1 2  9 0

UNIVERSIDADE D
COIMBRA

Adalberto Manuel Yambi Francisco

Integration of Optical and Radar Image Time Series with Supervised Learning and Geospatial Intelligence for Early Mapping of Rice and Maize Crops in the Baixo Mondego Region.

Dissertation in the context of the Master in Geospatial Information Engineering advised by Professor Gil Rito Gonçalves and presented to the Department of Mathematics of the Faculty of Sciences and Technology of University of Coimbra.

June 2024

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE D
COIMBRA

Adalberto Manuel Yambi Francisco

Integration of Optical and Radar Image Time Series with Supervised Learning and Geospatial Intelligence for Early Mapping of Rice and Maize Crops in the Baixo Mondego Region.

Dissertation in the context of the Master in Geospatial Information Engineering advised by Professor Gil Rito Gonçalves and presented to the Department of Mathematics of the Faculty of Sciences and Technology of Coimbra

June 2024

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE D
COIMBRA

Adalberto Manuel Yambi Francisco

Integração de Séries Temporais de Imagens Ópticas e Radar com Aprendizagem Supervisionada e Inteligência Geoespacial no Mapeamento Precoce de Culturas de Arroz e Milho na Região do Baixo Mondego.

Dissertação no âmbito do Mestrado de Engenharia de Informação Geoespacial orientada pelo Professor Gil Rito Gonçalves e apresentada ao Departamento de Matemática da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

Junho de 2024

Acknowledgements

This work was supported by the Portuguese Foundation for Science and Technology (FCT) and by the European Regional Development Fund (FEDER) through COMPETE 2020, Operational Program for Competitiveness and Internationalization (POCI) under the project grant UIDB 00308/2020 with the 10.54499/UIDB /00308/2020, the research project GreenBotics(PTDC/EEI-ROB/2459/2021) and the research grant CAASatBMondego 1/2024.

From a passion for Remote Sensing to an unrelenting search for the best solutions to scientific and life problems, I want to thank my advisor and professor Gil Rito Gonçalves for his academic and personal support during the course of this master's degree. Besides being certain that I will carry the knowledge and this contagious search for knowledge throughout my life, I am convinced that without his valuable help and trust, this journey would have been considerably more arduous.

I express my deepest gratitude to the professors of the Mathematics and Computer Engineering Departments at the University of Coimbra, who stood out for their patience, excellence, and humility in imparting their vast knowledge. Special thanks are due to Professor Cidália Maria Parreira da Costa Fonte and Professor José Paulo Elvas Duarte de Almeida, for their invaluable availability and the constant attention they always dedicated to me.

I could not fail to thank Ms. Drwisska Palone da Cruz Panzo for being my main emotional support during the final and most difficult phase of this Master's degree, giving all her time, even the time she didn't have, with the rarest loyalty one can experience.

To Jorge Edgar Marques Simões and Neidynilson Kennedy Kassindy, I am grateful for the companionship and support during this journey, which often seemed almost impossible. However, even the longest path becomes bearable when traveled with good and true friends.

To my parents, for life, constant encouragement, and investment in my education.

Finally, I wish to express deep and special gratitude to my beloved wife, Djamila Cassongo Gomes da Silva Francisco, whose relentless and unconditional support never waned, despite the distance, longing, and challenges that life placed in our path.

Thank you very much to everyone and for everything!

Abstract

The availability of information on the spatial distribution of agricultural crops is a critical factor in supporting the decision-making process. However, providing this information in the post-agricultural season compromises its utility in addressing technical issues related to agricultural production itself, as well as in administrative and policy decisions whose timing occurs before the crop harvest.

Despite this need, studies indicate that while the distinction between agricultural crops is most effective at the end of the season, rapid and automated early mapping techniques using satellite imagery and advanced methods like machine learning are enabling high classification accuracies and can significantly improve agricultural planning and optimization of practices, policies and resource allocation.

The objective of this study was to automate the early mapping of maize and rice agricultural plots using satellite imagery, with the utilization of time series of polarimetric and multispectral data, and the combination of geospatial intelligence and supervised learning in a cloud computing environment.

The study is situated within the context of Portugal, where the mapping of agricultural crops from satellite images has been in the experimental phase for less than five years. The study was applied to the agricultural plots of the Baixo-Mondego valley, which is located in the central region of Portugal.

To establish a robust and effective methodology for crop mapping from satellite data, a thorough review of 70 relevant articles was conducted. This review revealed that, in recent years, the most effective approach to agricultural mapping from satellite images involves the fusion of time series data from different sensors and the use of supervised computational learning.

The applied methodology consisted of the incremental classification of time series of satellite images throughout the summer agricultural season, using SVM and RF algorithms. The methodology combines geospatial intelligence and supervised learning in a approach that fuses SAR and multispectral data, and another in parallel that uses only multispectral data. This methodology permits the monitoring of the performance evolution of classification models (in accordance with the metrics accuracy and F1 score), the discriminative power of features derived from satellite images, and the contribution of the fusion of optical and radar images in the context of early agricultural mapping.

The findings indicated that the methodology employing fused data can enhance the efficacy of classification models by up to 6.26%. Moreover, the magnitude of classification improvement was observed to increase with the proximity to spring. This is supported by the paucity of high-quality optical images and the incorporation of radar data with geometric information about crops at their early stages.

The classification line with models trained using the RF algorithm proved to be more efficient in minimizing the time required to achieve acceptable performance

(equal to or greater than 90%), as evidenced by an accuracy of 91.45% on June 20th and an F1-Score of 90.88% on June 30th. In contrast, the classification line with models trained using the SVM algorithm only reached an accuracy of 92.33% on July 10th and an F1-Score of 91.86% on July 30th, resulting in a delay of approximately 20-30 days in achieving acceptable performance.

The proposed approach proved to be relevant, as the precise and timely classification of agricultural crops not only allows for the optimization of agricultural production processes and resources but also facilitates inclusive technological advancement. This advancement maximizes the economy and minimizes the environmental impact, thereby contributing to food security. Future research should continue to refine the use of data obtained through space technology and processed through cloud computing environments, with a view to improving inclusive support for precision agriculture.

Keywords

Early Crop Mapping, Remote Sensing, Supervised Learning, Time Series, Satellite data Fusion, Incremental Classification, Maize, Rice.

Resumo

A disponibilização de informação sobre a distribuição espacial das culturas agrícolas é um fator crítico no apoio ao processo de tomada de decisão. A disponibilização dessa informação no período pós-temporada agrícola compromete a sua utilização em problemas técnicos relacionados com a própria produção agrícola, bem como em decisões administrativas e políticas cujo momento de decisão ocorre antes da colheita das culturas.

Apesar desta lacuna, alguns estudos indicam que, embora a distinção entre culturas agrícolas seja mais eficaz no final da temporada, técnicas rápidas e automatizadas de mapeamento precoce, utilizando imagens de satélite e e métodos de classificação avançados, como a aprendizagem automática, permitido obter classificações com exatidões elevadas e podem, consequentemente, melhorar substancialmente o planeamento agrícola e a otimização de práticas, políticas e alocação de recursos.

O objetivo deste estudo foi automatizar o mapeamento precoce das parcelas agrícolas de milho e arroz a partir de imagens de satélite, utilizando séries temporais de dados polarimétricos e multiespectrais, e combinando inteligência geoespacial e aprendizagem supervisionada em ambiente de computação na nuvem.

O estudo insere-se no contexto de Portugal, onde o mapeamento de culturas agrícolas a partir de imagens de satélite está em fase experimental há menos de 5 anos, e foi aplicado às parcelas agrícolas do vale do Baixo-Mondego, localizado na região central de Portugal.

Para estabelecer uma metodologia robusta e eficaz para o mapeamento de culturas a partir de dados de satélite, foi realizada uma revisão aprofundada de 70 artigos relevantes. Esta revisão revelou que, nos últimos anos, a abordagem mais eficaz para o mapeamento agrícola a partir de imagens de satélite envolve a fusão de dados de séries temporais de diferentes sensores e o uso de aprendizagem computacional supervisionada.

A metodologia aplicada consistiu na classificação incremental de séries temporais de de imagens de satélite ao longo da temporada agrícola de verão, utilizando algoritmos SVM e RF. A metodologia combina inteligência geoespacial e aprendizagem supervisionada numa abordagem com fusão de dados SAR e multiespectral, e noutra, que utiliza apenas dados multiespectrais. Esta metodologia permite monitorizar a evolução do desempenho dos modelos de classificação (de acordo com as métricas de precisão e F1 score), o poder discriminativo das características derivadas de imagens de satélite e a contribuição da fusão de imagens óticas e de radar no contexto do mapeamento agrícola precoce.

Os resultados obtidos mostraram que a abordagem com dados combinados pode melhorar o desempenho dos modelos de classificação até 6,26% e que o grau de melhoria da classificação aumenta quanto maior for a proximidade da primavera, o que é justificado pela escassez de disponibilidade de imagens ópticas de qualidade e pelo complemento dos dados de radar com informação geométrica sobre as culturas em estado inicial.

A linha de classificação com modelos treinados com o algoritmo RF demonstrou ser mais eficiente ao minimizar o tempo necessário para alcançar um desempenho aceitável (igual ou superior a 90%), registando uma precisão de 91,45% a 20 de junho e uma pontuação F1-Score de 90,88% a 30 de junho. Em contraste, a linha de classificação com modelos treinados com o algoritmo SVM só alcançou uma precisão de 92,33% a 10 de julho, e uma pontuação F1-Score de 91,86% a 30 de julho, atrasando o alcance do desempenho aceitável em 20-30 dias.

A abordagem proposta revelou-se relevante, pois a classificação precisa e atempada das culturas agrícolas não só permite a otimização da gestão dos processos e recursos de produção agrícola, como também promove um avanço tecnológico inclusivo. Este avanço maximiza a economia e minimiza o impacto ambiental, contribuindo assim para a segurança alimentar. Investigações futuras devem continuar a aperfeiçoar a utilização de dados obtidos através de tecnologia espacial e processados por aprendizagem em ambientes computacionais na nuvem, visando melhorar o suporte inclusivo à agricultura de precisão.

Palavras-Chave

Mapeamento Prococe de Culturas Agrícolas, Deteção Remota, Aprendizagem Supervisionada, Séries Temporais, Fusão de Dados de Satélite, Classificação Incremental, Milho, Arroz.

Contents

1 Introduction **1**
- 1.1 Motivation . . . 1
- 1.2 Context and Justification of the Research . . . 3
- 1.3 Objectives . . . 4
- 1.4 Structure of the Dissertation . . . 4

2 Review Outcomes **7**
- 2.1 Agriculture Crop Mapping Using Remote Sensing . . . 7
  - 2.1.1 Overview of Remote Sensing in Agriculture . . . 7
  - 2.1.2 Key Challenges in Crop Mapping . . . 8
- 2.2 Systematic Review of Crop Mapping Approaches . . . 9
- 2.3 Article Selection Strategy . . . 9
- 2.4 Results from the Review . . . 9
  - 2.4.1 Spatial and Temporal Distribution of Selected Articles . . . 9
  - 2.4.2 Satellite Technology . . . 11
  - 2.4.3 Mapping Approaches . . . 13
  - 2.4.4 Satellite-Derived Features . . . 17
  - 2.4.5 Machine Learning Algorithms in Agricultural Crop Mapping 19
  - 2.4.6 Best Practices in Agricultural Crop Mapping . . . 20
  - 2.4.7 Fundamentals of the Support Vector Machine and Random Forest Algorithms . . . 21

3 Study Area and Data **29**
- 3.1 Study Area . . . 29
- 3.2 Data . . . 29
  - 3.2.1 Satellite Data . . . 30
  - 3.2.2 Ground Truth Data . . . 32

4 Methodology **33**
- 4.1 Overview of the Methodological Framework . . . 33
- 4.2 Feature Extraction . . . 35
- 4.3 Building the Reference Database . . . 35
  - 4.3.1 Determination of Sample Size . . . 35
  - 4.3.2 Splitting Data into Training and Test Sets . . . 36
  - 4.3.3 Oversampling of Training Data Partition . . . 36
- 4.4 Feature Selection . . . 38
- 4.5 Model Development . . . 39
  - 4.5.1 Training and Testing Setup . . . 39
  - 4.5.2 Model Training . . . 39

4.5.3 Model Evaluation Metrics . . . 39
4.6 Post-Classification with Stratified Random Sampling . . . 40
4.6.1 Post-Classification Sampling Design . . . 40
4.6.2 Post-Classification Response Design . . . 41
4.7 Implementation: System Architecture . . . 42

5 Results **45**
5.1 Proposed Approach . . . 45
5.1.1 Feature Selection Outcomes . . . 45
5.1.2 Model Performance Over Time . . . 53
5.1.3 Model Tuning . . . 58
5.1.4 Early Crop Mapping Identification . . . 59
5.1.5 Post-Classification Results . . . 61

6 Discussion **65**
6.1 Effectiveness of Selected Features . . . 65
6.2 Contribution of Optical and Radar Data Fusion . . . 66
6.3 Capability to Explore the Fused Data . . . 67
6.4 Comparison with Previous Studies . . . 68
6.4.1 Methodological Perspective . . . 68
6.4.2 Data Fusion Improvements Perspective . . . 69
6.4.3 Algorithm Performance Perspective . . . 69
6.4.4 Enhancements to the Surface Surveillance System in Portugal 69

7 Conclusions and Future Work **71**
7.1 Summary of Key Findings . . . 71
7.2 Implications for Agricultural Monitoring . . . 72
7.2.1 Agricultural Management Optimization . . . 72
7.2.2 Economic and Environmental Impacts . . . 73
7.2.3 Technological Advancements in Agriculture . . . 73
7.3 Recommendations for Future Research . . . 73

References **75**

Appendix A Glossary **81**

Appendix B Flowchart of the Methodological Framework in Landscape Orientation **83**

Appendix C Coding **85**
C.1 Coding in GEE . . . 85
C.1.1 Prepare Samples to Oversampling . . . 85
C.1.2 Classification . . . 87
C.2 Coding in Colab . . . 93
C.2.1 Oversampling . . . 93
C.2.2 Hyperparameters Optimization and Feature Selection . . . 93

Appendix D Reference Data Acquisition **99**

Appendix E Satellite Image Acquisition **101**

Acronyms

CAP Common Agricultural Policy.

DT Decision Tree.

FAO Food and Agriculture Organization.

GEE Google Earth Engine.

GNSS Global Navigation Satellite System.

IFAP Institute of Financing for Agriculture and Fisheries, Public Institute (Portugal).

IFOV Instantaneous Field of View.

k-NN k-Nearest Neighbors.

LIDAR Light Detection and Ranging.

MACAT Annual Map of Temporary Agricultural Crops from portuguese “Mapa Anual de Culturas Agrícolas".

MLE Maximum Likelihood Estimation.

MODIS Moderate Resolution Imaging Spectroradiometer.

NDVI Normalized Difference Vegetation Index.

NDWI Normalized Difference Water Index.

PCA Principal Component Analysis.

PF Parametric Feature.

PI Phenology Index.

RF Random Forest.

RFAI Random Forest Average Impurity.

RFAI Average Precision Reduction.

RFE Recursive Feature Elimination.

RFECV Recursive Feature Elimination with Cross Validation.

RTK Real Time Kinematic.

SAR Synthetic Aperture Radar.

SB Spectral Bands.

SI Spectral Index.

SMOS Land Use Monitoring System from portuguese “Sistema de Monitorização da Ocupação do Solo.

SMOTE Synthetic Minority Over-sampling Technique.

SVM Support Vector Machine.

SVS Surface Surveillance System.

T Texture.

