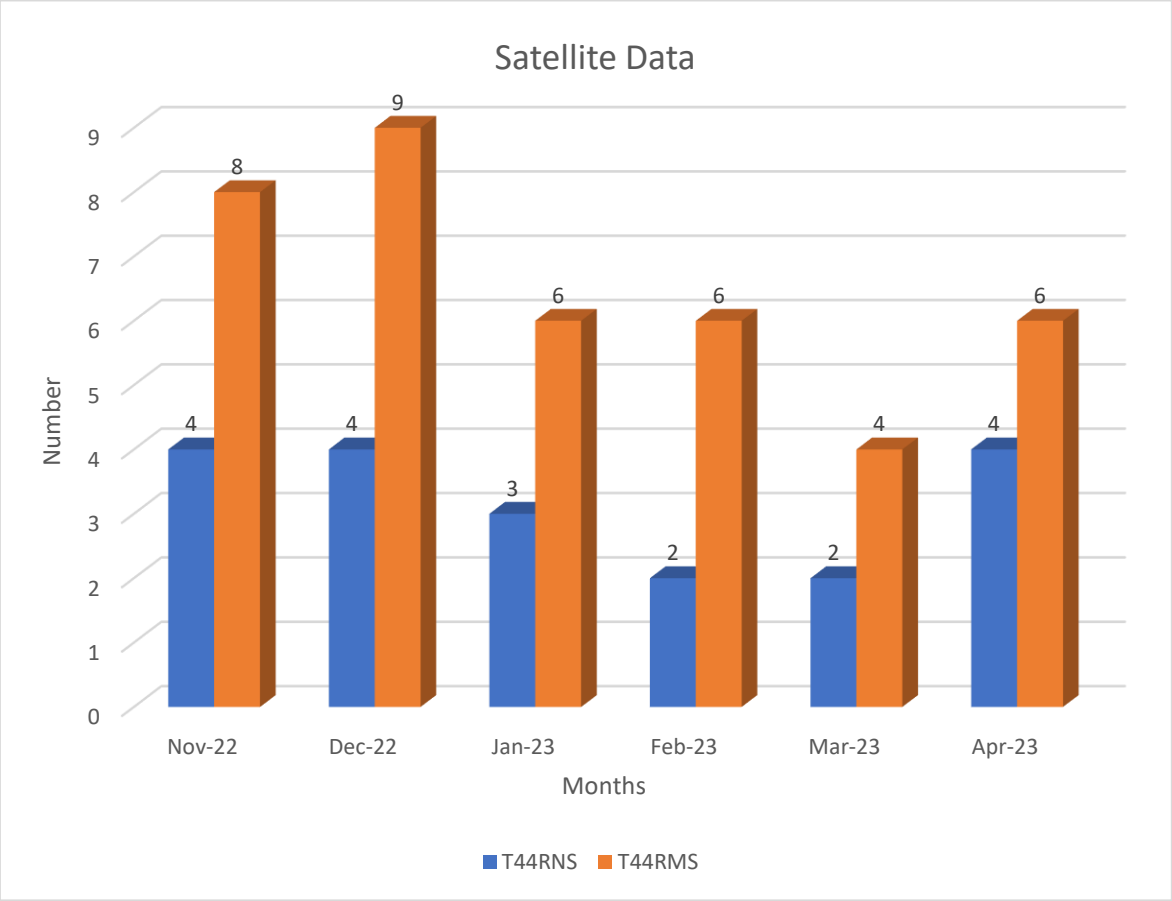


Figure 5: Sentinel 2 Imagery monthly availability chart

Table 3: Sentinel 2 imagery date availability base on T44RMS and T44RNS Tile

Months	T44RNS	Months	T44RMS
Nov-22	11/4/2022	Nov-22	11/4/2022
	11/19/2022		11/12/2022
	11/24/2022		11/17/2022
	11/29/2022		11/19/2022
Dec-22	12/4/2022		11/22/2022
	12/9/2022		11/24/2022
	12/14/2022		11/27/2022
	12/19/2022		11/29/2022
Jan-23	1/13/2023	Dec-22	12/2/2022
	1/18/2023		12/4/2022
	1/23/2023		12/7/2022
Feb-23	2/7/2023		12/9/2022
	2/22/2023		12/12/2022
Mar-23	3/4/2023		12/14/2022
	3/9/2023		12/17/2022
Apr-23	4/3/2023		12/19/2022
	4/13/2023		12/22/2022
	4/18/2023	Jan-23	1/13/2023
	4/23/2023		1/18/2023
			1/21/2023
			1/23/2023
			1/26/2023
			1/31/2023
		Feb-23	2/7/2023
			2/10/2023
			2/12/2023
			2/15/2023
			2/22/2023
			2/25/2023
		Mar-23	3/2/2023
			3/12/2023
			3/22/2023
			3/27/2023
		Apr-23	4/3/2023
			4/6/2023
			4/11/2023
			4/13/2023
			4/16/2023
			4/18/2023

It is found that the two Sentinel 2 tile have different Temporal revisit time. Tile T44RNS have around 5 days temporal revisit and tile T44RMS has 2 to 3 days temporal revisit. **Due to which 5 days temporal mosaic is taken for the study with reference in tile T44RMS**.

In situ data plays crucial role in training, validating, and enhancing the accuracy of our satellite-derived insights. Field surveys is conducted to gather crop information to compare with our satellite observations. Figure 5 maps out the field data distribution across the study area. This 'ground truth' data serves as a vital reference point, allowing us to calibrate and refine our satellite imagery for machine learning. By combining these two data sources satellite and ground truth data, study enhance the comprehensive and reliable analysis of land use and cropland dynamics within our study area. Following map represents the in-situ data spatially distributed in study area.

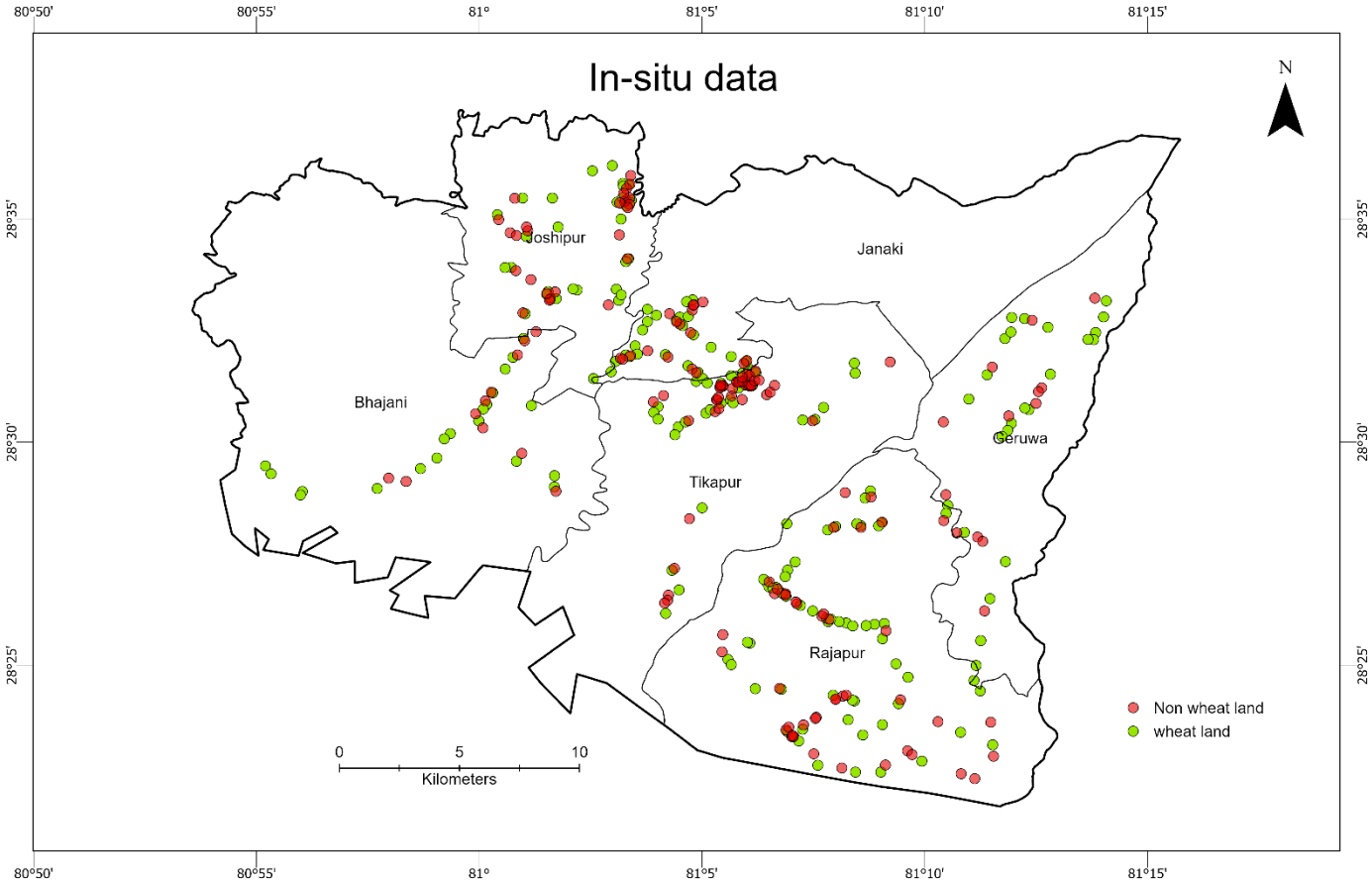


Figure 6: In-situ data map

2.2 Pre-processing and Feature Screening

The initial step in this study involved the acquisition of Sentinel satellite images spanning from November 2022 to April 2023, ensuring a comprehensive temporal coverage. Subsequently, an image collection was systematically curated from the obtained dataset. Cloud masking techniques were applied, capitalizing on the advantageous meteorological conditions during the winter season, which yielded cloud-free imagery. The pre-processed imagery was then used to compute vegetation indices, such as the Normalized Difference Vegetation Index (NDVI). Crop health and condition are measured by NDVI. This vegetation index, which measures greenness, has a significant relationship with green biomass, a measure of growth. Ground truth data were then fused with satellite imagery to plot NDVI charts for all the acquired images, facilitating the construction of a phenology timeline. This involved comprehensive statistical analyses and multiple runs within the GEE platform to identify the most pertinent features and bands for subsequent classification tasks as represented in figure 6.

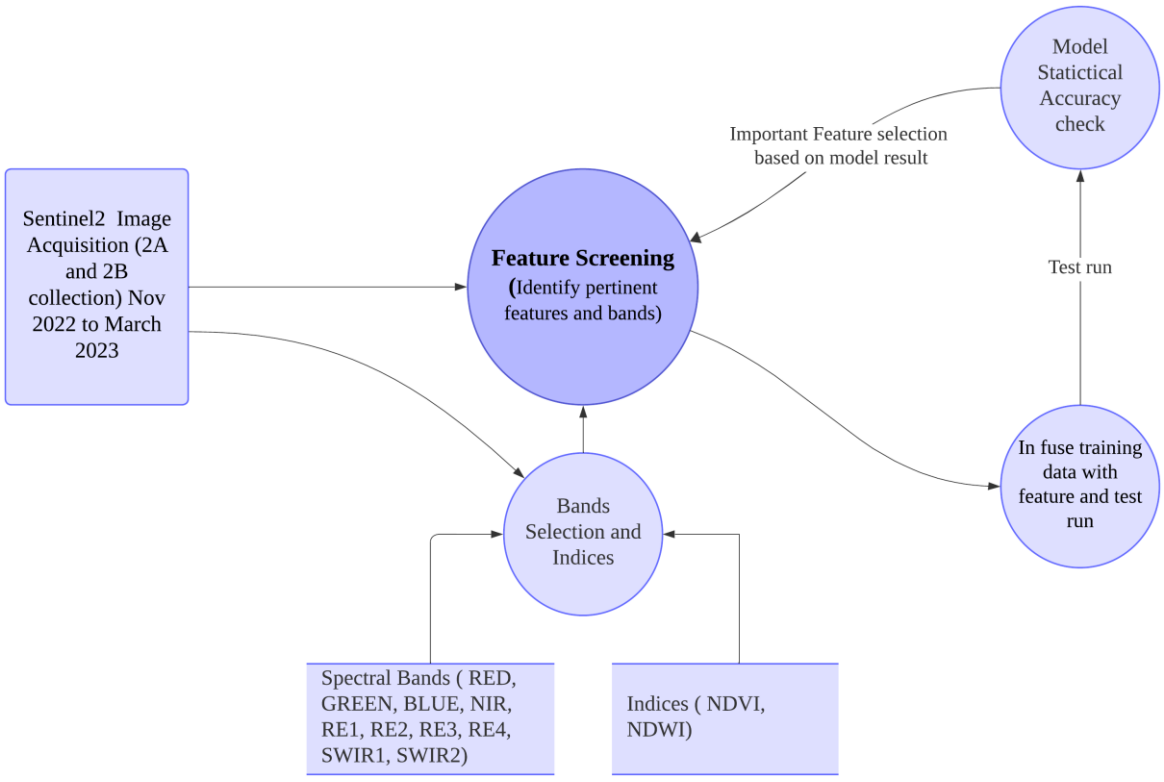


Figure 7: Feature screening chart

Plotted the NDVI chart from November 2022 to April 2023 for all the different crop in study area to get the temporal information. NDVI is a satellite-derived index that measures the amount of green vegetation in each area. NDVI values typically range from -1 to 1, with higher values indicating more green vegetation.

The NDVI chart for wheat in figure 7 shows that the NDVI values increase from November 2022 to April 2023. This indicates that the crops are growing and developing during this period. However, the rate of increase in NDVI values varies depending on the crop type. Wheat shows the most rapid increase in NDVI values, from around 0.3 in November to around 0.8 in February. This is because wheat is a winter crop that grows actively during the cool season. Wheat plants begin to emerge from the soil in November and December, and they reach their peak growth during the winter months. By April, the wheat plants are beginning to mature.

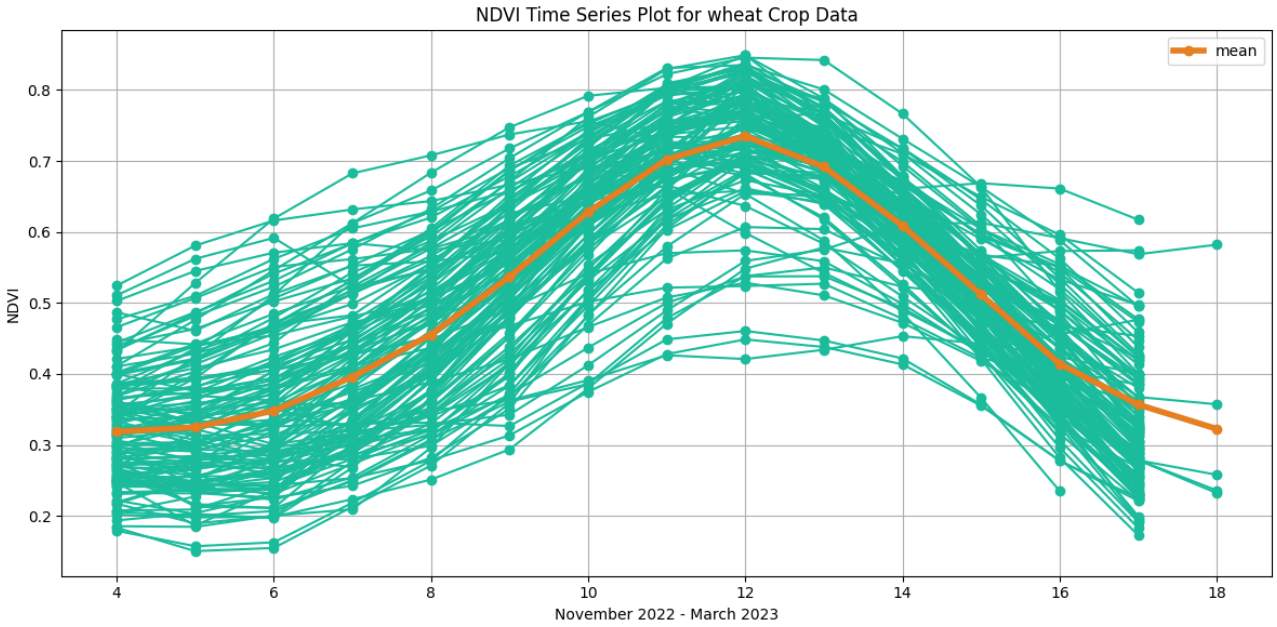


Figure 8: NDVI time series plot for wheat crop

Maize plot in figure 8 also shows a constant in NDVI values, from around 0.4 in November to around 0.5 in April. However, the rate of increase is slower than for wheat. This is because maize is a summer crop that does not begin to grow actively until the warm season begins. Maize plants are typically planted in March and April, and they reach their peak growth during the summer months. By April, the maize plants are still in their early stages of growth.

