Spectral Imaging for Crop Type Classification

MANISH CHAUHAN

Student No: 2321664

Swansea University
Prifysgol Abertawe

Master of Science

Department of Computer Science

SEP 2024

Declaration

Statement 1

This work has not been previously accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed	Manish Chauhan	Manish Chauhan (2321664)
Date	30/09/2024	Manish Chauhan (2321664)

Statement 2

This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by citations giving explicit references. A bibliography is appended.

Signed	Manish Chauhan	Manish Chauhan (2321664)
Date	30/09/2024	Manish Chauhan (2321664)

Statement 3

The University's ethical procedures have been followed and, where appropriate, ethical approval has been granted.

Signed	Manish Chauhan	Manish Chauhan (2321664)
Date	30/09/2024	Manish Chauhan (2321664)

Acknowledgments

I would like to express my profound gratitude to my supervisor, Dr. Sara Sharifzadeh, for her invaluable guidance and support throughout this project. Each meeting provided essential feedback and a critical review that significantly contributed to my research development.

I am also thankful to Swansea University for providing the necessary resources and a conducive learning environment. My gratitude extends to all my module professors whose teachings have been instrumental in my academic journey.

Special thanks to my parents, who have supported me emotionally and encouraged me throughout this journey. Their unwavering belief in my capabilities has been a constant source of strength. I am also grateful to my friends, who have been part of this academic journey. Our discussions, where we exchanged and developed ideas, have greatly enriched my experience and understanding.

Table of Contents

Abstract

Effective crop type classification is essential for enhancing agricultural productivity and management. This study utilizes the integrated capabilities of Sentinel-1 and Sentinel-2 satellite data to advance crop classification techniques in Brandenburg, Germany. By fusing ten multispectral bands from Sentinel-2 with a single Synthetic Aperture Radar (SAR) band from Sentinel-1, we exploit the complementary strengths of optical and radar data. This fusion addresses the individual limitations of each sensor type, enhancing the robustness of the classification results.

Employing a Self-Organizing Map (SOM), an unsupervised neural network, we classify various crop types based on this hybrid dataset. This innovative approach leverages the high-resolution optical data and the all-weather capabilities of radar data, yielding improved classification accuracy and reliability. Through this integrated approach, our study demonstrates the potential of combined satellite datasets to provide more reliable agricultural data. The findings contribute to the ongoing development of remote sensing methods for agriculture, offering insights that could inform further research in satellite data fusion and analysis techniques.

1 INTRODUCTION

Satellite imagery has become an essential tool in modern agriculture, transforming how farmers and researchers manage and analyze crop growth across vast landscapes. The data provided by these advanced technologies supports a variety of analytical applications, enhancing our ability to make informed decisions in agricultural management.

One critical application is the classification of crops, which plays an important role in sustainable agricultural practices and in addressing environmental challenges driven by climate change and other factors. Through advanced classification techniques, satellite image processing provides timely and accurate information on crop types, facilitating reliable estimates of crop production and supporting various decision-making processes for managing agricultural resources(Saini and Ghosh 2018).

The selection of appropriate satellite imagery for crop classification is a critical decision influenced by factors such as image availability, associated costs, the diversity of crop types, and the size of the study area. These elements determine not only the feasibility of the project but also the granularity and accuracy of the results obtained.(Zheng et al. 2015) Among these resources, the European Satellite Sentinel-2A stands out by offering multispectral data at medium spatial resolution with a fair revisit time of five days, addressing challenges posed by coarser spatial resolutions. The Multi-Spectral Instrument (MSI) on Sentinel-2 features thirteen spectral bands across three spatial resolutions, enhancing its utility across various remote sensing applications. (Table 1) (Revill et al. 2019) These bands provide essential data for monitoring vegetation health, moisture levels, and infrastructural changes, contributing significantly to agricultural and environmental research.

Transitioning to radar-based imaging, Sentinel-1 complements the optical data acquired by Sentinel-2 with its ability to penetrate cloud cover and collect data regardless of lighting conditions. Its radar instrument operates primarily in the C-band and is crucial for consistent data collection, particularly in regions prone to frequent cloud cover or where nighttime data acquisition is necessary. The versatility of Sentinel-1's data proves invaluable for continuous landscape monitoring, soil moisture estimation, and emergency management during natural disasters.

While Sentinel-2 provides invaluable multispectral data for monitoring vegetation and environmental changes, it is not without limitations, particularly when facing atmospheric challenges such as cloud cover, which severely reduces the usability of optical satellite imagery (Asner 2001).

In contrast, Sentinel-1, equipped with Synthetic Aperture Radar (SAR), operates independently of weather conditions and daylight. This radar technology enables the penetration of cloud cover and the acquisition of data regardless of the time of day or season, ensuring reliable and uninterrupted surveillance of the Earth's surface. However, despite its robust data collection capabilities, Sentinel-1 also presents unique challenges. The SAR data it generates can be affected by speckle noise, a granular disturbance that can obscure fine details in images. Moreover, while SAR is excellent for detecting physical changes and surface types, it lacks the spectral resolution provided by optical sensors like Sentinel-2, which is critical for detailed vegetation analysis and discrimination among different crop types. This interplay of strengths and weaknesses underscores the need for integrating these data sources to overcome the individual limitations and enhance overall data utility.

To harness the strengths and mitigate the weaknesses of both satellite types, data fusion emerges as a strategic approach. By integrating the high spectral resolution of Sentinel-2 with the all-weather operational capabilities of Sentinel-1, the fusion process creates a more comprehensive and accurate tool for crop type classification and environmental monitoring. This integration not only enhances the quality and reliability of the data but also extends its applicability to a broader range of environmental conditions. Such a synergistic approach leverages the unique strengths of each satellite system, ultimately leading to improved decision-making capabilities in agriculture and enhanced environmental stewardship.

The objective of this study is to explore the potential of combining Sentinel-1 and Sentinel-2 data using advanced fusion techniques and to assess the improvements in crop classification accuracy that such integration can achieve. By doing so, this research aims to provide a robust framework for agricultural monitoring that enhances the precision and efficiency of crop type identification. This approach is particularly vital for optimizing resource allocation and improving agricultural productivity, which are essential for meeting the increasing food demands without expanding agricultural footprints. Additionally, the study examines crop type classification across different temporal windows, specifically targeting peak agricultural seasons to capture the dynamic growth stages of crops. Such temporal focus is crucial for accurately understanding crop behavior and enhancing classification outcomes.

1.1 Aims & Objectives

This study aims to enhance crop type classification accuracy by integrating Sentinel-1 and Sentinel-2 data through advanced fusion techniques. This integration seeks to leverage the complementary strengths of radar and optical data to provide a more robust framework for agricultural monitoring. The project aims to improve crop classification when fused data is used compared to solo sensors. This will provide more informed agricultural planning and intervention strategies by monitoring crops and variations effectively. This study will demonstrate how fused data can support more informed agricultural planning and intervention strategies by monitoring crops and variations effectively.

Preprocess and Synchronize data: Preprocess Sentinel-1 & Sentinel-2 datasets to ensure optimal data quality and synchronization for effective fusion.

Develop Fusion Techniques: Explore and implement advanced techniques for fusing radar-based and optical-based data, aiming to enhance the dataset's comprehensive utility for accurate classification.

Implement Classification Algorithms: Accurately categorize crop types, apply self-organizing maps (SOM) and potentially other relevant machine learning algorithms to the fused dataset.

Evaluate Model Performance: Assess the effectiveness of the fusion and classification processes by comparing the accuracy of the fused data model against models using single-source data.

Analyze Results: Conduct a detailed analysis of the classification results to identify key factors contributing to the success or limitations of the fusion approach in crop type classification.

2 LITERATURE REVIEW

2.1 Crop Classification Techniques

Crop classification has been a fundamental aspect of agricultural remote sensing, facilitating efficient land use management, crop monitoring, and yield estimation. Traditionally, crop type classification relied heavily on statistical methods that utilized basic spectral features extracted from satellite imagery. These methods, while pioneering for their time, often struggled with the high variability and similar spectral signatures between different crop types, leading to less accurate classifications.

2.1.1 Traditional Methods

Early approaches to crop classification were primarily based on supervised and unsupervised statistical techniques. These methods included basic classifiers like the Maximum Likelihood Classifier (MLC) which, despite their simplicity, were widely used due to their ease of implementation and interpretation. However, the accuracy of these classifiers was generally limited by the spectral resolution of the satellite sensors and the lack of ability to capture complex patterns within the crop fields (Zheng et al. 2015).

2.1.2 Shift to Machine Learning Models

The advent of advanced satellite systems like Sentinel-1 and Sentinel-2 has been accompanied by significant advancements in data processing techniques. Machine learning models have progressively taken the forefront in crop classification tasks, owing to their ability to handle large datasets and extract complex spectral and spatial patterns that are often invisible to traditional methods.

Support Vector Machines (SVM) and Random Forests (RF) are among the first wave of machine learning techniques applied to crop classification. These methods provide better accuracy by learning non-linear boundaries between different crop types and incorporating multiple data sources, including optical and radar data. SVM, for instance, has been particularly noted for its robustness in high-dimensional spaces, making it suitable for handling the multi-spectral data from Sentinel-2 (Eisfelder et al. 2024).

2.2 Remote Sensing Data in Agriculture

Remote sensing technology, particularly through the use of Sentinel-1 (SAR) and Sentinel-2 (optical) satellites, has become a cornerstone in agricultural applications. These satellites provide critical data that supports a wide range of activities, from basic crop monitoring to advanced crop type classification and environmental impact assessments.

2.2.1 Utilization of Sentinel-1 SAR Data in Agriculture

Sentinel-1's SAR technology is instrumental in agriculture, especially in areas frequently obscured by cloud cover or during nighttime. SAR data's ability to penetrate cloud cover and its independence from sunlight enable continuous monitoring of agricultural fields, a crucial advantage during critical planting or harvesting periods. This data is particularly valuable for assessing soil moisture levels and tracking changes in field conditions, which are vital for irrigation management and assessing crop health during unexpected weather events. Despite these advantages, SAR data's complexity and the presence of speckle noise often require sophisticated processing techniques to ensure accurate interpretations and useful application in crop classification tasks.

2.2.2 Integration with Sentinel-2 Optical Data

Sentinel-2 complements the capabilities of Sentinel-1 by providing high-resolution optical data, capturing details at a finer scale with its multispectral capabilities. This data is paramount for identifying crop types, assessing vegetation health through indices like NDVI, and monitoring changes over time with its short revisit cycle. The optical data, however, faces limitations under poor weather conditions and during nighttime, where no data capture is possible. These limitations underscore the necessity of integrating optical data with SAR data to mitigate gaps caused by environmental conditions.

2.2.3 Synergistic Applications for Enhanced Agricultural Insight

The integration of Sentinel-1 and Sentinel-2 data harnesses the strengths of both SAR and optical modalities, providing a more robust dataset for agricultural monitoring. Studies such as those by (Felegari et al. 2021) and (Van Tricht et al. 2018) highlight how combining these data types improves the accuracy and reliability of crop type classifications and agricultural mappings. This synergy allows for more detailed and comprehensive agricultural assessments, supporting the development of strategies to optimize crop yield, improve crop health monitoring, and manage resources more efficiently.

This integrated approach not only overcomes individual limitations of each satellite system but also enhances the temporal and spatial resolution of the observations, making it possible to conduct detailed analyses and make informed decisions to support sustainable agricultural practices.

2.3 Data Fusion Techniques in Remote Sensing

Data fusion techniques in remote sensing have increasingly become crucial in agricultural applications, enhancing the accuracy and reliability of monitoring systems. The integration of Synthetic Aperture Radar (SAR) and optical data from satellites like Sentinel-1 and Sentinel-2 provides a comprehensive approach to agricultural monitoring, overcoming individual limitations of each data type.

2.3.1 Pixel-Level and Feature-Level Fusion

Recent advancements have shown significant improvements in crop classification through the use of pixel-level and feature-level fusion techniques. The pixel-level fusion directly combines the information from SAR and optical images before any classification process, allowing for a more detailed representation of the surface features. This method is particularly effective in enhancing the resolution of the features captured, thus providing richer information for subsequent classification algorithms (Anon n.d.-a).

Feature-level fusion, on the other hand, involves extracting and combining features from SAR and optical images, which are then used to train machine learning models. This approach leverages the strengths of both data types—optical data provides detailed spectral information about crop health and type, while SAR data contributes with its all-weather capabilities and structural information. Studies have demonstrated that feature-level fusion effectively addresses the limitations of each sensor type, such as optical sensors' susceptibility to atmospheric conditions and SAR's sensitivity to surface roughness (Quan et al. 2024).