List of Figures

2.1 Literature Review Strategy 10
2.2 Spatial Distribution Map of Selected Articles 11
2.3 Temporal Distribution of Selected Articles 11
2.4 Percentage of Satellite Technology Application in Selected Articles 13
2.5 Mapping Approaches Based on the Data Acquisition Timing 15
2.6 Mapping Approaches Based on the Minimum Data Classification Unit 16
2.7 Mapping Approaches Based on the Data Acquisition Timing 17
2.8 Ranking of the 5 Most Utilized Features in Selected Articles 18
2.9 Efficiency of the 5 Most Utilized Features in Selected Articles 19
2.10 Ranking of the 5 Most Used Algorithms in Selected Articles 20
2.11 Efficiency of the 5 Most Utilized Algorithms in Selected Articles . . 21
2.12 Best Practices for Mapping Agricultural Crops Using Remote Sensing Data with Machine Learning 22
2.13 Generic Illustration of SVM (Lorena & De Carvalho, 2007) 23
2.14 Illustration of a Soft Margin SVM (Lorena & De Carvalho, 2007) . . 24
2.15 Kernel mapping from a 2D space to a 3D space: (a) Nonlinear dataset; (b) Non-linear Boundary in the Transformed Feature Space; (c) Linear Boundary in the Transformed Feature Space (Lorena & De Carvalho, 2007). 25
2.16 Diagram of a Random Forest (Liu et al., 2012) 26

3.1 Geographic Framework of the Region of Interest 30
3.2 Sample Acquisition Scheme for Model Training 32

4.1 Flowchart of the Methodological Framework. The letters 'i', 'a', A, and B represent the number of classifications, accuracy, and accuracy for approach A, end accuracy for approach B , respectively. . . 34
4.2 Sampling Points In Plots 37
4.3 Feature Selection Using the RFECV Algorithm 38
4.4 Classification System Architecture 42
4.5 Data workflow 43

5.1 Univariate Contribution of Radar Features in RF Models Based on Gini Importance 46
5.2 Impact of Radar Features on RF Performance Model (Red Dot Represents a Value Equal to or Below 0) 47

5.3 Ranking of Radar Feature Effectiveness for RF Models Based on the Number of Times Each Band's Features are Selected or Not Selected in the RFECV Process (Red Dot Represents a Value Equal to 0) . . . 47
5.4 Univariate Contribution of Optical Features in RF Models Based on Gini Importance . . . 48
5.5 Impact of Optical Features on RF Performance Models (Red Dot Represents a Value Equal to or Below 0) . . . 48
5.6 Ranking of Optical Feature Effectiveness for RF Models Based on the Number of Times Each Band's Features are Selected or Not Selected in the RFECV Process . . . 49
5.7 Univariate Contribution of Radar Features in SVM Models Based on the Mean Value of the Support Vector Coefficients (Magnitude) . 50
5.8 Impact of Radar Features on SVM Models . . . 50
5.9 Ranking of Radar Feature Effectiveness for SVM Models Based on the Number of Times Each Band's Features are Selected or Not Selected in the RFECV Process . . . 51
5.10 Univariate Contribution of Optical Features On SVM Models Based on the Mean Value of the Support Vector Coefficients (Magnitude) . 51
5.11 Impact of Optical Features on SVM Models Performance (Red Dot Represents a Value Equal to or Below 0) . . . 52
5.12 Ranking of Radar Feature Effectiveness for SVM Models Based on the Number of Times Each Band's Features are Selected or Not Selected in the RFECV Process . . . 52
5.13 Evolution of RF Models' Accuracy Over Time . . . 54
5.14 Evolution of RF Models' F1-Score Over Time . . . 56
5.15 Evolution of SVM Models' Accuracy Over Time . . . 56
5.16 Evolution of SVM Models' F1-Score Over Time . . . 58
5.17 Comparison of Classification Models Where an F1-Score of 90% or Higher is First Achieved in Both Approaches, Considering the RF and SVM Classification Lines . . . 59
5.18 Early Crop Mapping of the Mondego Valley as of July 30th . . . 60
5.19 Phenological Stages of Rice and Maize . . . 60
5.20 Post-Classification Sampling Design . . . 61
5.21 Post-Classification Response Design . . . 62

6.1 Cumulative Availability of Sentinel-1 and Sentinel-2 Images Over Time . . . 66
6.2 Development of the Improvement of the Classification Resulting from Data Fusion in the Two Classification Lines . . . 67

D.1 Reference Data Acquisition Equipment . . . 100

List of Tables

2.1 Kernel Types, Functions and Parameters 25

3.1 Description of Monitored Crops 32

4.1 Imbalanced Sampling (Sample Points Stratified by Each Class) . . . 36
4.2 Balanced Sampling (Sample Points Balanced by Each Class+ Synthetic Points) 38
4.3 Description of Crop Statistics 40

5.1 Accuracy of RF Models 53
5.2 F1-Score of RF Models 55
5.3 Accuracy of SVM Models 57
5.4 F1-Score of SVM Models 57
5.5 Confusion Matrix 62
5.6 Post-Classification Metrics 63
5.7 Class Area Adjusted 63

6.1 Difference Between F1-Score Ratings in Two Approaches (A and B) Mapped According to Classification Lines (RF and SVM) 67
6.2 Time Series Increment Across Various Studies 68
6.3 Enhanced classification through the fusion of optical and radar images 69
6.4 Best Machine Learning Algorithm in Different Studies 70

Chapter 1

Introduction

1.1 Motivation

In light of the rapid population growth and the subsequent need to increase agricultural production, humanity is confronted with the challenge of increasing food production in a sustainable manner. The 2017 Food and Agriculture Organization (FAO) report highlighted that the food production of that year would not be sufficient to meet the needs of over 10 billion people by 2060, particularly in light of the extreme climate variations affecting ecosystems, water resources, and arable lands, which could potentially jeopardize food security.

In response to this issue, the FAO's 2022 report emphasizes the importance of digital automation and its inclusion in precision agriculture as innovative solutions that significantly contribute to the sustainable production and distribution of food. This process involves both technical and political aspects, and among other considerations, it relies on the spatial distribution of agricultural crops.

From a technical standpoint, information about the spatial distribution of agricultural crops allows for the development of sustainable strategies to manage production costs. These strategies include optimizing irrigation or energy, adapting crops to climate changes, and protecting ecosystems. Politically, the data facilitates the implementation of policies related to the market pricing of agricultural products, disaster management, regulation of foreign trade, management and control of subsidies to farmers, and the construction of infrastructure for storing products.

As stated in Rahmati et al. (2022), the spatial distribution of crop types and their respective areas can be collected in three distinct ways:

- Expert Estimates
- Field Surveys
- Use of new technologies such as Remote Sensing Surveys

In the realm of new technologies, computational learning techniques are increasingly utilized in various engineering fields to automate processes that previously required significant time and energy. This is also true of advances in precision agriculture. Consequently, in the domain of remote sensing surveys applied to precision agriculture, the production of spatial distribution maps of agricultural crops from intelligent classification models trained with satellite data has seen significant growth over the past decade. This growth is justified by two factors: the increase in the number of Earth observation satellites and the popularity and advancement of artificial intelligence techniques.

In this context, the scientific community has been confronted with three significant challenges:

1. The first challenge, as previously noted by Santos (2021), pertains to the processing of voluminous data (on the order of terabytes) generated by observation satellites. One solution to this challenge is the utilization of cloud computing technologies.

2. In the context of machine learning algorithms, the second challenge, pertains to the Hughes effect, also known as the "curse of dimensionality." This phenomenon indicates that an increase in the number of features does not necessarily lead to an increase in classification accuracy. This issue has led to the common practice of feature engineering, which involves reducing or selecting the most relevant variables for classifying crops (Yi et al., 2022).

3. The third challenge pertains to the temporality of the production and dissemination of spatial information about crops. The timing of this production, occurring post-agricultural season, can impede its application in technical issues related to agricultural production itself, as well as in administrative and political decision-making processes whose timing precedes the crop harvest moment.

In this context, Google Earth Engine's cloud computing technology was employed to automate the preliminary mapping of maize and rice plots in the Baixo-Mondego region. This was achieved through the use of intelligent models trained on a time series of polarimetric and multispectral data from Sentinel-1 and Sentinel-2 satellites, respectively.

The development of these intelligent models entails the integration of supervised learning and geospatial intelligence. Geospatial intelligence is a field of intelligence that employs the analysis and interpretation of events and patterns in georeferenced data to inform decision-making processes. This multidisciplinary field encompasses techniques from data science, geographic information systems, and remote sensing. On the other hand, supervised learning is a machine learning technique that develops algorithms enabling computers to identify patterns in data and map inputs (data) to feature space and their respective known outputs (classes).

1.2 Context and Justification of the Research

With regard to the spatial distribution of agricultural crops in the study area (described in chapter 3), as well as in Portugal in general, the (MACAT) Annual Map of Temporary Agricultural Crops from portuguese “Mapa Anual de Culturas Agrícolas" (MACAT) produced by the Directorate-General for Territory is available for consultation. This map can be accessed through Land Use Monitoring System from portuguese “Sistema de Monitorização da Ocupação do Solo (SMOS).

The MACAT represents an innovative cartographic product developed for periodic release, occurring annually (following the agricultural season) and comprising 30 distinctive classes delineated by the type of agricultural crop. This novel approach combines advanced spatial technologies and machine learning algorithms to address the challenge of supervised classification of time series of multispectral Sentinel-2 imagery.

From the perspective of decision-making, the availability of MACAT in the post-production phase renders it unsuitable for use in the context of issues such as water and fertilizer management by farmers, production estimates and construction of storage structures, market price control and imposition of new export and import rules by policy makers, among others, which require access to information in a timely manner (during production). The topic has been widely discussed in the literature, and among the proposed solutions are early mapping (in the early stages of the growth cycle) of agricultural crops, as suggested by authors such as Hao et al. (2018), yu HAO et al. (2020) and Inglada et al. (2016).

Conversely, as indicated in its datasheet, MACAT is manufactured exclusively with optical data from Sentinel-2. The production of cartographic information from optical images is constrained by rainy seasons (winter and spring), during which the quality of the images deteriorates due to the presence of clouds. The presence of clouds affects the images due to the reflection and scattering of visible light (400nm-700nm) and near-infrared (700nm-1400nm) light, the electromagnetic regions in which multispectral satellites operate. This effect results from interactions with cloud particles of a similar size, which can range from micrometers to millimeters.

This issue has also been discussed in the literature, and authors such as Asam et al. (2022) and Inglada et al. (2016) propose integrating polarimetric and multispectral data as a potential solution. This proposal is based on the ability of polarimetric data to undergo minimal or no dispersion by water droplets or ice crystals in clouds. This is due to the fact that SAR satellites operate in the microwave band (radio waves with shorter wavelengths), where the wavelength (not less than 1 cm) is larger than the size of cloud particles (usually ranging from micrometers to millimeters).

In this context, the justification for this research is twofold. Firstly, there is a need to develop a supervised classification approach that allows for the assessment of how early maize and rice can be classified with considerable performance met-

rics (performance $\geq$ 90%), thus enabling the timely availability of information. Secondly, there is a need to explore the contribution of radar data in the early agricultural crop mapping.

1.3 Objectives

The primary objective of the study was to automate the early mapping of maize and rice plots in the Baixo-Mondego region. This was achieved through the use of cloud computing technology and intelligent classification models, which were trained on a time series of polarimetric and multispectral data from Sentinel-1 and Sentinel-2 satellites, respectively.

The study focused on the following interests:

1. Identify best practices for classifying agricultural crops using satellite data and artificial intelligence;
2. Determine the earliest time at which the accuracy curve of the classification reaches an accuracy of 90% or higher;
3. Explore the contribution of SAR data to the accuracy of classification models;
4. Identify the more effective satellite-derived features that allow distinguishing maize and rice crops throughout the spring-summer agricultural season;
5. Explore the capabilities of Google Earth Engine (GEE) for constructing intelligent models for cartographic mapping of land use.

1.4 Structure of the Dissertation

The dissertation is comprised of seven chapters. Introduction; 2. Review Outcomes; 3. Study Area and Data; 4. Methodology; 5. Results; 6. Discussion; 7. Conclusion and Future Works. The content of each of these chapters is outlined below:

- In the Introduction, the research is introduced and contextualized. This includes a presentation of the context, justification, and objectives of the investigation.;
- The Review Outcomes presents a literature review of remote sensing applied to agriculture, crop mapping, and key challenges. Additionally, the chapter presents a strategy and results of a review of 70 publications on the mapping of agricultural crops using satellite data and artificial intelligence over the past 20 years. This review identifies the most effective practices and an understanding of the principal concepts and challenges;

- The third, Study Area and Data, provides an overview of the study area and the data utilized in this study;
- The methodology chapter outlines the applied methodology;
- The findings on the monitoring of the performance evolution of the classification models, the identification of the early mapping model with acceptable performance, feature efficiency, and the contribution of data fusion are presented in the results chapter;
- The fifth chapter presents a discussion of the aforementioned results, contextualizing them within the existing literature;
- The seventh chapter presents the principal findings of the study, together with their implications for agricultural monitoring.

Chapter 2

Review Outcomes

This chapter presents a review surrounding the main topics, methodologies, and results found in recent literature on agricultural crop mapping, using a combination of remote sensing data and computational learning techniques. Divided into two parts, the first aims to explain the process of obtaining and selecting articles for the literature review. The second focuses on presenting the results and the main discussions of the authors on the key topics, with the objective of identifying best practices.

2.1 Agriculture Crop Mapping Using Remote Sensing

This section provide an overview of agricultural mapping with remote sensing and the main challenges are presented.

2.1.1 Overview of Remote Sensing in Agriculture

The field of remote sensing has played a pivotal role in the advancement of precision agriculture, a contemporary approach to agricultural practice. This field necessitates the utilization of sophisticated technologies to model field conditions and optimize resource management.

In the context of precision agriculture, remote sensing represents a significant advancement in the collection of data on the presence and health status of crops, obviating the need for direct contact. This process allows for the optimization of time and resources.

The underlying physical principle of this process is based on the capture and processing of electromagnetic energy reflected or emitted by crops using sensors mounted on aerial platforms (such as drones) or satellites. This electromagnetic energy is transformed into digital signals, which are then compared to standard digital signals for each crop in order to identify the presence or condition of the

crop. The integration of remote sensing in precision agriculture enables the implementation of applications such as the mapping of crop distribution, the monitoring of crop health, and the management of irrigation.

2.1.2 Key Challenges in Crop Mapping

The principal challenges encountered in the process of crop mapping using satellite images can be classified into two main categories. The first category of challenges pertains to data-related issues, while the second category encompasses processing-related challenges.

2.1.2.1 Data-Related Challenges

One of the most significant data-related challenges is that the spatial, temporal, spectral, and radiometric resolutions of the sensors limit the accuracy of mapping and identification. For example, low spatial resolution may result in inaccuracies in mapping in areas with small agricultural plots. With regard to temporal resolution, the temporal resolution of satellites is frequently insufficient for time series mapping, where the minimum temporal unit is the crop cycle. In such instances, the cycles are aggregated into a single stage of the time series. The radiometric resolution of sensors determines the type of crop information that can be captured and the range of values that can be sensed (the sensitivity of the information).

Moreover, Joshi et al. (2023) identify the availability and cost of ground truth data for use in training and validation as another significant challenge in the field of data science.

2.1.2.2 Processing-Related Challenge

With regard to processing-related challenges, the findings of Joshi Joshi et al. (2023) indicate that the model architecture, which depends on the choice of features and algorithm, can also be a limiting factor in mapping accuracy. Moreover, the increasing prevalence of integrating data from disparate satellites and employing time series analysis has led to the integration of these data sets and the availability of computing platforms to reduce the associated computational costs. Finally, Joshi et al. (2023) also discusses the interpretability of certain processing models, such as deep learning algorithms. These models may be perceived as opaque, obscuring valuable insights about the data.

2.2 Systematic Review of Crop Mapping Approaches

2.3 Article Selection Strategy

To initially narrow the scope of the bibliographic survey, the first step involved defining generic search strings according to the research interest. These search strings were employed primarily to identify meta-analyses on the identification of agricultural crops using remote sensing and machine learning in databases such as Web of Science, Science Direct, and Google Scholar. Three meta-analysis articles were then selected that met the research interest expressed in the initial search strings. The meta-analysis articles served to identify the principal topics of discussion on the subject. This process allowed the identification and choice of more specific strings that, when applied to the same databases, directed the location of over 100 publications containing the principal topics. A temporal filter (selection of publications with a publication date between 2000 and 2023) and a filter related to the sensor's deployment platform (selection of publications where the sensor was deployed on a satellite) were then applied to these publications. After the exclusion of publications that did not meet both criteria, 70 publications were selected for consultation in the State of the Art Database [1]. These publications concern the use of satellite data and machine learning for crop mapping, and were published between 2000 and 2023.

The literature review strategy is illustrated in Figure 2.1.

2.4 Results from the Review

This section aims to present the results of the literature review and will be divided into the following subsections: 2.4.1. Spatial and Temporal Distribution of Selected Articles; 2.4.2. Satellite Technology; 2.4.3. Mapping Approaches; 2.4.4. Satellite-Derived Features; 2.4.5. Machine Learning Algorithms in Crop Mapping; 2.4.6. Best Practices in Agricultural Crop Mapping and 2.4.7. Fundamentals of the Support Vector Machine and Random Forest Algorithms.

2.4.1 Spatial and Temporal Distribution of Selected Articles

To initially narrow the scope of the bibliographic survey, the first step involved defining generic search strings according to the research interest. These search strings were employed primarily to identify meta-analyses on the identification of agricultural crops using remote sensing and machine learning in databases such as Web of Science, Science Direct, and Google Scholar. Three meta-analysis articles were then selected that met the research interest expressed in the initial search strings. The meta-analysis articles served to identify the principal topics of discussion on the subject. This process allowed the identification and choice

[1] Connection to State of the Art Database (click on word to open the link)

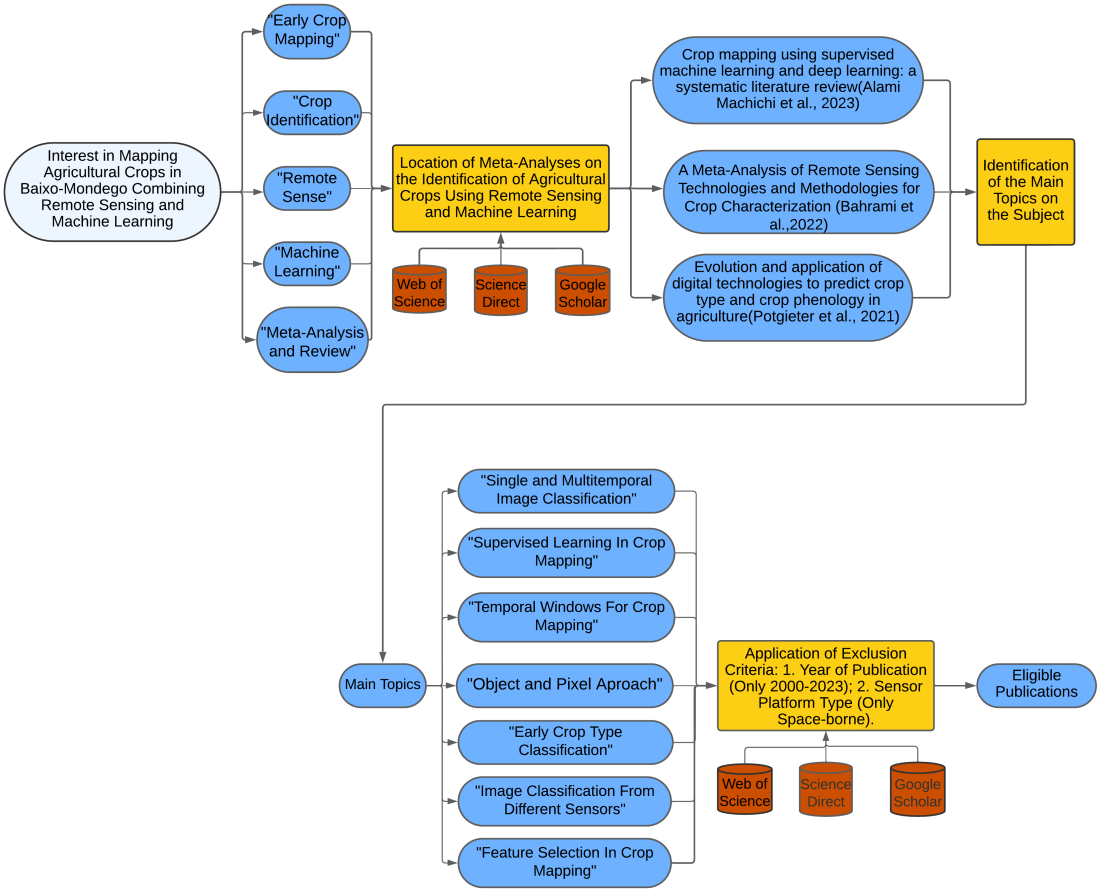

Figure 2.1: Literature Review Strategy

of more specific strings that, when applied to the same databases, directed the location of over 100 publications containing the principal topics. A temporal filter (selection of publications with a publication date between 2000 and 2023) and a filter related to the sensor's deployment platform (selection of publications where the sensor was deployed on a satellite) were then applied to these publications. After the exclusion of publications that did not meet both criteria, 70 publications were selected for consultation in the State of the Art Database. These publications concern the use of satellite data and machine learning for crop mapping, and were published between 2000 and 2023.