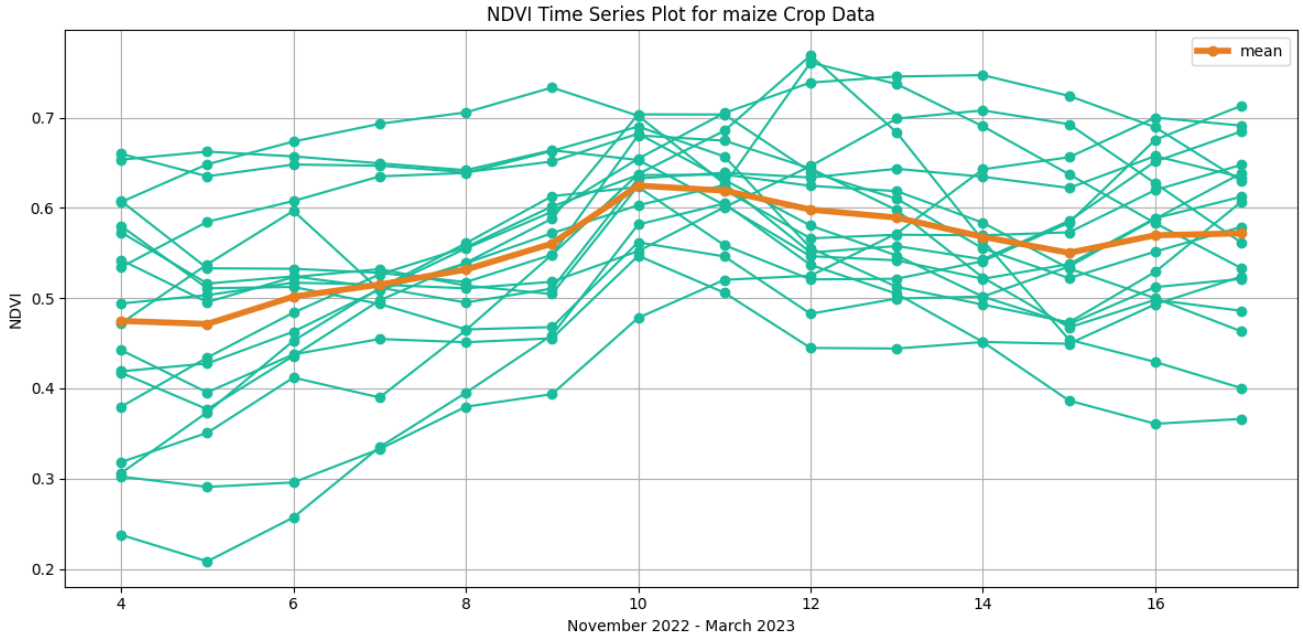


Figure 9: NDVI time series plot for maize crop

The NDVI chart for mustard in figure 9 shows that the NDVI values increase from November 2022 to April 2023. Mustard shows a similar pattern to wheat, with a steady increase in NDVI values from November to April. However, the overall NDVI values are lower than for wheat. This is because mustard is a relatively low-growing crop with a low leaf area index. Mustard plants are typically planted in October, and they reach their peak growth during the winter months.

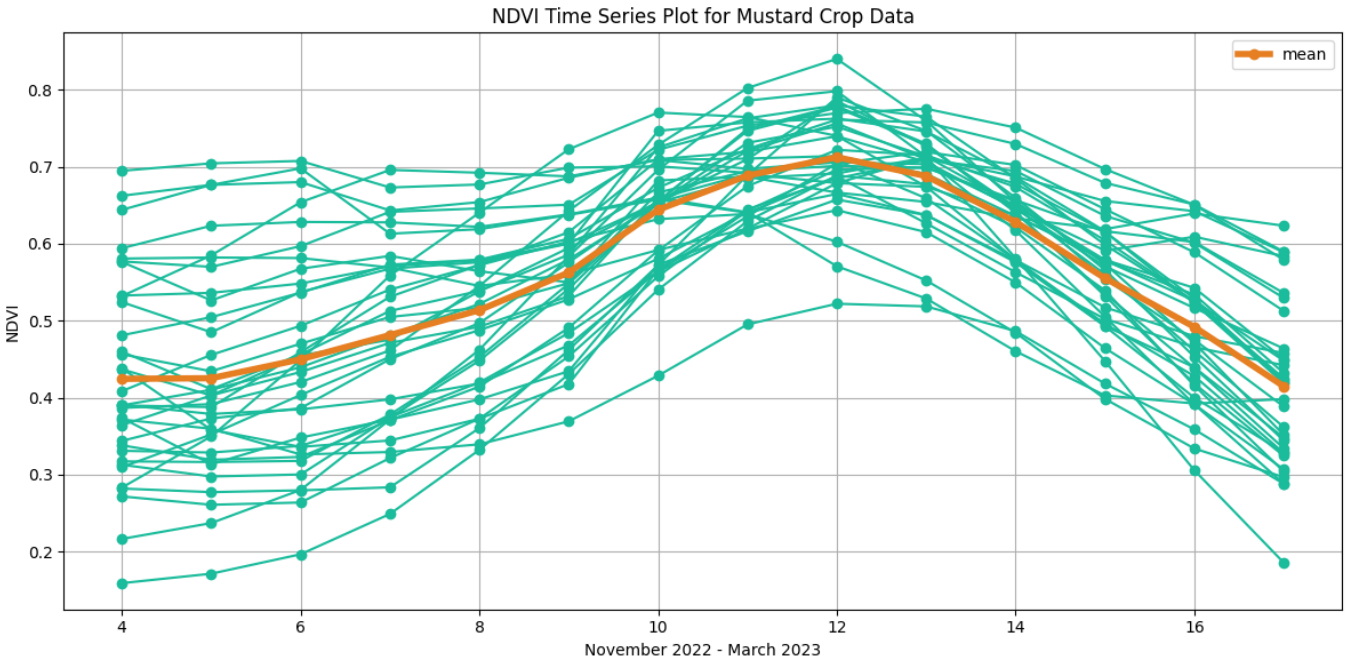


Figure 10: NDVI time series plot for mustard crop data

Potato shows a more gradual increase in NDVI values than the other crops, with the NDVI values increasing from around 0.4 in November to around 0.6 in April as shown in figure 10. This is because potatoes are a tuber crop, and the NDVI values are less influenced by the growth of the above-ground vegetation. Potato plants are typically planted in November and December, and they reach their peak growth during the spring months.

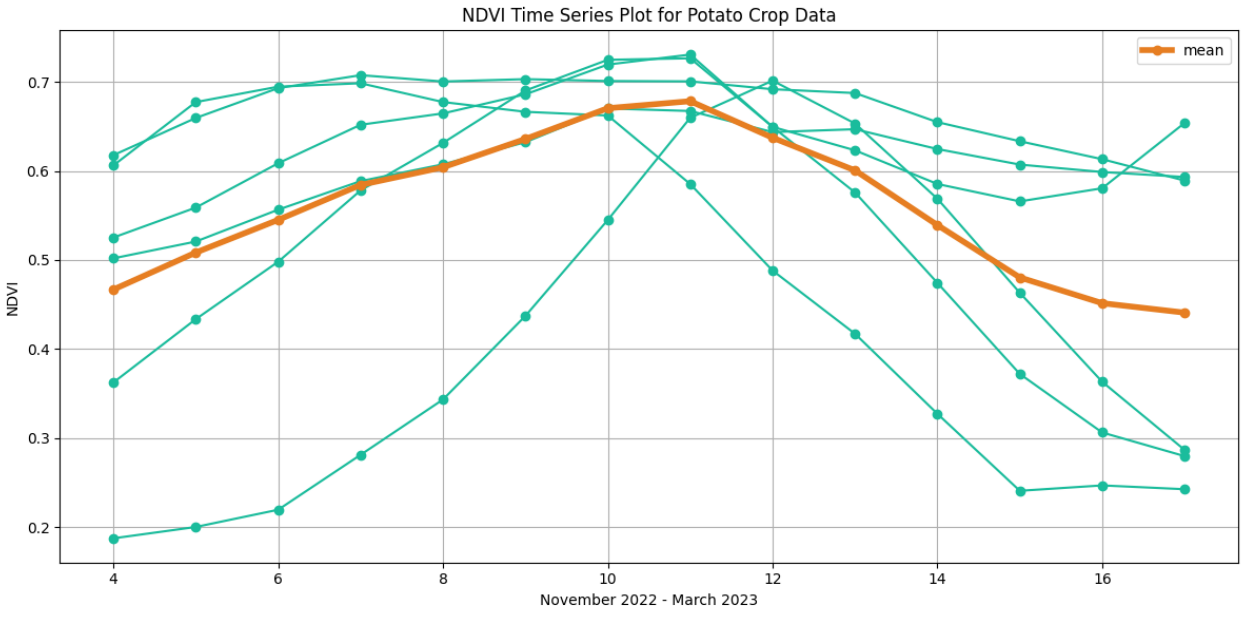


Figure 11: NDVI time series plot for potato data

Lentil shows a similar pattern to Mustard as shown in figure 11, with a gradual increase in NDVI values from November to April. However, the overall NDVI values are lower than for wheat. This is because lentils are a relatively low-growing crop with a low leaf area index. Lentil plants are typically planted in October and November, and they reach their peak growth during the winter months. By April, the lentil plants are beginning to flower and produce seed.

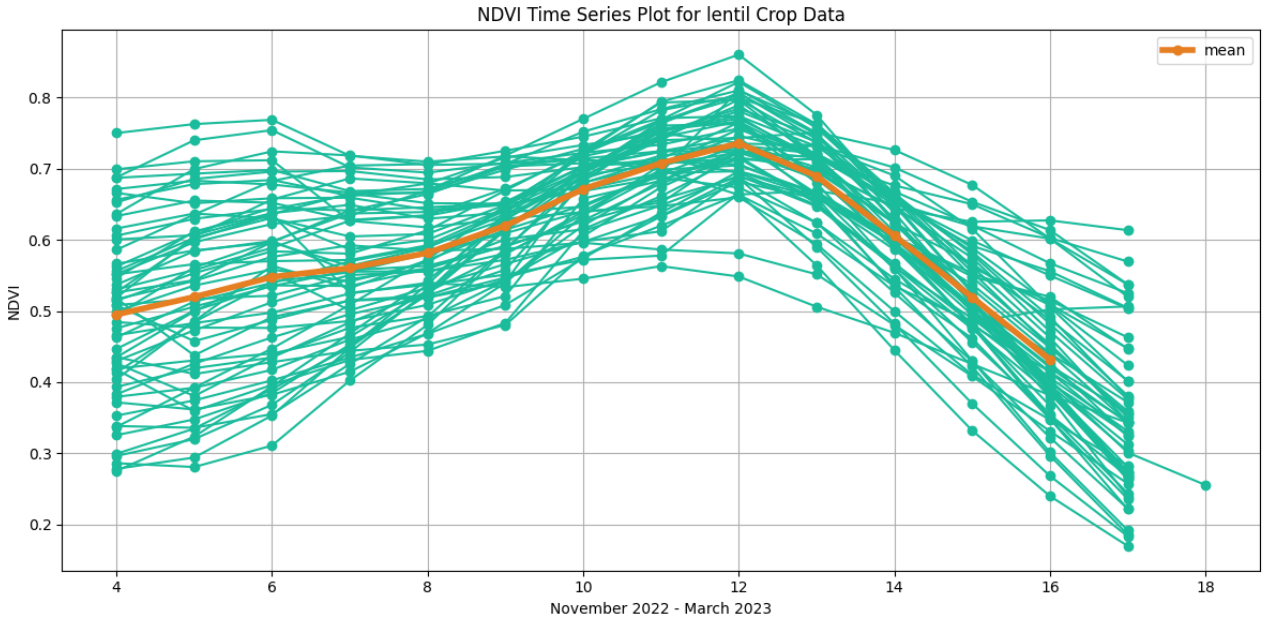


Figure 12: NDVI time series plot for lentil crop data

Shrubs show a relatively small increase in NDVI values from November and decrease on winter month like December, January and increase on April as shown figure 12. This is because shrubs are evergreen plants that maintain their leaves throughout the year with fluctuating NDVI values based on the season.

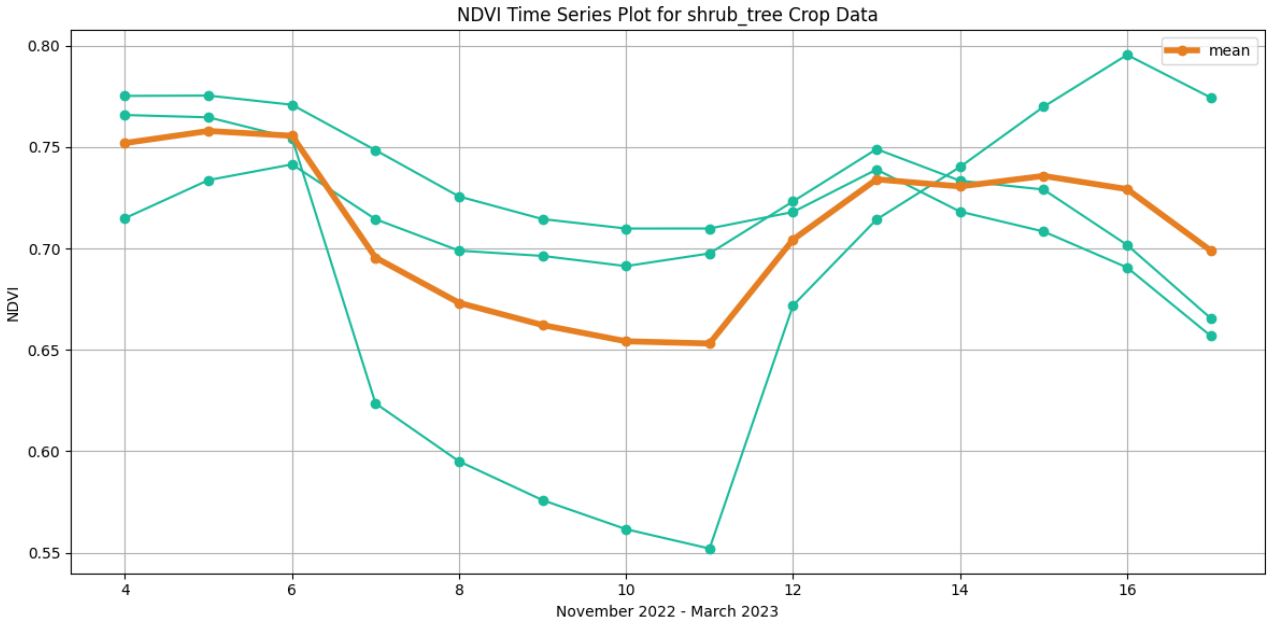


Figure 13: NDVI time series plot for shrub tree

Vegetables show a variety of different patterns, depending on the type of vegetable. For example, leafy vegetables such as spinach and lettuce show a rapid increase in NDVI values, while root vegetables such as carrots and beetroot show a more gradual increase in NDVI values as shown in figure 13.

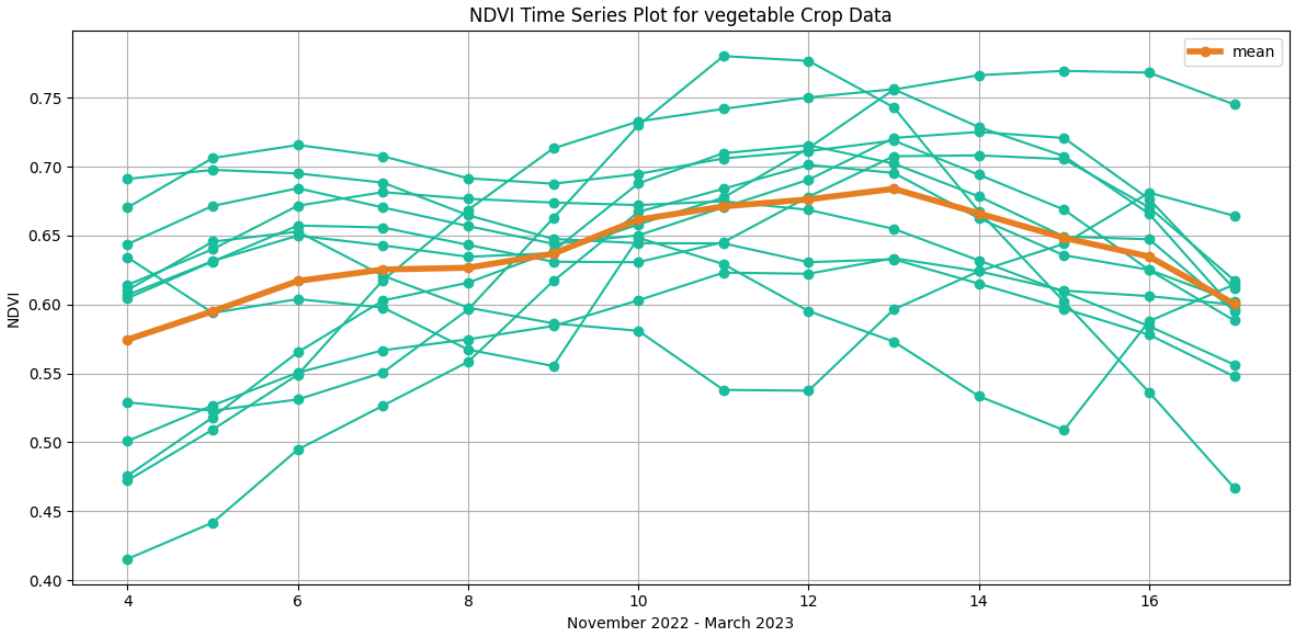


Figure 14: NDVI time series plot for vegetable

Banana plants are growing and developing during November to April period. However, the rate of increase in NDVI values is slower than for some other crops, such as wheat and maize as shown in figure 14. This is because banana plants are perennial plants that grow continuously throughout the year. The NDVI chart shows that the NDVI values for banana are relatively high, even in November. This is because banana plants have a large leaf area index, which means that they reflect a lot of sunlight in the near-infrared region of the electromagnetic spectrum.