2.3.2 Decision-Level Fusion

Decision-level fusion involves making separate classifications from SAR and optical data and then integrating these classifications to produce a final decision. This technique allows for the exploitation of different characteristics of the data sources, where each classification can be optimized independently before fusion. The integration at the decision level has been shown to improve the robustness of the classification results, particularly in complex agricultural environments where different crops and land cover types present unique challenges.

2.4 Multispectral Data Utilization

Multispectral imaging from Sentinel-2 plays a pivotal role in modern agricultural monitoring, offering a nuanced view of vegetation health and dynamics through its array of spectral bands ranging from the visible to the near-infrared. These bands are instrumental in assessing various physiological states of crops, such as chlorophyll concentration and water stress, enhancing the precision of agricultural management strategies.

The utilization of Sentinel-2's multispectral data is significantly enhanced when complemented by SAR data from Sentinel-1. This combination leverages the strengths of both optical and radar technologies to provide a more comprehensive understanding of agricultural landscapes. For instance, while Sentinel-2 excels in detailed vegetation analysis through indices like NDVI (Normalized Difference Vegetation Index), Sentinel-1's SAR data adds value by offering reliable data acquisition under any weather conditions, crucial for consistent monitoring (Anon n.d.-b) (Quan et al. 2024).

In practical applications, the integration of these data types supports advanced agricultural tasks such as soil moisture estimation and crop type differentiation. Recent studies have demonstrated the effectiveness of combining Sentinel-1 and Sentinel-2 for tasks like soil moisture retrieval, where the multispectral data from Sentinel-2 help in refining the interpretations made from SAR data, especially in differentiating surface characteristics and improving the accuracy of moisture estimates (Ma, Li, and McCabe 2020).

Moreover, the fusion of SAR and optical data has been applied successfully in comprehensive land cover classification frameworks. This approach not only improves the classification accuracy but also enhances the capability to monitor subtle changes in land cover over time, crucial for long-term environmental and agricultural planning (De Fioravante et al. 2021).

2.5 Self-Organizing Maps (SOM)

Self-Organizing Maps (SOM) represent a class of unsupervised neural networks that excel in clustering and visualizing high-dimensional data. Developed by Teuvo Kohonen in the 1980s, SOMs operate by organizing data into a two-dimensional grid where similar data points are mapped close to each other, thus preserving the topological properties of the original dataset. This process involves iteratively adjusting the weights of a predefined grid of neurons to reflect the features of input vectors, a method effective in handling the complexities of large datasets often encountered in fields like remote sensing and bioinformatics. SOMs are particularly noted for their ability to reduce dimensionality and create a discretized representation of input data, which facilitates both the visualization and analysis of data characteristics that are not immediately apparent in the original high-dimensional space. This makes them an invaluable tool for feature extraction and data exploration, highlighting subtle variations within the data that might be missed by other analytical methods (Kohonen 2001).

In the context of remote sensing for agricultural applications, SOMs are particularly valuable due to their ability to organize large datasets without prior labeling, revealing intrinsic patterns that are often not immediately obvious. The use of SOMs helps in the classification of land cover, detection of agricultural patterns, and monitoring of crop health, all without the need for extensive and often laborious labeling of training data. This capability not only saves significant time and resources but also reduces the subjectivity involved in manual classification, leading to more objective and reproducible results.

2.5.1 Advantages of SOM in Agricultural Remote Sensing

1. **Reduction of Label Dependency**: One of the primary advantages of using SOM in remote sensing is its unsupervised nature, which eliminates the need for extensively labeled datasets. Labeling satellite images is notoriously time-consuming, subject to subjectivity, and often impractical for large-scale studies due to the sheer volume of data and variability of terrain (Wan and Fraser 2000) (Ruß et al. 2009).

2. **Exploratory Data Analysis**: SOM provides a powerful tool for exploratory data analysis, offering visual insights into the data structure. This capability is crucial for identifying patterns, anomalies, and clusters within complex agricultural datasets, which can inform further analysis and decision-making (Ruß et al. 2009).

2.5.2 Drawbacks and Challenges

While SOMs offer significant advantages, they are not without their challenges:

1. **Parameter Sensitivity**: The performance of SOMs heavily relies on the correct setting of parameters such as learning rate, grid size, and neighborhood function. Inappropriate parameter settings can lead to poor model convergence and suboptimal clustering outcomes (Ruß et al. 2009).

2. **Generalization Difficulty**: Although SOMs are excellent for exploring and clustering data, they can struggle with generalization across different datasets. This is due to their reliance on the specific statistical properties of the data they were trained on, which might not be representative of other datasets (Yuan, Van Der Wiele, and Khorram 2009).
3. **Computational Complexity**: For very large datasets, such as those typical in satellite imagery, the computational demand of training SOMs can be significant. This factor can limit their applicability in operational settings where quick turnaround is required (Wan and Fraser 2000).

2.5.3 Unsupervised vs. Supervised Learning

Compared to supervised learning methods, which require precise labels and often involve considerable preprocessing of data, unsupervised methods like SOM offer a more flexible and less labor-intensive approach (Kaski 2017; Ruß et al. 2009). However, supervised methods typically achieve higher accuracy and are more straightforward in their application to prediction tasks since they are trained explicitly to optimize performance on such tasks (Ruß et al. 2009).

Despite these challenges, the adoption of unsupervised learning techniques like SOM in remote sensing has been growing, particularly in applications where rapid, large-scale analysis of unlabelled data is necessary (Wan and Fraser 2000). The ability of SOMs to provide valuable insights without the extensive labeling required by supervised methods makes them especially useful in ongoing monitoring of agricultural environments where conditions change rapidly and new data constantly flows in (Yuan et al. 2009).

2.6 Challenges and Opportunities in Satellite Data Fusion

In the analysis of challenges and opportunities in satellite data fusion, several technical difficulties arise, particularly concerning the integration of disparate data sources with varying resolutions and alignment issues. The fusion of remote sensing data, such as from optical, infrared, and SAR (Synthetic Aperture Radar) sources, offers significant opportunities for enhanced observation capabilities but also presents notable challenges:

1. **Data Misalignment**: Misalignment between data from different sensors can be significant, particularly when combining data with different spatial and temporal resolutions. This misalignment must be corrected to ensure accurate data integration and analysis (Wang 2024) (Taylor 2024).

2. **Resolution Differences**: The differences in resolution between data sources can lead to complications in data integration. High-resolution data might capture fine details that are absent in lower-resolution data, complicating the fusion process and potentially leading to information loss or misrepresentation in the fused product (Almohsen 2024).
3. **Integration of Disparate Data Sources**: Each type of satellite data be it optical, SAR, or others has its own unique characteristics and information content. Combining these effectively to leverage their strengths without losing critical information poses significant challenges. The need for innovative fusion techniques that can handle the inherent differences in data types is crucial for successful integration (Declaro and Kanae 2024).
4. **Complexities in Processing**: The fusion process itself can be computationally intensive, especially when dealing with large datasets typical in satellite imagery. This often involves sophisticated data transformation techniques which need to be both efficient and effective to be practical for large-scale applications (Declaro and Kanae 2024).
5. **Opportunities for Enhanced Monitoring**: Despite these challenges, the fusion of multisource satellite data opens up vast opportunities for improved environmental monitoring, disaster management, and urban planning. By combining different data types, researchers and practitioners can gain a more comprehensive understanding of the observed areas, leading to better-informed decision-making and resource management (Wang 2024).

Addressing these challenges involves developing advanced algorithms that can effectively handle the complexities of data fusion. Research and innovation in areas such as machine learning, artificial intelligence, and advanced statistical models are critical to overcoming these obstacles and maximizing the potential of satellite data fusion for various applications (Almohsen 2024).

2.7 Advancements in Satellite Technologies and Their Impact on Agriculture

Recent technological innovations in satellite imaging and data processing have revolutionized the field of precision agriculture. Enhanced sensor capabilities and sophisticated computational methods now allow for more detailed and frequent monitoring of crop health, optimization of agricultural inputs, and effective management of resources.

1. **Enhanced Sensor Technologies**: Platforms like Sentinel-2 have significantly advanced with improved temporal, spatial, and spectral resolutions, enabling precise detection and management of various crop conditions (Radočaj, Plaščak, and Jurišić 2023; Segarra et al. 2020).
2. **Global Navigation Satellite Systems (GNSS)**: Systems such as GPS, GLONASS, Galileo, and BeiDou offer precise field mapping and machinery guidance, essential for implementing efficient farming techniques and optimizing resource allocation (Radočaj et al. 2023).
3. **Integration of IoT and AI**: The convergence of Internet of Things (IoT) and Artificial Intelligence (AI) technologies with satellite data enriches farming practices by providing real-time insights and enhancing decision-making processes (Anon n.d.-d).
4. **Machine Learning and Big Data Analytics**: These technologies play a crucial role in analyzing complex datasets from satellites, leading to actionable agricultural insights and improved predictive capabilities (Anon n.d.-d; Soussi et al. 2024).
5. **Challenges and Opportunities**: Despite these advancements, integrating various satellite data types remains challenging, requiring ongoing innovation in data processing and analysis methods to fully realize the potential of these technologies in agriculture (Mulla 2021).

These advancements not only contribute to increased agricultural productivity but also support sustainable practices by ensuring better resource management and reducing environmental impacts.

3 METHODOLOGY

3.1 Data Description

3.1.1 Sentinel-1 Data

Sentinel-1, part of the Copernicus program, provides Earth observation data in the form of C-band Synthetic Aperture Radar (SAR) imagery. SAR can collect data under any weather conditions, day or night. This is particularly important for agricultural applications, such as crop health or soil moisture observations that might be impeded by the presence of clouds or nighttime conditions.

The Sentinel-1 satellites operate in dual-polarization mode (VV and VH), which are particularly adept at detecting changes in surface textures and moisture content. Crop canopies can partially block microwaves, but some signals can still penetrate, giving us significant information about the underlying soils and crop structural properties.

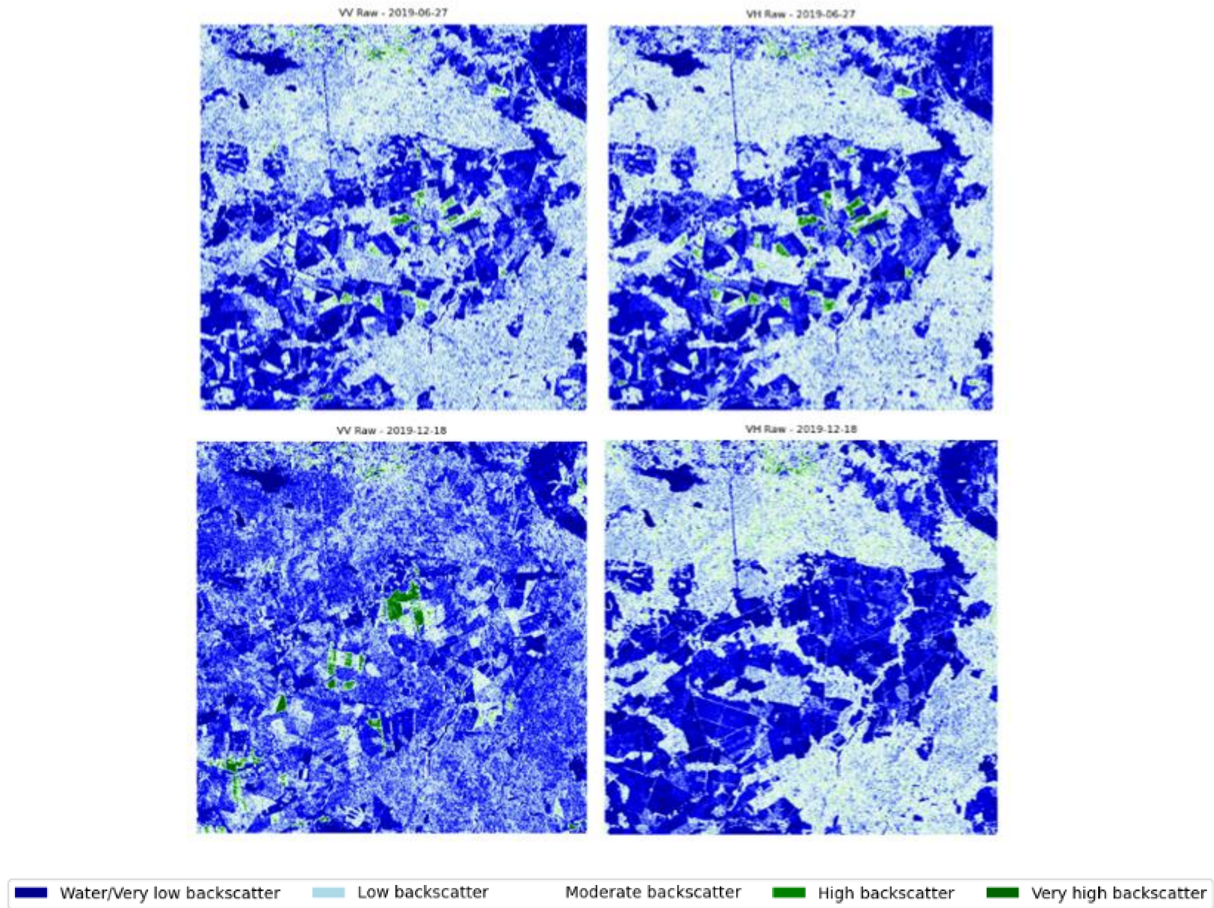

Figure 1: Seasonal Variability in Sentinel-1 SAR Imagery: Comparative Analysis of VV and VH Polarizations

These images Figure 1 show raw Sentinel-1 radar data of the Vegetation-to-Vegetation polarisation (VV) and Vegetation-to-Horizontal polarisation (VH) on 27 June 2019 and 18 December 2019. The first date clearly shows the summer vegetation growth, while the second shows the duller landscape in winter.