The spatial distribution map of selected articles (Figure 2.2) illustrates that the mapping of agricultural cultures using Earth observation data and processing through artificial intelligence is dominated by major powers and is under-explored globally. This is also evident in Portugal, where there remains much to be done.

With regard to the temporal distribution of the articles in Figure 2.3, it is observed that the highest frequencies are associated with the last decade. Machichi et al. (2023) propose that the increasing popularity of machine learning algorithms, population growth, food scarcity, and climate change are explanations for this surge, particularly in the second half of the decade.

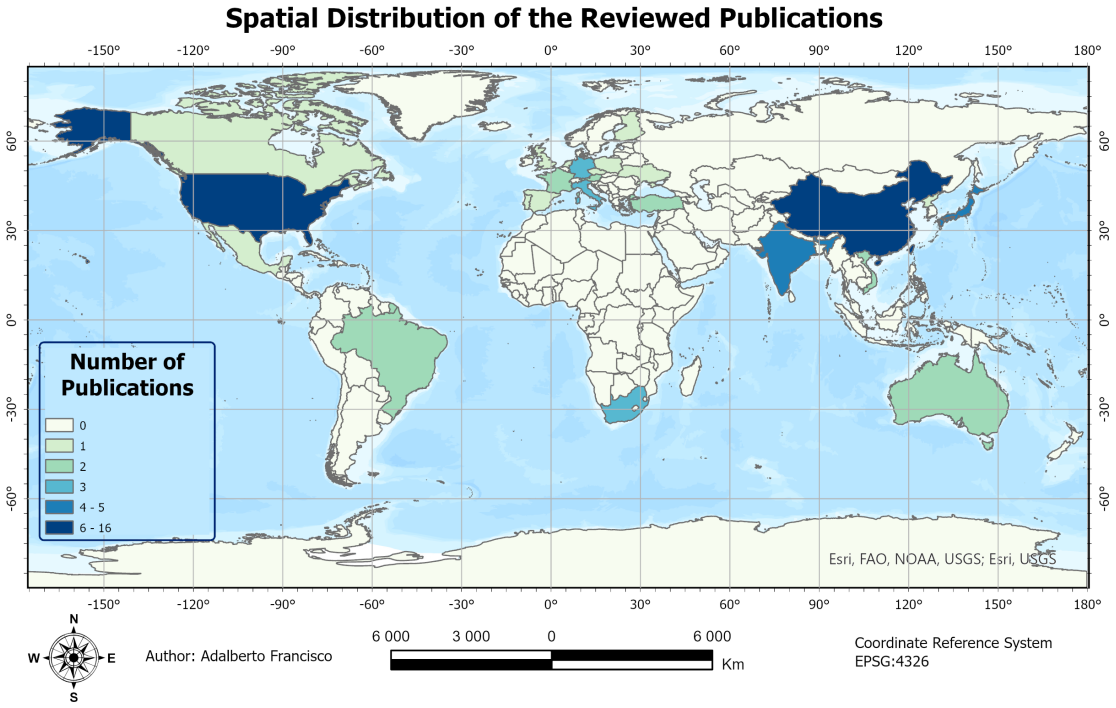

Figure 2.2: Spatial Distribution Map of Selected Articles

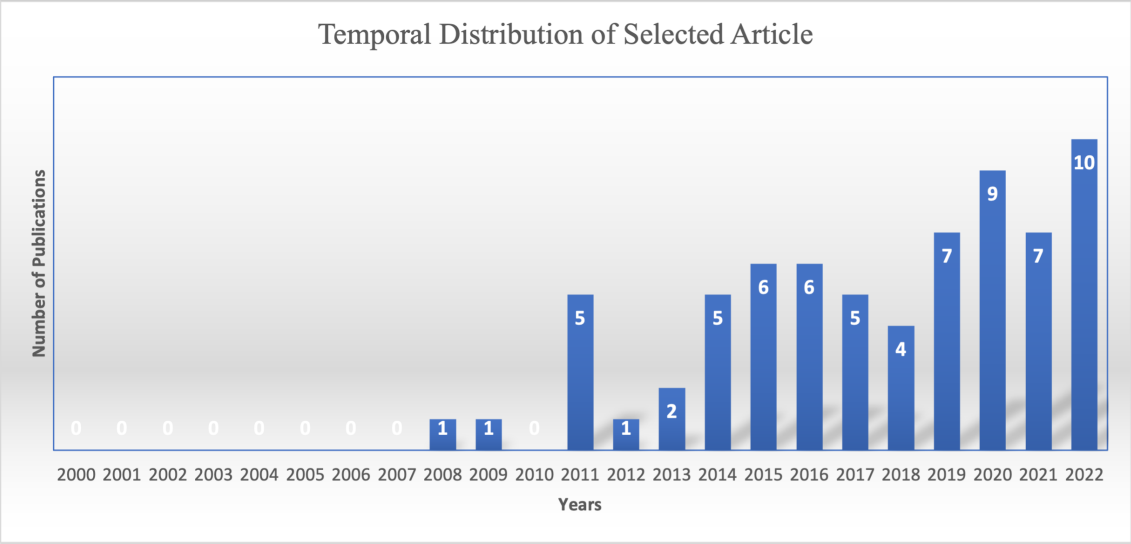

Figure 2.3: Temporal Distribution of Selected Articles

2.4.2 Satellite Technology

The selection of satellite technology represents one of the most pivotal elements in the conceptualization of a satellite-based agricultural crop mapping project, as its specifications directly influence the project's overall efficacy. As evidenced by the literature, the two most pertinent parameters regarding satellite technology

are spectral and spatial resolutions. Spectral resolution is a parameter that is unique to the sensor, while spatial resolution is a parameter that is relative to the platform-sensor system, depending on the Instantaneous Field of View (IFOV), which is a magnitude related to the sensor (detector diameter and focal distance), but also the flight altitude, which is related to the platform (generally considered negligible for space platforms).

With regard to the wavelengths at which the sensor is capable of detecting electromagnetic energy, the spectral resolution of the sensor in question determines the feasibility of implementing it in a given project. This is because the sensor's ability to acquire electromagnetic energy in the bands reflected by the objects to be mapped ultimately determines its suitability for use in the project. This parameter allows for the differentiation of four technologies:

1. Synthetic Aperture Radar (SAR) Technology: This technology is implemented in active sensors that detect only the radio wave band of the electromagnetic spectrum reflected by objects on the surface;
2. Multispectral Technology: This technology is implemented in passive sensors that capture and record electromagnetic energy reflected by objects in multiple wavelengths;
3. Hyperspectral Technology: This technology is implemented in passive sensors capable of capturing and recording electromagnetic energy reflected by objects in various wavelengths;
4. Light Detection and Ranging (LIDAR) Technology: This technology is implemented in active sensors that capture the laser reflected by objects on the surface.

A review of the literature from the past twenty years has permitted an assessment of the most commonly used satellite technologies in the mapping of agricultural crops in recent years. This is illustrated in Figure 2.4.

The prevalence of multispectral technology can be attributed to the fact that vegetation reflects electromagnetic energy in the visible, near-infrared, and shortwave infrared bands. Nevertheless, certain authors, such as Santos (2021), Kussul et al. (2016), and Verma et al. (2019), have combined SAR and multispectral technology in vegetation mapping and demonstrated that the incorporation of SAR data, due to its insensitivity to clouds, can enhance classification accuracy by approximately 5%, particularly in instances where images were captured during rainy seasons. In agreement, Blickensdörfer et al. (2022) and Asam et al. (2022) conducted studies that demonstrated improvements in classification accuracy of between 6% and 10% and 6% and 9%, respectively.

Spatial resolution is defined as the minimum size of an object that a satellite is capable of distinguishing. Consequently, given a specific spatial resolution, only objects with a size equal to or greater than that of the spatial resolution of the platform-sensor system can be identified. It is therefore evident that spatial resolution can be adjusted to suit the specific requirements of a given study, area, and

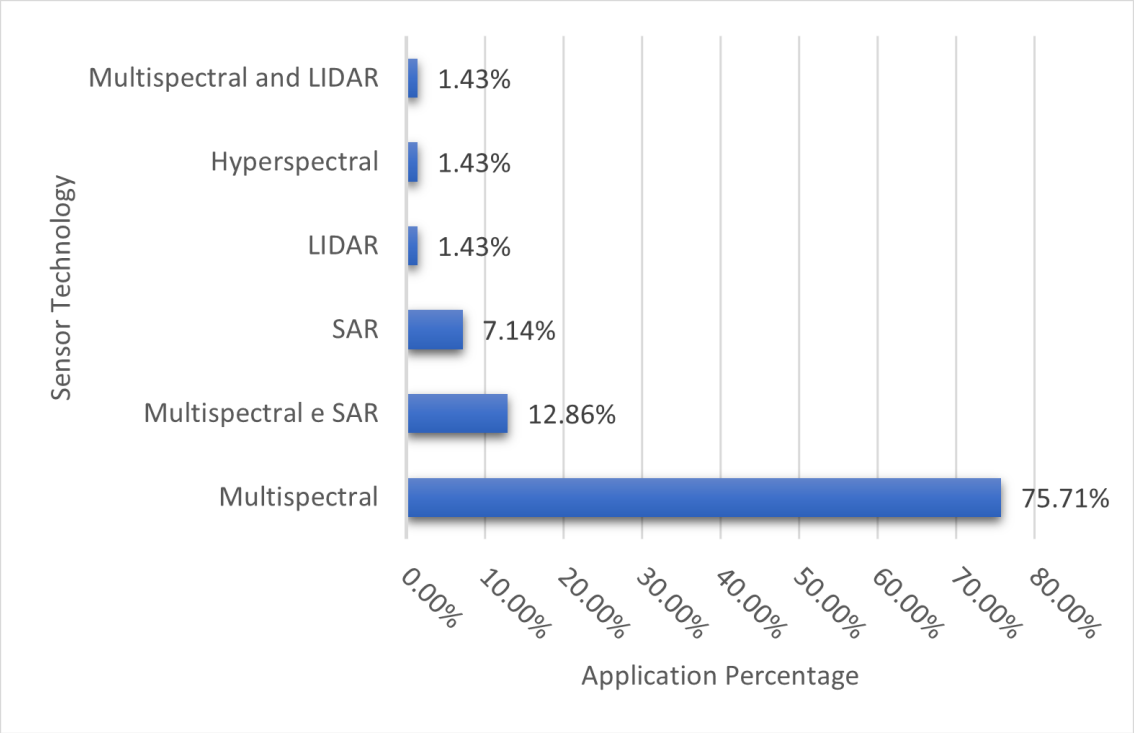

Figure 2.4: Percentage of Satellite Technology Application in Selected Articles

agricultural plot complexity. It is crucial to exercise sound judgment to ensure that classification is not compromised by selecting an acquisition system with an inadequate spatial resolution, one that is insensitive to the agricultural complexity of the region, or a system with an excessively high spatial resolution, which would result in an increased computational cost for classification.

In the majority of studies, the vast extent of agricultural areas necessitated the utilisation of satellites equipped with low spatial resolution sensors (such as Moderate Resolution Imaging Spectroradiometer (MODIS) with a resolution of 250 metres). Nevertheless, the utilization of satellites equipped with sensors of 10, 20, and 60 meters of spatial resolution, such as those employed by the Sentinel and Landsat missions, is a prevalent practice.

In conclusion, the principal findings of this literature review are as follows: 1. The spectral resolution of the sensor is the most crucial parameter to be taken into account when designing a satellite-based agricultural crop mapping project. 2. Over the past 20 years, multispectral technology has been the most prevalent, followed by SAR technology. 3. The combination of the two technologies has demonstrated the potential to achieve high levels of accuracy in cloud-covered areas.

2.4.3 Mapping Approaches

Mapping approaches can be distinguished according to three different criteria: (1) Temporal range of data acquisition; (2) Minimum image classification unit; (3) Moment of mapping. The three different criteria are described below.

2.4.3.1 Mapping Approaches Based on Temporal Range of Data Acquisition

Over the past two decades, two distinct approaches have emerged in the analysis of satellite images based on the temporal range of the images:

Single-Date Image Classification

This approach is based on single-date images and is less commonly used due to the difficulty in identifying the optimal point of distinction between the phonologies of different crops over time. Different crops may exhibit the same spectral response at certain phenological stages, leading to confusion in classification models. Regarding the causes of spectral confusion, Potgieter et al. (2021) cites the bandwidths around the central wavelength of some satellites (such as Landsat, for example). Despite the difficulties, the fact that some authors, such as Jayakumari et al. (2021), Heinl et al. (2009), and Immitzer et al. (2016), have achieved maximum overall accuracies of 81%, 94.4%, 89.2%, and 83%, respectively, demonstrates that moderately high accuracies can be achieved with single-date images. The results demonstrate that moderately high accuracies can be achieved with single-date images, with overall accuracies of 81%, 94.4%, 89.2%, and 83% being achieved, respectively.

Multi-temporal Satellite Image Series Classification

This is based on the classification of a set of images acquired throughout the plant's phenological cycle. Currently, it has been the most recurrent and promising approach due to the capacity of feature densification following the phenological cycle to resolve the problem of identifying the optimal temporal point to distinguish agricultural crops, a challenge present in the single-date image approach. However, Langley et al. (2001) and Niel and McVicar (2004) warn that the success of this approach depends on the use of limited observations that include the main phenological stages, risking leading the model to overfitting through the addition of redundant (auto-correlated) information.

This approach is based on the classification of a set of images acquired throughout the plant's phenological cycle or agricultural season. Currently, it has been the most recurrent and promising approach due to the capacity of feature densification following the phenological cycle to resolve the problem of identifying the optimal temporal point to distinguish agricultural crops, a challenge present in the single-date image approach. Nevertheless, as observed by Langley et al. (2001) and Niel and McVicar (2004), the efficacy of this approach is contingent upon the utilization of a limited number of observations encompassing the principal phenological stages. This may potentially lead to the model becoming overly fitted through the introduction of superfluous (auto-correlated) information, thereby compromising its accuracy.

A review of the literature revealed that in 83% of the 70 studies, multitemporal series were utilized, while in only 17% of the studies single-date images were

employed. This indicates a clear predominance of multitemporal images, as illustrated in Figure 2.5.

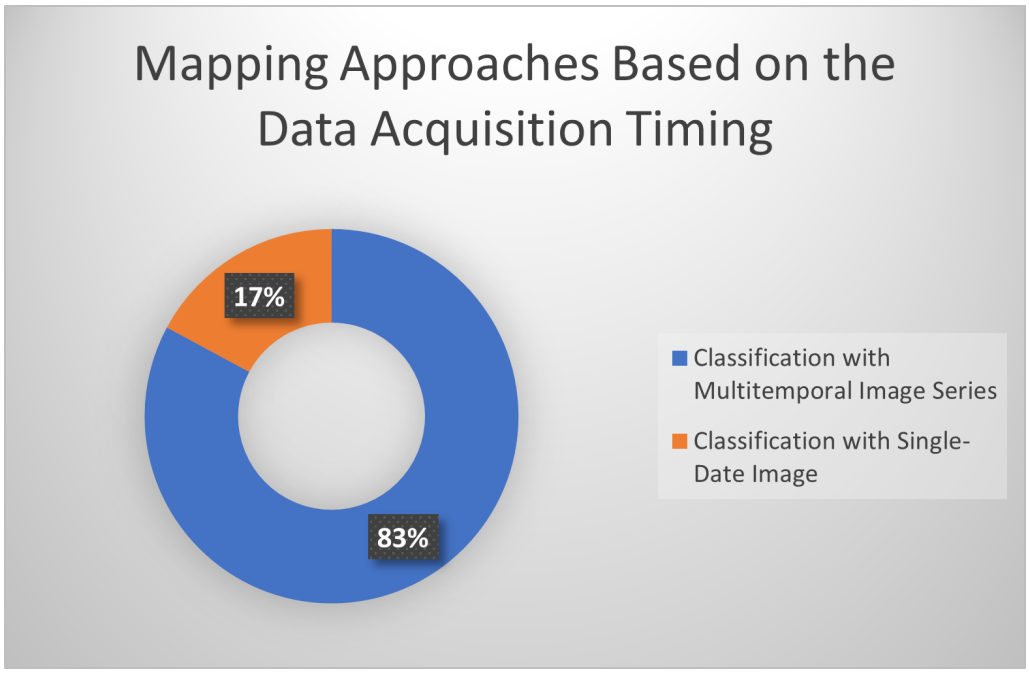

Figure 2.5: Mapping Approaches Based on the Data Acquisition Timing

2.4.3.2 Mapping Approaches Based on Minimum Unit of Image Classification

The minimum unit of classification is defined as the smallest disaggregation unit of the data to be classified. According to this criterion, three distinct approaches can be identified: 1. Pixel-Based Approach; 2. Object-Based Approach and 3. Hybrid Pixel and Object-Based Approach.