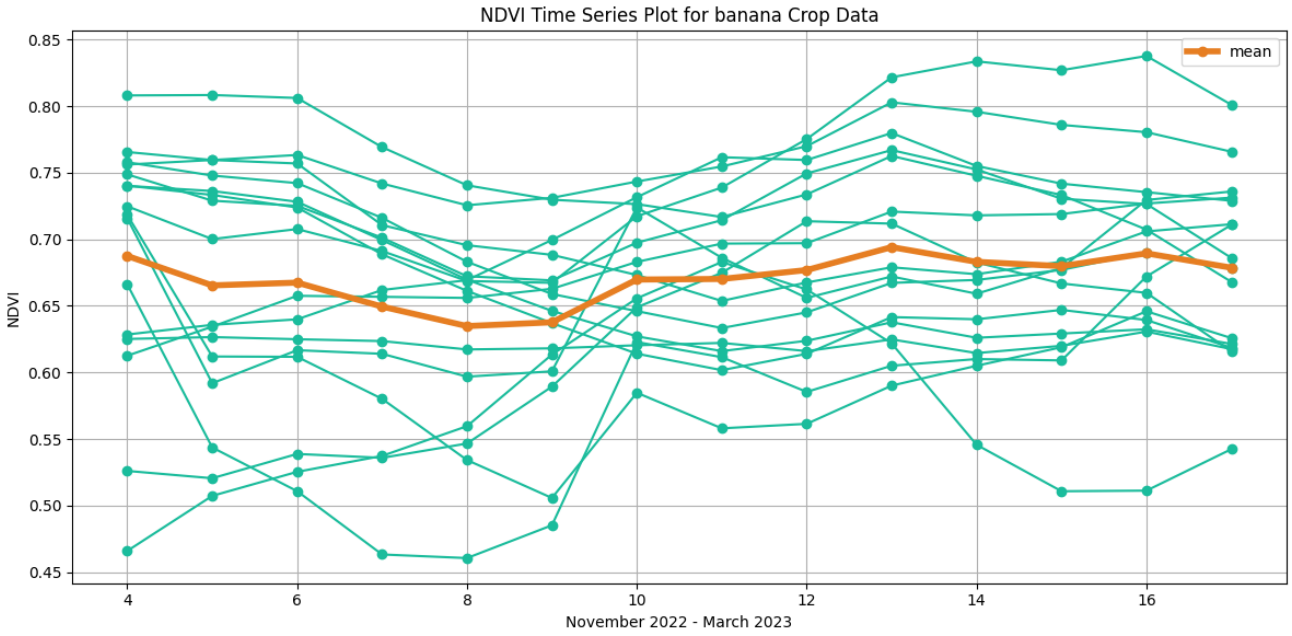


Figure 15: NDVI time series plot for banana

Overall, the NDVI chart provides valuable insights into the growth and development of different crops in from November 2022 to April 2023. The NDVI chart can be used to classify different types of crops. For example, crops with high NDVI values, such as wheat, can be classified as cereal crops. With the help of NDVI chart the temporal information can be extracted and which later can be significate for the crop classification. Also, it can be useful to select the temporal image which will play vital role in the crop identification for any machine learning algorithm.

The snippet of code to use Sentinel-2 (S2) Earth observation satellite data to produce a time series chart of normalized difference vegetation index (NDVI) values for a given area of interest (ROI). The NDVI is a widely used indicator of vegetation health, with values ranging from -1 to 1. Below code first calculates the mean NDVI value for each pixel within the ROI over a specified time. This mean NDVI value is then plotted against the corresponding date, resulting in a time series chart that shows how the NDVI has changed over time. The chart is titled "NDVI time series" and the axes are labeled accordingly.

```
// code for the NDVI chart plot
var plot_NDVI = ui.Chart.image.seriesByRegion(S2, geometry,ee.Reducer.mean(),
        'nd',500,'system:time_start', 'system:index')
                .setChartType('LineChart').setOptions({
                  title: 'NDVI time series',
                  hAxis: {title: 'Date'},
                  vAxis: {title: 'NDVI'}
        });
```

2.3 Classification

The basic aim is to use a multi-pronged strategy to accurately classify wheat land and non-wheat land by utilizing the capabilities of two different but complimentary classification algorithms: the Random Forest Classifier and the Support Vector Machine (SVM) and comparing the outcome.

2.3.1 Random Forest

A random forest is created by combining several tree predictors so that every tree in the forest is reliant on the values of a random vector that is randomly sampled and has the same distribution for every tree (Breiman, 2001). The Random Forest Classifier represents a versatile and robust tool in the classification of crops. It functions as an ensemble learning

method, capitalizing on the wisdom of crowds. Within this framework, multiple decision trees are trained independently on the spectral and vegetation index data acquired from various crop fields. Each decision tree is a specialist in discerning the unique spectral signatures and vegetation patterns associated with different crop types. By aggregating the predictions from these individual decision trees, the Random Forest Classifier excels in capturing the intricate relationships between the input features and crop classes. This ensemble approach not only enhances the accuracy of classification but also mitigates overfitting, making it an invaluable asset in your research.

2.3.2 Support Vector Machine (SVM)

The Support Vector Machine provides a different approach to crop categorization in addition to the Random Forest Classifier. SVM functions by identifying the best hyperplanes in the feature space to serve as borders between various crop classes. The SVM algorithm is the optimal decision boundary or line that can divide n-dimensional space into classes, making it simple to place fresh data points in the appropriate category later on (Evgeniou & Pontil, 2001). In the context of research, the feature space encompasses the spectral and vegetation index information. By identifying these hyperplanes, SVM effectively carves out decision boundaries that maximize the separation between crop classes. This attribute makes SVM particularly adept at handling datasets with complex, non-linear relationships. Furthermore, SVM's flexibility in using kernel functions enables it to map data into higher-dimensional spaces, enhancing its capacity to discern intricate patterns in the input data.

These two classification algorithms broaden the scope of investigation but also offers a unique opportunity for comparison. By subjecting the same dataset to both the Random Forest Classifier and SVM, you can discern the nuances in their performance characteristics. This comparative analysis can shed light on which algorithm is better suited for specific crop classification scenarios, potentially providing valuable insights for practitioners in the field of agriculture and remote sensing. The comparative evaluation of these algorithms not only enhances understanding of their applicability in agriculture but

also contributes to the ever-evolving landscape of precision farming and remote sensing technologies.

2.4 Accuracy Assessment

To evaluate the performance of the classification models, various accuracy metrics are computed. These include overall accuracy, Kappa coefficient, producer accuracy, and user accuracy.

2.4.1 Overall Accuracy

Overall accuracy is a fundamental metric used to gauge the overall correctness of your classification models. In relation to the total number of samples in dataset, it indicates the percentage of accurately identified samples (crop instances). A high overall accuracy score indicates that your models are proficient at classifying crops accurately across all classes.

Table 4: Random Forest and Support Vector Machine comparison

Characteristic	**Random Forest**	**Support Vector Machine**
Algorithm type	Ensemble learning	Discriminative learning
Training time	Relatively fast	Relatively slow
Prediction time	Very fast	Very fast
Interpretability	Good	Fair
Robustness to noise	Very good	Very good
Accuracy	High	High

2.4.2 Kappa Coefficient

Reliability between the model's predictions and the actual class labels is evaluated using the Kappa coefficient, also called Cohen's Kappa, which takes chance agreement into consideration. When working with unbalanced datasets or attempting to gauge the model's performance while accounting for random chance, it is very helpful. Kappa values vary from -1 to 1, where a value of 1 denotes complete agreement, a value of 0 denotes agreement by chance, and a value of negative indicates disagreement. A Kappa coefficient greater than 0 suggests that the model is performing better than random chance.

2.4.3 Producer Accuracy

Producer accuracy (or Sensitivity) measures the model's ability to correctly classify samples belonging to a specific class. It assesses how well the model identifies a particular crop class among all the instances of that class in the dataset. Producer accuracy is particularly valuable for assessing the model's performance on individual classes and identifying potential areas of improvement.

2.4.4 User Accuracy

User accuracy measures the precision of the model's predictions for a specific class. It assesses how well the model's predictions match the actual class labels for a specific crop class. High user accuracy suggests that the model is probably right when it predicts a given class.

In conclusion, it is critical to evaluate classification models using metrics such as overall accuracy, Kappa coefficient, producer accuracy, and user accuracy to fully evaluate their performance, guarantee the accuracy of classification results, and obtain knowledge about areas in which your crop classification research needs to be improved.

Chapter-3: Processes and Results

3.1 Classification Results

3.1.1 Random Forest results

The Random Forest algorithm successfully separated wheat land and non-wheat land with highest overall accuracy 0.99 as represented in table 6.

Table 5: Random Forest classification result

Training accuracy	Validation accuracy	Consumer accuracy (wheat)	Consumer accuracy (non-wheat)
99%	86%	87.8%	86%

We tested different scenarios with different sets of data and satellite bands to address all the possible cases for the best accuracy. Most of scenarios effected the model in different result. In all the case NDVI is compute and other indices like EVI and NDWI is also take in account.

Following code snippet is used for the random forest classification in GEE.

```
//  Random Forest classifier
var classifier = ee.Classifier.smileRandomForest({
                    numberOfTrees: 1000, // You can adjust this number
                    variablesPerSplit: 5, // You can adjust this number
                    minLeafPopulation: 1 // You can adjust this number
                  });
```

Scenario 1: Four bands with all temporal images

We conducted a study to investigate the performance of a classifier for wheat mapping using four bands: red, blue, green, and NIR, as well as NDVI. The classifier was trained on a temporal image collection from November 2022 to April 2023. Unfortunately, the classifier performed poorly, with an overall accuracy of only 76.19% as represented in table 7 and 8. This suggests that the classifier is not able to generalize well to new data, and that additional features or training data may be needed to improve its performance.

Table 6: Random Forest classification prediction table for case four bands with all temporal images

Predicted class	True class	Count
Wheat	Wheat	32
Wheat	Non-wheat	13
Non-wheat	Wheat	12
Non-wheat	Non-wheat	48

Table 7: Random Forest classification accuracy table for case four bands with all temporal images

Training accuracy	Validation accuracy	Consumer accuracy (wheat)	Consumer accuracy (non-wheat)
99%	76%	72.7%	78.6%

Overall, the validation error matrix shows that the classifier is performing well on the validation dataset. However, there is still some room for improvement, especially in terms of distinguishing between wheat and non-wheat pixels.

Scenario 2: Ten bands with all temporal images

In another case we trained a classifier to classify wheat pixels from non-wheat pixels using all 10 bands of Sentinel-2 data (red, blue, green, re1, re2, re3, re4, swir1, swir2 and NIR, as well as NDVI. The classifier was trained on a temporal image collection from November 2022 to April 2023. The consumer accuracy for wheat is 63.6%, and the consumer accuracy for non-wheat is 90.9% as represented in table 9 and 10. This means that the classifier is better at identifying non-wheat pixels than wheat pixels. One possible explanation for the difference in consumer accuracy is that the classifier is trained on more non-wheat pixels than wheat pixels. This can cause the classifier to be biased towards non-wheat pixels, which can lead to more wheat pixels being misclassified as non-wheat.

Table 8: Random Forest classification prediction table for case ten bands with all temporal images

Predicted class	**True class**	**Count**
Wheat	Wheat	28
Wheat	Non-wheat	15
Non-wheat	Wheat	5
Non-wheat	Non-wheat	50

Table 9: Random Forest classification accuracy table for case ten bands with all temporal images

Training accuracy	**Validation accuracy**	**Consumer accuracy (wheat)**	**Consumer accuracy (non-wheat)**
96%	83%	63.6%	90.9%

Overall, the classifier has a good overall accuracy, but there is still some room for improvement, especially in terms of distinguishing between wheat from non-wheat pixels.

Scenario 3: Four bands with top ten important images

Next case we trained a classifier to classify wheat pixels from non-wheat pixels using all 4 bands of Sentinel-2 data (red, blue, green and NIR, as well as NDVI. The classifier was trained on a ten important image collection from November 2022 to April 2023 which was determined by the classifier. The number of pixels properly identified for each class divided by the total number of pixels projected to be in that class yields the consumer accuracy for each class. In this case, the consumer accuracy for wheat is 36 / (36 + 7) = 87.8, and the consumer accuracy for non-wheat is 43 / (43 + 5) = 86 % as represented in table 11 and 12. By dividing the total number of pixels in the validation dataset by the number of properly identified pixels, the classifier's overall accuracy is determined. In this case, the overall accuracy is (36 + 43) / (36 + 7 + 5 + 43) = 86%. This means that the classifier performed better at identifying wheat pixels than non-wheat pixels.

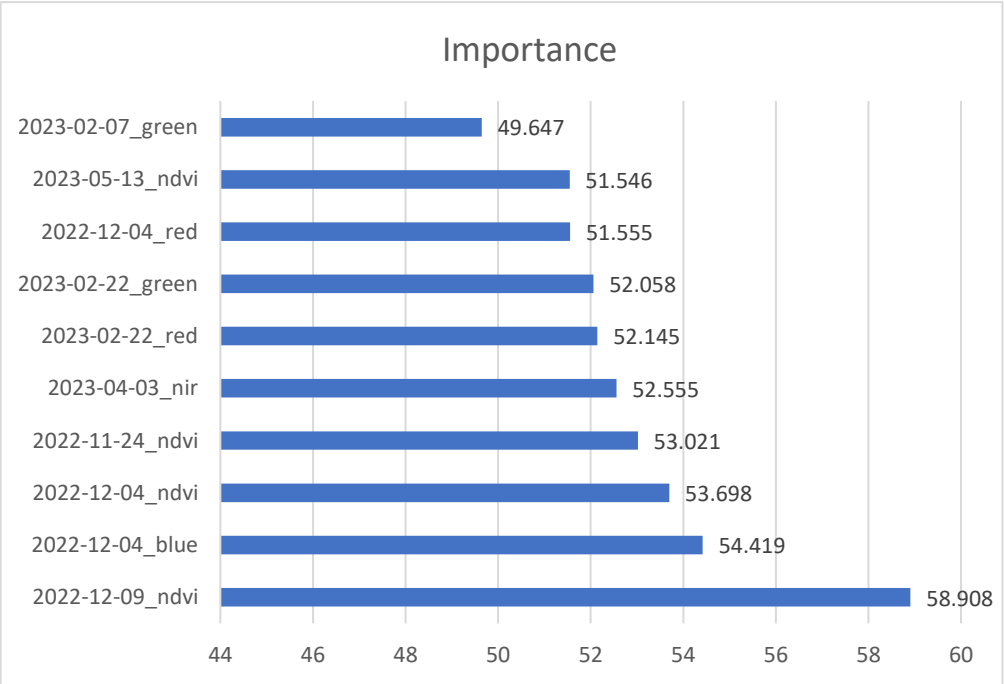


Figure 16: Top ten important feature from RF classifier based on four bands.

Table 10: Random Forest classification prediction table for case four bands with top ten images

Predicted class	True class	Count
Wheat	Wheat	36
Wheat	Non-wheat	7
Non-wheat	Wheat	5
Non-wheat	Non-wheat	43

Table 11: Random Forest classification accuracy table for case four bands with top ten images

Training accuracy	Validation accuracy	Consumer accuracy (wheat)	Consumer accuracy (non-wheat)
99%	86%	87.8%	86%

Overall, the results you provided show that the classifier is a good model for wheat mapping. It has a high training accuracy and a good validation accuracy. It is also slightly better at identifying wheat pixels than non-wheat pixels.

Scenario 4: Ten bands with top ten important images

In another case we trained a classifier to classify wheat pixels from non-wheat pixels using all 10 bands of Sentinel-2 data (red, blue, green, re1, re2, re3, re4, swir1, swir2 and NIR, as well as NDVI. The classifier was trained on a ten important image collection from November 2022 to April 2023 which was determined by the classifier. The number of pixels that are properly identified is divided by the total number of pixels in the validation dataset

to determine the classifier's overall accuracy. In this case, the overall accuracy is (39 + 47) / (39 + 14 + 7 + 47) = 80.37%. The number of pixels properly identified for each class divided by the total number of pixels projected to be in that class yields the consumer accuracy for each class. In this case, the consumer accuracy for wheat is 39 / (39 + 14) = 84.78%, and the consumer accuracy for non-wheat is 47 / (47 + 7) = 77.05% as represented in table 13 and 14.

Table 12: Random Forest classification prediction table for case ten bands with top ten images:

Predicted class	**True class**	**Count**
Wheat	Wheat	39
Wheat	Non-wheat	14
Non-wheat	Wheat	7
Non-wheat	Non-wheat	47

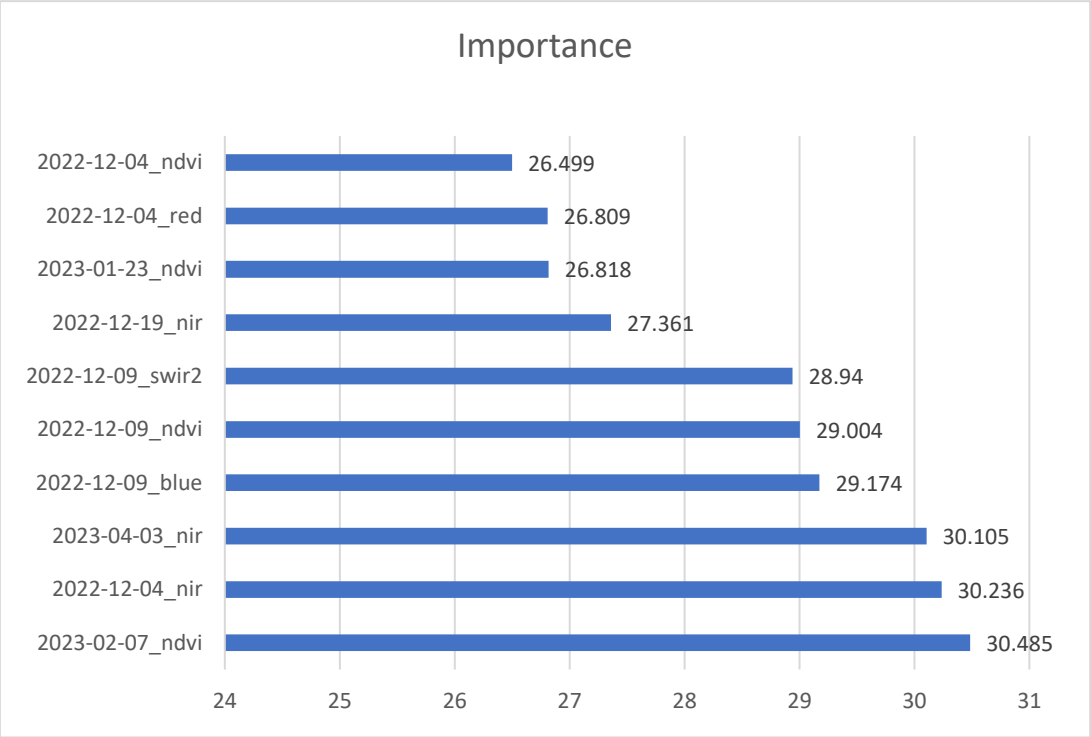


Figure 17: Top ten important feature from RF classifier based on ten bands.