3.1.2 Sentinel-2 Data

Sentinal-2, operated by the European Space Agency (ESA) as part of the Copernicus Programme, offers high-resolution optical imagery from the Earth's surface. The Multi-Spectral Instrument (MSI) on Sentinel-2 captures data across thirteen spectral bands in visible, near-infrared, and short-wave infrared parts Table 1

Spatial Resolution (m)	Bands	Central Wavelength (nm)	Bandwidth (nm)
10	Band 2 - Blue	492.4	66
	Band 3 - Green	559.8	36
	Band 4 - Red	664.6	31
	Band 8 - NIR	832.8	106
20	Band 6 – Red edge	740.5	15
	Band 7 – Red edge	782.8	20
	Band 8A – Narrow NIR	864.7	21
	Band 11 - SWIR	1613.7	91
	Band 12 - SWIR	2202.4	175
60	Band 1 – Coastal aerosol	4442.7	21
	Band 9 – Water vapour	945.1	20
	Band 10 – SWIR - Cirrus	1373.5	31

Table 1: Spectral Band Characteristics of Sentinel-2's Multi-Spectral Instrument (MSI)

Sentinal-2 revisits the equator every five days. It offers timely data that is vital for monitoring seasonal changes in vegetation and environmental transformation over time. This imagery is also valuable for monitoring crop cycles.

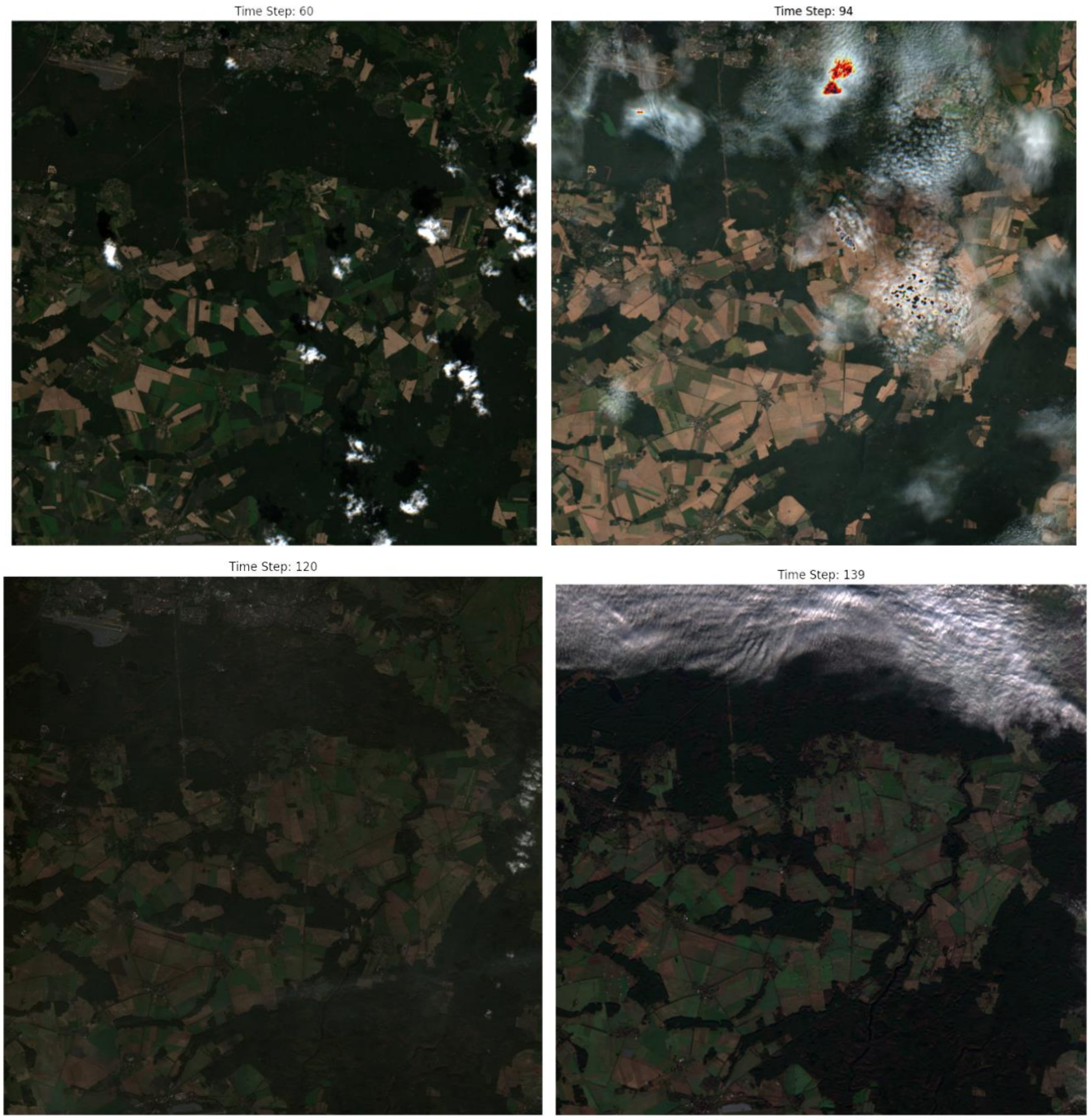

Figure 2: Sentinel-2 Images at Time Steps: 60 (May 29, 2019), 94 (August 27, 2019), 120 (October 14, 2019), and 139 (December 30, 2019)

The Figure 2 show seasonal changes in land cover from May to December, with a noticeable shift in vegetation patterns. The August image shows distinct areas of bare soil or harvested fields, and there is increased cloud cover. Images in October and December show a gradual reduction in green vegetation, with more exposed soil visible and varying cloud patterns, reflecting the progression from late summer through autumn into winter.

3.1.3 Temporal Coverage in Sentinel-1 and Sentinel-2

The Sentinal-1 dataset spans from 2019, with data collected at regular intervals from January 4 to December 31; this allows for an analysis of phenological changes throughout the crop growth stages from Seedling to harvest. Sensinal-2 captures high-resolution optical imagery across 144 timestamps within the same year, from January 2 to December 30, frequently revisiting every phenological stage in multiple spectral bands. The detailed temporal coverage of both Sentinal-1 and Sentinal-2 offers a dynamic and robust dataset essential for temporal analysis in agricultural applications.

3.1.4 Temporal Window Selection for Crop Analysis in Germany

The May to July period was selected as the optimal temporal window for crop type classification using the Self-Organizing Map (SOM). During May to July, several key crops undergo crucial growth phases. For instance, Spring Barley and Corn are in their mid-season, Winter Barley and Winter Wheat, on the other hand, are nearing harvest, Rapeseed, another significant crop, remains in mid-season through May.

Crop Name	Growth Stage	Key Characteristics
Barley (Spring)	Mid-Season	Active growth, distinct spectral signatures
Corn	Mid-Season	Vigorous growth phase
Barley (Winter)	Approaching Harvest	Maturing crop, easier to distinguish crop type as it prepares for harvest.
Wheat (Winter)	Approaching Harvest	Similar to winter barley
Rapeseed	Mid-Season	Flowering stage

Table 2: Characteristics of Crops in Germany (May to July)

The May to July window typically experiences stable weather conditions in Germany, reducing the likelihood of cloud cover interfering with satellite observations. This stability ensures high-quality, consistent imagery across the study area, essential for leveraging multi-spectral satellite data to its fullest potential.

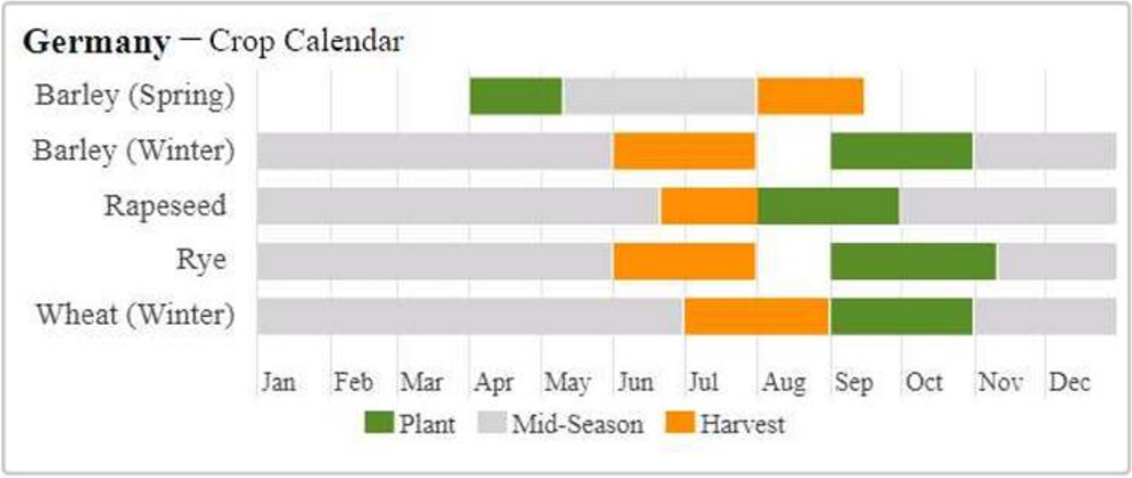

Figure 3:Germany Crop Calendar (Anon n.d.-c)

3.2 Data Preprocessing

Sentinal-1 Preprocessing includes thermal noise removal; reducing speckle noise is crucial to enhance the clarity and usability of Synthetic Aperture Radar (SAR) imagery.

The primary features recorded by Sentinel-1 are the VV and VH polarized backscatter. For Sentinel-1, a Lee filter (Lee 1981) was applied to reduce radar speckles in the images. This filter is specifically designed to surpass noise while preserving the essential details and edges of the image, making it best for enhancing SAR imagery. Lee filters work by averaging pixels within a moving window and adjusting the weights according to variance. With this adaptive approach, the filter preserves high-resolution features in areas with low variance while smoothing out areas with high variance, indicative of speckle noise.

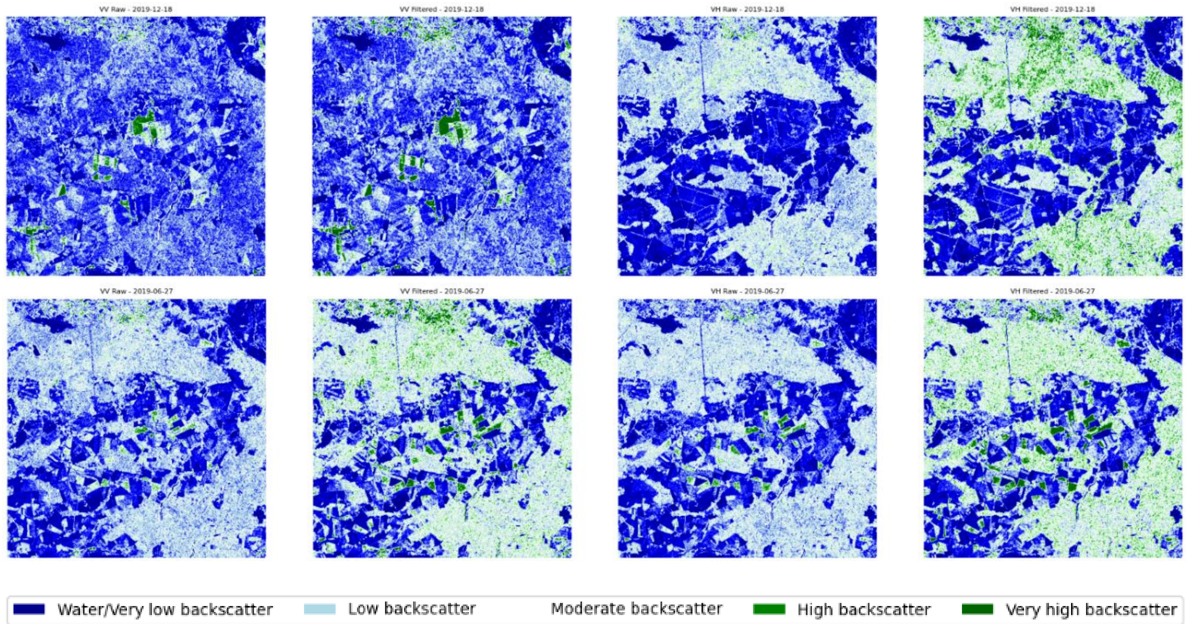

Figure 4: Comparative Visualization of Sentinel-1 SAR Imagery: Raw vs. Filtered Across VH and VV Polarizations for June and December 2019, Highlighting Backscatter Intensity Levels

Analyses showed that after applying the Lee filter to Sentinel-1 SAR data, there was a significant reduction in speckle noise, enhancing the clarity of both the June and December images. The filtered images exhibit more defined landscape features, demonstrating the filter's effectiveness in maintaining essential geographical details while minimizing noise-induced distortions.