Pixel-Based Approach

In this approach, the minimum unit of classification is identical to the image's minimum unit, namely the pixel. In the discourse among authors, the principal disadvantage of this technique is attributed to salt-and-pepper noise resulting from the absence of contextual information about the pixel, which impairs the model's performance. This is in accordance with the findings of Potgieter et al. (2021). As a solution, the application of spatial filtering techniques, such as the Savitzky-Golay filter, is a common approach. This was demonstrated by Chen et al. (2004). The primary advantage is the enhanced capacity to identify minute and slender objects in settings devoid of significant topographical intricacies.

Object-Based Approach

This approach employs the object as the fundamental unit of classification, defined as a cluster of pixels with spatial relation. This approach overcomes the

limitations of the pixel-based approach by allowing for a more nuanced understanding of the spatial relationships between objects. The primary challenge in implementing this methodology is the accurate determination of segmentation size. As elucidated by Xue et al. (2023), this parameter is contingent upon the dimensions of the study area, the spatial resolution of the sensor, and the topography. To ensure the integrity of the plot, the segmentation size must be calibrated to avoid errors due to insufficient resolution or over-segmentation. Accordingly, Xue et al. (2023)) propose a gradual increase in the size of segmentation, guided by the factors on which it depends and the objectives of the research.

Hybrid Pixel and Object-Based Approach

The review of publications in which both approaches were applied (Kussul et al. (2016), Immitzer et al. (2016), Qadeer et al. (2021), and Xue et al. (2023)) proved useful in that it allowed the understanding that it is not possible to absolutely rank the two approaches, as their influence does not play a deterministically isolated role in the overall classification accuracy. The relative accuracy of the two approaches may vary depending on a number of factors, including the quality of the segmentation, the choice of algorithms, the quantity and representativeness of the training data, and the feature selection.

In Figure 2.6, a summary of the reviewed articles is presented based on the minimum classification unit.

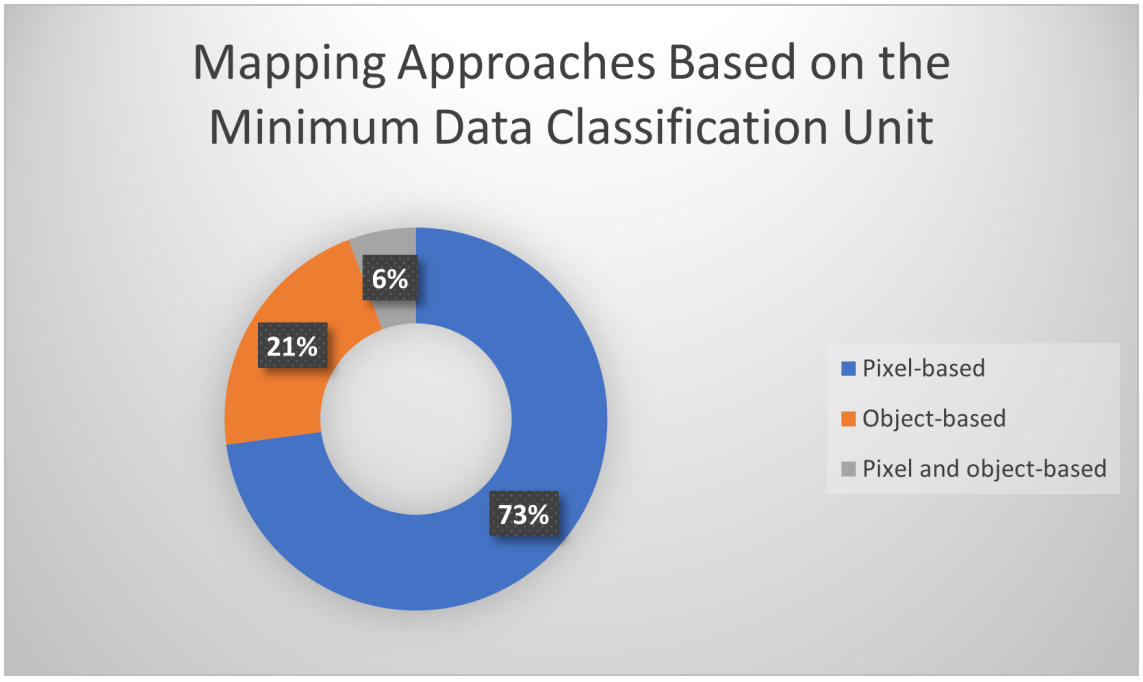

Figure 2.6: Mapping Approaches Based on the Minimum Data Classification Unit

2.4.3.3 Mapping Approaches Based on Mapping Time

With regard to the mapping of temporal data, two distinct methodologies may be identified: end-of-season mapping and early mapping, which occurs at the inception of the agricultural season. As illustrated in Figure 2.7, a review of the literature revealed that only five of the 70 authors employed an early mapping approach, which involves the identification of crops at the beginning of the agricultural season. As indicated by I Inglada et al. (2016), this is due to the fact that Sentinel 2 images demonstrate greater efficacy at the end of the season when the crops are mature, thereby exhibiting more distinct spectral signatures.

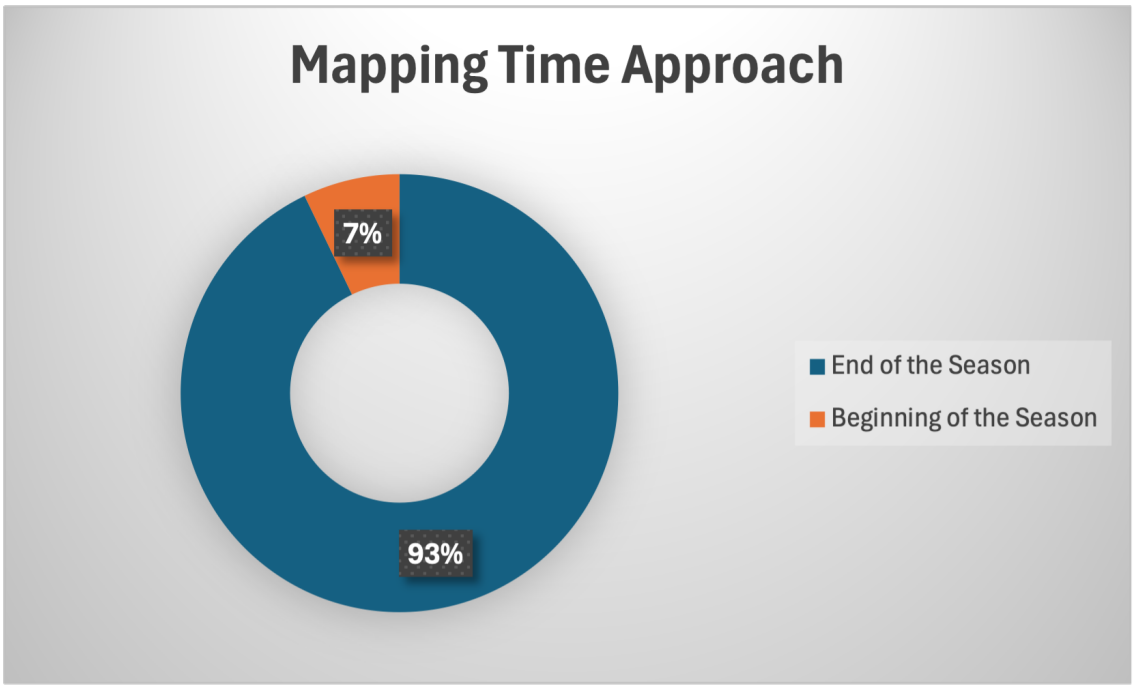

Figure 2.7: Mapping Approaches Based on the Data Acquisition Timing

However, given the significance of the timing of agricultural crop distribution in the decision-making process, there has been a growing necessity in recent years to develop methodologies that permit the early mapping of agricultural crops.

2.4.4 Satellite-Derived Features

Feature selection represents a pivotal stage in the model, as it determines the data from which the machine learning model will learn patterns. As stated in He et al. (2022), there is a correlation between the quantity of features, their type, and the model's performance. It is important to note that if a large number of features are utilized randomly during the training process, the model's prediction accuracy may be negatively impacted by the redundancy of the features. Conversely, if only a few features are selected at random, there is a risk of selecting features with minimal discriminatory power, which could result in a decline in classification accuracy. Consequently, the objective of optimizing models is to select non-redundant features with strong data discrimination power and guide the

model towards accurate learning.

The discriminatory power of a feature in crop mapping through machine learning is directly linked to its sensitivity in describing the changes in the phenological characteristics of the crops throughout the phenological cycle. A variety of techniques for identifying features with greater discriminatory power, such as Recursive Feature Elimination (RFE), Random Forest Average Impurity (RFAI), and Average Precision Reduction (RFAI), are described in the literature.

In this sense, according to the systematic review of the literature, the most common features in crop mapping are the Spectral Bands (SB) from the VIS, NIR, and SWIR regions, backscattering data or Parametric Feature (PF), Spectral Index (SI), Phenology Index (PI), and Texture (T). The percentages of feature type usage are presented in Figure 2.8.

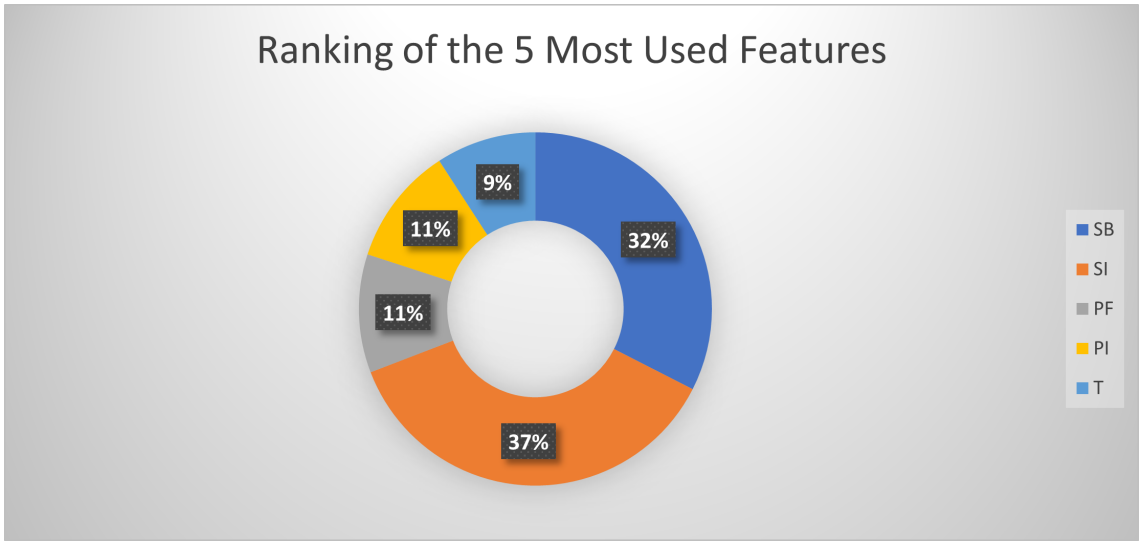

Figure 2.8: Ranking of the 5 Most Utilized Features in Selected Articles

The identification of the most effective features is a contentious topic in the literature, as it appears to fluctuate contingent on the agricultural crop, study context, and the specific characteristics of the classification. Despite the controversy surrounding this topic, an attempt was made to identify which of the five most frequently occurring features were associated with the highest efficacy rates. In order to achieve this, the methodology proposed by Machichi et al. (2023) was employed. In accordance with this methodology, efficacy is defined as the ratio between the frequency of use of a feature and the number of times this feature was part of the set of features that generated the best result. Consequently, the efficiency ranking of the five most frequently utilized feature types, as determined by the application of this concept to the reviewed publications, is presented in Figure 2.9.

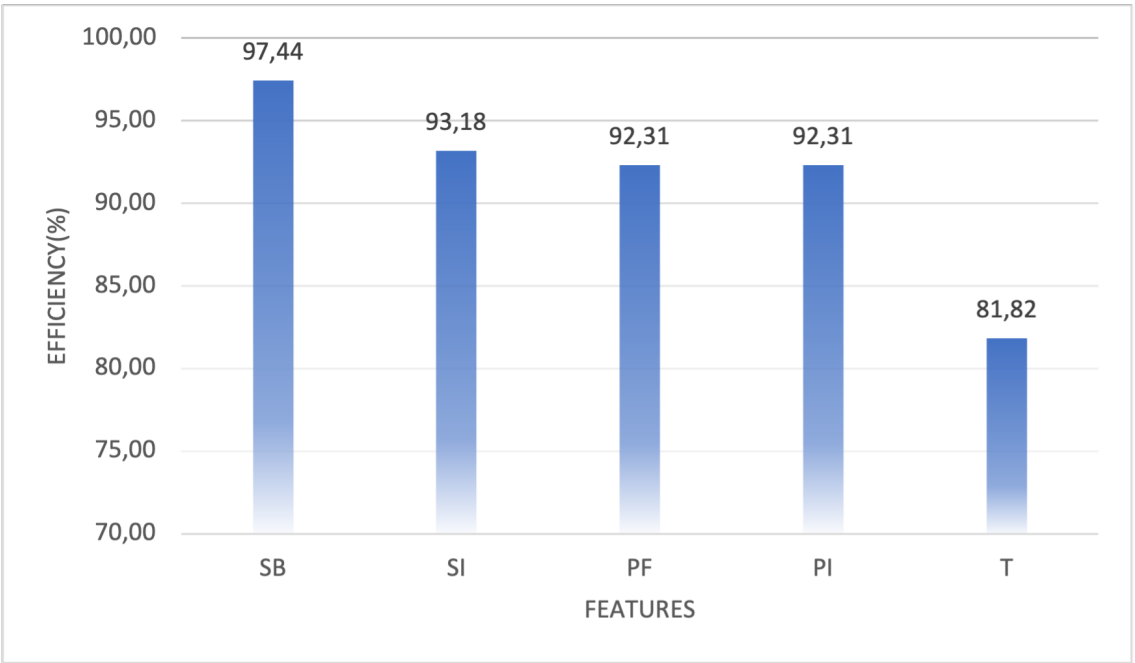

Figure 2.9: Efficiency of the 5 Most Utilized Features in Selected Articles

2.4.5 Machine Learning Algorithms in Agricultural Crop Mapping

From the literature review, it is evident that, although there is a growth in the use of deep learning algorithms, the majority trend still leans towards the application of machine learning algorithms - Random Forest (RF), Support Vector Machine (SVM), k-Nearest Neighbors (k-NN), Decision Tree (DT), Maximum Likelihood Estimation (MLE) - as demonstrated in Figure 2.10.

A review of the literature reveals that, although there has been a notable increase in the utilization of deep learning algorithms, the majority of studies still employ machine learning algorithms, including RF, SVM, k-NN, DT, MLE. This is evidenced by Figure 2.10, which ranks the five most commonly used algorithms in selected articles.

As with features, there is no clear consensus in the literature regarding the ranking of the most effective algorithms. However, as was done in the analysis of features, the methodology of Machichi et al. (2023) was applied to construct a ranking of the efficiency rate of the five most used algorithms. This analysis was based on the relationship between the frequency of use of an algorithm and the number of times it generated the optimal classification model in the same context (with the same data). Consequently, the efficiency ranking of the five most utilized algorithms is presented in Figure 2.11. The efficiency of the five most utilized algorithms in selected articles is presented in the following Figure.