Table 13: Random Forest classification accuracy table for case ten bands with top ten images

Training accuracy	**Validation accuracy**	**Consumer accuracy (wheat)**	**Consumer accuracy (non-wheat)**
98%	80%	77%	84%

Compared to the previous result, the classifier in this case has a lower overall accuracy (80.37% vs 86%) and a lower consumer accuracy for non-wheat (77.05% vs 86%). However, the consumer accuracy for wheat is still high (84.78% vs 87.8%).

Total area

Based on the RF classification, the total area of wheatland is 253.872 square kilometers, and the total area of non-wheat land is 201.847 square kilometers. This means that wheatland accounts for approximately 56% of the total land area, while other land accounts for approximately 44%.

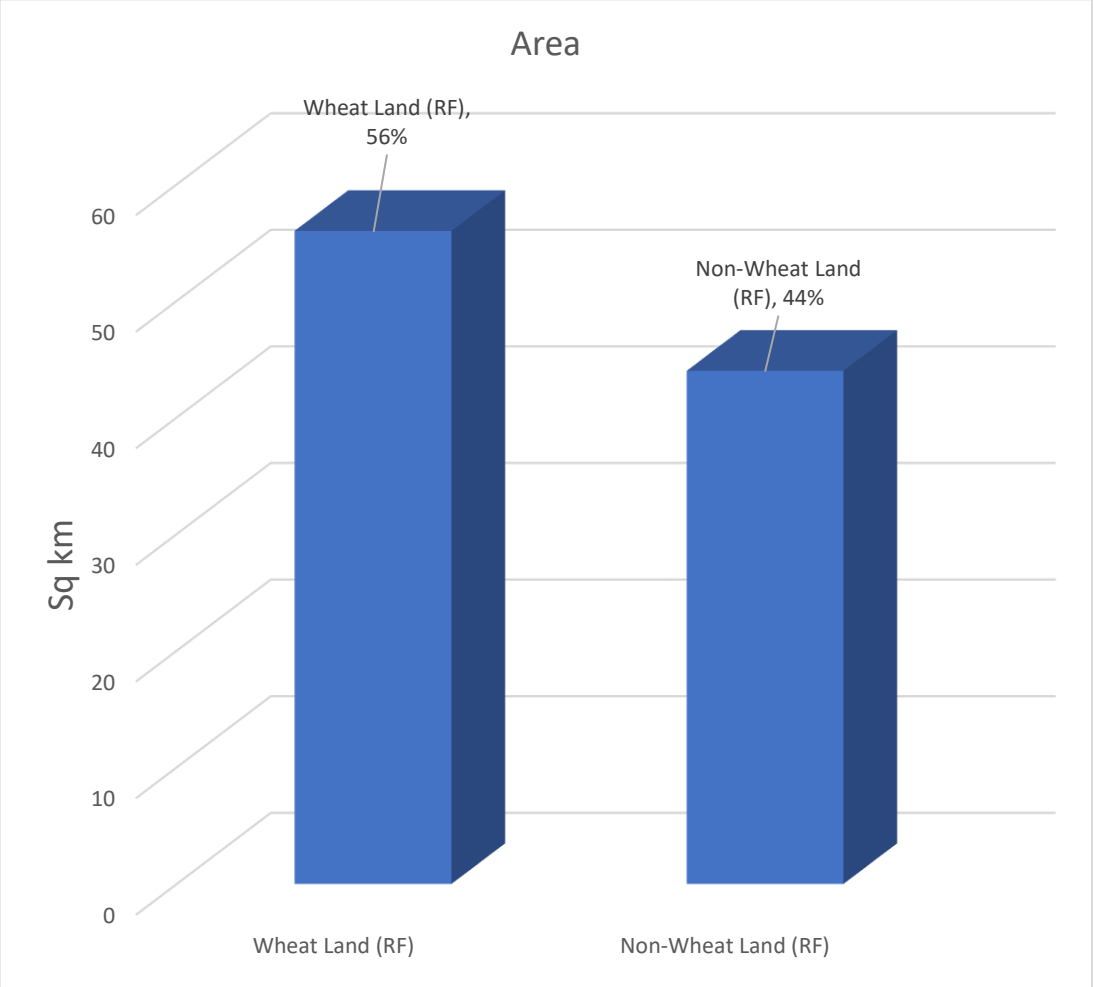


Figure 18: Random Forest result wheat land and not wheat land chart

Result map

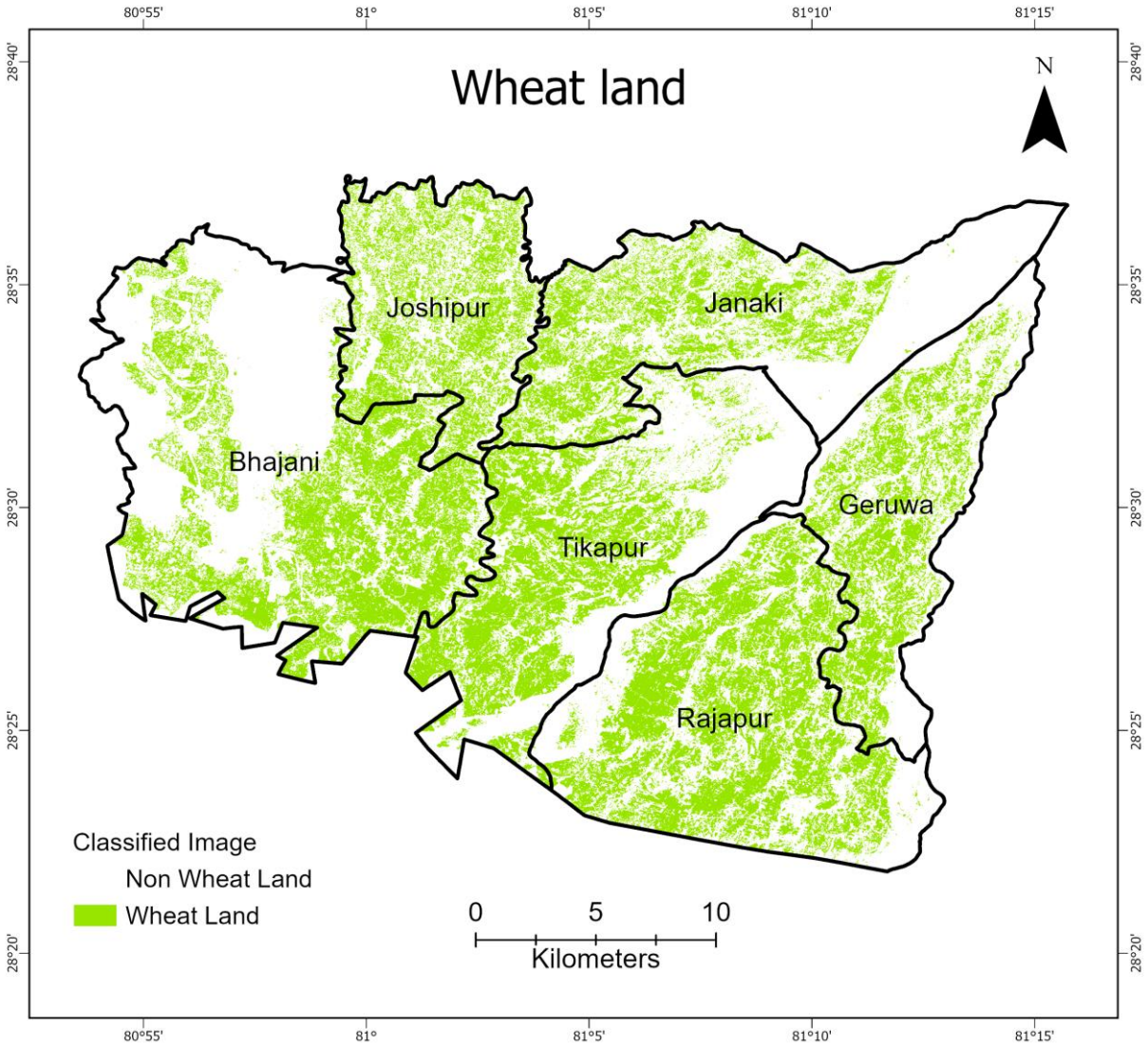


Figure 19: Random Forest resulted wheat map.

3.1.2 Support Vector Machine results

The Support Vector Machine algorithm successfully separated wheat land and non-wheat land with overall accuracy 90% as represented in table 15.

Table 14: SVM classification result

Training accuracy	Validation accuracy	Consumer accuracy (wheat)	Consumer accuracy (non-wheat)
90%	72.83%	70.97%	73.77%

We tested different scenarios with different sets of data and satellite bands to address all the possible cases for the best accuracy. Most of scenarios effected the model in different result. In all the case NDVI is compute and other indices like EVI and NDWI is also take in account.

Following code snippet is used for the SVM classification in GEE.

```
// SVM classifier
var svm = ee.Classifier.libsvm({
      kernelType: 'RBF', // You can choose different kernel types (e.g., 'linear', 'poly')
      gamma: 0.5, // Adjust gamma as needed
      cost: 10 // Adjust cost as needed
    });
```

Scenario 1: Four bands with all temporal image

We conducted a study to investigate the performance of a classifier for wheat mapping using four bands: red, blue, green, and NIR, as well as NDVI. The classifier was trained on a temporal image collection from November 2022 to April 2023. In this case, the overall accuracy is (22 + 45) / (22 + 16 + 9 + 45) = 72.83%. Consumer accuracy: The number of pixels properly identified for each class divided by the total number of pixels projected to be

in that class yields the consumer accuracy for each class. In this case, the consumer accuracy for wheat is 22 / (22 + 16) = 70.97%, and the consumer accuracy for non-wheat is 45 / (45 + 9) = 73.77% as represented in table 16 and 17.

Table 15: SVM classification prediction table for case four bands with all temporal images

Predicted class	**True class**	**Count**
Wheat	Wheat	22
Wheat	Non-wheat	16
Non-wheat	Wheat	9
Non-wheat	Non-wheat	45

Table 16: SVM classification accuracy table for case four bands with all temporal images

Training accuracy	**Validation accuracy**	**Consumer accuracy (wheat)**	**Consumer accuracy (non-wheat)**
90%	72.83%	70.97%	73.77%

Scenario 2: Ten bands with all temporal image

We conducted a study to investigate the performance of a classifier for wheat mapping using all 10 bands of Sentinel-2 data (red, blue, green, re1, re2, re3, re4, swir1, swir2 and NIR, as well as NDVI. The classifier was trained on a temporal image collection from November 2022 to April 2023. By dividing the total number of pixels in the validation dataset by the number of properly identified pixels, the overall accuracy of the SVM classifier is determined. In this case, the overall accuracy is (30 + 44) / (30 + 16 + 8 + 44) = 78.95%. In

this case, the consumer accuracy for wheat is 30 / (30 + 16) = 65.22%, and the consumer accuracy for non-wheat is 44 / (44 + 8) = 84.62% as represented in table 18 and 19. Compared to the previous results for the RF classifiers, the SVM classifier has a lower overall accuracy (78.95% vs 80.37% and 86%), but it has a higher consumer accuracy for non-wheat (84.62% vs 77.05% and 86%).

Table 17: SVM classification prediction table for case ten bands with all temporal images

Predicted class	**True class**	**Count**
Wheat	Wheat	30
Wheat	Non-wheat	16
Non-wheat	Wheat	8
Non-wheat	Non-wheat	44

Table 18: SVM classification accuracy table for case ten bands with all temporal images

Training accuracy	**Validation accuracy**	**Consumer accuracy (wheat)**	**Consumer accuracy (non-wheat)**
94%	78.95%	65.22%	84.62%

Scenario 3: Four bands with top ten important images

We conducted a study to investigate the performance of a classifier for wheat mapping using four bands: red, blue, green, and NIR, as well as NDVI. The classifier was trained on a ten important image collection from November 2022 to April 2023 which was determined

by the RF. The overall accuracy of the classifier is good, but it could be better. The consumer accuracy for both wheat and non-wheat is also fair, but the consumer accuracy for wheat is slightly lower. This means that the classifier is better at identifying non wheat pixels than wheat pixels However, it is important to note that the consumer accuracy for non-wheat is still relatively high (81.54%) as represented in table 20 and 21.

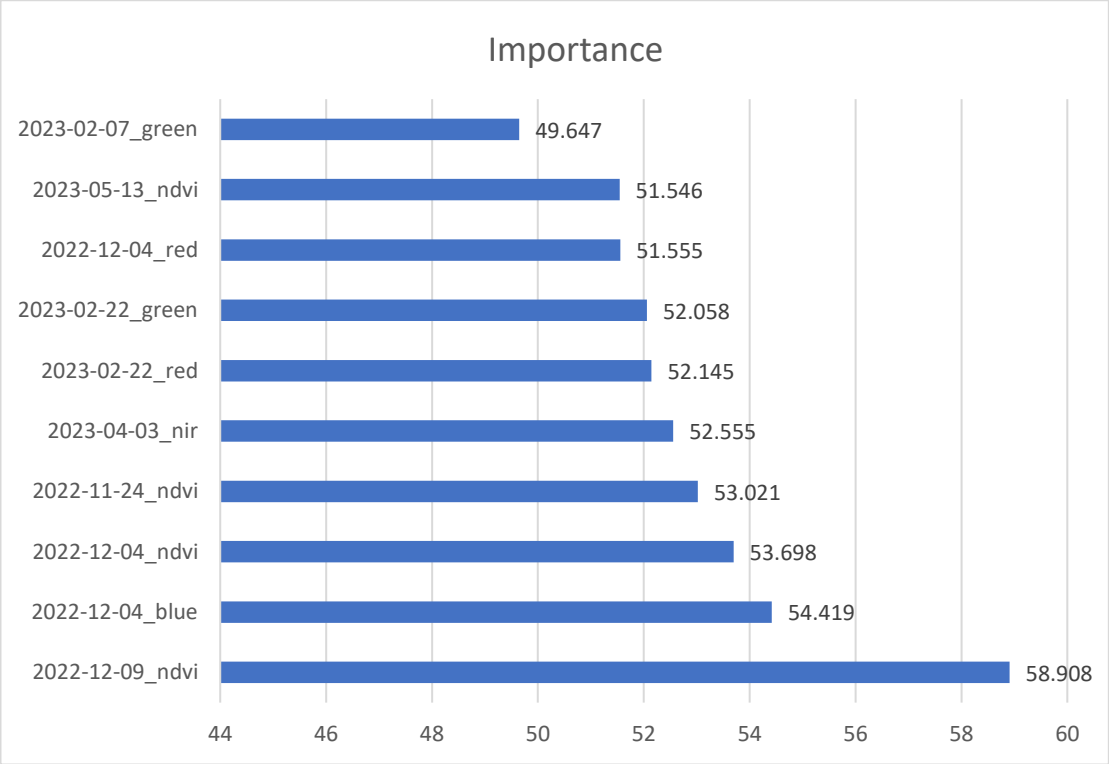


Figure 20: Top ten important feature for SVM based on four bands.

Table 19: SVM classification prediction table for case four bands with top ten images

Predicted class	**True class**	**Count**
Wheat	Wheat	33
Wheat	Non-wheat	25
Non-wheat	Wheat	6
Non-wheat	Non-wheat	52

Table 20: SVM classification accuracy table for case four bands with top ten images

Training accuracy	Validation accuracy	Consumer accuracy (wheat)	Consumer accuracy (non-wheat)
74.8%	73%	67%	81%

Scenario 4: Ten bands with top ten important images

In another case we trained a classifier to classify wheat pixels from non-wheat pixels using all 10 bands of Sentinel-2 data (red, blue, green, re1, re2, re3, re4, swir1, swir2 and

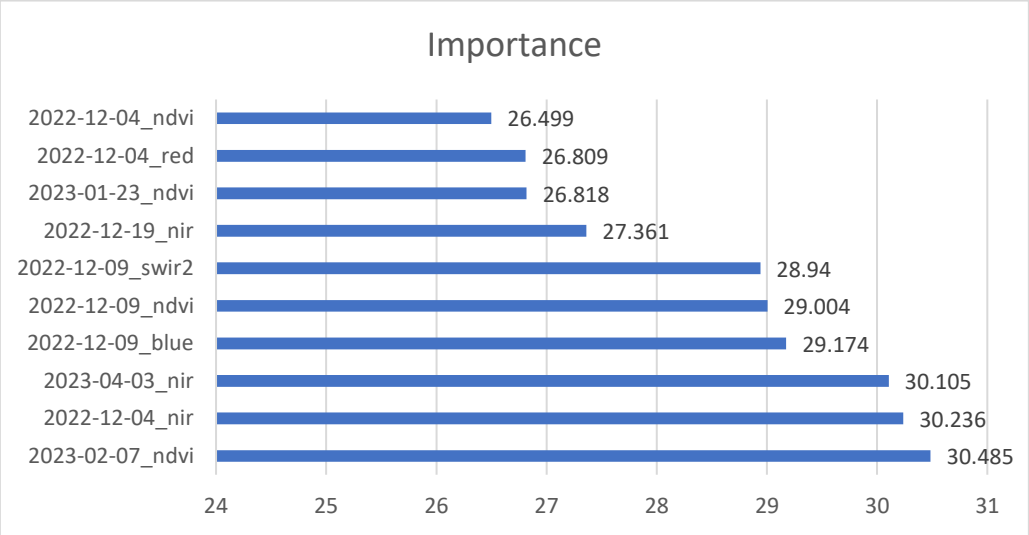


Figure 21: Top ten important feature for SVM based on Ten bands.