Preprocessing of sentinel 2 data has several vital aspects. These steps include normalization, cloud masking, and handling missing data.

3.2.1 Handling Missing Data & Normalization

Handling missing data, including NaNs (Not a Number) and infinite values, is essential to prevent errors in computation and analysis. Once these data anomalies are addressed properly, normalization is applied as a crucial preprocessing step. This involves scaling the pixel values of each spectral band to a standard range, typically 0-255, converting them to 8-bit unsigned integers. The normalization process rescales the raw pixel values of each spectral band to a standardized range of 0-255, converting them to 8-bit unsigned integers. This process enhances the performance of classification algorithms by reducing skewness in data distribution.

3.2.2 Cloud Masking

Cloud cover is widespread in Sentinel-2 data (Frantz et al. 2018). A cloud mask is applied where cloud pixels are set to zero, and clear pixels are set to one. The cloud masking Figure 5 process employs a threshold-based approach utilizing Sentinel-2's visible bands. Predefined brightness thresholds are applied to bands 2, 3, and 4 to identify and mask potential cloud pixels. This binary mask effectively removes cloud-affected areas and ensures that only cloud-free pixels are processed.

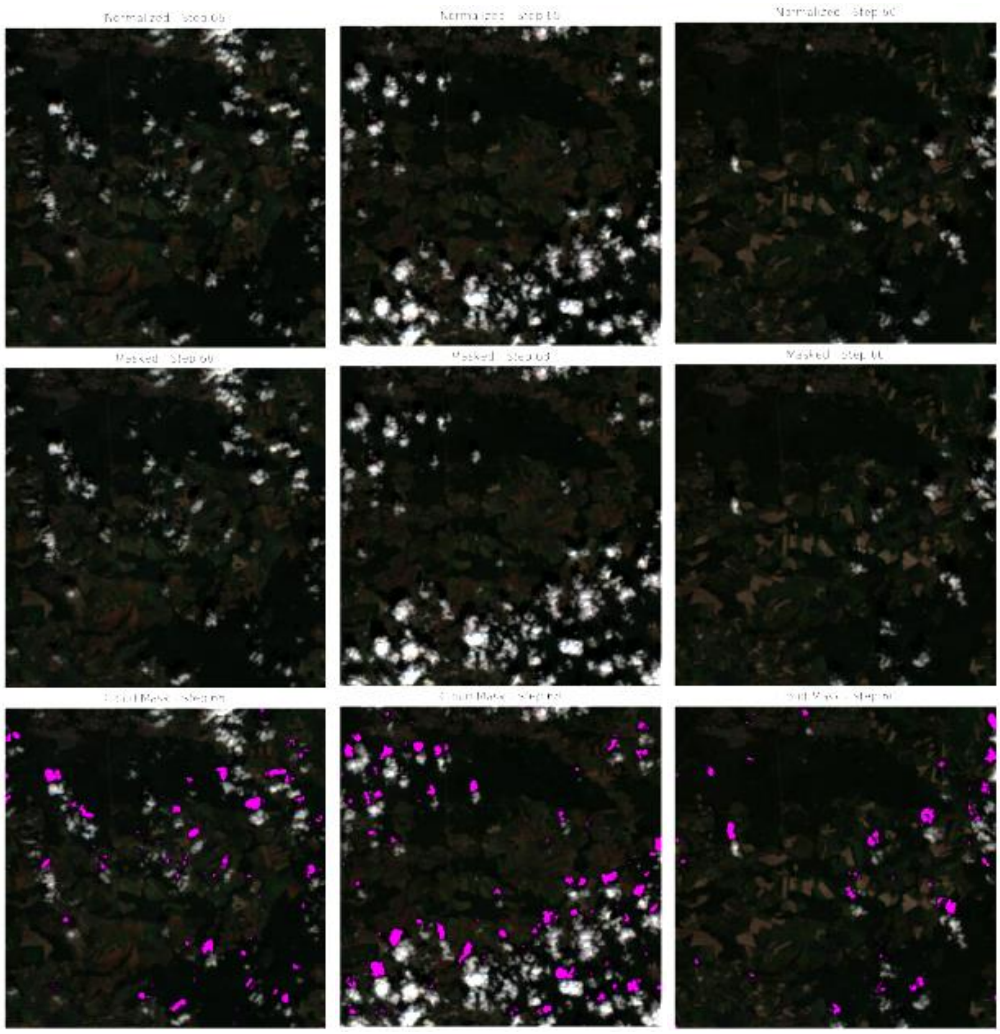

Figure 5: Progression of Sentinel-2 Image Processing: From Normalization to Cloud Masking Highlighted in Pink

Table 3: Data Integrity Post-Cloud Masking Across Sentinel-2 Spectral Bands

Band 0 - Data lost: 1.73%, Data remaining: 98.27%
Band 1 - Data lost: 0.64%, Data remaining: 99.36%
Band 2 - Data lost: 0.18%, Data remaining: 99.82%
Band 3 - Data lost: 0.21%, Data remaining: 99.79%
Band 4 - Data lost: 0.16%, Data remaining: 99.84%
Band 5 - Data lost: 0.13%, Data remaining: 99.87%
Band 6 - Data lost: 0.13%, Data remaining: 99.87%
Band 7 - Data lost: 0.12%, Data remaining: 99.88%
Band 8 - Data lost: 0.13%, Data remaining: 99.87%
Band 9 - Data lost: 0.18%, Data remaining: 99.82%
Band 10 - Data lost: 0.18%, Data remaining: 99.82%
Band 11 - Data lost: 0.23%, Data remaining: 99.77%

Percentage of data lost due to cloud masking: 0.33%, Percentage of data remaining after cloud masking: 99.67%

3.3 Data Fusion

The temporal process involved handling the different timestamps of Sentinal-1 and Sentinal-2 to synchronize them; both satellites do not capture images at the same time due to their different orbits,

For each data and time that Sentinal-1 captured data, the Sentinal-2 data set is examined to identify the nearest available data both before and after this date/time; for e.g, if Sentinel-1 captured data on June 10th and Sentinel-2's data from June 10th was obscured by clouds, the nearest clear sentinel-2 data points around this data such as 9th June or June 11th are considered. Following the preprocessing steps, temporal alignment is to synchronize the data times between the two satellite systems. The final stage involves data fusion using a feature stacking technique.

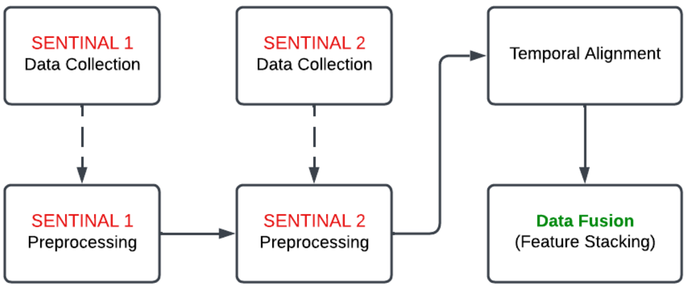

Figure 6: illustrates this process, providing a visual representation of the sequential steps involved in the data handling and analysis procedure.

3.3.1 Linear Interpolation

Once clear data points are identified a linear interpolation method is applied; the interpolation estimates the data values at missing times using this formula Equation $y = y1 + (x - x1) \cdot (x2 - x1y2 - y1)$ 1 where $x1$ and $x2$ are the times for the available data points, $y1$ and $y2$ are the data values at these times, x is the time at which data is missing, with y being the interpolated value.

$$y = y1 + (x - x1) \cdot (x2 - x1y2 - y1) \quad 1$$

For the implementation of this interpolation, libraries such as NumPy and SciPy in Python were employed. Specifically, the *numpy.interp* function was used to interpolate Sentinel-2 data values at the required Sentinel-1 timestamps, ensuring data continuity and alignment.

3.3.2 Feature stacking

Feature stacking is a data fusion technique used to combine multiple sets of data features into a single dataset. In the context of satellite data, feature stacking involves layering the different spectral bands and radar data from Sentinel-1 and Sentinel-2 satellites to create a comprehensive feature set.

Ten specific bands from the Sentinel-2 dataset were selected for fusion. These bands correspond to indices 1, 2, 3, 4, 5, 6, 7, 8, 10, and 11 in the original dataset. The single-band Sentinel-1 VH polarization data was then combined with these 10 Sentinel-2 bands. The fusion process was executed on a per-pixel basis. For each pixel location, the corresponding Sentinel-1 value was stacked with the 10 selected Sentinel-2 values, resulting in an 11-dimensional feature vector.

3.3.3 Feature Distribution

Following the stacking of features, the next step involved a detailed examination of the distribution of all 11 Figure 7 bands within the fused dataset; the initial distribution analysis of the fused data unveiled some concerning patterns. Histograms for each band showed highly skewed distributions, with many exhibiting sharp peaks and long tails. This non-uniform distribution suggested that the raw data might not be optimal for subsequent analysis and crop classification.

To address these distribution issues, a series of preprocessing steps were implemented. A logarithmic transformation was applied to the Sentinel-1 band (Band 1) to handle its typically wide range of values, while a square root transformation was used for the Sentinel-2 bands (Bands 2-11) to moderate their skewness. The result of preprocessing showed an improvement in data distribution. Histograms of all bands displayed are balanced, transformation, particularly in band 1, where the extreme skewness was significantly reduced. The other bands also showed a more uniform distribution with a peak closer to the center and a reduced tail effect.

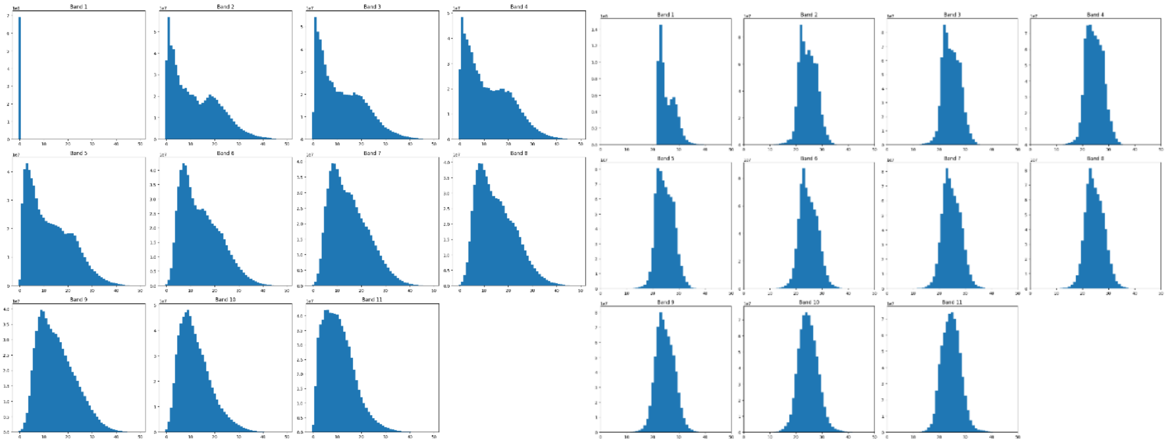

Figure 7: Distribution Histograms

The first image Figure 8 is a multispectral RGB composite capturing visible light and some near-infrared light, providing a natural color representation of the terrain. The second image Figure 8 captures data in a single band, offering higher resolution. The third image Figure 8 is a fused image presenting all spectral data.

Additionally, the third image Figure 8 shows some apparent discontinuities or "seem lines" that create a mosaic-like effect; these lines, visible as subtle changes in tone or texture across different sections of the image, likely resulting from the data fusion process or from merging of multiple satellite bands.