In addition to representing 74% of the usage of the five most recurrent algorithms, this review also notes that RF and SVM stand out as the two most efficient algorithms. In favor of the RF algorithm, Qadir and Mondal (2020) justifies its success due to its robustness against overfitting, as it can handle a large amount of data

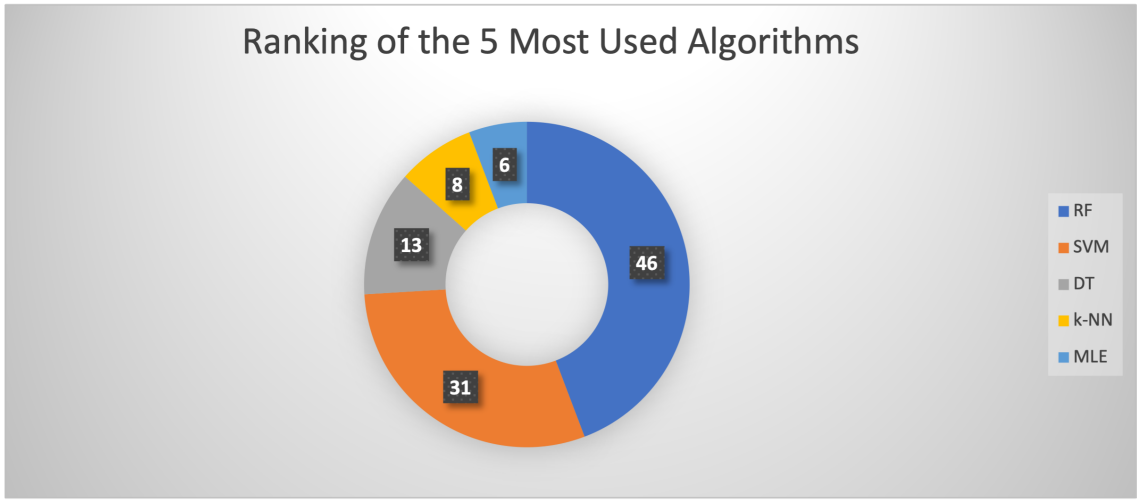

Figure 2.10: Ranking of the 5 Most Used Algorithms in Selected Articles

from various sources. Furthermore, Chabalala et al. (2022) asserts that the algorithm is well-suited to heterogeneous landscapes due to its capacity to effectively manage noise and intrinsic correlations inherent in satellite data. In contrast, the SVM is also a viable option. As demonstrated by HU et al. (2017), it performs well in high-dimensional problems and has good generalization ability from a few training data, making it suitable for mapping various crops.

2.4.6 Best Practices in Agricultural Crop Mapping

A systematic review of the literature has enabled the identification of best practices for mapping agricultural crops using satellite data and processing through machine learning algorithms over the past 20 years (Figure 2.12). These practices integrate criteria related to multimodal data and temporal dimensions (which pertain to the information present in the data and the timing of acquisition, respectively), and the computational dimension concerning the data processing through computational learning.

It can be concluded that, over the past 20 years, the most effective approach to mapping agricultural crops using machine learning from satellite data in the multimodal data dimension is one that combines the spectral bands from the visible, near-infrared, and short-wave infrared (SB) regions, spectral indices (SI), phenological indices (PI), texture information (T), and radar backscatter data (PF).

In the temporal dimension, the use of multitemporal data has yielded the most favorable outcomes. With regard to the computational dimension, the random forest (RF), support vector machine (SVM), and k-nearest neighbor (KNN) algorithms have demonstrated superior performance in machine learning, while the convolutional neural network (CNN) and long short-term memory (LSTM) algorithms have exhibited exceptional capabilities in deep learning.

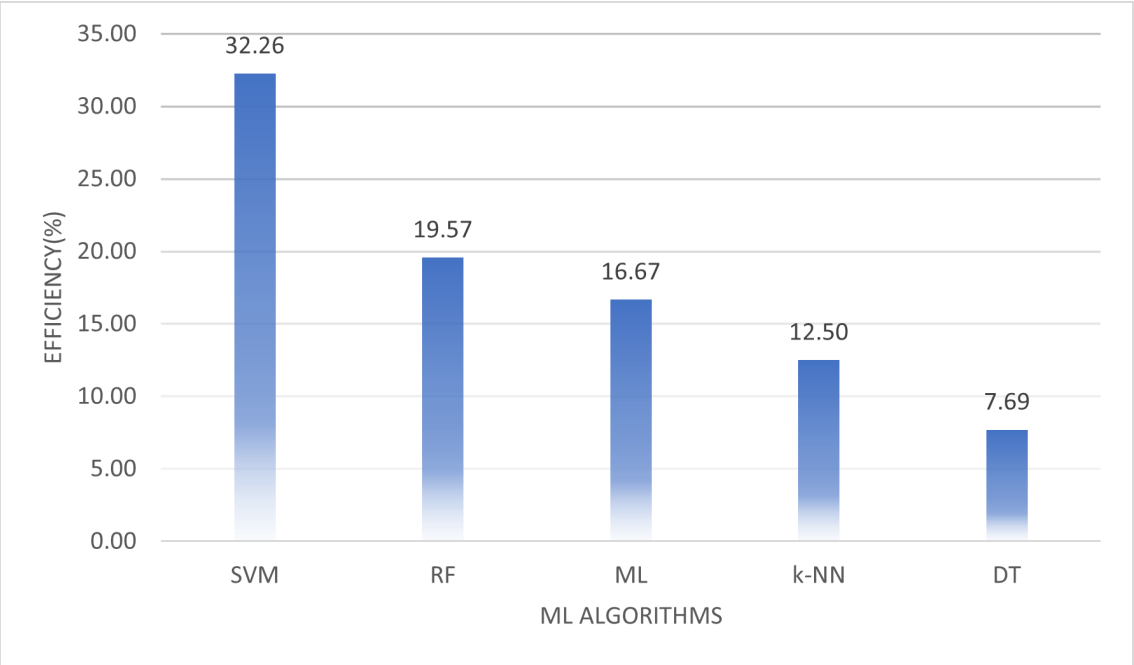

Figure 2.11: Efficiency of the 5 Most Utilized Algorithms in Selected Articles

2.4.7 Fundamentals of the Support Vector Machine and Random Forest Algorithms

2.4.7.1 Support Vector Machines (SVM)

As outlined by Cristianini & Shawe-Taylor (2000), Support Vector Machines (SVMs) represent a supervised learning method employed for classification and regression problems. The underlying principle of SVMs is to identify a hyperplane that optimizes the margin between classes while minimizing the classification error, as illustrated in 2.13.

Given a binary problem, where x_n are feature vectors and $y_i \in \{-1, 1\}$ are the classes corresponding to the vectors, the objective of an SVM for a binary problem is defined as follows:

$$w \cdot x + b = 0 \tag{2.1}$$

where:

- w is the weight vector;
- x is the feature vector;
- b is the bias.

Prior to any further explication, it is essential to clarify that the term "Support Vector Machines" encompasses a broader range of applications than those traditionally associated with support vector classifiers. SVMs can be classified as

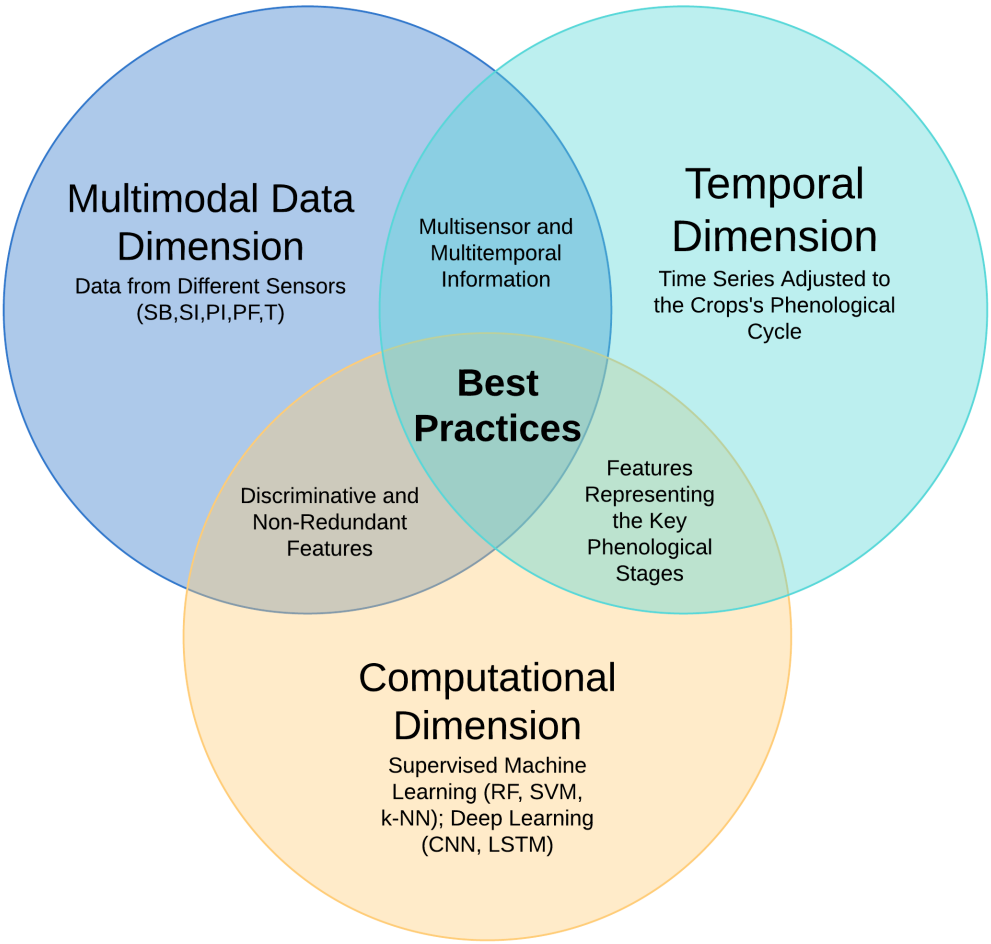

Figure 2.12: Best Practices for Mapping Agricultural Crops Using Remote Sensing Data with Machine Learning

either linear or non-linear, depending on the presence of outliers in the data or the nature of the problem itself.

Both linear and non-linear SVMs can employ either fixed-margin or soft-margin classifiers. In the context of SVMs, the margin is defined as the distance between the hyperplane and the support vector of each class (sample of the class closest to it).

Since the margin is inversely proportional to the norm of w, the general optimization problem is defined by:

$$\min_{w,b} \frac{1}{2}\|w\|^2 \tag{2.2}$$

$$\text{subject to } y_i(w \cdot x_i + b) \geq 1, \text{ for } i = 1, \ldots, n. \tag{2.3}$$

During the training phase, the presence of samples positioned between the margins (samples incorrectly classified in the training data) results in the use of a soft-margin support classifier. In the absence of margin violations, one is dealing

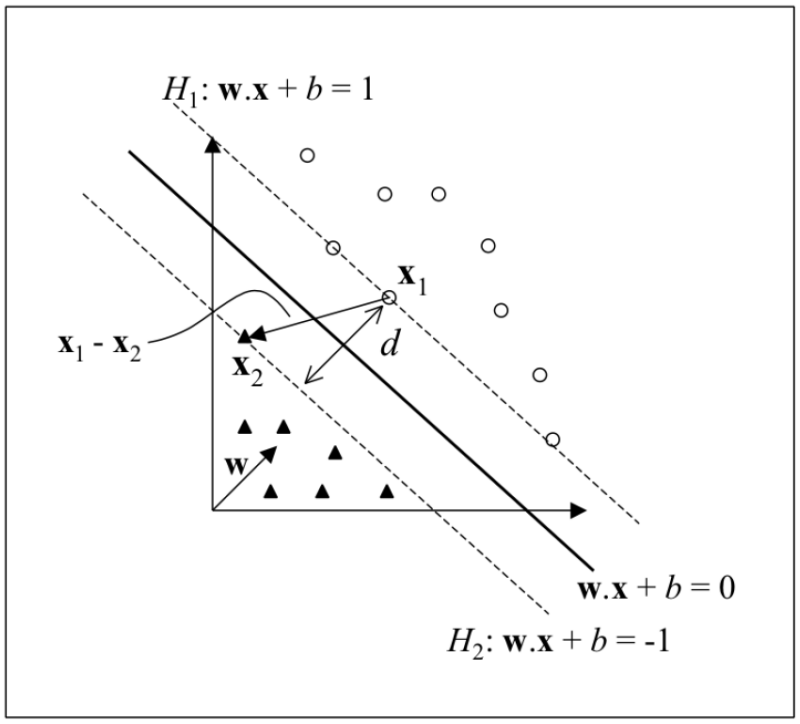

Figure 2.13: Generic Illustration of SVM (Lorena & De Carvalho, 2007)

with a fixed-margin support classifier. Consequently, fixed-margin support classifiers posit that the data are perfectly linearly separable and devoid of outliers.

Given that real-world problems rarely exhibit a perfectly linear and unidimensional structure, it is necessary to consider the use of soft-margin SVMs. Mathematically, with α_i, y_i, x_i and C representing the Lagrange multipliers associated with each point, the labels, feature vectors, and the parameter that balances margin maximization and model complexity, respectively, a soft-margin SVM is formulated for a linear binary problem (not perfectly linear).

$$\max_{\alpha} \left(\sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \right) \tag{2.4}$$

$$\sum_{i=1}^{n} \alpha_i y_i = 0 \quad \text{e} \quad 0 \leq \alpha_i \leq C, \text{ para } i = 1, \ldots, n. \tag{2.5}$$

where:

- α_i and $\alpha_i$are the Lagrange multipliers for the feature vectors xi and xj;
- C, as previously mentioned, is the parameter that controls the trade-off between the complexity and the generalization capability of the model;
- The first condition allows the hyperplane to be positioned neutrally. The second condition requires each Lagrange multiplier to adhere to the constraint within the interval from 0 to C.

In contrast to fixed-margin classifiers, soft-margin classifiers permit a certain proportion of training samples to transgress the margins through the use of a slack variable (ϵ_i) and a penalty parameter (C) for misclassifications. These parameters enable the optimization of the margin while minimizing classification error. Graphically, a soft-margin SVM is illustrated in Figure 2.14.

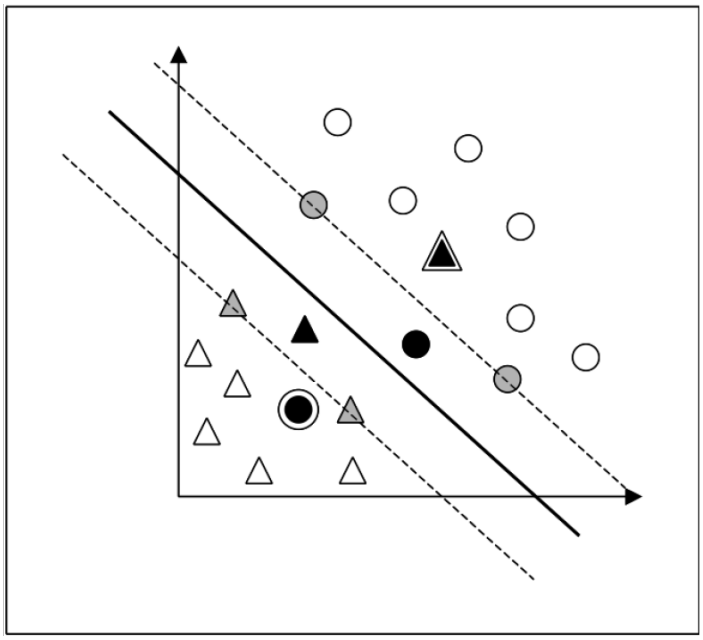

Figure 2.14: Illustration of a Soft Margin SVM (Lorena & De Carvalho, 2007)

Consequently, the Lagrange multipliers permit the identification of the hyperplane vectors (support vectors), where α is greater than 0, as well as the orientation of the hyperplane based on the values of the multipliers' coefficients that satisfy the constraint.

In the event that the data is not linearly separable, Support Vector Machines employ kernel functions that "map" the data into a higher-dimensional space. This allows them to utilize the relationship between variables in order to identify support vectors that are capable of linearly separating the data. This process is illustrated in Figure 2.15.

The essence of the kernel functions magic can be described as a method of working with data that avoids any direct transformation of the data to the mentioned higher-dimensional space. This means that the coordinates of the points in this space are never calculated. Instead, it is sufficient to map how these data points would relate to each other in the higher-dimensional space. This is achieved by calculating the inner product between the feature vectors in the higher-dimensional space.

Once the inner product has been calculated, it provides information about the similarity of the feature vectors during the training phase. This calculation is performed between all pairs of sample combinations. The patterns in the data are thus elucidated, and a hyperplane that separates the classes in the high-dimensional space is identified. Consequently, for a nonlinear problem, the decision function from equation 1 assumes the following form:

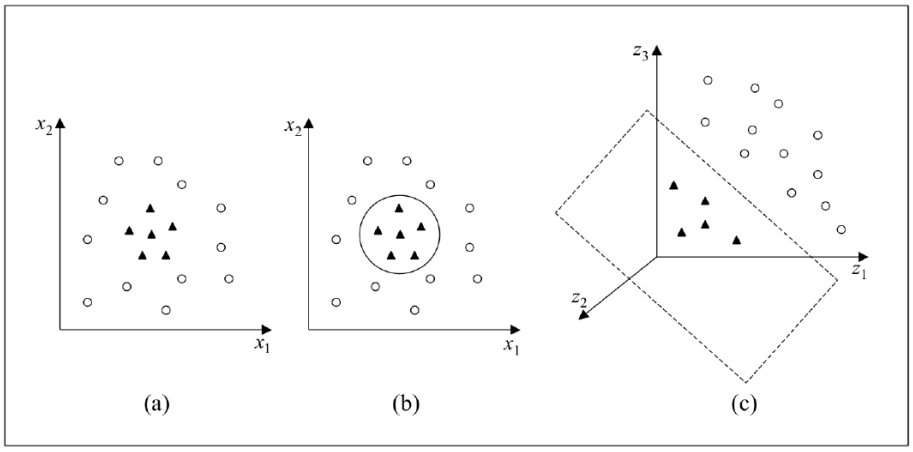

Figure 2.15: Kernel mapping from a 2D space to a 3D space: (a) Nonlinear dataset; (b) Non-linear Boundary in the Transformed Feature Space; (c) Linear Boundary in the Transformed Feature Space (Lorena & De Carvalho, 2007).

$$f(x) = \sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b \tag{2.6}$$

where:

- $K(x, x_i)$ is the kernel function;
- $x_i$xi is the new feature vector to be classified;
- x s the support vector.