NIR, as well as NDVI. The classifier was trained on a ten important image collection from November 2022 to April 2023 which was determined by the RF. The overall accuracy of the classifier is good, and it is slightly higher than the overall accuracy of the classifier in four band runs as represented in table 22 and 23. The consumer accuracy for wheat is also slightly higher than in the previous run, but the consumer accuracy for non-wheat is slightly lower.

Figure 20: Top ten important feature for SVM classifier based on ten bands.

Table 21: SVM classification prediction table for case ten bands with top ten images

Predicted class	True class	Count
Wheat	Wheat	30
Wheat	Non-wheat	8
Non-wheat	Wheat	23
Non-wheat	Non-wheat	59

Table 22: SVM classification accuracy table for case ten bands with top ten images

Training accuracy	Validation accuracy	Consumer accuracy (wheat)	Consumer accuracy (non-wheat)
77.8%	75%	78.95%	71.95%

Total area

Based on the SVM classification you provided, the total area of wheatland is 298.02 square kilometers, and the total area of non-wheat land is 157.69 square kilometers. This means that wheatland accounts for approximately 65.39% of the total land area, while other land accounts for approximately 34.60% as shown in figure 21.

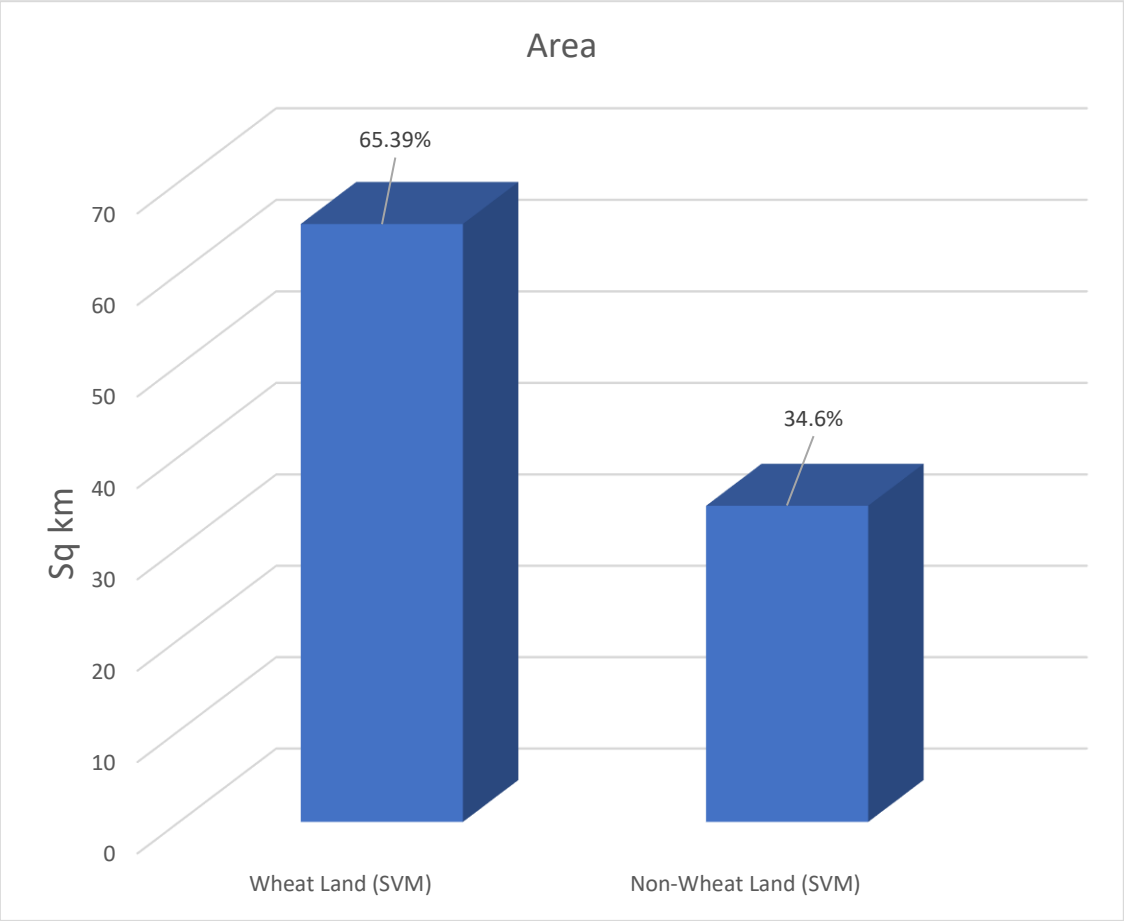


Figure 22: SVM result wheat land and not wheat land chart.

Result Map

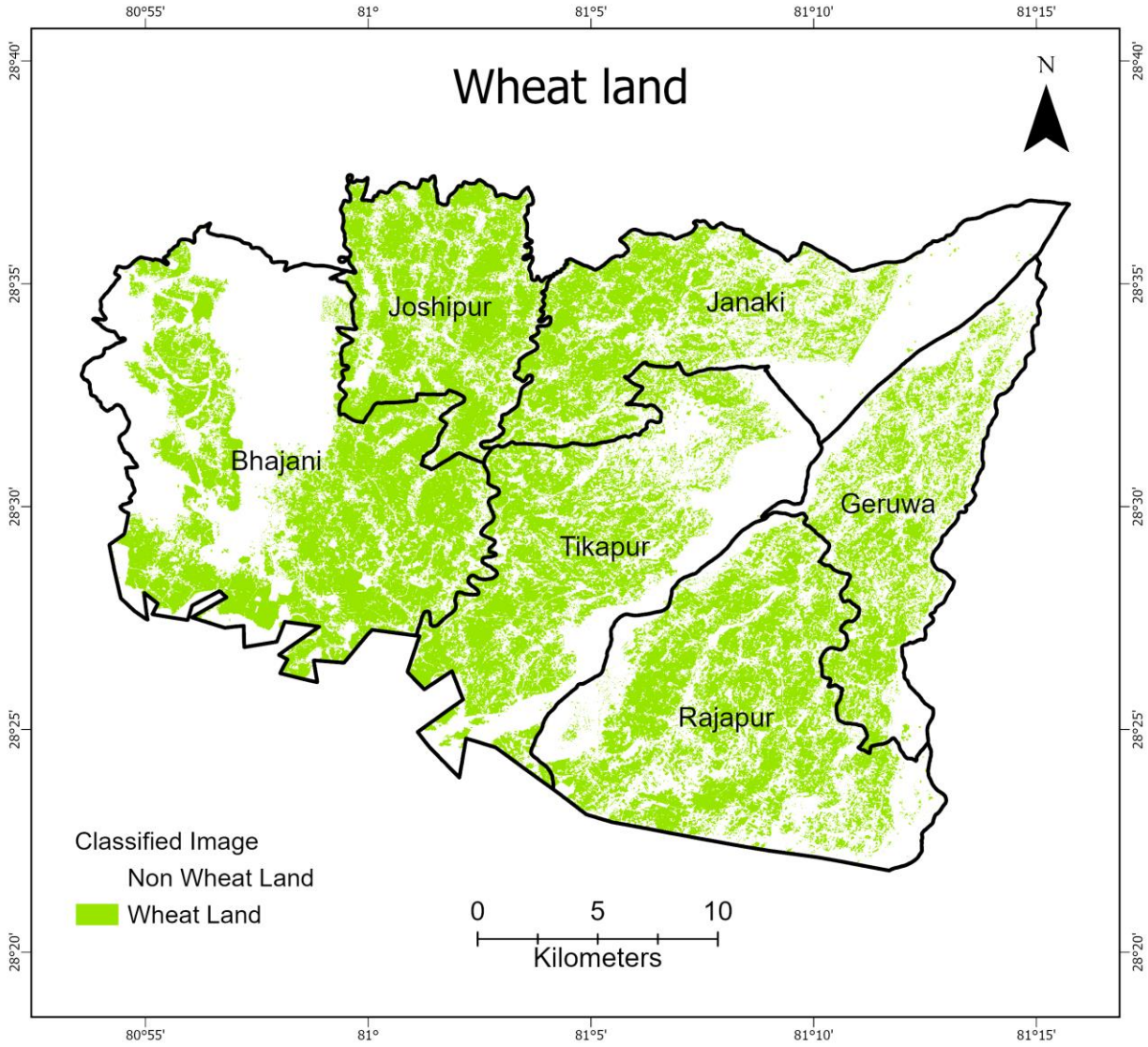


Figure 23: SVM resulted wheat map.

3.2 Comparison between the RF and SVM

RF achieved an overall accuracy of 86%, while SVM achieved an overall accuracy of 78% with considering best scenarios. This is a significant difference in performance. Additionally, RF models are more interpretable than SVM models. This means that it is easier to understand how the Random Forest model is making its predictions. This is important for wheat land classification, as it allows researchers to identify the features that are most important for distinguishing wheat land from other types of land cover.

RF algorithm predicted that 56% of the total area is wheat land, and 44% of the total area is non-wheat land. This is different from the SVM algorithm, which predicted that 65.39% of the total area is wheat land, and 34.6% of the total area is non-wheat land as shown in figure 23. Possible explanation for the difference in results is that the RF algorithm is more conservative than the SVM algorithm. This means that the RF algorithm is less likely to classify a pixel as wheat land unless it is very confident in its prediction. The RF algorithm produced a higher accuracy than the SVM algorithm for the two fields, predicting that 56% of the total area is wheat land, and 44% of the total area is non-wheat land.

RF is a better choice for wheat land classification than SVM due to its superior classification performance and interpretability. RF is a more ensemble learning algorithm that is more robust to overfitting and better able to handle complex data with many features, such as satellite imagery data. This is because RF builds a set of decision trees and averages their predictions to produce a final prediction. This makes RF more resilient to outliers and noise in the data.

SVM, on the other hand, is a discriminative learning algorithm that finds a hyperplane that separates the data into two classes. This makes SVM more efficient computationally and better able to handle high-dimensional data. However, SVM is more susceptible to overfitting and can be less accurate on complex data with many features.

In addition to its superior classification performance, RF is also more interpretable than SVM. This means that it is easier to understand how the RF model is making its predictions. This is important for wheat land classification, as it allows researchers to identify the features that are most important for distinguishing wheat land from other types of land cover.

From a prospective standpoint, RF is expected to continue to be the preferred algorithm for wheat land classification due to its advantages in classification performance and interpretability. As the availability of satellite imagery data continues to grow, RF will be able to leverage this data to produce even more accurate and reliable wheat land classifications.

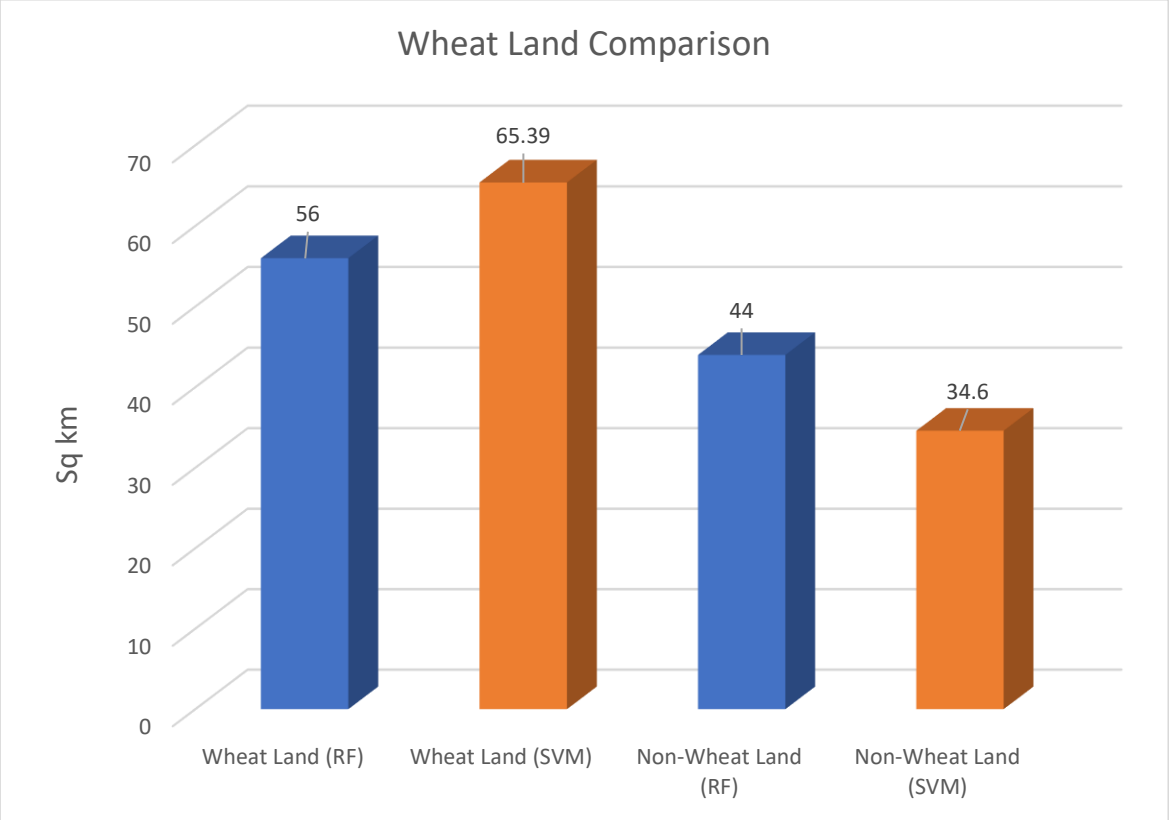


Figure 24: RF and SVM wheat land comparison chart

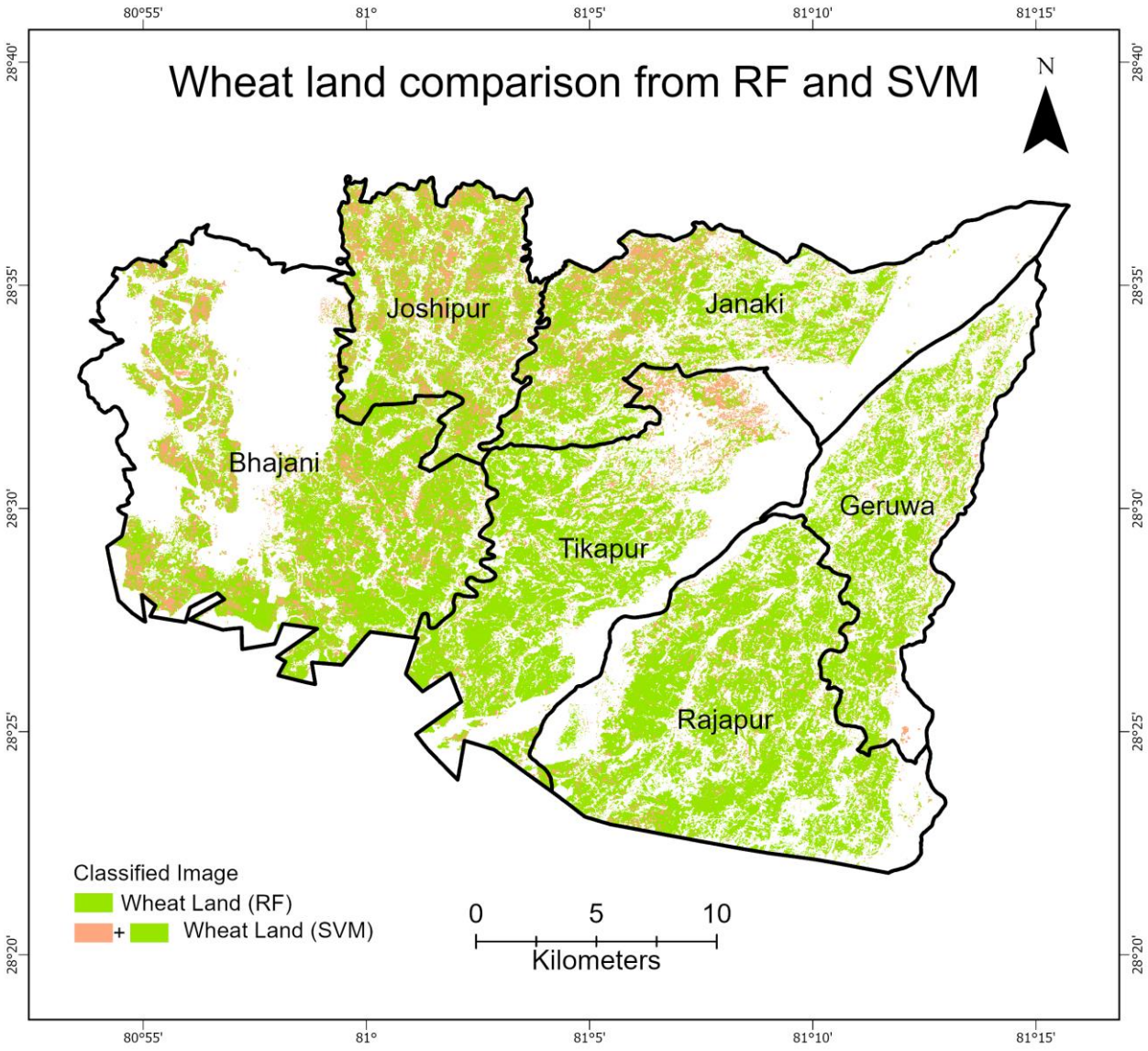


Figure 25: RF and SVM wheat land comparison map

Chapter-4: Conclusion

4.1 Conclusion

In this study, We examined the efficacy of SVM and RF, two machine learning methods, in the categorization of crop land. The results showed that the RF algorithm produced a higher accuracy than the SVM algorithm, predicting that 56% (253 sq km) of the total area is wheat land, and 44% of the total area is non-wheat land. According to the Ministry of Agriculture & Livestock Development report for 2021 agricultural report, the total wheat land is approx. 240 sq km which proves RF results produce higher accuracy.