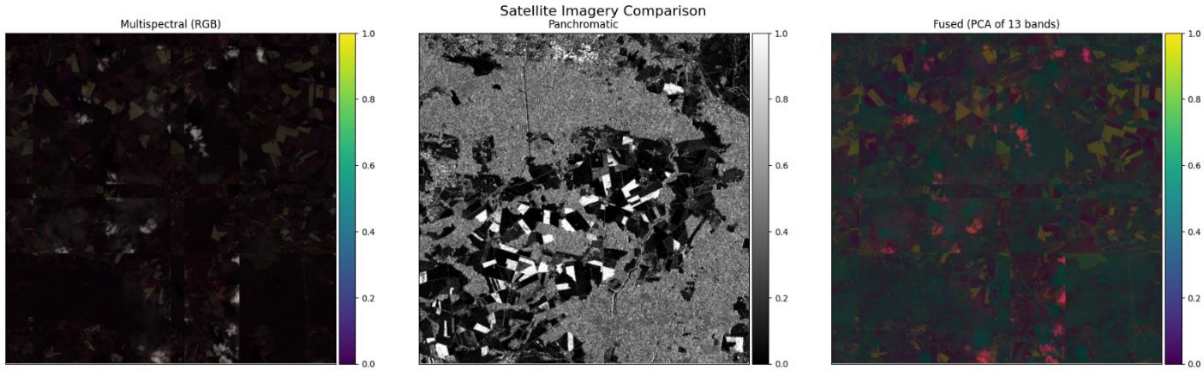

Figure 8: Comparative Display of Multispectral, Panchromatic, and Fused PCA Imagery

3.4 Dimensionality Reduction

Principal component Analysis (PCA) is applied to reduce dimensionality of the data while preserving essential information (Chitwong et al. n.d.). This technique is useful for handling high-dimensional datasets common in remote sensing and hyperspectral imaging applications. Prior to applying PCA, data undergoes a standardization process. The fused data consists of multiple bands is standardized such that each band has a mean of zero and a standard deviation of one. Standardization ensures that PCA treats all bands equally.

The PCA transformation is performed on flattened data where each pixel across all bands is treated as a separate observation; the flatting process involves reshaping the three-dimensional data into two-dimensional, enabling the PCA algorithm to process the data effectively. After flatting, PCA identifies the direction of maximum variance in the data, creating new uncorrelated variables called principal components. These components are ranked by their ability to explain variance in the data, with the first few typically capturing the most significant information.

3.5 Deep Learning Model

The study implements a Self-Organizing Map (SOM), an unsupervised learning technique, to analyze and classify Sentinel satellite datasets. Unlike supervised methods, SOMs do not require pre-labeled training data, making them suitable for exploring complex, high-dimensional remote sensing data. This unsupervised approach allows for the discovery of inherent patterns and structures within the data without prior knowledge of the output classes.

3.5.1 SOM architecture

The research aimed at classifying crop types. The SOM architecture is structured to effectively manage and interpret complex satellite imagery data from Sentinel-1, Sentinel-2, and their fused dataset. Utilizing a 90 x 90 neuron grid provides high resolution for capturing intricacies of the input data. The SOM is set with a learning rate of 0.5 and a sigma of 1.5; these settings ensure that the model dynamically adapts to the varied features presented by different crop types. This architecture not only uncovers spectral and spatial relationships in the data but also significantly boosts the accuracy and reliability of the crop classification results, all without the need for pre-labeled training datasets.

3.5.2 Unsupervised Training Process

The learning rate and the sigma parameter play pivotal roles in the training dynamics. During training, random samples from the respective dataset are presented to the network without associated labels. The algorithm identifies the best matching unit (BMU) – the node whose weight vector is most similar to the input vector. The BMU and its neighboring nodes are then updated, with the magnitude of updates decreasing with distance from the BMU and as training progresses.

This iterative, unsupervised process gradually organizes each map, with similar input patterns activating nearby regions on the SOM grid. The result is a series of self-organized topological representations, each capturing the unique characteristics of its respective dataset (Sentinel-1, Sentinel-2, or fused). This approach allows for the discovery of dataset-specific natural clusters and patterns that might not be apparent through traditional supervised methods, providing insights into the distinct information captured by radar, optical, and fused data.

3.5.3 Clustering and Label

After the Self-Organizing Map (SOM) has completed its training phase, the structured grid of neurons represents different clusters, each corresponding to distinct patterns identified within the satellite imagery data. The next critical phase in the process involves assigning labels to these clusters, a step that transforms the unsupervised learning achievements of the SOM into actionable, supervised outcomes.

To map these labels to the predicted clusters, the process begins by identifying the Best Matching Unit (BMU) for each input vector within the dataset. The clusters formed are analyzed to identify the most common label among the data points within each cluster. This is done by cross-referencing the geographical coordinates and corresponding labels in the GeoJSON file with the data points assigned to each cluster. For each cluster, the process involves examining the labels of all the data points it contains, as provided by the GeoJSON file. The label that appears most frequently is selected as the representative label for that cluster.

This label assignment process integrates the unsupervised learning results from the SOM with the supervised information provided by the GeoJSON file, effectively bridging the gap between discovering patterns in data (unsupervised) and applying meaningful labels to those patterns (supervised).

3.5.4 Performance Evaluation

Evaluating the performance of a SOM in unsupervised learning scenarios involves assessing how well the model has identified and grouped similar data points. While traditional metrics like accuracy, precision, recall, and F1-score can provide insight, These metrics help determine the accuracy of the clustering by comparing the SOM's output to the actual data categories.

4 RESULT & DISCUSSION

4.1 Overview of SOM Applications

Self-Organizing Maps (SOM) were applied to three distinct datasets: Sentinel-1, Sentinel-2, and a fusion of both. The primary objective of using SOM for each dataset was to effectively classify and visualize the complex spatial patterns of satellite data. The study focused on four main scenarios:

- Sentinel-1 temporal window data
- Sentinel-2 temporal window data
- Single-day fused Sentinel-1 and Sentinel-2 data
- Temporal window fused Sentinel-1 and Sentinel-2 data

Expected outcomes included improved cluster formation, higher classification accuracy, and the potential for unique insights that may not be apparent from single-sensor data.

4.2 Results from SOM on Sentinel-1 Data

The application of Self-Organizing Maps (SOMs) on Sentinel-1 data has yielded insightful results that significantly enhance our understanding of crop type classifications based on radar imagery. Sentinel-1, with its Synthetic Aperture Radar (SAR) capabilities, provides robust data that is not affected by cloud cover or lighting conditions, making it an invaluable resource for consistent and reliable remote sensing, particularly in agricultural monitoring.

4.2.1 Model Performance

The Self-Organizing Map (SOM) was applied to Sentinel-1 data, focusing on VV and VH polarizations for the period from May to July. This temporal focus captured the key growing season for crops in the region. The SOM was configured with a 90x90 grid, a learning rate of 0.5, and a neighborhood function sigma of 1.5. The model was trained for 30,000 iterations, ensuring thorough exploration of the data space and stable cluster formation.

4.2.2 Accuracy and Validation

The accuracy of the SOM was evaluated by comparing the clustered results with ground truth labels, The overall accuracy of the SOM classification was 70.49%, which is promising for crop classification using SAR data alone. A detailed classification report revealed varying performance across different crop types:

Crop ID	Crop Type	Precision	Recall	F1-Score	Support
0	Wheat	83.85%	91.86%	87.67%	3,761,685
1	Rye	28.13%	79.24%	41.52%	412,906
2	Barley	57.76%	2.96%	5.63%	349,826
3	Oats	69.18%	17.09%	27.41%	271,474
4	Corn	46.20%	2.10%	4.02%	30,137
5	Oil Seeds	62.70%	13.47%	22.18%	284,708
6	Root Crops	58.47%	44.31%	50.41%	252,216
7	Meadows	49.85%	10.44%	17.26%	3,162
8	Forage Crops	53.33%	26.14%	35.08%	219,891
9	No Data	42.21%	7.09%	12.14%	173,995
Overall Accuracy					70.49%

Table 4: Classification Results Table for SOM on Sentinel-1 Data

4.2.3 Visual Analysis

Predicted Crop Types Map

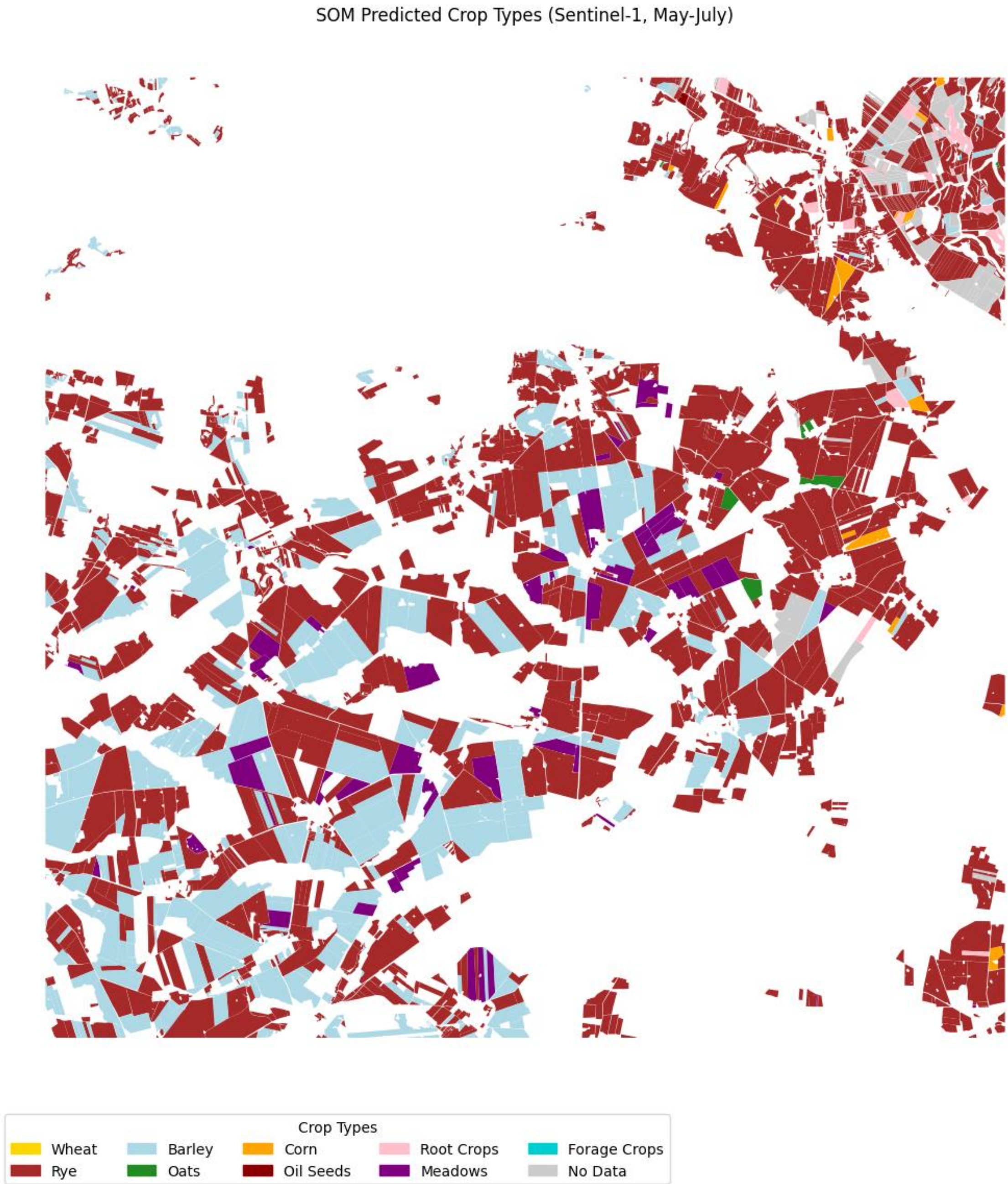

Figure 9: Displays the classification results from the SOM, where each color represents a different crop type predicted by the model.