In the testing phase, each sample is mapped to the high-dimensional space, and the inner product between the sample's feature vector and the support vectors is calculated. The classifier is then defined by the following sign function:

$$f(x) = \text{sgn}\left(\sum_{i=1}^{n} \alpha_i y_i K(x, x_i) + b\right) \tag{2.7}$$

Regarding the parameters of an SVM (C and kernel parameters), listed in the table below, these can be determined through various methods such as Convex Optimization and Approximate Methods. However, a common practice has been to search for them using cross-validation.

Kernel Type	**Function** $K(x_i, x_j)$	**Parameters**
Polynomial	$(\delta\langle x_i \cdot x_j\rangle + \kappa)^d$	δ, κ e d
Gaussian	$\exp\left(-\sigma\|x_i - x_j\|^2\right)$	σ
Sigmoidal	$\tanh\left(\delta\langle x_i \cdot x_j\rangle + \kappa\right)$	δ e κ

Table 2.1: Kernel Types, Functions and Parameters

2.4.7.2 Random Forest (RF)

As outlined by Breiman (2001), the Random Forest represents a supervised machine learning approach. Its operational framework entails the construction of multiple decision trees based on training data, with the subsequent identification of the mode or mean of each tree for the classification and regression of new data, respectively.

Thus, the decision function of this algorithm is given by:

$$H(x) = \arg\max \sum_{i=1}^{K} I(h_i(x) = Y) \tag{2.8}$$

where:

- $H(x)$ is the final classification of a new feature vector;
- I is an indicator function that quantifies how many trees classified x as Y;
- *hi* is the ith classifier (decision tree).

Graphically, this is illustrated in Figure 2.16:

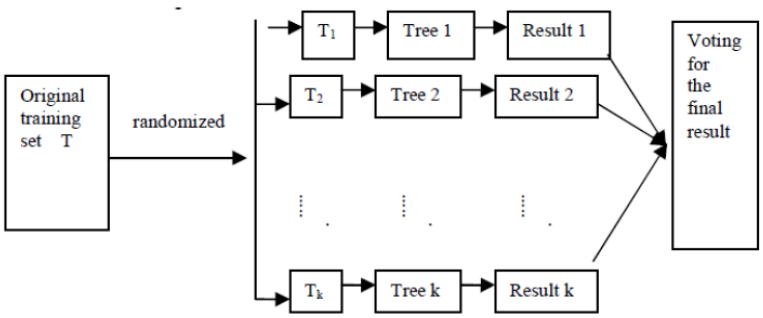

Figure 2.16: Diagram of a Random Forest (Liu et al., 2012)

Two fundamental aspects of this algorithm are the selection of samples and feature vectors for training each tree. Regarding sampling, the Random Forest (RF) employs Bootstrap Sampling, a random sampling of the same size with replacement of samples for each tree. This method of sampling allows a sample to appear multiple times in the training data subset for a tree. This sampling allows each tree to train with distinct data, thereby reducing model overfitting by capturing diverse information about the data.

In terms of feature selection, for each tree, features are chosen randomly from a set specified by the user, with the number of features chosen for each tree determined randomly. The selection of features at each node of the decision tree is commonly performed using the Gini impurity or information gain based on entropy. Both criteria aim to quantify the quality of the split.

Mathematically, for a set S with n classes, the Gini impurity is defined by:

$$\text{Gini}(S) = 1 - \sum_{i=1}^{n} p_i^2 \tag{2.9}$$

where:

- pi is the proportion of samples belonging to class in the set.

The Gini value ranges from 0 to 1, with values closer to 0 indicating lower impurity in the feature. For a feature, the calculation of information gain is performed as follows:

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^{k} \left(\frac{|S_i|}{|S|} \right) \tag{2.10}$$

with,

$$\text{Entropy}(S) = -\sum_{i=1}^{n} p_i \log_2(p_i) \tag{2.11}$$

where:

- pi represents the proportion of instances of class i relative to the total number of instances in the training set (S);
- Entropy (Si) refers to the entropy of the training set;
- Si/Si s the ratio between the number of samples in the subset Si and the total number of samples in training set;
- Entropy (S) is the entropy of the feature.

The most important hyper-parameters of this algorithm are:

1. Number of Trees;
2. Number of Features;
3. Maximum Depth of Trees.

Similar to the hyper-parameters of SVM, among many techniques, these can also be determined through cross-validation.

Chapter 3

Study Area and Data

3.1 Study Area

The study area corresponds to the Baixo-Mondego Valley, located in Coimbra, in the central region of Portugal. It covers an area of 13.34 hectares and extends through the municipalities of Coimbra, Figueira da Foz, and Montemor-o-Velho.

As it is traversed by the Mondego River, the fifth largest river in Portugal with a course entirely within the country, the valley of the Baixo-Mondego sub-region (see Figure 3.1) is distinguished by its rich water potential, making it suitable for agricultural practices. Indeed, according to Soares (2011), it has been the target of development projects by the Ministry of Agriculture such as the "General Plan for the Utilization of the Baixo-Mondego" in 1962 or the "Secondary Irrigation Network" in 2010. These projects not only highlight its fertile potential but also its consequent importance as a regional input to the food and economic sectors, particularly in its main productions—maize of the species Zea mays L. and rice of the Japonica subspecies of Oryza Sativa L., as described in the product datasheet of the Montemor-o-Velho Cooperative [2].

3.2 Data

This section, where the data used and the corresponding acquisition setups are described, is divided into two subsections: 3.2.1 Satellite Data, and 3.2.2 Ground Truth Data.

[2]Connection to Product Datasheet of the Montemor-o-Velho Cooperative

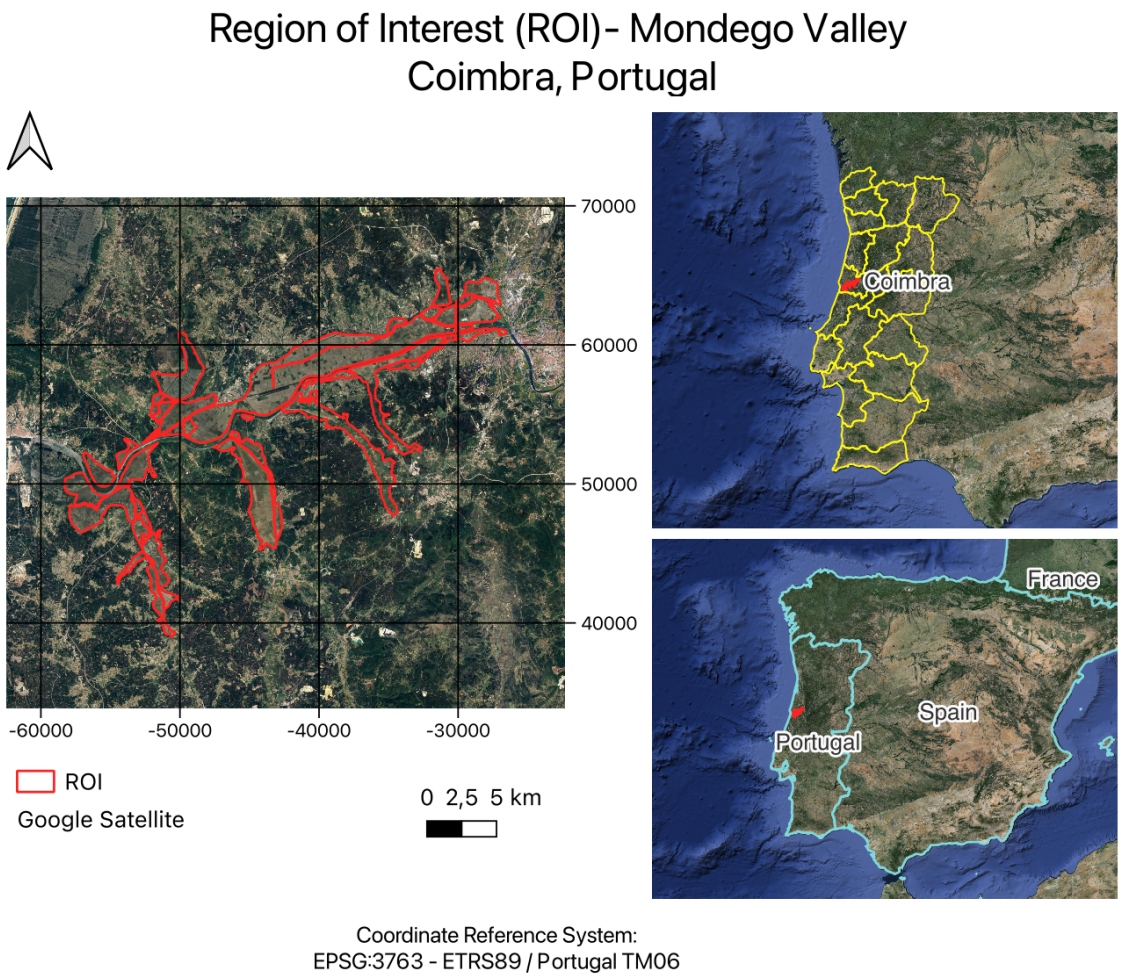

Figure 3.1: Geographic Framework of the Region of Interest

3.2.1 Satellite Data

3.2.1.1 Description of Satellite Data

Backscatter Data

Given the moderate density found in rice and maize crops, C-band SAR float data from Sentinel-1 were used to balance depth and sensitivity. Since the leaves of rice and maize crops have a narrow and vertical configuration, and broad and more robust, respectively, to map this geometric difference, a configuration of cross-polarizations VV and HV was utilized. This configuration enhances the polarimetric information of the models, as VV polarization (Vertical Transmit, Vertical Receive) is sensitive to surface roughness, capturing information about the density of objects, and VH polarization (Vertical Transmit, Horizontal Receive) is sensitive to orientation and vertical structure. The acquisition of Sentinel-1 data was done in IW mode (Interferometric Wide swath), which balances resolution and coverage.

Multispectral Data

Regarding spectral data, bands were used in regions where vegetation shows significant reflectance—Visible (400-700nm), Near-Infrared (700-1300nm), and Short-Wave Infrared (1300-2500nm), which in Sentinel-2 correspond to bands B2 (Blue), B3 (Green), B4 (Red), B8 (NIR), B11 (SWIR), and B12 (SWIR).

In addition to the bands, two spectral indices (Normalized Difference Vegetation Index (NDVI) and Normalized Difference Water Index (NDWI)) were used because they allow for the analysis of the presence and condition of vegetation, and the moisture content of vegetation, respectively, variables that make them useful in distinguishing agricultural crops. Although these are calculated by combining visible and near-infrared bands, their combination with the bands has been a recurrent practice as seen in section 2.4.4.

To reduce complexity in the data processing stage, Sentinel-2 data processed at SR Harmonized Level (L2A) with a cloud cover of less than 10%.

3.2.1.2 Sampling Interval of the Satellite Data Time Series

The sampling interval of the satellite data time series is a critical issue in the early mapping of agricultural crops using satellite data and artificial intelligence due to the trade-off of the "curse of dimensionality,". The curse of dimensionality also known as the Hughes effect describes the deterioration of classifier performance when the number of features is increased beyond an optimal value for a fixed training set size.

On one hand, shorter intervals that allow for a finer temporal series capture subtle temporal variations between crops and, theoretically, maximize classification accuracy. However, in models that do not alter the nature of the data during the learning phase (such as SVM and RF) and seek deep complex relationships as deep neural networks do due to their deep architecture, shorter intervals may introduce noise to the data due to the informative redundancy of features.

On the other hand, as demonstrated by Yi et al. (2022), longer intervals may fail to capture important details and delay early mapping, thus necessitating a balance between maximizing classification accuracy and reducing noise.

In fact, the findings of Yi et al. (2022) indicated that the optimal results for classical machine learning models (SVM and RF) were achieved with a 10-day sampling interval, utilising data interpolation to address the issue of missing observations. This represented the optimal trade-off between model complexity and temporal detail.

Therefore, for this research, a 10-day satellite data sampling interval was used within the bounds of the spring-summer agricultural season, which according to the Regional Directorate of Agriculture and Fisheries of Central Portugal, in a specialized consultation, runs from April/May to September/October.

3.2.2 Ground Truth Data

These data correspond to the ground truth obtained by delineating and monitoring 31 plots in the Baixo Mondego from May to September. The delineation was performed using two GNSS receivers (Geomax Zenith 10 from GeoMax and the rtkREP—a homemade Global Navigation Satellite System (GNSS) receiver), both in Real Time Kinematic (RTK) mode, with corrections from the observations of the Montemor Correction Station of the ReNEP network (a public geoprocessing service in Portugal, provided by the Directorate General of Territory) which allowed for achieving positioning accuracy of 2 centimeters in the coordinate systems EPSG: 3763 and EPSG: 25829 for Zenith 10 and rtkREP, respectively [3]

These plots, sampled according to their prevalence in the field, are described in the table that follows:

Table 3.1: Description of Monitored Crops

Description of the Crop	Number of Monitored Plots
Maize	11
Rice	12
Potatoes	3
Chestnut Trees	1
Nurseries	2
Bare Soil	2

In the Figure 3.2, the general scheme for acquiring satellite and field data throughout the spring-summer agricultural season in the Baixo-Mondego valley is summarized.

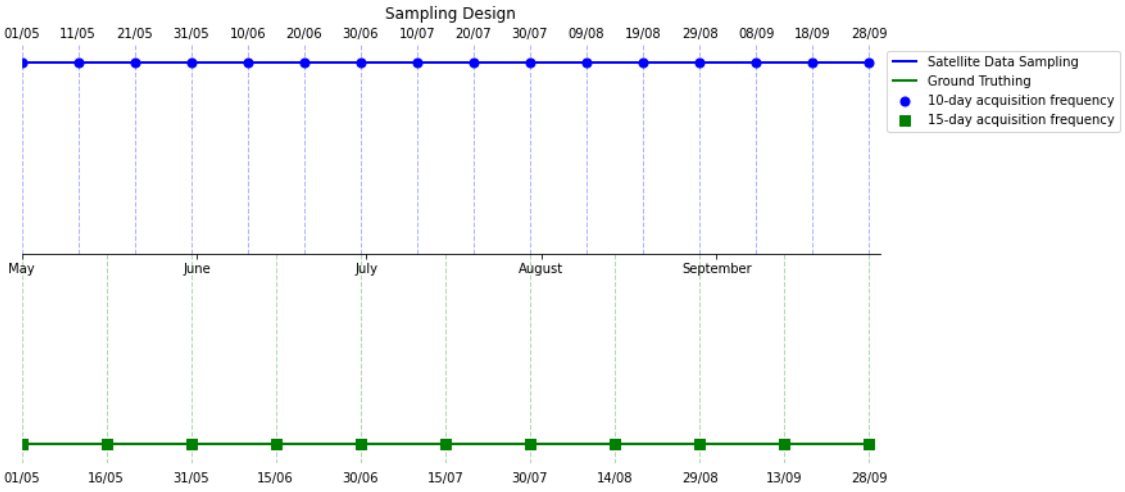

Figure 3.2: Sample Acquisition Scheme for Model Training

[3] See appendix D.1

Chapter 4

Methodology

This chapter describes the methodology applied in this study. The chapter is divided into seven sections: 4.1. Overview of the Methodological Framework; 4.2. Feature Extraction; 4.3. Building the Reference Database; 4.4. Feature Selection; 4.5. Model Development; 4.6. Post-Classification with Stratified Random Sampling; 4.7. Implementation: System Architecture.

4.1 Overview of the Methodological Framework

Aligned with the best practices for automation of crop mapping from satellite data, identified in section 2.4.6, and the research objectives outlined in section 1.3, an incremental classification methodology of satellite data was applied throughout the spring-summer agricultural season in the Baixo-Mondego Valley.

In each iteration, this methodology involved the use of geospatial intelligence and supervised machine learning for training classification models with the RF (Random Forest) and SVM (Support Vector Machine) algorithms. Thus, for each 10-day throughout the spring-summer agricultural season, each model was trained according to two approaches:

1. **Approach A** - involved supervised classification with time series of fused data from Sentinel-1 and Sentinel-2 satellites;
2. **Approach B** - employed supervised classification with time series of exclusive data from the Sentinel-2 satellite. The term "incremental" refers to the iterative growth of 10 days in the length of the time series data in both approaches.

From May to September, this methodology permitted the training of 15 models across two classification lines (RF Classification Line and SVM Classification Line). In each classification line, each model was trained on one version based on Approach A with fused data and another version based on Approach B with only optical data.

To determine the most effective features for training each model at each 10-day interval of the agricultural season, feature selection with the Recursive Feature Elimination with Cross Validation (RFECV) algorithm was employed. Subsequently, the performance of each model was evaluated in terms of overall accuracy and F1-score. Consequently, the identification of the early mapping model entailed the determination of the model that minimizes the time at which performance reaches an acceptable level (set at 90% in this study), as demonstrated in Eq. 4.1.

Once an performance of 90% or above had been achieved in the initial stages of the season, a post-classification was conducted using the IFAP monitoring map as a reference. This allowed for the evaluation of the model's performance on a sample that reflected the actual distribution of classes in the data.

$$t_{\min} = \min\{t \mid P(t) \geq 0.9\} \tag{4.1}$$

with

- t time in units of 10 days;
- P is model performance.

This methodology permitted the identification of the early stage at which accuracy reached or exceeded 90%, the mapping of the significance of features derived from satellites throughout the agricultural season, and the assessment of the contribution of SAR data in optimizing early mapping.