This suggests that the RF algorithm is better suited for wheat land classification tasks in this area. One possible explanation for the higher accuracy of the RF algorithm is that it is more robust to noise in the image and is better at learning the complex relationships between the spectral characteristics of the two fields and their respective crop types.

The high accuracy of the RF algorithm also suggests that it could be used to create a reliable wheat land map for the area. This map could be used for a variety of purposes, such as agricultural planning, environmental monitoring, and urban development. For example, agricultural planners could use the map to identify areas that are best suited for growing wheat, and to develop strategies for improving crop yields. Environmental monitors could use the map to track changes in land cover over time, and to identify areas that are at risk of desertification or deforestation. Urban developers could use the map to plan new development projects in a way that minimizes the impact on the environment.

Overall, the results from this study suggest that the RF algorithm is a promising tool for crop classification and land cover classification tasks. The high accuracy of the RF algorithm suggests that it could be used to create reliable crop type maps that can be used for a variety of purposes.

4.2 Recommendation

Strong machine learning algorithms like the random forest (RF) algorithm have shown to be particularly successful in classifying crops and land covers. In this study, we demonstrated that the RF algorithm can be used to create highly accurate crop type maps for a variety of purposes, including agricultural planning, environmental monitoring, and urban development.

Agricultural planners can use RF-generated crop type maps to identify areas that are best suited for growing wheat and other crops. This information can be used to develop strategies for improving crop yields and increasing agricultural productivity. For example, agricultural planners can use crop type maps to identify areas with optimal soil conditions, water availability, and climate conditions for growing wheat. They can also use crop type maps to identify areas that are at risk of pests and diseases, and to develop strategies for mitigating these risks. Environmental monitors can use RF-generated crop type maps to track changes in land cover over time. This information can be used to identify areas that are at risk of desertification, deforestation, or other environmental problems. For example, environmental monitors can use crop type maps to track the spread of invasive plant species, or to identify areas that are at risk of drought or flooding. Urban developers can use RF-generated crop type maps to plan new development projects in a way that minimizes the impact on the environment. For example, urban developers can use crop type maps to avoid developing areas that are important for agriculture or that are at risk of environmental problems. Urban developers can also use crop type maps to design new development projects that incorporate green spaces and other features that can help to mitigate the impacts of climate change.

The RF algorithm could also be used to develop new applications in agriculture, environmental monitoring, and urban development. For example, the RF algorithm could be used to develop systems that can:

- Predict crop yields
- Identify pests and diseases
- Monitor the impact of climate change on agricultural production.
- Identify areas that are suitable for afforestation.
- Map the distribution of biodiversity
- Monitor the impact of urbanization on the environment.

References

Ashourloo, D., Nematollahi, H., Huete, A., Aghighi, H., Azadbakht, M., Shahrabi, H. S., & Goodarzdashti, S. (2022). A new phenology-based method for mapping wheat and barley using time-series of Sentinel-2 images. *Remote Sensing of Environment*, *280*. https://doi.org/10.1016/j.rse.2022.113206

Bhatta, R. D., Amgain, L. P., Subedi, R., & Kandel, B. P. (2020). Assessment of productivity and profitabilty of wheat using Nutrient Expert®-Wheat model in Jhapa district of Nepal. *Heliyon*, *6*(6). https://doi.org/10.1016/j.heliyon.2020.e04144

Breiman, L. Random Forests. *Machine Learning* **45**, 5–32 (2001). https://doi.org/10.1023/A:1010933404324

Evgeniou, T., & Pontil, M. (2001). Support vector machines: Theory and applications. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *2049 LNAI*, 249–257. https://doi.org/10.1007/3-540-44673-7_12

Feng, S., Zhao, J., Liu, T., Zhang, H., Zhang, Z., & Guo, X. (2019). Crop Type Identification and Mapping Using Machine Learning Algorithms and Sentinel-2 Time Series Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *12*(9), 3295–3306. https://doi.org/10.1109/JSTARS.2019.2922469

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

Guo, Y., Xia, H., Pan, L., Zhao, X., Li, R., Bian, X., Wang, R., & Yu, C. (2021). Development of a new phenology algorithm for fine mapping of cropping intensity in complex planting areas using sentinel-2 and google earth engine. *ISPRS International Journal of Geo-Information*, *10*(9). https://doi.org/10.3390/ijgi10090587

Hunt, M. L., Blackburn, G. A., Carrasco, L., Redhead, J. W., & Rowland, C. S. (2019). High resolution wheat yield mapping using Sentinel-2. *Remote Sensing of Environment*, *233*. https://doi.org/10.1016/j.rse.2019.111410

Joshi, B., Mudwari, A., & Bhatta, M. (1970). Wheat Genetic Resources in Nepal. *Nepal Agriculture Research Journal*, *7*, 1–10. https://doi.org/10.3126/narj.v7i0.1859

Khadka, K., Torkamaneh, D., Kaviani, M., Belzile, F., Raizada, M. N., & Navabi, A. (2020). Population structure of Nepali spring wheat (Triticum aestivum L.) germplasm. *BMC Plant Biology*, *20*(1). https://doi.org/10.1186/s12870-020-02722-8

Li, G., Cui, J., Han, W., Zhang, H., Huang, S., Chen, H., & Ao, J. (2022). Crop type mapping using time-series Sentinel-2 imagery and U-Net in early growth periods in the Hetao irrigation district in China. *Computers and Electronics in Agriculture*, *203*. https://doi.org/10.1016/j.compag.2022.107478

Li, Y., Qin, Y., Ma, L., & Pan, Z. (2020). Climate change: vegetation and phenological phase dynamics. *International Journal of Climate Change Strategies and Management*, *12*(4), 495–509. https://doi.org/10.1108/IJCCSM-06-2019-0037

Mishra, B., Bhandari, R., Bhandari, K. P., Bhandari, D. M., Luintel, N., Dahal, A., & Poudel, S. (2023). High-Resolution Mapping of Seasonal Crop Pattern Using Sentinel Imagery in Mountainous Region of Nepal: A Semi-Automatic Approach. *Geomatics*, *3*(2), 312–327. https://doi.org/10.3390/geomatics3020017

Muruganantham, P., Wibowo, S., Grandhi, S., Samrat, N. H., & Islam, N. (2022). A Systematic Literature Review on Crop Yield Prediction with Deep Learning and Remote Sensing. In *Remote Sensing* (Vol. 14, Issue 9). MDPI. https://doi.org/10.3390/rs14091990

Pan, Z., Huang, J., Zhou, Q., Wang, L., Cheng, Y., Zhang, H., Blackburn, G. A., Yan, J., & Liu, J. (2015). Mapping crop phenology using NDVI time-series derived from HJ-1 A/B

data. *International Journal of Applied Earth Observation and Geoinformation*, *34*(1), 188–197. https://doi.org/10.1016/j.jag.2014.08.011

Phiri, D., Simwanda, M., Salekin, S., Nyirenda, V. R., Murayama, Y., & Ranagalage, M. (2020). Sentinel-2 data for land cover/use mapping: A review. In *Remote Sensing* (Vol. 12, Issue 14). MDPI AG. https://doi.org/10.3390/rs12142291

Qamer, F. M., Shah, S. N. P., Murthy, M. S. R., Baidar, T., Dhonju, K., & Hari, B. G. (2014). Operationalizing crop monitoring system for informed decision making related to food security in Nepal. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, *40*(8), 1325–1330. https://doi.org/10.5194/isprsarchives-XL-8-1325-2014

Sakamoto, T., Yokozawa, M., Toritani, H., Shibayama, M., Ishitsuka, N., & Ohno, H. (2005). A crop phenology detection method using time-series MODIS data. *Remote Sensing of Environment*, *96*(3–4), 366–374. https://doi.org/10.1016/j.rse.2005.03.008

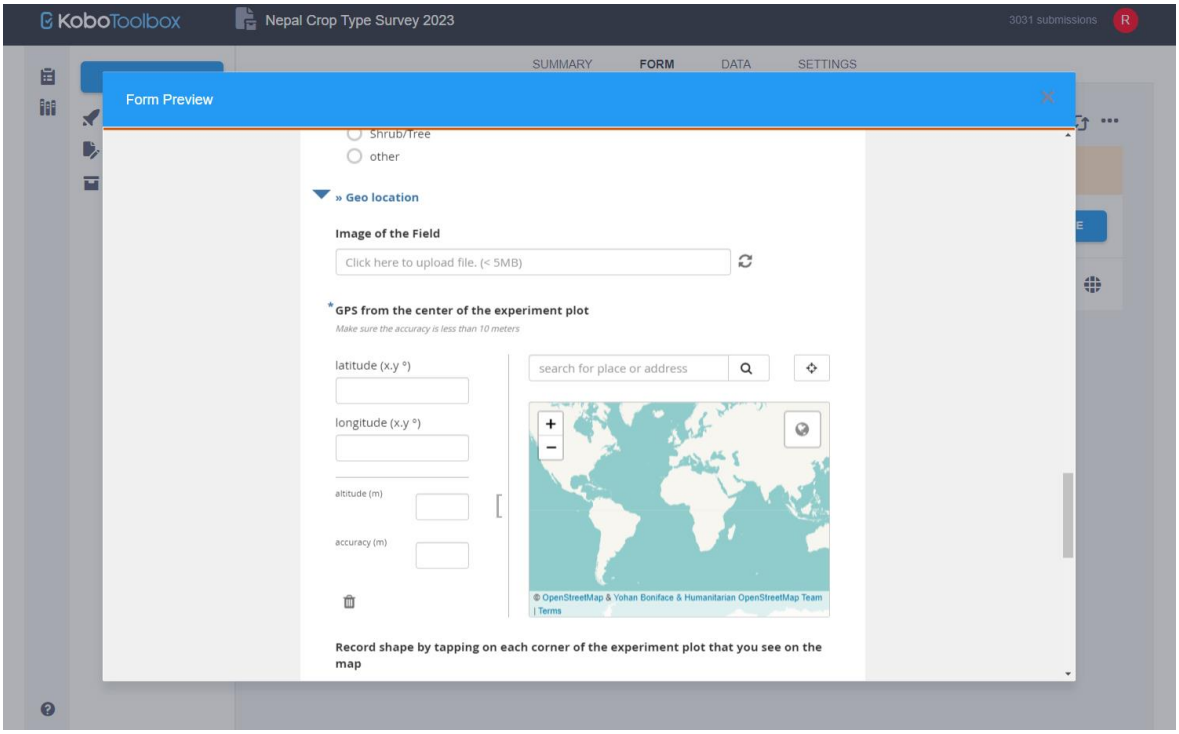
TAO, J., ZHANG, X., WU, Q., & WANG, Y. (2022). Mapping winter rapeseed in South China using Sentinel-2 data based on a novel separability index. *Journal of Integrative Agriculture*. https://doi.org/10.1016/j.jia.2022.10.008

Xie, Y., Sha, Z., & Yu, M. (2008). Remote sensing imagery in vegetation mapping: a review. *Journal of Plant Ecology*, *1*(1), 9–23. https://doi.org/10.1093/jpe/rtm005ssss

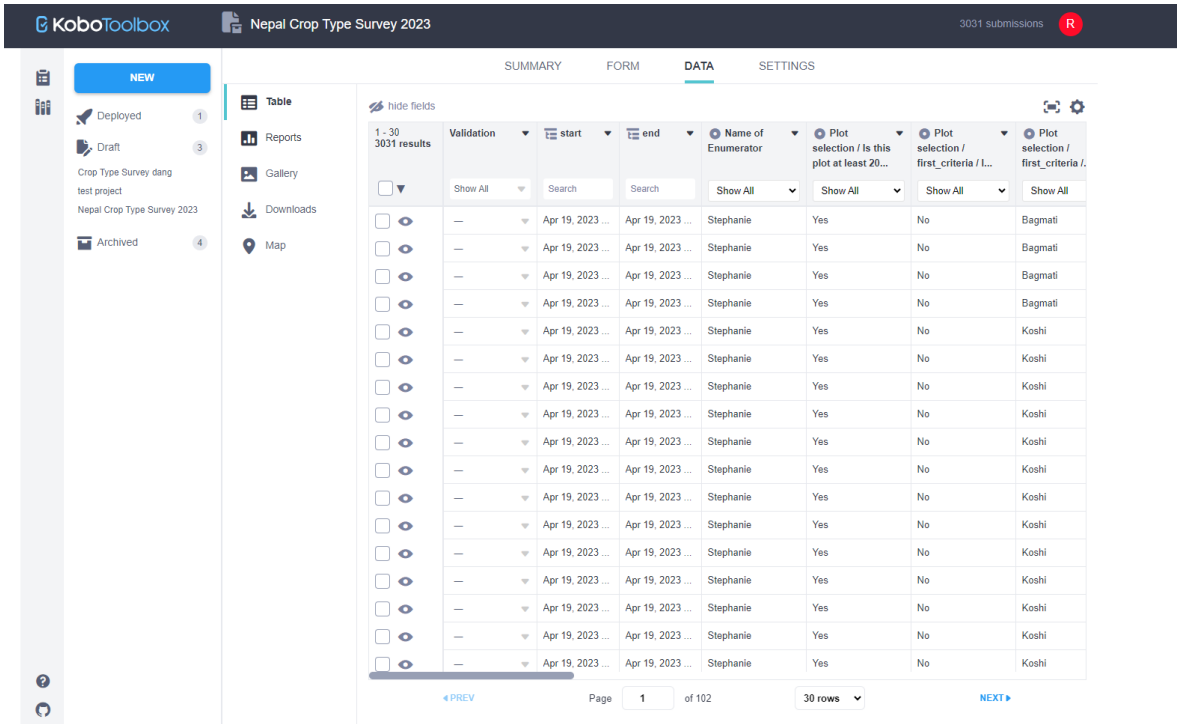
Zhao, Y., Wang, X., Guo, Y., Hou, X., & Dong, L. (2022). Winter Wheat Phenology Variation and Its Response to Climate Change in Shandong Province, China. *Remote Sensing*, *14*(18). https://doi.org/10.3390/rs14184482