Ground Truth Map

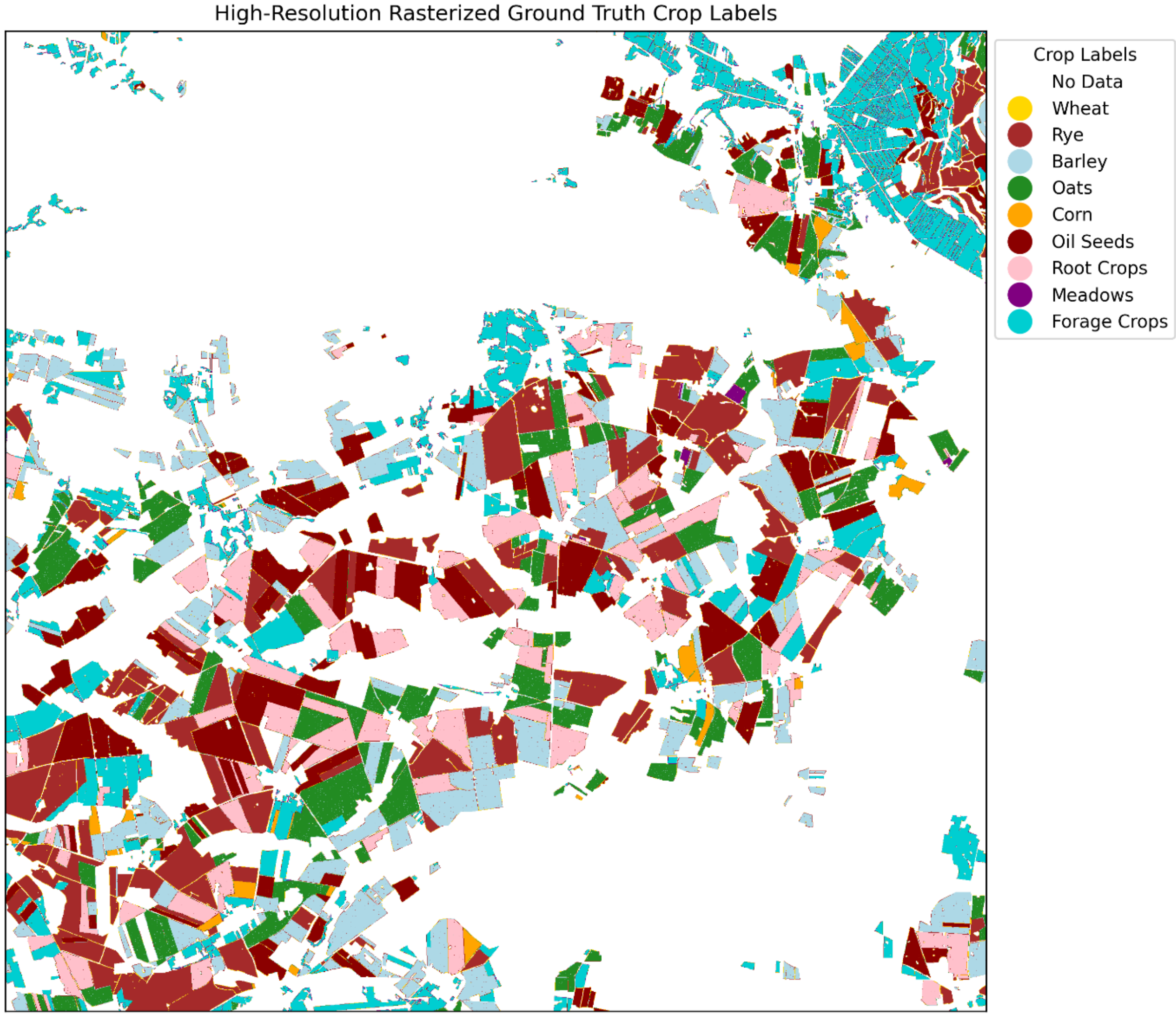

Figure 10: Actual distribution of crop types, as verified by ground truth data

4.2.4 Comparative Analysis

The analysis between the predicted crop Figure 9 and the ground truth Figure 10 data reveals the effectiveness and precision of the model applied to Sentinel-1 data during the May to July temporal window. The comparison, visually represented through the provided maps, allows for an assessment of how well the SOM managed to classify and differentiate between various crop types based on radar data.

The ground truth data shows a rich mosaic of various crops, with significant representation of forage crops (cyan), oats (green), and oil seeds (dark red). In contrast, the SOM prediction heavily favors rye (brown) and barley (light blue), with a noticeable absence of some crop types like corn and oats.

Rye classification in the SOM prediction appears to be overestimated. Large areas predicted as rye in Image 1 correspond to a mix of different crops in the ground truth, including forage crops, oats, and oil seeds. This overclassification aligns with the high recall (79.24%) but low precision (28.13%) for rye in the classification report.

Barley prediction shows interesting patterns. While some areas match well with the ground truth, others seem to be misclassified. The SOM tends to predict barley in larger, more continuous patches compared to the more fragmented distribution in the ground truth. This could explain the moderate precision (57.76%) but very low recall (2.96%) for barley.

The SOM struggled to identify oats, which are clearly visible in the ground truth data but almost entirely absent in the prediction. This aligns with the low recall (17.09%) for oats in the classification report. The model seems to confuse oats with other cereal crops, particularly rye and barley.

Wheat classification shows yellow patches in the SOM prediction corresponding to wheat areas in the ground truth. However, the prediction underestimates wheat coverage compared to the ground truth.

The SOM's performance in identifying meadows (purple) and root crops (pink) appears limited. These crops are sparsely represented in the prediction compared to their more significant presence in the ground truth.

Notably, the SOM prediction shows large areas of no data (gray), particularly in the image's corners and edges. These areas correspond to various crop types in the ground truth, shows difficulties in classification for certain regions or field conditions.

4.2.5 Specific Observations

The model has a substantial capability to identify crop types. The predicted map mirrors the ground truth in some areas, showing that the model is capable of discerning between crop types with significant accuracy. Discrepancies are visible, particularly where the model may confuse crop types with similar spectral signatures or those affected by mixed pixel effects in boundary areas.

The model performs well in regions where a single crop type dominates. Areas with a mix of crop types or transitional zones between different crops pose challenges. Incorporating more diverse training data or increasing the resolution of input data might help in reducing classification errors, particularly in complex agricultural landscapes.

4.3 Results from SOM on Sentinel-2 Data

Self-Organizing Maps (SOM) applied to Sentinel-2 data revealed significant insights into the clustering of crop type classification. This analysis focused on multispectral imagery captured during the specific temporal window from May to July. This focused period captures critical growth stages of various crops, enhancing the model's ability to utilize the rich, high-dimensional data provided by Sentinel-2's multispectral imagery.

4.3.1 Model Performance

The model's configuration was designed to leverage the detailed spectral data from Sentinel-2 which includes visible and near-infrared wavelengths. This spectral range is particularly useful for capturing crops at different growth stages. A detailed review of the model's output with ground truth data indicated some areas where performance could be improved. The classification accuracy, while generally good, showed discrepancies in certain crop types where the model's predictions did not fully align with the ground truth labels. The visual analysis supported a detailed examination of the model's cluster, providing insights into both its strengths and the areas where the classification was less precise.

4.3.2 Accuracy and Validation

The Self-Organizing Map (SOM) applied to Sentinel-2 data demonstrated a high accuracy rate of 99.21% in identifying and classifying different crop types. For a thorough evaluation, the model's performance was measured using precision, recall, and F1-score for each type of crop. These metrics help understand how precisely and reliably the model can identify each crop:

Crop ID	Crop Type	Precision	Recall	F1-Score	Support
1	Wheat	100.00%	98.68%	99.34%	152
2	Rye	100.00%	99.49%	99.74%	196
3	Barley	97.74%	98.48%	98.11%	132
4	Oats	96.77%	100.00%	98.36%	30
5	Corn	99.02%	100.00%	99.51%	101
6	Oil Seeds	99.01%	100.00%	99.50%	100
7	Root Crops	91.67%	100.00%	95.65%	11
8	Meadows	100.00%	98.98%	99.49%	196
9	Forage Crops	99.10%	99.10%	99.49%	196

Table 5: Classification Accuracy of SOM for Different Crop Types Using Sentinel-2 Data

4.3.3 Visual Analysis

Predicted Crop Types Map

Figure 11: SOM Predicted Crop Types Visualization with Data Exclusions

Ground Truth Map

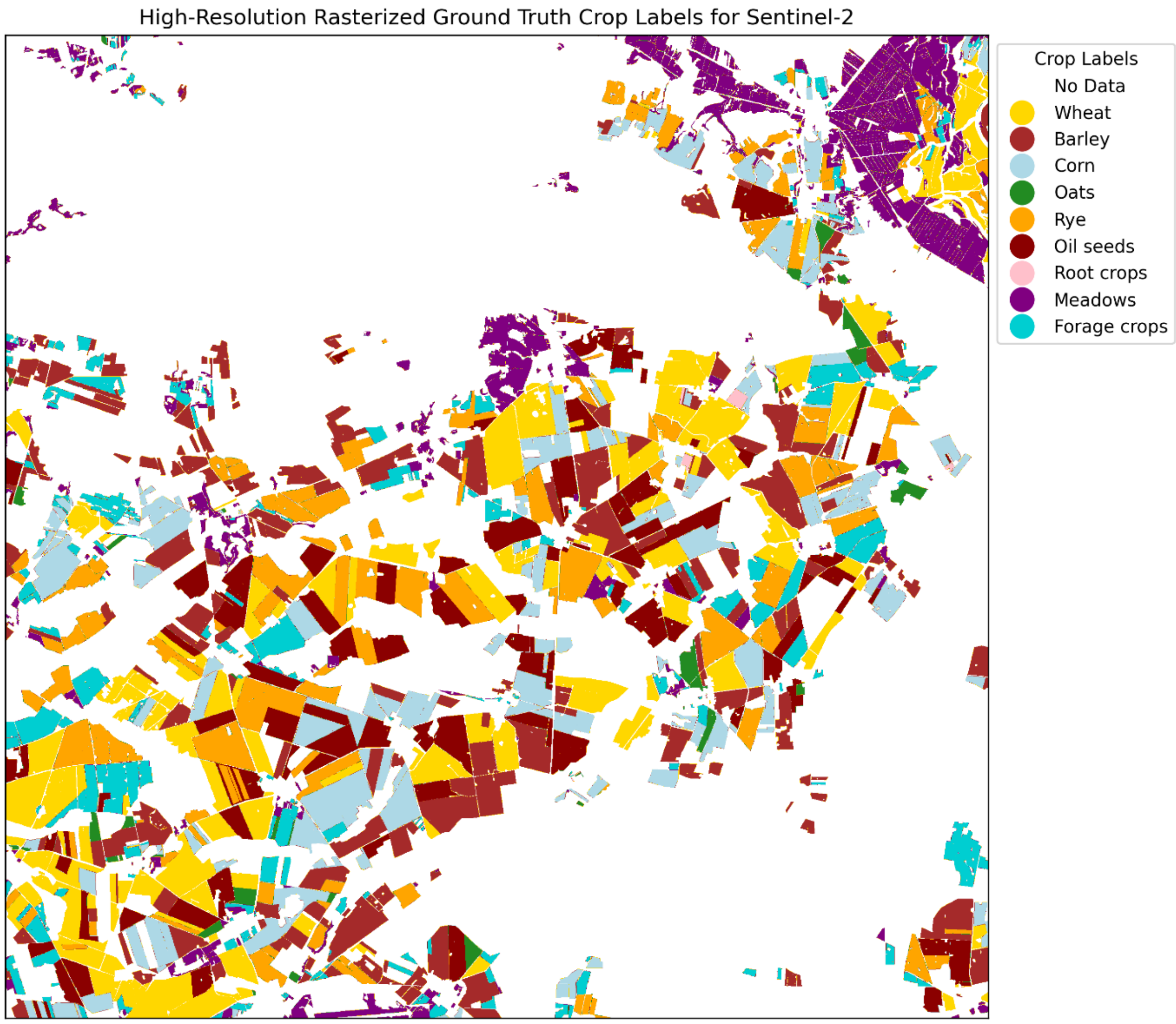

Figure 12: Sentinal 2 Ground Truth data

4.3.4 Comparative Analysis

In the analysis of Sentinel-2 data using the Self-Organizing Map (SOM), it's essential to note a critical aspect regarding the comparison with ground truth data. The SOM model was specifically executed on Sentinel-2 data collected within a temporal window from May to July Figure 11. However, the ground truth data Figure 12 used for comparison covered the entire year and not just the specified temporal window. This discrepancy led to a lack of direct correspondence between the model's predictions and the available ground truth data.

The ground truth dataset contains 2,064 geometries, while the model predictions cover only 1,141 data points. This difference of 923 data points, or about 45% of the total area, accounts for the significant "No Data" regions in the prediction map. This discrepancy in data coverage explains why certain crops or stages might appear 'missing' or underrepresented in the SOM's output.

Large portion of the map is dominated by "No Data" areas, represented in gray. These areas likely represent fields or regions where crop data was not available during the specific temporal window used for the Sentinel-2 analysis. This could be due to various factors such as cloud cover during satellite passes, data quality issues, or simply because certain crops were not present or identifiable during that particular time frame.

Despite the lack of direct comparability between the current predictions and the ground truth, the high accuracy within the analyzed subset remains encouraging. It suggests that the SOM model, when applied to temporally aligned data, has potential for crop classification.