An overview illustration is presented in the flowchart in Figure 4.1. [4]

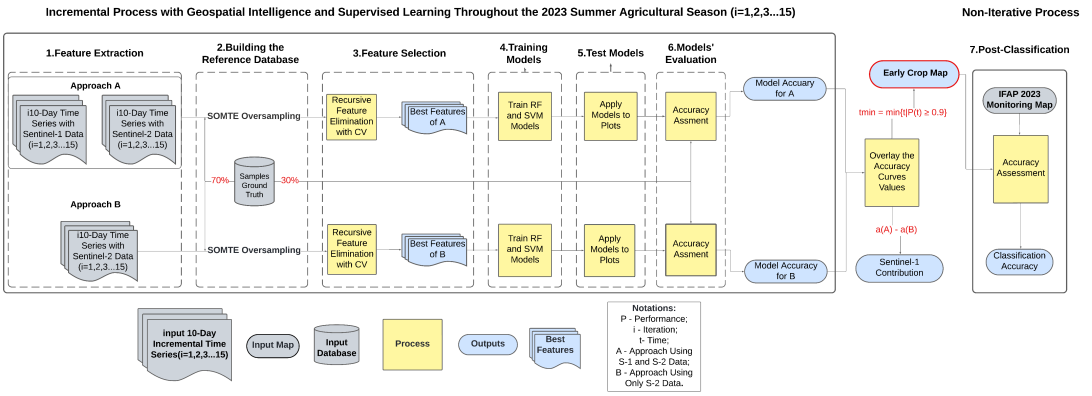

Figure 4.1: Flowchart of the Methodological Framework. The letters 'i', 'a', A, and B represent the number of classifications, accuracy, and accuracy for approach A, end accuracy for approach B , respectively.

The following section provides a detailed account of the specific elements pertaining to each stage of the methodology, as outlined in the preceding sections. These

[4] **Tap Here to See the Flowchart in a Larger Scale.** A landscape visualization of this flowchart is available in the Appendix B.

stages include feature extraction, the construction of the reference database, feature selection, the training of models, the testing of models, the evaluation of models, and post-classification with stratified random sampling.

4.2 Feature Extraction

The stage involved the filtration and reduction of data from the Sentinel-1 and Sentinel-2 satellites, as detailed below.

4.2.0.1 Filtering

In this task, an iterative temporal filter was applied to Sentinel-1 and Sentinel-2 data, incremented every 10 days, in conjunction with a spatial filter that restricted the search to the study area. Consequently, each iteration yielded as many Sentinel-1 and Sentinel-2 images as were available during the preceding 10-day period and the subsequent periods. The overlap in data has the advantage of densifying the time series and capturing detailed temporal variations in the crops, thereby enhancing the richness of information available for model training.

A detailed account of the number of Sentinel-1 and Sentinel-2 images utilised in each iteration and the respective capture dates of each image can be found in Data Extraction Report in [5].

4.2.0.2 Reduction

In order to mitigate the presence of redundant features and outliers in the data, a reduction was applied to the filtered data every 10 days. This involved calculating the median of each band individually based on the satellite observations in that band over the 10-day period.

Consequently, the data structure at each time point in the time series corresponds to a multi-band image comprising the medians of the optical bands, radar bands, and spectral indices (described in Section 3.2), sampled every 10 days. This reduction achieved a balance between the robustness and detail of the time series data.

4.3 Building the Reference Database

4.3.1 Determination of Sample Size

In this task, random sampling of points within the reference plots described in section 3.1.2 was conducted, with each point being assigned to the class of the

[5] See Appendix E

surrounding plot. Since the number of plots for each crop type in the field was sampled according to their prevalence, this sampling incorporates a stratified design by class and preserves the probability of their occurrences in the field.

The number of samples per plot was determined based on the convergence calculation of classification accuracy with data corresponding to the maximum duration of the time series (150 days). Four options for the number of points per plot (10, 20, 30, 50) were tested, and it was found that while increasing the number of points from 20 per plot increases computational cost, it does not enhance the accuracy of the models. Therefore, considering the computational power of the Google Earth Engine (GEE), a value of 50 points per plot was adopted, resulting in a total of 1.550 reference points. This testing with data corresponding to the maximum length of the time series (150 days) ensures that none of the classifications with shorter temporal sub-series suffer from decreased accuracy due to an insufficient number of samples.

Consequently, the distribution of the 1,550 reference points across each class is presented in Table 4.1 and can be seen in Figure 4.2.

Table 4.1: Imbalanced Sampling (Sample Points Stratified by Each Class)

Description of the Crop	**Number of Sampled Points**
Maize	550
Rice	600
Potatoes	150
Chestnut Trees	50
Nurseries	100
Bare Soil	100

4.3.2 Splitting Data into Training and Test Sets

Given the possibility of spatially correlated samples within the same plot, which could bias the models, a method was introduced to randomize the data by shuffling the samples. This was accomplished by incorporating a column of randomly generated values, ranging from 0 to 1, with 1 representing an exclusive value. Based on the column of random values added in the previous step, the reference database was divided into two partitions: a training partition (values less than 0.70) and a test partition (values greater than 0.70).

4.3.3 Oversampling of Training Data Partition

With the objective of reducing model bias (better learning of the majority classes), giving the model the opportunity to learn the patterns of all classes, thereby maximizing its generalization capability on new data, oversampling was applied to the training set using the Synthetic Minority Over-sampling Technique (SMOTE)

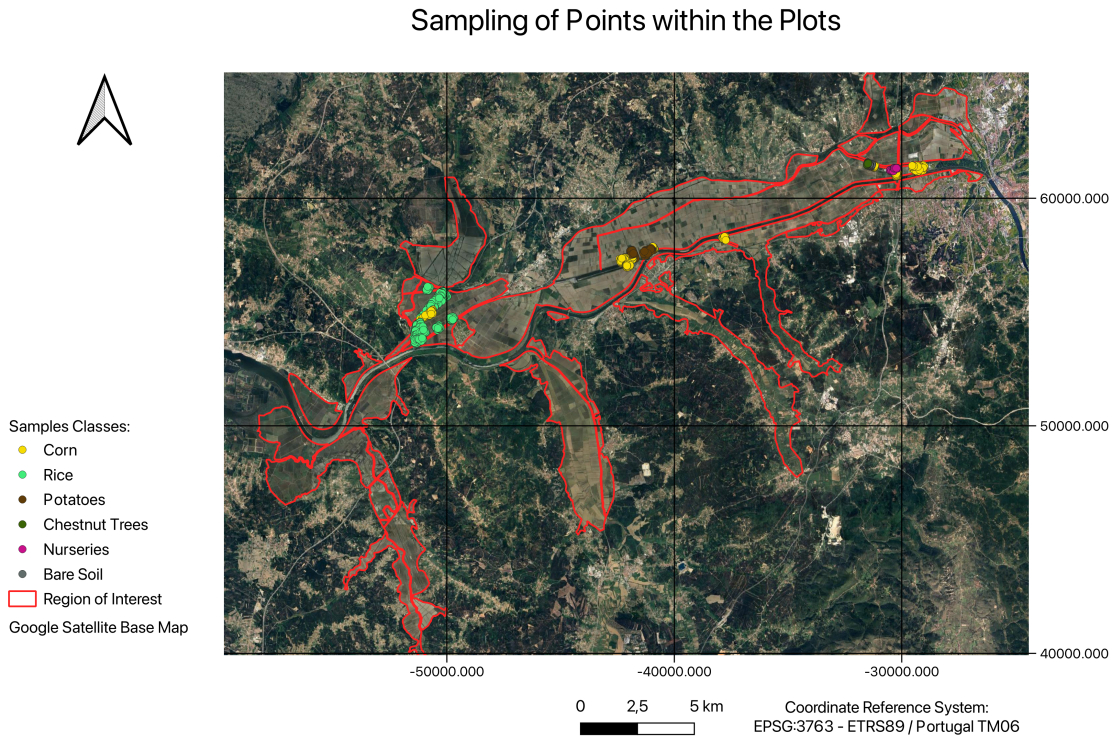

Figure 4.2: Sampling Points In Plots

algorithm from the Python imbalanced-learn library [6].

The logic of SMOTE lies in using linear interpolation to create synthetic points for the minority classes iteratively until the dataset is balanced. For each point in the minority class (x), its k-nearest neighbors belonging to the same class in the feature space are identified. For a new synthetic point (xs) in the feature space, the distance between x and one of the chosen nearest neighbors is calculated. A random number between 0-1 is used, and the synthetic point (xs) is calculated based on the following expression:

$$x_s = x + \lambda(x_{\text{neighbor}} - x) \tag{4.2}$$

Thus, a k of 5 was defined, and from the application of SMOTE, a balanced training dataset with 3624 training points was obtained (Table 4.2).

The objective of this task was to create a database for training and evaluating models. Each instance in the database is a vector of satellite features associated with a label (agricultural land use class). It is essential to clarify that each value in this feature vector is not an observation from a single band. Rather, it represents the median of reflectance or backscatter values observed in a band over a 10-day period.

[6]Oversampling was applied only to the training data to avoid affecting the real probability occurrence of classes in the test data, which reflects the reality of the study area.

Table 4.2: Balanced Sampling (Sample Points Balanced by Each Class+ Synthetic Points)

Description of the Crop	Number of Sampled Points
Maize	604
Rice	604
Potatoes	604
Chestnut Trees	604
Nurseries	604
Bare Soil	604

4.4 Feature Selection

Feature selection was conducted using the RFECV algorithm from the scikit-learn library. The algorithm employed an RF or SVM classifier to select the optimal features from the time series in each approach, as illustrated in Figure 4.3 [7].

The rationale behind this algorithmic approach was to generate a ranking of the relative importance of individual features within the initial dataset, and then to assess the performance of the model using this initial set of features.

Subsequently, the features were removed one by one in a recursive manner, and the impact on the model's performance was evaluated. If the removal of a feature resulted in a decline in model performance, that feature was included in the optimal feature set. In the event that the feature did not meet the aforementioned criteria, it was eliminated. The following image presents this approach.

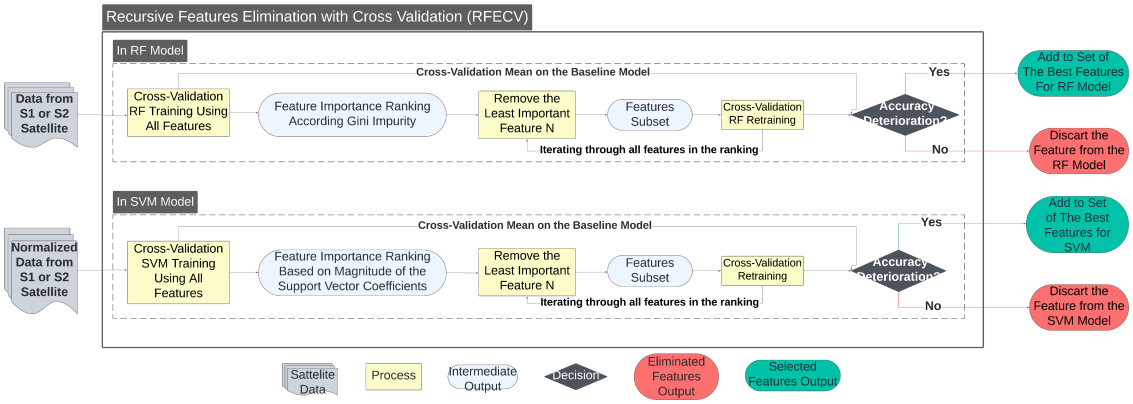

Figure 4.3: Feature Selection Using the RFECV Algorithm

In the case of RF, the feature importance was based on Gini Importance (that expresses how well each feature separates the data relative to the target classes. This information is derived from quantifying the reduction in impurity in the bootstrap data whenever a feature is selected during the feature bagging process). In contrast, for SVM, which does not provide a direct measure of importance, the magnitude and direction of each feature's contribution in determining the hyper-

[7] **Tap Here to See the Flowchart in a Larger Scale**

planes that separate the classes in the data were used. The principal advantage of this technique is that it combines three distinct analyses.

The analysis of feature importance was conducted on an individual basis and within the context of the entire dataset. Additionally, the evaluation of feature correlation and the impact on model performance were considered.

4.5 Model Development

This subsection explains the model's training and testing setup and performance evaluation metrics.

4.5.1 Training and Testing Setup

The data from the reference database were divided into training and test sets with a ratio of 70% and 30%, respectively. This ratio is applied to the column of random values ranging from 0-1 that was added during the reference database construction process.

4.5.2 Model Training

In each approach, each model was trained on the GEE with the best feature set for a time t of the agricultural season. The training parameters described in Section 2.4.7 were computed using the randomized search algorithm from the scikit-learn library distributed in Python.

As in the feature selection process, for the models trained with the SVM algorithm, a Z-score normalization was applied to the training partition, where the S-1 and S-2 data were normalized independently.

Consequently, 30% was used to evaluate the performance of the model. The test data for the SVM models were normalized separately from the training data. This approach reduces what is known in the machine learning literature as data leakage.

4.5.3 Model Evaluation Metrics

Each model was evaluated using the Accuracy and F1 Score metrics. The accuracy metric is calculated according to the following expression:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{4.3}$$

where:

- TP means True Positive;
- TN means True Negative;
- FP means False Positive;
- FN means False Negative.

The F1-Score metric is calculated according to the following expression:

$$F1\text{-Score} = 2 \times \left(\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \right) \tag{4.4}$$

where precision and recall is given by the following equations:

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}} \tag{4.5}$$

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}} \tag{4.6}$$

4.6 Post-Classification with Stratified Random Sampling

In this phase, a post-classification process was performed on the most accurate early map to evaluate its accuracy across all parcels in the study area. The sampling and validation strategy used is detailed below.

4.6.1 Post-Classification Sampling Design

Accordingly, the overall expected standard error parameter was defined in the AcATaMa plugin of QGIS with a default value of 0.0050. This parameter facilitated the generation of 396 randomly stratified points across the three classes throughout the entire Baixo Mondego valley area, thereby enabling the early agricultural map to be validated with 95% confidence (z-score = 1.96). The number of points per class is presented in the table below.

Table 4.3: Description of Crop Statistics

Class	Number of Samples	Standard Deviation
Maize	193	0.01
Rice	188	0.01
Other Occupations	15	0.01

As stated in Fonte et al. (2024), this stratified sampling design considers the diversity of the landscape and the probability of occurrence of the classes, is cost-effective and practical due to the reduction in the number of sampling points, and allows for the collection of data from samples in rare classes.

4.6.2 Post-Classification Response Design

As a reference, the 2023 agricultural land monitoring map from Institute of Financing for Agriculture and Fisheries, Public Institute (Portugal) (IFAP) was employed, which was derived through a supervised classification of Sentinel- 2 images and corrected using information from field visits and geo-referenced photographic evidence provided by farmers. The AcATaMa plugin of QGIS was employed to cross-reference the data from the most accurate early classification map and the IFAP agricultural land monitoring map, thereby constructing a confusion matrix and calculating producer and user accuracies.

4.7 Implementation: System Architecture

At the architectural level, there are three components: Google Earth Engine (GEE), Colab, and QGIS. The latter is the only platform accessed on a local machine and is used solely for the creation of points with the crop classes. However, the majority of the processing is carried out through scripting in Python and JavaScript on GEE and Colab, both of which are part of Google LLC's suite of cloud computing services.

Below, the relationship between these three components is described in the Figure 4.4 below:

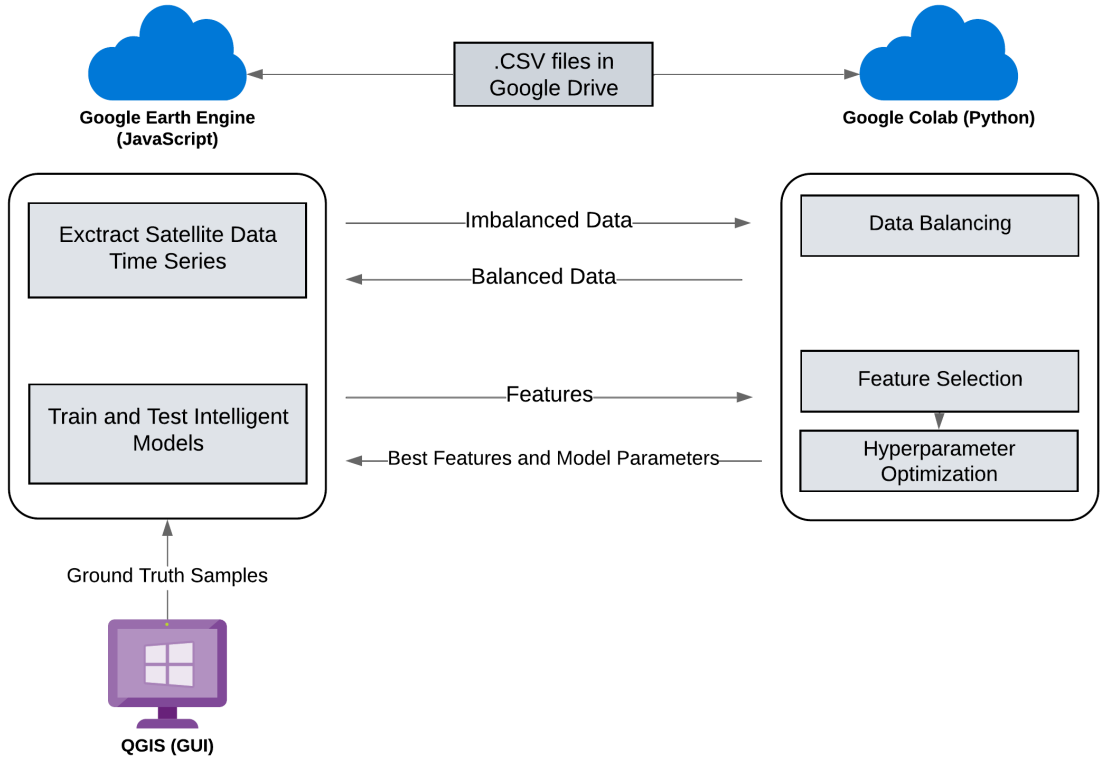

Figure 4.4: Classification System Architecture

The GEE, designed for cloud analysis and visualization of large amounts of geospatial data, is the primary component of the system where satellite data extraction, filtering, reduction, and the training and validation of intelligent models are developed. The system then receives the points from the stratified sampling within the plots, produced in QGIS, through an upload of a shapefile (which must be zipped). Once the satellite data has been collected, it is exported to Google Drive in the form of a CSV file, thereby making it accessible from Colab, a cloud-based Python notebook. Two essential functions of the system that are not available in GEE can be applied in Colab: oversampling of data with synthetic points and hyperparameter estimation, using the imbalanced-learn and scikit-learn libraries. This integration, illustrated in Figure 4.5 exploits the processing capabilities of Google Earth Engine and the adaptability of these libraries.