Appendix

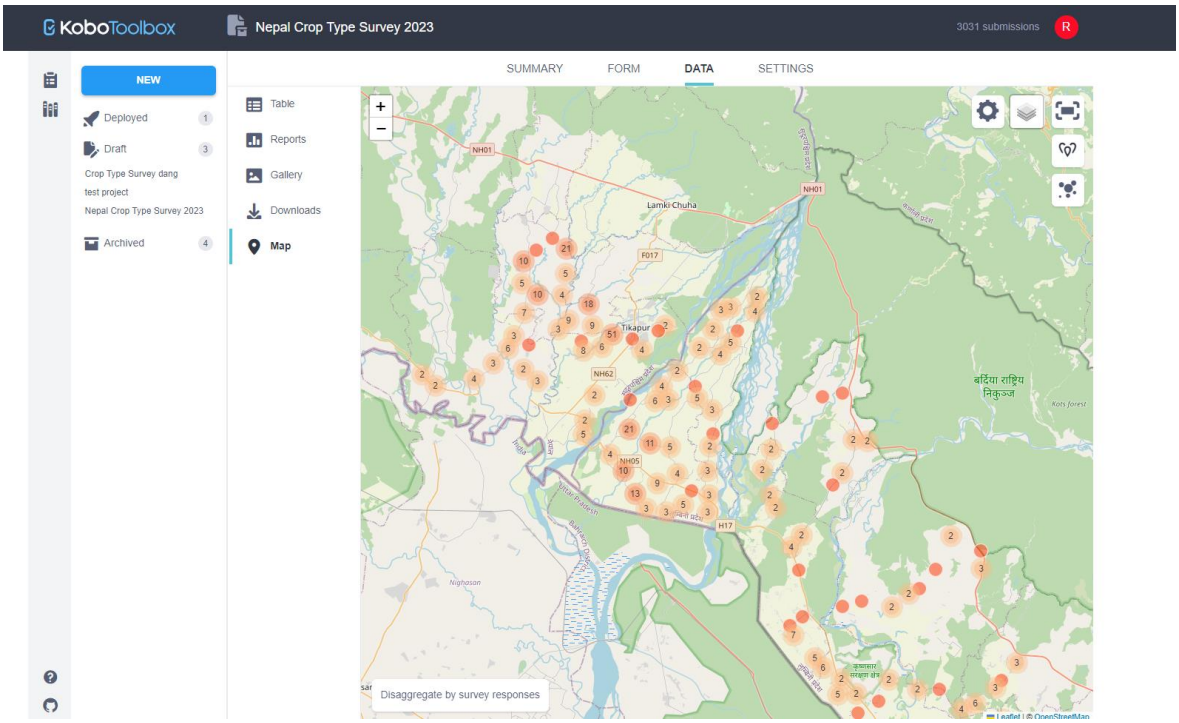
1. Kobo Forms



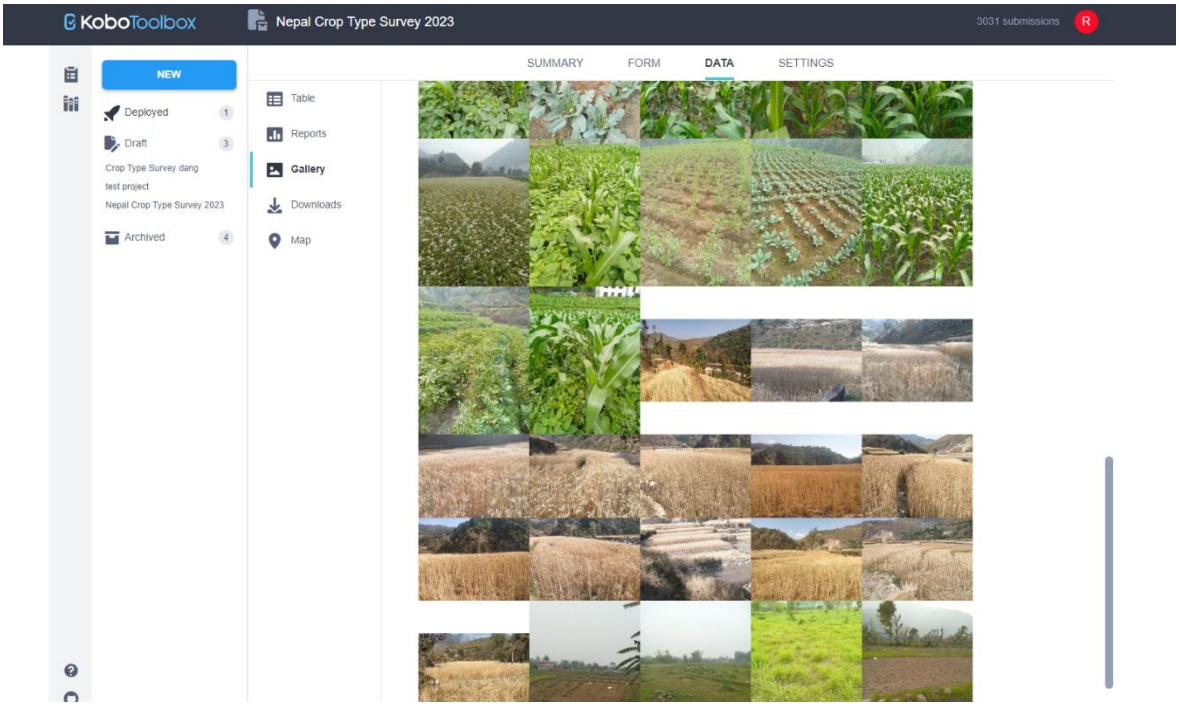
2. Kobo Data



3. Kobo Spatial Data

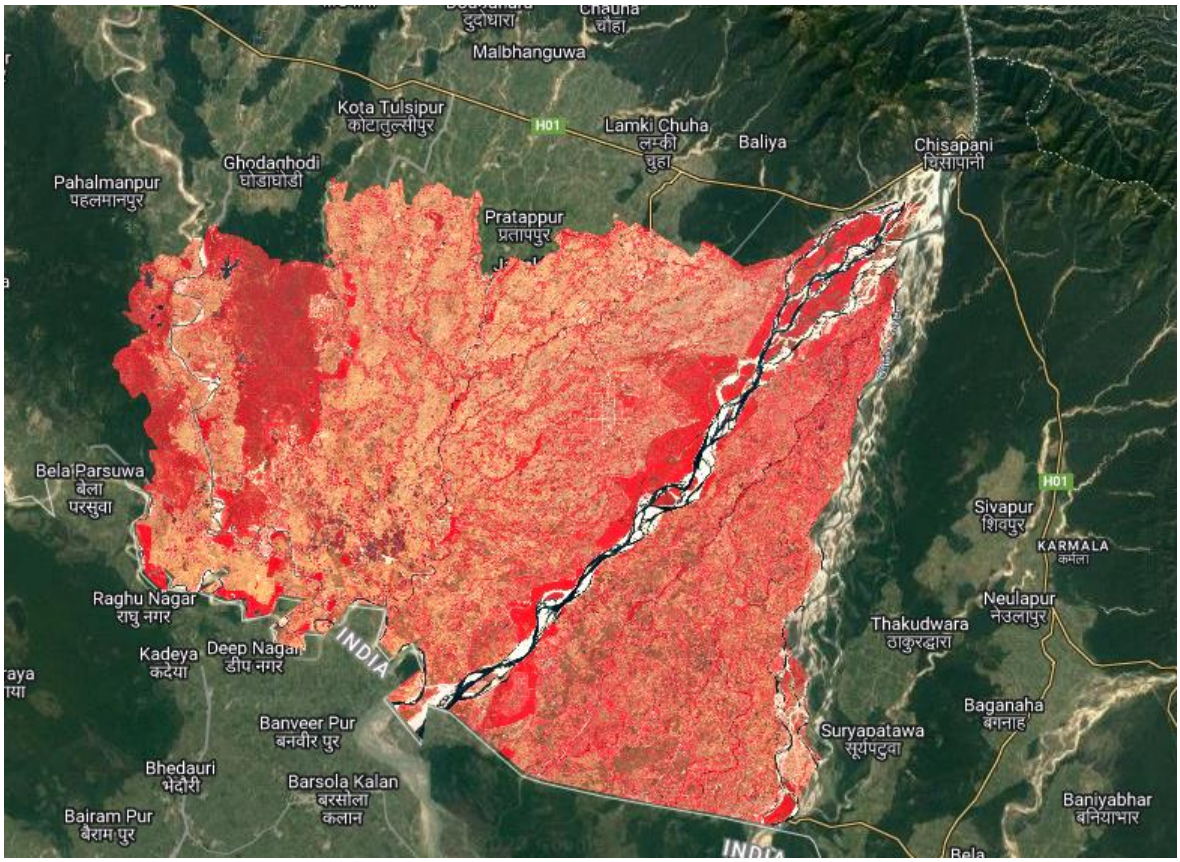


4. Survey Images

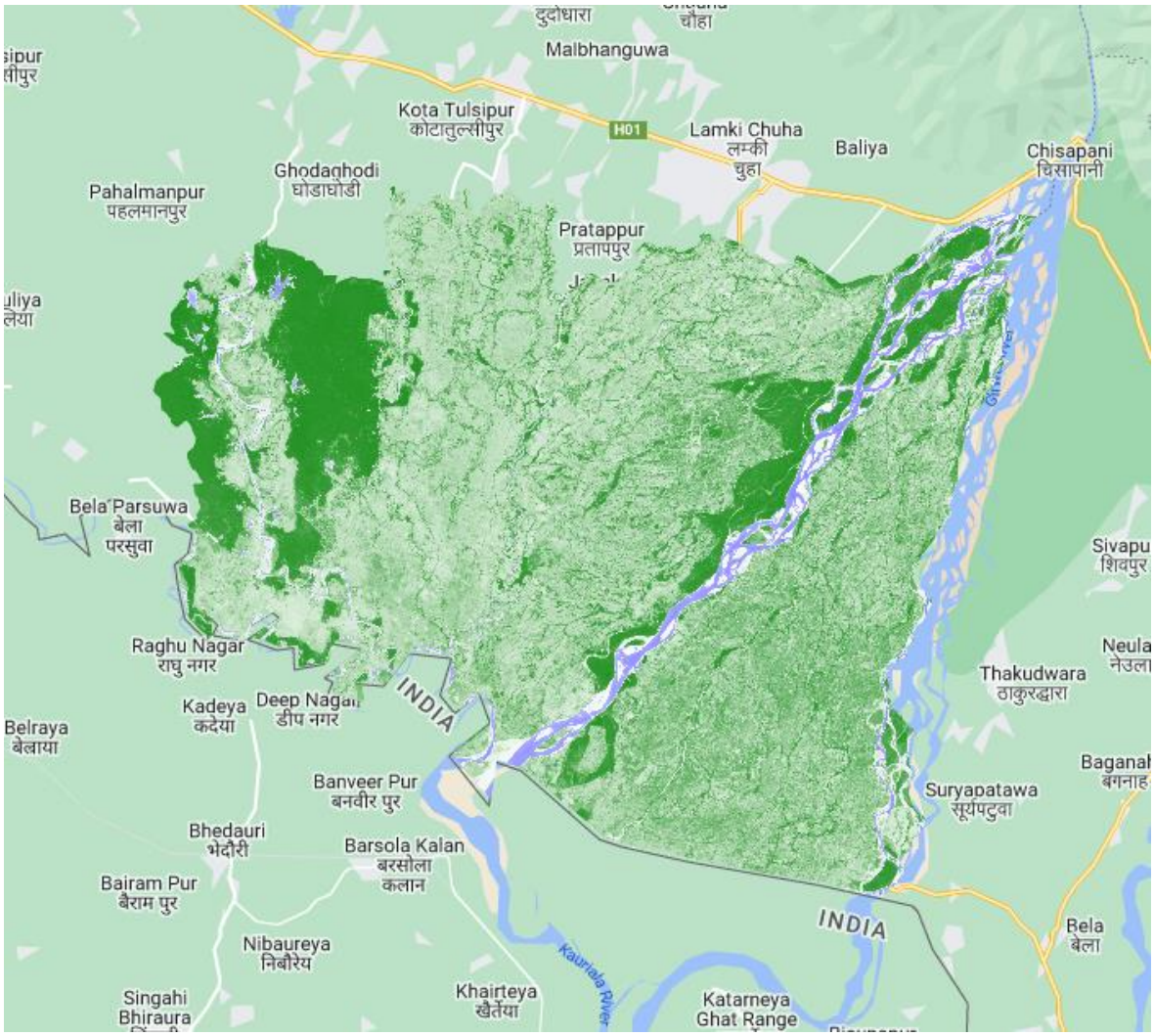


5. Study area satellite images (Natural color)

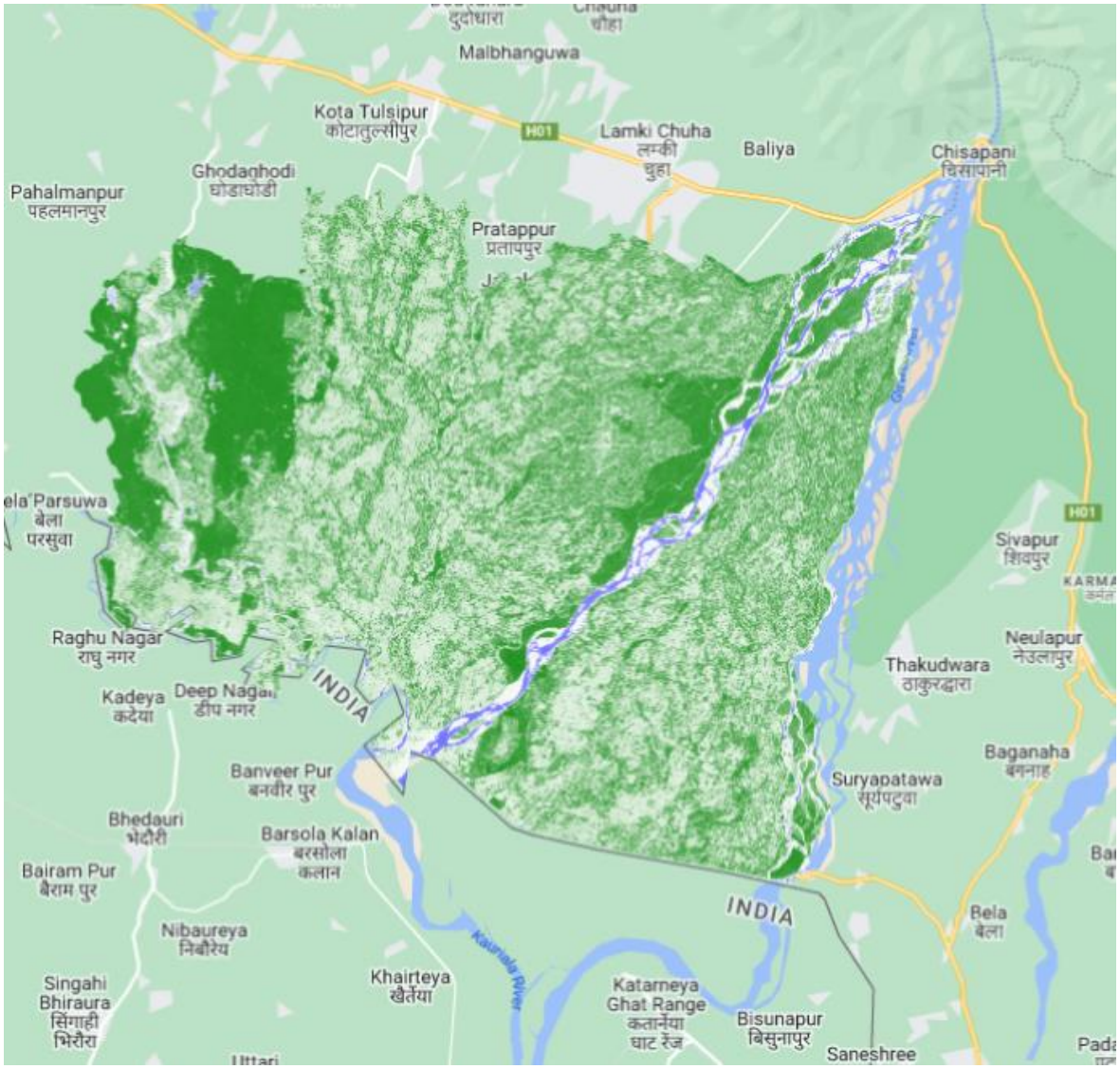
6. Study area satellite images (False color)



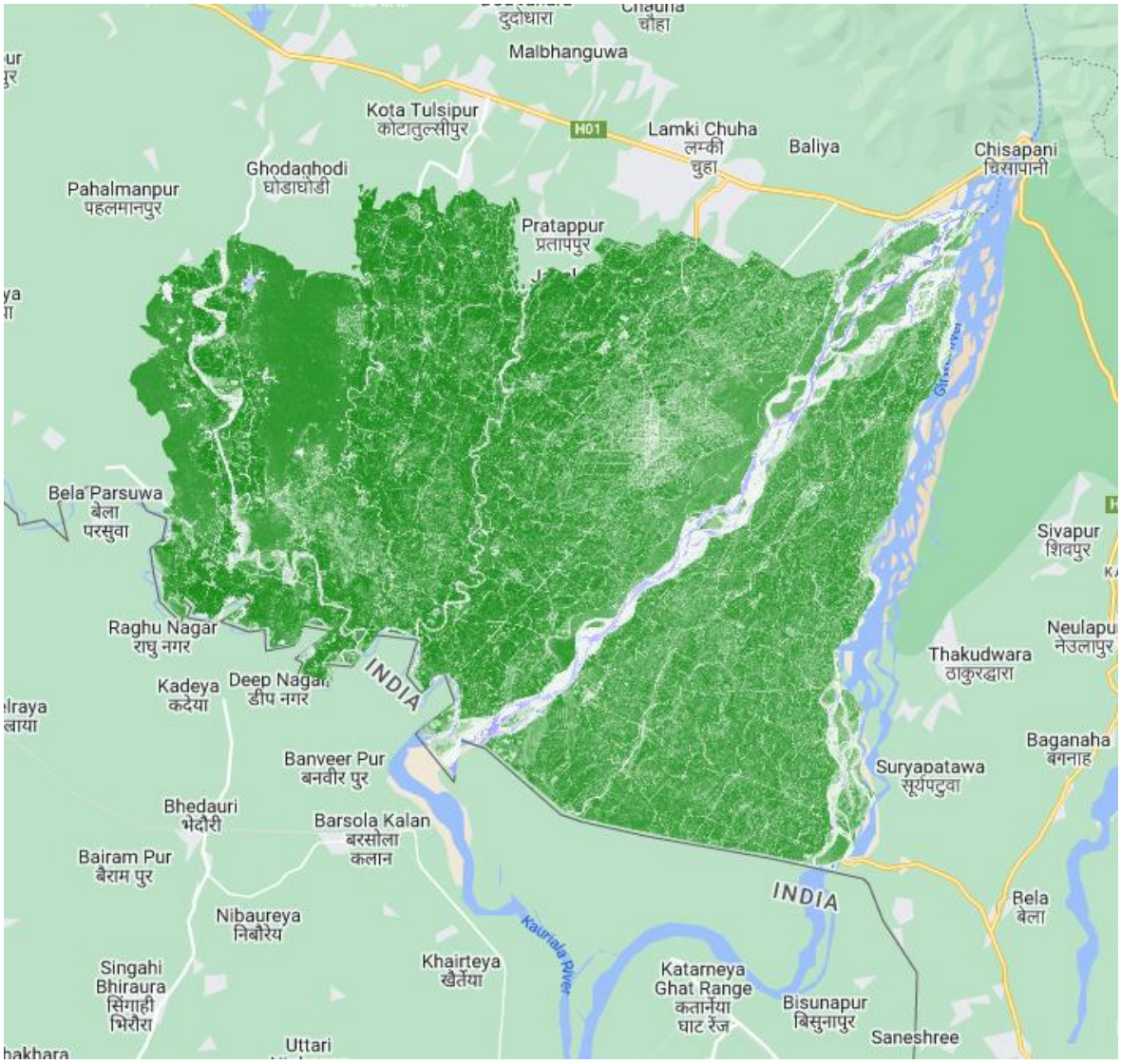
7. Study area satellite images (NDVI at Germination)