4.4 Fused Sentinel-1 and Sentinel-2 data

Analyzing crop types using fused data from Sentinel-1 and Sentinel-2 through Self-Organizing Maps (SOM) harnesses the unique strengths of both satellite systems to create a dynamic and precise classification model. Sentinel-1, primarily known for its radar capabilities, offers valuable data regardless of weather conditions. Sentinel-2, on the other hand, contributes high-resolution optical imagery that captures a wide spectrum of visible to near-infrared light.

The fusion of these two datasets was engineered to create a richer, more detailed set of inputs for the SOM. This process involved interpolating Sentinel-2 data to align temporally with Sentinel-1, followed by feature stacking to integrate the diverse spectral and radar data into a single, comprehensive dataset. This fusion aims to leverage the radar's ability to penetrate cloud cover and the optical sensor's detailed spectral resolution, providing dataset that enhances the SOM's ability to discern subtle differences between crop types.

4.4.1 Model Performance on Fused Data

The accuracy rate of 80.23% Table 6 indicates that the model correctly identified and classified a high percentage of the crop types based on the fused data. The model showed high precision in most categories, particularly for crop types represented by Classes 2, 5, 6, and 8, with precision scores above 90%. Recall varied more significantly across classes, with some like Class 9 (91.17%)

showing high recall. However, some classes like Class 4 showed lower recall, suggesting the model missed identifying a significant number of instances of this crop type. The F1-score, which balances precision and recall, was generally high, with standout scores for Class 3 (82.90%) and Class 8 (90.45%).

Crop ID	Crop Type	Precision	Recall	F1-Score	Support
1	Wheat	89.09%	69.34%	77.98%	212
2	Barley	91.33%	69.11%	78.68%	259
3	Oats	87.73%	78.57%	82.90%	182
4	Corn	81.48%	55.00%	65.67%	40
5	Rye	91.41%	75.97%	82.98%	154
6	Oil Seeds	93.51%	66.06%	77.42%	109
7	Root Crops	70.59%	80.00%	75.00%	15
8	Meadows	94.43%	86.79%	90.45%	742
9	Forage Crops	52.55%	91.17%	66.67%	351
Overall Accuracy					80.23%

Table 6: Classification Results Table for SOM on Fused Sentinel-1 and Sentinel-2 Data

Class ID	Label Name	Distribution
1	Wheat	7.99%
2	Barley	9.50%
3	Oats	7.90%
4	Corn	1.31%
5	Rye	6.20%
6	Oil Seeds	3.73%
7	Root Crops	0.82%
8	Meadows	33.04%
9	Forage Crops	29.51%

Table 7: Prediction Distribution Table for SOM on Fused Sentinel-1 and Sentinel-2 Data

4.4.2 Visual Analysis

Figure 13: Distribution of Predicted Crop Types from Fused Sentinel-1 and Sentinel-2 Data Using Self-Organizing Maps (May to July Temporal Window)

4.4.3 Comparative Analysis

One of the main difficulties in analysing the fused SOM model is that there isn't a set of ground truth data made just for this mixed dataset. Comparing the predictions of the Self-Organizing Map (SOM) with ground truth data is essential to evaluate both the strengths and the potential areas for improvement in the model. It allows for an objective assessment of how well the model can identify and classify different crop types by comparing model predictions against known, verified outcomes. Due to the lack of specifically tailored ground truth data for the fused dataset, the available ground truth data from Sentinel-1 and Sentinel-2 is utilized.

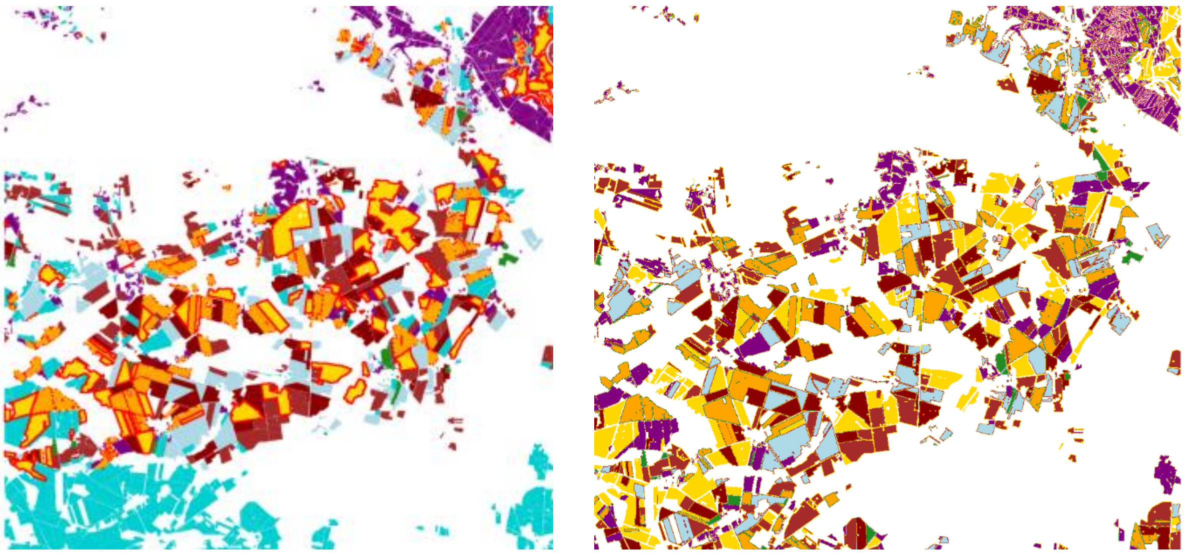

Figure 14: Comparative Visualization of Fused SOM Predicted and Sentinal-2 Ground Truth Crop Types for Fused Sentinel Data (May-July)

The visual comparisons Figure 14 reveal good alignment in many areas, indicating that the SOM predictions generally match well with the high-resolution ground truth for many crop types. Certain crop types such as wheat and barley show high levels of accuracy and alignment between the SOM predictions and the ground truth. However, there are noticeable discrepancies in some areas where the predicted crop types do not match. These mismatches are particularly evident in regions with complex agricultural patterns or where crop types have similar spectral signatures, leading to potential misclassifications.

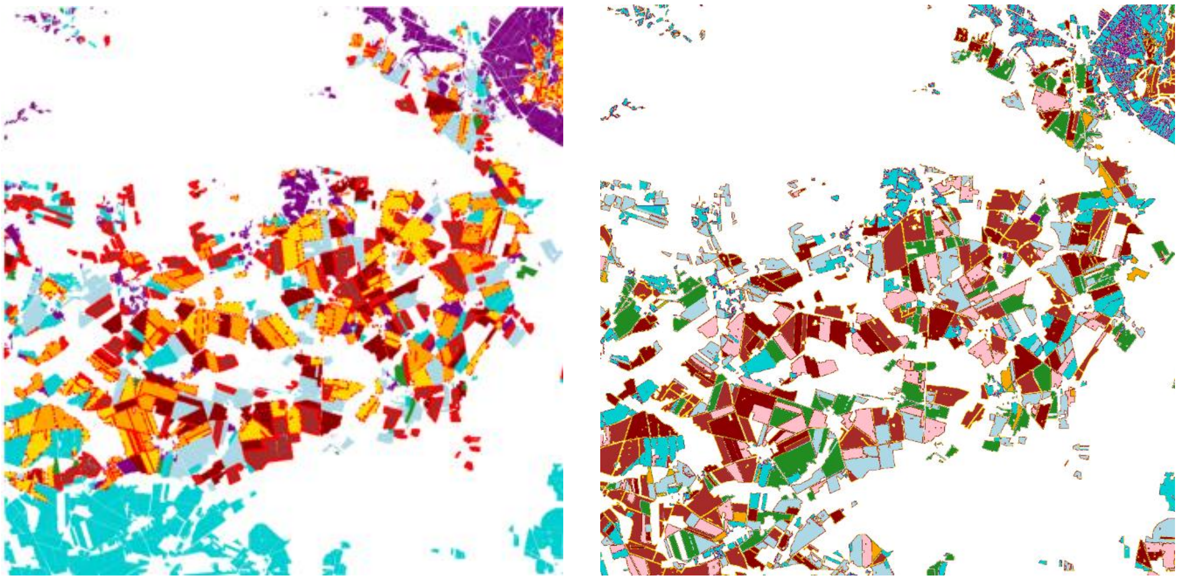

Figure 15: Comprehensive Visualization of SOM Predicted Crop Types Against Sentinel-1 Ground Truth Data (May – July)

This analysis Figure 15 specifically focuses on the May to July temporal window, aligning both the SOM predictions and the Sentinel-1 ground truth data within the same period. The visual mappings show significant alignment in some areas, there are noticeable areas where discrepancies appear. Not only wheat and barley but also rye shows a good degree of matching between the predicted crop types and ground data. Expanding the temporal window of analysis or increasing the frequency of data acquisition might capture more detailed growth dynamics, potentially improving classification accuracy.

5 SOM Data Set Analysis

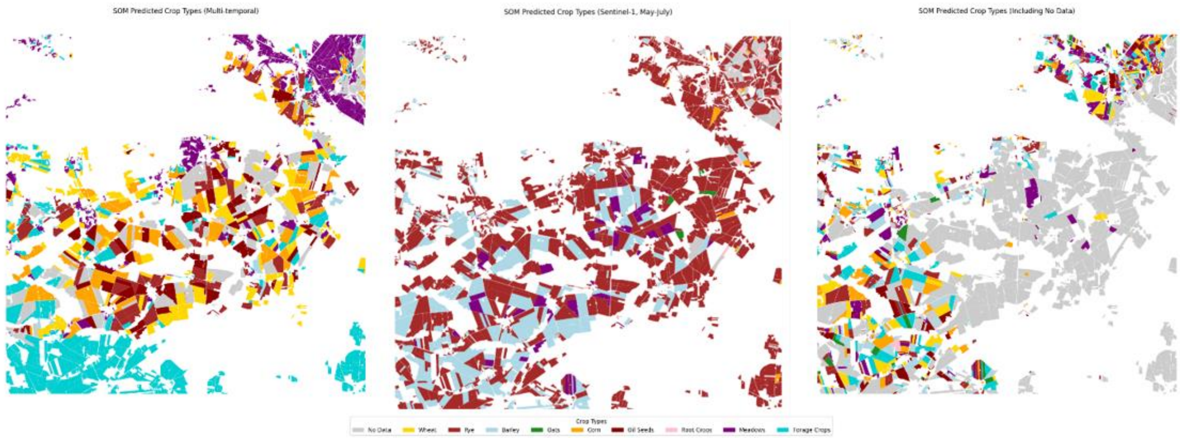

Figure 16: Comparative Visualization of SOM Predicted Crop Types Across Fused, Sentinel-1, and Sentinel-2 Data

Analysis of Self-Organizing Map (SOM) predictions, focusing on the comparison of results from fused Sentinel-1 and Sentinel-2 data with individual Sentinel datasets. This multidimensional comparison provides insights into the effectiveness and nuances of crop type classification using different satellite data configurations within the May to July temporal window.

The SOM predictions from the fused data often show more refined classifications compared to individual datasets. This is likely due to the enriched feature set available from combining radar and optical images, which provides a more comprehensive view of crop types under various conditions.

Sentinel-1 data, with its focus on radar imagery, provides consistent year-round observations but may lack the detailed spectral information necessary for distinguishing closely related crop types or young crops with less developed features. Sentinel-2-based classifications are particularly effective in well-defined agricultural zones with clear crop segregation.

Wheat fields (yellow) show interesting variations across the three predictions. In the multi-temporal analysis, wheat appears more widespread, forming larger contiguous areas in the central and southern parts of the map. This differs notably from the Sentinel-1 and Sentinel-2 predictions, where wheat is less prominent. The multi-temporal model's success in identifying wheat likely stems.

The Sentinel-1 prediction (center image) presents a starkly different picture, dominated by rye (dark red) classifications. This overwhelming prevalence of rye in the SAR-based analysis suggests that the backscatter characteristics of rye are particularly distinct in Sentinel-1 data. The Sentinel-1 prediction shows very little diversity in crop types compared to the other two analyses. The absence of certain crops, such as meadows and forage crops, which are prominent in the multi-temporal prediction, highlights the limitations of using SAR data.

The Sentinel-2 prediction (right image) offers a middle ground between the other two approaches. It shows more diversity in crop types compared to the Sentinel-1 analysis but less than the multi-temporal prediction. The significant presence of "No Data" areas (gray) in the Sentinel-2 prediction, particularly in the eastern and southern portions, illustrates a common challenge with optical data – gaps due to cloud cover or other data quality issues.

The distribution of barley (light blue) and oats (green) is notably different across the three predictions. These crops are more visible in the multi-temporal and Sentinel-2 analyses but almost absent in the Sentinel-1 prediction. This discrepancy underscores the challenge of distinguishing cereal crops using SAR data alone.

Oil seeds (dark brown) and root crops (pink) show limited presence across all three predictions, with slightly better representation in the multi-temporal analysis. This consistent underrepresentation suggests that these crops may be particularly challenging to classify accurately, regardless of the data source or temporal resolution.

The multi-temporal approach appears to produce smoother, more cohesive field boundaries, while the Sentinel-1 and Sentinel-2 predictions show more fragmented field edges. This difference suggests that the multi-temporal analysis may be better at handling mixed pixels at field boundaries.

This comparative analysis underscores the complementary nature of different remote sensing data sources and temporal resolutions in crop classification. The multi-temporal approach clearly offers advantages in terms of crop diversity representation and spatial coverage. However, the unique strengths of Sentinel-1 in identifying certain crop types (like rye) and Sentinel-2 in providing detailed spectral information suggest that an integrated approach, combining SAR and optical data across multiple time points, could yield comprehensive and accurate crop type maps.

6 DISCUSSION AND CONCLUSION

6.1 Conclusion

This study demonstrates the vast potential of remote sensing technology in agricultural applications, particularly in crop classification. The analysis of Sentinel-1 and Sentinel-2 data, both individually and in combination, reveals the richness of information these platforms provide. Each iteration of the research process, from single-sensor to multi-temporal fused approaches, has yielded valuable insights and opened new avenues for exploration.

This research not only advances our understanding of crop classification methodologies but also illuminates the challenges inherent in agricultural remote sensing. The variability in performance across different crop types and the impact of factors such as cloud cover and field boundary delineation emphasize the complexity of the task at hand.

Moreover, this study serves as a springboard for future research directions. The insights gained pave the way for more sophisticated data fusion techniques, advanced machine learning algorithms tailored to agricultural applications, and the integration of additional data sources to enhance classification accuracy.

The integration of multi-sensor and multi-temporal data, coupled with advanced machine learning techniques, holds great promise for enhancing the accuracy and reliability of crop mapping. These advancements have far-reaching implications. As methodologies evolve and new technologies emerge, the field of agricultural remote sensing is poised for further innovations, promising more accurate, timely, and comprehensive crop assessments.

6.2 Failures And Further Improvements

Data Definition and Understanding: Insufficient initial comprehension of the dataset's temporal and spatial characteristics led to a trial-and-error approach in the early stages. This underscores the crucial need for thorough data exploration and metadata analysis prior to model development.

Unsupervised Learning Complexities: The application of SOM, an unsupervised technique, presented significant challenges in the post-classification phase. The task of accurately labeling and interpreting SOM-generated clusters proved to be complex and time-consuming. This process required extensive domain knowledge and data engineering to align cluster outputs with actual crop types. Future work should focus on developing more robust methods for cluster interpretation and labeling, possibly integrating semi-supervised approaches or advanced feature engineering techniques to enhance the interpretability of unsupervised classifications.

Crop Type Misclassification: Persistent issues in distinguishing between similar crop types, especially cereals, indicate the need for more sophisticated spectral and temporal feature extraction methods.

7 RISK ANALYSIS

7.1 Technical Challenges

I encountered significant challenges in data alignment and temporal range documentation, highlighting the critical importance of these aspects in satellite data analysis and ground truth collection.

7.2 Lack of Domain knowledge

The journey of this research project began with curiosity and inexperience in machine learning, evolving into a significant learning experience. Throughout the process, numerous challenges were encountered, requiring extensive reading, experimentation, and perseverance. While the final work may not claim 100% accuracy, the process itself proved invaluable in building competence and insight into the complexities of remote sensing.

7.3 Time Constraints

Executing a project of this nature demands considerable time, patience, and access to expert review channels to ensure quality and relevance. While higher-level colleague reviews were available, time constraints limited the ability to further clarify and refine the work. I have plans to improve the work by creating a standard approach for handling unsupervised data. This will make future projects easier and more consistent. The goal is to have a simple, step-by-step process provides a robust framework for future research in agricultural remote sensing.

8 BIBLIOGRAPHY

Almohsen, Abdulmohsen S. 2024. “Challenges Facing the Use of Remote Sensing Technologies in the Construction Industry: A Review.” *Buildings* 14(9):2861. doi: 10.3390/buildings14092861.

Anon. n.d.-a. “[2103.05543] Self-Supervised SAR-Optical Data Fusion and Land-Cover Mapping Using Sentinel-1/-2 Images.” Retrieved September 19, 2024 (https://ar5iv.labs.arxiv.org/html/2103.05543).

Anon. n.d.-b. “Enhancing Vegetation Indices from Sentinel-2 Using Multispectral UAV Data, Google Earth Engine and Machine Learning | SpringerLink.” Retrieved September 19, 2024 (https://link.springer.com/chapter/10.1007/978-981-19-2519-1_24).

Anon. n.d.-c. “Europe - Crop Calendars.” Retrieved September 30, 2024 (https://ipad.fas.usda.gov/rssiws/al/crop_calendar/europe.aspx).

Anon. n.d.-d. “Satellite IoT Is the Key to Success in Precision Agriculture | AgriTechTomorrow.” Retrieved September 19, 2024 (//agritechtomorrow.com/article/2022/05/satellite-iot-is-the-key-to-success-in-precision-agriculture/13783).

Asner, Gregory. 2001. “Cloud Cover in Landsat Observations of the Brazilian Amazon.” *Int. j. Remote Sensing* 22:3855–62. doi: 10.1080/01431160010006926.

Chitwong, S., S. Witthayapradit, S. Intajag, and F. Cheevasuvit. n.d. “MULTISPECTRAL IMAGE CLASSIFICATION USING BACK-PROPAGATION NEURAL NETWORK IN PCA DOMAIN.”

De Fioravante, Paolo, Tania Luti, Alice Cavalli, Chiara Giuliani, Pasquale Dichicco, Marco Marchetti, Gherardo Chirici, Luca Congedo, and Michele Munafò. 2021. “Multispectral Sentinel-2 and SAR Sentinel-1 Integration for Automatic Land Cover Classification.” *Land* 10(6):611. doi: 10.3390/land10060611.

Declaro, Alexis, and Shinjiro Kanae. 2024. “Enhancing Surface Water Monitoring through Multi-Satellite Data-Fusion of Landsat-8/9, Sentinel-2, and Sentinel-1 SAR.” *Remote Sensing* 16(17):3329. doi: 10.3390/rs16173329.

Eisfelder, Christina, Bruno Boemke, Ursula Gessner, Patrick Sogno, Genanaw Alemu, Rahel Hailu, Christian Mesmer, and Juliane Huth. 2024. “Cropland and Crop Type Classification with Sentinel-1 and Sentinel-2 Time Series Using Google Earth Engine for Agricultural Monitoring in Ethiopia.” *Remote Sensing* 16(5):866. doi: 10.3390/rs16050866.

Felegari, Shilan, Alireza Sharifi, Kamran Moravej, Muhammad Amin, Ahmad Golchin, Anselme Muzirafuti, Aqil Tariq, and Na Zhao. 2021. “Integration of Sentinel 1 and Sentinel 2 Satellite Images for Crop Mapping.” *Applied Sciences* 11(21):10104. doi: 10.3390/app112110104.

Frantz, David, Erik Haß, Andreas Uhl, Johannes Stoffels, and Joachim Hill. 2018. “Improvement of the Fmask Algorithm for Sentinel-2 Images: Separating Clouds from Bright Surfaces Based on Parallax Effects.” *Remote Sensing of Environment* 215:471–81. doi: 10.1016/j.rse.2018.04.046.

Kaski, Samuel. 2017. “Self-Organizing Maps.” Pp. 1129–32 in *Encyclopedia of Machine Learning and Data Mining*, edited by C. Sammut and G. I. Webb. Boston, MA: Springer US.

Kohonen, Teuvo. 2001. *Self-Organizing Maps*. Vol. 30. Berlin, Heidelberg: Springer.

Lee, Jong-Sen. 1981. “Refined Filtering of Image Noise Using Local Statistics.” *Computer Graphics and Image Processing* 15(4):380–89. doi: 10.1016/S0146-664X(81)80018-4.

Ma, Chunfeng, Xin Li, and Matthew F. McCabe. 2020. “Retrieval of High-Resolution Soil Moisture through Combination of Sentinel-1 and Sentinel-2 Data.” *Remote Sensing* 12(14):2303. doi: 10.3390/rs12142303.

Mulla, David J. 2021. “Satellite Remote Sensing for Precision Agriculture.” Pp. 19–57 in *Sensing Approaches for Precision Agriculture*, edited by R. Kerry and A. Escolà. Cham: Springer International Publishing.

Quan, Yujun, Rongrong Zhang, Jian Li, Song Ji, Hengliang Guo, and Anzhu Yu. 2024. “Learning SAR-Optical Cross Modal Features for Land Cover Classification.” *Remote Sensing* 16(2):431. doi: 10.3390/rs16020431.

Radočaj, Dorijan, Ivan Plaščak, and Mladen Jurišić. 2023. “Global Navigation Satellite Systems as State-of-the-Art Solutions in Precision Agriculture: A Review of Studies Indexed in the Web of Science.” *Agriculture* 13(7):1417. doi: 10.3390/agriculture13071417.

Revill, Andrew, Anna Florence, Alasdair MacArthur, Stephen P. Hoad, Robert M. Rees, and Mathew Williams. 2019. “The Value of Sentinel-2 Spectral Bands for the Assessment of Winter Wheat Growth and Development.” *Remote Sensing* 11(17):2050. doi: 10.3390/rs11172050.

Ruß, Georg, Rudolf Kruse, Martin Schneider, and Peter Wagner. 2009. “Visualization of Agriculture Data Using Self-Organizing Maps.” Pp. 47–60 in *Applications and Innovations in Intelligent Systems XVI*, edited by T. Allen, R. Ellis, and M. Petridis. London: Springer London.

Saini, R., and S. K. Ghosh. 2018. “CROP CLASSIFICATION ON SINGLE DATE SENTINEL-2 IMAGERY USING RANDOM FOREST AND SUPPOR VECTOR MACHINE.” *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLII–5:683–88. doi: 10.5194/isprs-archives-XLII-5-683-2018.

Segarra, Joel, Maria Luisa Buchaillot, Jose Luis Araus, and Shawn C. Kefauver. 2020. “Remote Sensing for Precision Agriculture: Sentinel-2 Improved Features and Applications.” *Agronomy* 10(5):641. doi: 10.3390/agronomy10050641.

Soussi, Abdellatif, Enrico Zero, Roberto Sacile, Daniele Trinchero, and Marco Fossa. 2024. “Smart Sensors and Smart Data for Precision Agriculture: A Review.” *Sensors* 24(8):2647. doi: 10.3390/s24082647.

Taylor, Gage. 2024. “First-of-Its-Kind Satellite Data Fusion Project Sets Stage for Future Surface Topography and Vegetation Missions.” *NASA Earth Science and Technology Office*. Retrieved September 19, 2024 (https://esto.nasa.gov/first-of-its-kind-satellite-data-fusion-project-sets-stage-for-future-surface-topography-and-vegetation-missions/).

Van Tricht, Kristof, Anne Gobin, Sven Gilliams, and Isabelle Piccard. 2018. "Synergistic Use of Radar Sentinel-1 and Optical Sentinel-2 Imagery for Crop Mapping: A Case Study for Belgium." doi: 10.20944/preprints201808.0066.v1.

Wan, Weijian, and Donald Fraser. 2000. "A Multiple Self-Organizing Map Scheme for Remote Sensing Classification." Pp. 300–309 in *Multiple Classifier Systems*. Vol. 1857, *Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Wang, Zirui. 2024. "Extraction and Fusion of Geographic Information from Multi-Source Remote Sensing Images Based on Artificial Intelligence." Pp. 17–27 in *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, edited by R. Kountchev, S. Patnaik, K. Nakamatsu, and R. Kountcheva. Singapore: Springer Nature.

Yuan, Hui, Cynthia F. Van Der Wiele, and Siamak Khorram. 2009. "An Automated Artificial Neural Network System for Land Use/Land Cover Classification from Landsat TM Imagery." *Remote Sensing* 1(3):243–65. doi: 10.3390/rs1030243.

Zheng, Baojuan, Soe W. Myint, Prasad S. Thenkabail, and Rimjhim M. Aggarwal. 2015. "A Support Vector Machine to Identify Irrigated Crop Types Using Time-Series Landsat NDVI Data." *International Journal of Applied Earth Observation and Geoinformation* 34:103–12. doi: 10.1016/j.jag.2014.07.002.