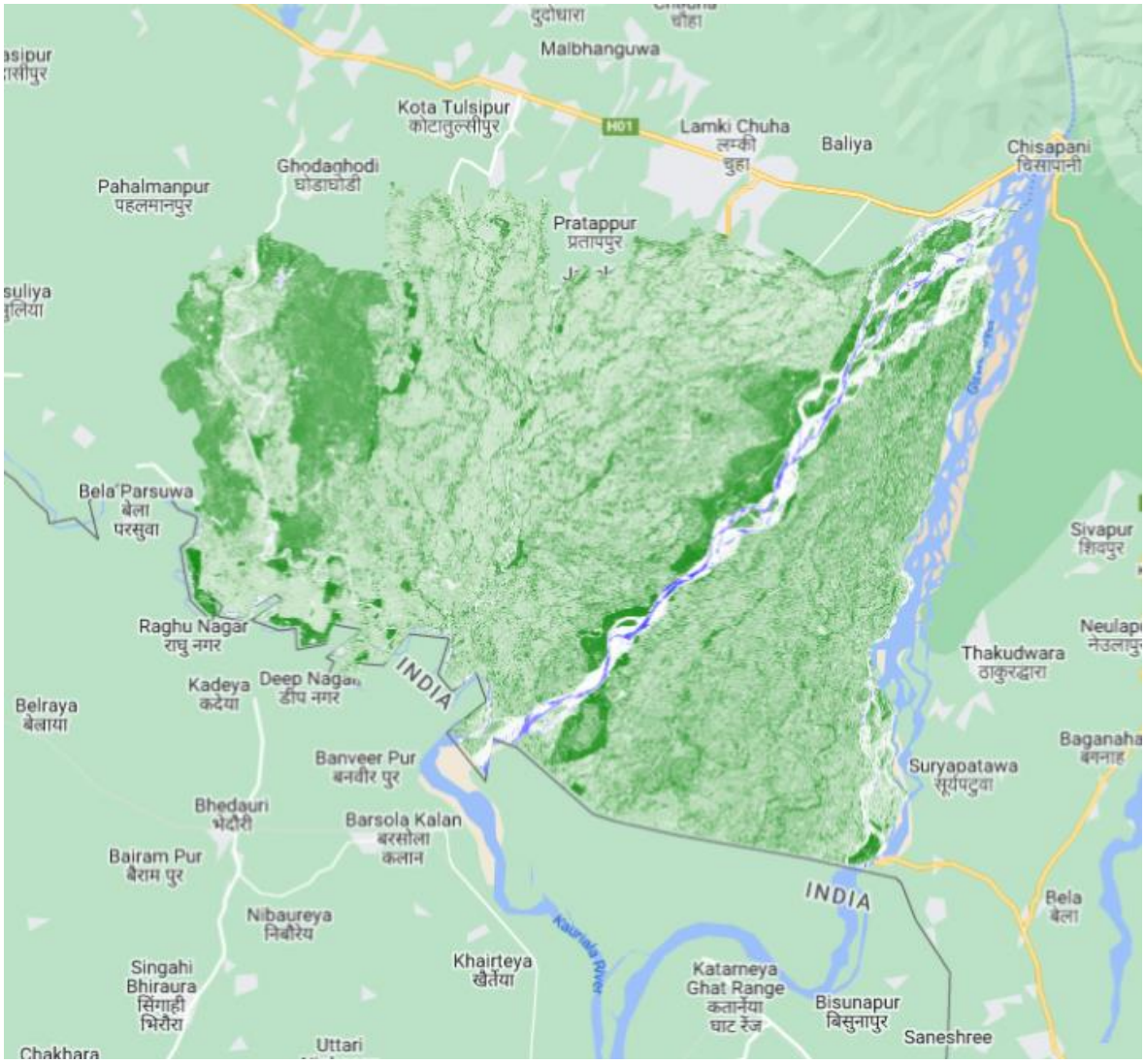
8. Study area satellite images (NDVI at Tillering)



9. Study area satellite images (NDVI at Flowering)



10. Study area satellite images (NDVI at Ripening)



11. Satellite image tile id T44RNS

Image ID	Tile	Date
2_20221020T050829_20221020T051951_T44RNS	T44RNS	10/20/2022
2_20221025T050901_20221025T051100_T44RNS	T44RNS	10/25/2022
2_20221030T050929_20221030T051554_T44RNS	T44RNS	10/30/2022
2_20221104T051001_20221104T051001_T44RNS	T44RNS	11/4/2022
2_20221119T051109_20221119T051111_T44RNS	T44RNS	11/19/2022
2_20221124T051121_20221124T051122_T44RNS	T44RNS	11/24/2022
2_20221129T051149_20221129T051146_T44RNS	T44RNS	11/29/2022
2_20221204T051201_20221204T051202_T44RNS	T44RNS	12/4/2022
2_20221209T051209_20221209T051209_T44RNS	T44RNS	12/9/2022
2_20221214T051221_20221214T051220_T44RNS	T44RNS	12/14/2022
2_20221219T051229_20221219T051223_T44RNS	T44RNS	12/19/2022
2_20230113T051151_20230113T051152_T44RNS	T44RNS	1/13/2023
2_20230118T051139_20230118T051134_T44RNS	T44RNS	1/18/2023
2_20230123T051111_20230123T052031_T44RNS	T44RNS	1/23/2023
2_20230207T050959_20230207T051145_T44RNS	T44RNS	2/7/2023
2_20230222T050821_20230222T051156_T44RNS	T44RNS	2/22/2023
2_20230304T050711_20230304T051951_T44RNS	T44RNS	3/4/2023
2_20230309T050659_20230309T052025_T44RNS	T44RNS	3/9/2023
2_20230403T050651_20230403T051535_T44RNS	T44RNS	4/3/2023
2_20230413T050651_20230413T051512_T44RNS	T44RNS	4/13/2023
2_20230418T050659_20230418T051910_T44RNS	T44RNS	4/18/2023
2_20230423T050651_20230423T051305_T44RNS	T44RNS	4/23/2023
2_20230513T050651_20230513T051520_T44RNS	T44RNS	5/13/2023
2_20230518T050659_20230518T051514_T44RNS	T44RNS	5/18/2023
2_20230523T050651_20230523T051519_T44RNS	T44RNS	5/23/2023

12. Satellite image tile id T44RMS

Image ID	Tile	Date
2_20221015T050801_20221015T051837_T44RMS	T44RMS	10/15/2022
2_20221018T051821_20221018T052148_T44RMS	T44RMS	10/18/2022
2_20221020T050829_20221020T051951_T44RMS	T44RMS	10/20/2022
2_20221023T051849_20221023T052729_T44RMS	T44RMS	10/23/2022
2_20221025T050901_20221025T051100_T44RMS	T44RMS	10/25/2022
2_20221028T051931_20221028T052450_T44RMS	T44RMS	10/28/2022
2_20221030T050929_20221030T051554_T44RMS	T44RMS	10/30/2022
2_20221102T051949_20221102T052958_T44RMS	T44RMS	11/2/2022
2_20221104T051001_20221104T051001_T44RMS	T44RMS	11/4/2022
2_20221112T052049_20221112T052220_T44RMS	T44RMS	11/12/2022
2_20221117T052111_20221117T052416_T44RMS	T44RMS	11/17/2022
2_20221119T051109_20221119T051111_T44RMS	T44RMS	11/19/2022
2_20221122T052119_20221122T052433_T44RMS	T44RMS	11/22/2022
2_20221124T051121_20221124T051122_T44RMS	T44RMS	11/24/2022
2_20221127T052151_20221127T052145_T44RMS	T44RMS	11/27/2022
2_20221129T051149_20221129T051146_T44RMS	T44RMS	11/29/2022
2_20221202T052159_20221202T052600_T44RMS	T44RMS	12/2/2022
2_20221204T051201_20221204T051202_T44RMS	T44RMS	12/4/2022
2_20221207T052211_20221207T052214_T44RMS	T44RMS	12/7/2022
2_20221209T051209_20221209T051209_T44RMS	T44RMS	12/9/2022
2_20221212T052219_20221212T052632_T44RMS	T44RMS	12/12/2022
2_20221214T051221_20221214T051220_T44RMS	T44RMS	12/14/2022
2_20221217T052231_20221217T052227_T44RMS	T44RMS	12/17/2022
2_20221219T051229_20221219T051223_T44RMS	T44RMS	12/19/2022

2_20221222T052229_20221222T052753_T44RMS	T44RMS	12/22/2022
2_20230113T051151_20230113T051152_T44RMS	T44RMS	1/13/2023
2_20230118T051139_20230118T051134_T44RMS	T44RMS	1/18/2023
2_20230121T052129_20230121T052439_T44RMS	T44RMS	1/21/2023
2_20230123T051111_20230123T052031_T44RMS	T44RMS	1/23/2023
2_20230126T052111_20230126T052757_T44RMS	T44RMS	1/26/2023
2_20230131T052039_20230131T052214_T44RMS	T44RMS	1/31/2023
2_20230207T050959_20230207T051145_T44RMS	T44RMS	2/7/2023
2_20230210T051949_20230210T052635_T44RMS	T44RMS	2/10/2023
2_20230212T050931_20230212T051514_T44RMS	T44RMS	2/12/2023
2_20230215T051911_20230215T052932_T44RMS	T44RMS	2/15/2023
2_20230222T050821_20230222T051156_T44RMS	T44RMS	2/22/2023
2_20230225T051801_20230225T052114_T44RMS	T44RMS	2/25/2023
2_20230302T051729_20230302T052948_T44RMS	T44RMS	3/2/2023
2_20230312T051659_20230312T052443_T44RMS	T44RMS	3/12/2023
2_20230322T051659_20230322T052301_T44RMS	T44RMS	3/22/2023
2_20230327T051651_20230327T052835_T44RMS	T44RMS	3/27/2023
2_20230403T050651_20230403T051535_T44RMS	T44RMS	4/3/2023
2_20230406T051651_20230406T052526_T44RMS	T44RMS	4/6/2023
2_20230411T051649_20230411T052709_T44RMS	T44RMS	4/11/2023
2_20230413T050651_20230413T051512_T44RMS	T44RMS	4/13/2023
2_20230416T051651_20230416T052735_T44RMS	T44RMS	4/16/2023
2_20230418T050659_20230418T051910_T44RMS	T44RMS	4/18/2023
2_20230506T051651_20230506T052521_T44RMS	T44RMS	5/6/2023
2_20230511T051649_20230511T052839_T44RMS	T44RMS	5/11/2023
2_20230513T050651_20230513T051520_T44RMS	T44RMS	5/13/2023
2_20230516T051651_20230516T052334_T44RMS	T44RMS	5/16/2023

2_20230521T051659_20230521T052824_T44RMS	T44RMS	5/21/2023
2_20230523T050651_20230523T051519_T44RMS	T44RMS	5/23/2023

Google Earth Engine Code

```
// Define the list of available image dates
var dates = [
  '2022-10-20', '2022-10-25', '2022-10-30',
  '2022-11-04', '2022-11-19', '2022-11-24', '2022-11-29',
  '2022-12-04', '2022-12-09', '2022-12-14', '2022-12-19',
  '2023-01-13', '2023-01-18', '2023-01-23',
  '2023-02-07', '2023-02-22',
  '2023-04-03', '2023-04-13', '2023-04-18',
  '2023-05-13', '2023-05-23'
];

// Set cloud cover threshold
var cloud = 10;

// Define input and output bands
var inBands = ee.List(['B2', 'B3', 'B4', 'B5', 'B6', 'B7', 'B8', 'B8A',
'B11', 'B12', 'ndvi']);
var outBands = ee.List(['blue', 'green', 'red', 're1', 're2', 're3', 'nir',
're4', 'swir1', 'swir2', 'ndvi']);

var inBands = ee.List(['B2', 'B3', 'B4',  'B8', 'ndvi']);
var outBands = ee.List(['blue', 'green', 'red',  'nir',  'ndvi']);


// Initialize an empty image collection
var sentinel2Collection = ee.ImageCollection([]);

// Function to mask clouds in Sentinel-2 images
function maskS2clouds(image) {
  var qa = image.select('QA60');
  var cloudBitMask = 1 << 10;
  var cirrusBitMask = 1 << 11;
  var mask = qa.bitwiseAnd(cloudBitMask).eq(0)
    .and(qa.bitwiseAnd(cirrusBitMask).eq(0));
  return image.updateMask(mask).divide(10000);
}

// Function to calculate NDVI for an image
function calculateNDVI(image) {
  var ndvi = image.normalizedDifference(['B8', 'B4']).rename('ndvi');
```

```
  return image.addBands(ndvi);
}
// Function to rename bands based on date
var rename = function (oldBand) {
  return ee.String(date).cat('_').cat(oldBand); // Concatenate date variable
and band name
};

// Function to rename bands
var renameBands = function(name) {
  var parts = ee.String(name).split('_');
  return ee.String(parts.get(-2)).cat('_').cat(ee.String(parts.get(-1))); //
Get the last part (date string)
};

// Function to downscale image resolution to 10 meters
function downScaleImageTo10m(image) {
  var newResolution = 10;
  var downScaleImage = image.reproject({
    crs: 'EPSG:32644',
    scale: newResolution,
  });
  return downScaleImage;
}

// Loop through date list and create the image collection
for (var i = 0; i < dates.length; i++) {
  var date = dates[i];
  var startdate = date;
  var enddate = ee.Date(date).advance(1, 'day');

  var updated_outBands = outBands.map(rename);

  var image = ee.ImageCollection('COPERNICUS/S2_SR_HARMONIZED')
    .filterDate(startdate, enddate)
    .filter(ee.Filter.lt('CLOUDY_PIXEL_PERCENTAGE', cloud))
    .filterBounds(bardiya)
    .map(maskS2clouds)
    .map(downScaleImageTo10m)
    .map(calculateNDVI)
    .max()
    .clip(cropland2)
    .select(inBands, updated_outBands);

  if (image) {
    sentinel2Collection =
sentinel2Collection.merge(ee.ImageCollection([image]));
  } else {
    print('No image available for ' + date);
```

```
  }
}

// Print the resulting image collection
print('Sentinel-2 Collection:', sentinel2Collection);

// Define the label for classification
var label = 'wheatland';

// Sort the image collection by date
var sortedSentinel2Collection = sentinel2Collection.sort('DATE_ACQUIRED',
false);

// Convert the image collection to bands
var sentinel2Bands = sortedSentinel2Collection.toBands();
var bandNames = sentinel2Bands.bandNames();
var renamedBands = bandNames.map(renameBands);
var sentinel2Bands = sentinel2Bands.select(bandNames, renamedBands);


var featurebackselection = [
    '2022-12-09_ndvi',
    '2022-12-04_blue',
    '2022-12-04_ndvi',
    '2022-11-24_ndvi',
    '2023-04-03_nir',
    '2023-02-22_red',
    '2023-02-22_green',
    '2022-12-04_red',
    '2023-05-13_ndvi',
    '2023-02-07_green'

];

var sentinel2Bands = sentinel2Bands.select(featurebackselection);

// Reduce regions to obtain training data
var trainings = sentinel2Bands.reduceRegions({
  collection: g2,
  reducer: ee.Reducer.max(),
  scale: 10,
  tileScale: 16
});

print(trainings)

// Define the seed for reproducibility
var seed = 153;
```

```
// Split the data into calibration (70%) and validation (30%)
var split = trainings.randomColumn('split', seed);
var calibration = split.filter(ee.Filter.lt('split', 0.7)); // 70% for
calibration
var validation = split.filter(ee.Filter.gte('split', 0.7)); // 30% for
validation

// Get the list of property names
var propertyList = validation.first().propertyNames();

// Remove unwanted properties from the list
var propertyList_unwanted = ['split','Shape_Leng', 'Shape_Area',
'wheatland', 'submission', 'image_widg', 'AOI', 'crop', 'system:index'];
propertyList = propertyList.removeAll(propertyList_unwanted);

// Print the list of property names
print("Property Names:", propertyList);

// Check the size of the resulting datasets
print('Calibration Size:', calibration.size());
print('Validation Size:', validation.size());

// Create a Random Forest classifier
var classifier = ee.Classifier.smileRandomForest({
  numberOfTrees: 1000, // You can adjust this number
  variablesPerSplit: 5, // You can adjust this number
  minLeafPopulation: 1 // You can adjust this number
});

// Train the classifier
var trainedClassifier = classifier.train({
  features: calibration,
  classProperty: label,
  inputProperties: propertyList
});

// Get information about the trained classifier
var classifierResult = trainedClassifier.explain();
print('Results of trained classifier', classifierResult);

// Get the confusion matrix and overall accuracy for the training sample
var trainAccuracy = trainedClassifier.confusionMatrix();
print('Training error matrix', trainAccuracy);
print('Training overall accuracy', trainAccuracy.accuracy());

// Apply the classifier to the validation dataset
var classifiedValidation = validation.classify(trainedClassifier);
```

```
// Calculate the confusion matrix and consumer accuracy for the validation
dataset
var validationAccuracy = classifiedValidation.errorMatrix(label,
'classification');
print('Validation error matrix', validationAccuracy);
print('Consumer accuracy', validationAccuracy.consumersAccuracy());

// Classify the entire Sentinel-2 image collection
var classifiedCollection = sentinel2Bands.classify(trainedClassifier);

// Define visualization parameters
var visParams = {
  min: 0,
  max: 1,
  palette: ['FF0000', '00FF00']
};

// Add the classified image to the map
Map.centerObject(studyarea);
Map.addLayer(classifiedCollection, visParams, 'Classified Image');
Map.addLayer(g2, { color: '#177245' }, 'Ground Data');
```

Also, can be access on link
https://code.earthengine.google.com/1a56fdbb98aca25a6c9f7f5dda248f9a

Python script to create the NDVI time series chart.

```
import pandas as pd
import matplotlib.pyplot as plt
import statistics
Crop = 'wheat'
secondcrop = 'mustard_lentil'
thirdcrop = 'wheat_mustard'
# Replace 'your_file.csv' with the path to your CSV file.
file_path = 'trainings.csv'

# Read the CSV file into a DataFrame
try:
    df = pd.read_csv(file_path)
except FileNotFoundError:
    print(f"Error: File '{file_path}' not found.")
    exit(1)

wheat_df = df[(df['crop'] == Crop)]

# Group the filtered data by 'Shape_Area' and 'Shape_Leng' and plot 'ndvi'
for each group
grouped = wheat_df.groupby(['Shape_Area', 'Shape_Leng'])
plt.figure(figsize=(12, 6))
count = 0
smoothed_ndvi_list = []
for (area, length), group in grouped:
    if group['ndvi'].iloc[0] >= 0.4:
        continue  # Skip this group
    count+=1
    smoothed_ndvi = group['ndvi'].rolling(window=5).mean()
    plt.plot(range(len(group)), smoothed_ndvi, marker='o', linestyle='-
',color='#1abc9c')
    smoothed_ndvi_list.append(smoothed_ndvi.values.tolist())
print(count)

means = [statistics.mean(item) for item in zip(*smoothed_ndvi_list)]
plt.plot(range(len(means)), means, marker='o', linestyle='-',
label='mean',color='#e67e22', linewidth=4)
plt.xlabel('November 2022 - March 2023')
plt.ylabel('NDVI')
plt.title(f'NDVI Time Series Plot for {Crop} Crop Data')
plt.grid(True)
plt.legend()
plt.tight_layout()
plt.show()
```