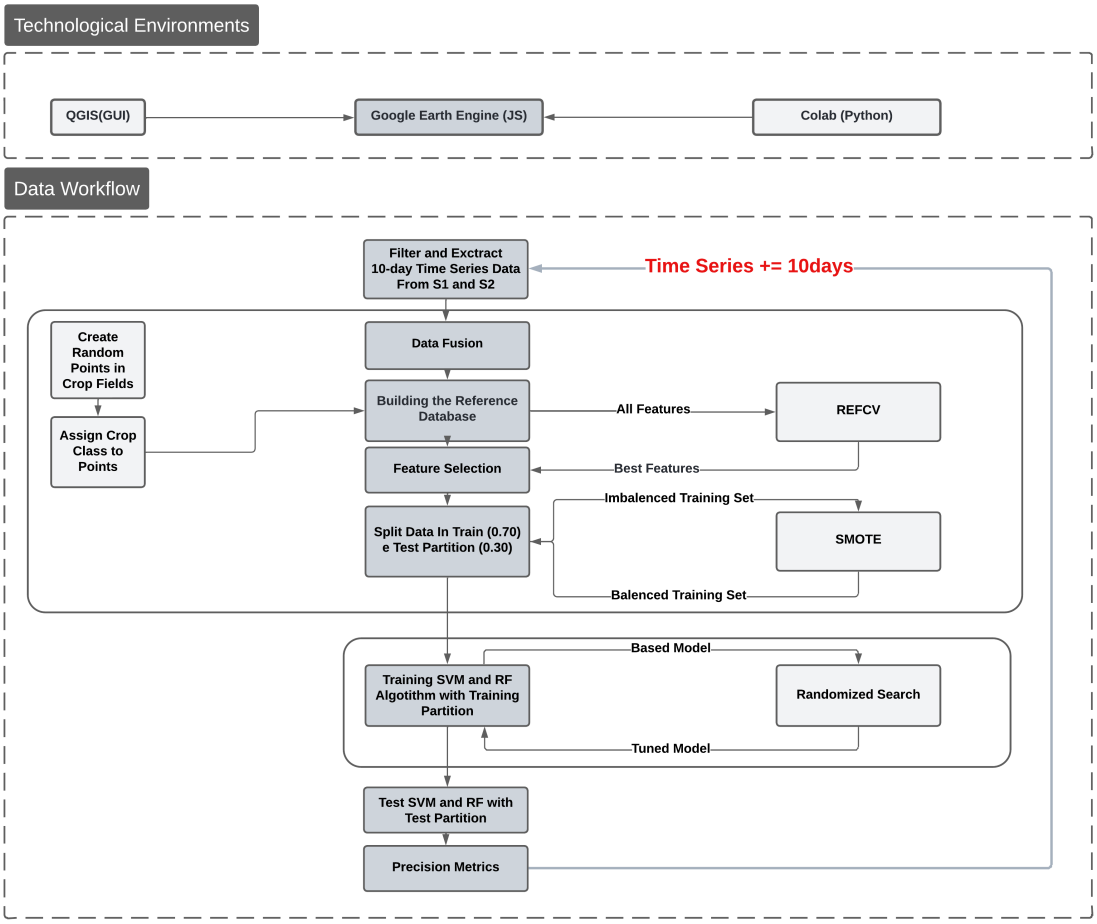

Figure 4.5: Data workflow

Chapter 5

Results

5.1 Proposed Approach

The presentation of the results in this subsection is divided into 5.1.1. Feature Selection Results; 5.1.2. Model Performance Over Time, 5.1.3. Early Crop Mapping Identification, and 5.1.4. Post-Classification Results.

5.1.1 Feature Selection Outcomes

Before presenting these results, it is important to clarify that the information chosen to train the models was extracted from features (observations in each satellite band over a given time period). To make the analysis easier and to assess the usefulness of each satellite band, the results are grouped together and presented based on each individual band.

In essence, the key to understanding these results has to do with the frequency and number of times each type of data is selected for the set of best features, such that the more frequently a type of data is selected, the more efficient it is for crop mapping.

This information is presented for each algorithm used in RFECV.

5.1.1.1 Feature Effectiveness Using Random Forest Models

In this section the average graphs of Gini importance present the univariate contribution of features from each band in the automatic classification of crops.

As the Gini Importance values of each feature (observation in each band) are originally normalized between 0 and 1, where values close to 0 represent higher informative power and values close to 1 represent lower informative power, the normalization of the observations from each band by the total number of observations, despite reducing the scale of the values, preserves the original interpretation of the Gini Importance.

The impact graphs on model performance in this section illustrate the mean variation in model performance when the features from each band were either selected or excluded during the training of RF models. The performance evaluation metric employed in RECV was classification accuracy. Given that RFECV was applied to the training set with balanced data, the evaluation metric was also classification accuracy, without prejudice to potential data bias towards the majority classes.

The frequency with which features from each band are selected by RFECV indicates a balance between three factors:

1. The individual informative power of the band;
2. How the band relates to other bands;
3. The overall impact of the band on the classification accuracy.

In simpler terms, if a band is selected frequently, it means that it provides useful information on its own, works well in combination with other bands, and significantly improves the model's accuracy when used. This frequent selection robustly represents the overall importance and contribution of that band when using the Random Forest algorithm.

Radar Features

The results of the multivariate analysis (Figure 5.1) of radar features derived from the Sentinel-1 satellite using the Random Forest algorithm indicated that the VV and VH polarizations exhibited Gini importance scores of 0.065 and 0.066, respectively. This result suggests that the univariate contribution of these bands to the mapping of agricultural crops is highly comparable.

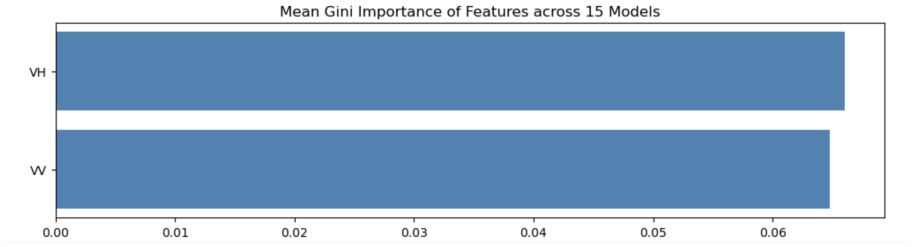

Figure 5.1: Univariate Contribution of Radar Features in RF Models Based on Gini Importance

However, the results of the impact on model performance (Figure 5.2) demonstrated that the impact caused by the introduction of VH polarization features was significantly greater than the impact caused by the introduction of VV polarization features, with average Gini importance scores of 0.042 and 0.006, respectively.

These graphs also demonstrated that when the VV features are not selected, there is no significant decay in model performance.

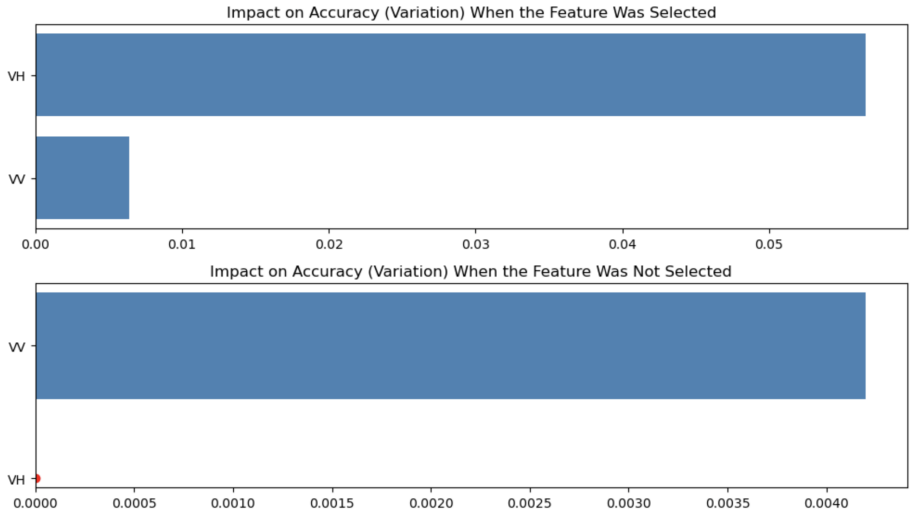

Figure 5.2: Impact of Radar Features on RF Performance Model (Red Dot Represents a Value Equal to or Below 0)

The balance of these two scenarios is summarized by the number of times each radar feature was selected or discarded from the dataset optimized by RFECV using the RF algorithm, as illustrated in Figure 5.3. These results demonstrate that VH polarization features were chosen 107 times for the set of best features, while VV features were chosen 100 times.

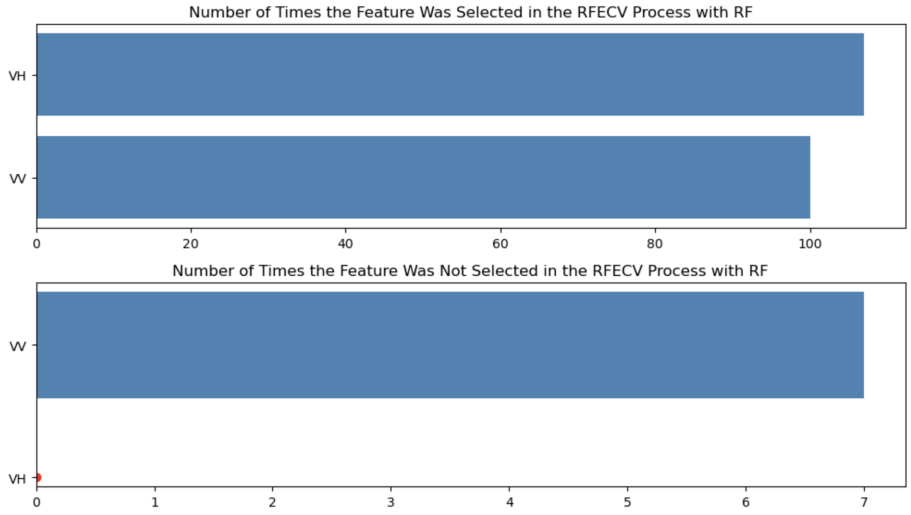

Figure 5.3: Ranking of Radar Feature Effectiveness for RF Models Based on the Number of Times Each Band's Features are Selected or Not Selected in the RFECV Process (Red Dot Represents a Value Equal to 0)

Optical Features

The univariate importance analysis (Figure 5.4) revealed that the univariate informative power of the features of each band decreases in the following sequence: B8 (0.0510), NDVI (0.04142), NDWI (0.0383), B3 (0.0259), B4 (0.0254), B2 (0.0189), B1 (0.0182), B12 (0.0131), and B11 (0.0089).

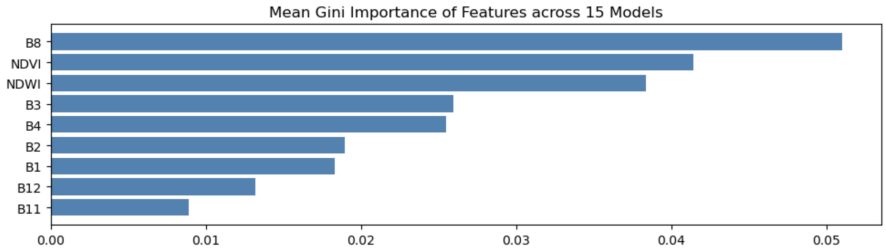

Figure 5.4: Univariate Contribution of Optical Features in RF Models Based on Gini Importance

The graphs of impact on model performance (Figure 5.5) demonstrated that the sequence from the features of bands with the greatest impact to those with the least impact can be defined as follows: The results indicated that the following bands exhibited the greatest impact on model performance: B11 (0.067), B12 (0.0374), B1 (0.0064), B2 (0.0020), B3 (0.0011), B8 (0.0004), NDVI (0.0003), B4 (0.0001), and NDWI (-0.000050393).

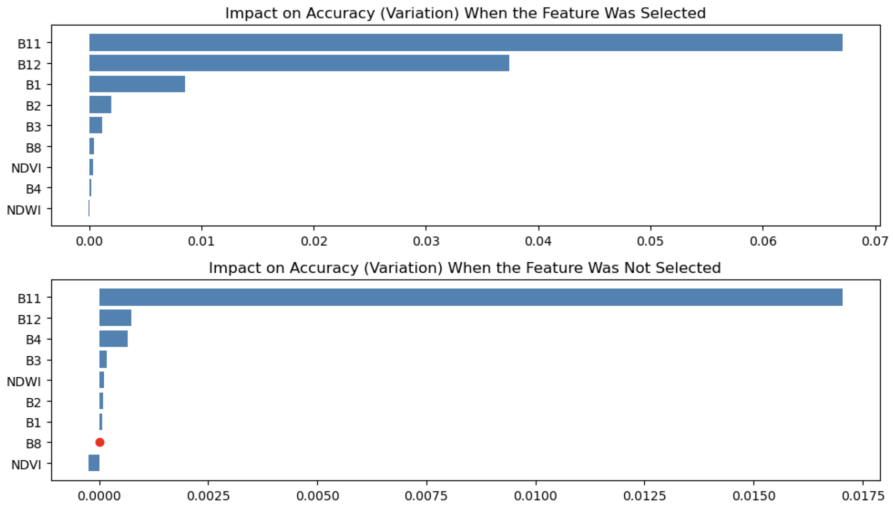

Figure 5.5: Impact of Optical Features on RF Performance Models (Red Dot Represents a Value Equal to or Below 0)

However, the ranking of the efficiency of the optical band features (Figure 5.6) revealed that the sequence from the most effective to the least effective features is as follows: The following order of efficiency was observed for the optical band features: B8 (56), NDWI (48), B3 (45), B11 (40), NDVI (39), B1 (33), B2 (33), B12 (33), B4 (30).

The discrepancy between the ranking sequences, which consider only the univariate relationship (Figure 5.4) and those that take into account the multivariate relationships (Figure 5.6), underscores the significance of striking a balance between the individual ability of each attribute to discriminate data and its impact on the performance of the classification model when it interacts with other attributes in the dataset.

Moreover, the selection of attributes through RFECV has the advantage of being carried out within the specific context of the machine learning algorithm, result-

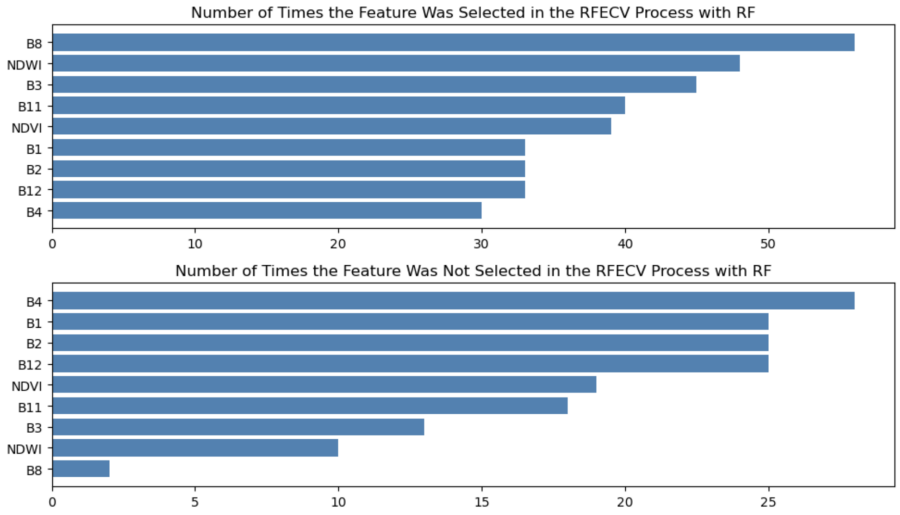

Figure 5.6: Ranking of Optical Feature Effectiveness for RF Models Based on the Number of Times Each Band's Features are Selected or Not Selected in the RFECV Process

ing in outputs whose interpretation is more intuitive compared to dimensional reduction techniques such as principal component analysis Principal Component Analysis (PCA). It is also important to highlight that this method incorporates cross-validation, providing more robust and reliable solutions.

5.1.1.2 Feature Effectiveness in Support Vector Machine Models

In this subsection, the importance graphs illustrate the univariate contribution of each band's features in defining the hyperplane that separates the classes throughout the classification process. This value is extracted from the average of the sum of the coefficients of each feature in the support vectors. These coefficients were normalized by the number of observations (features) of each data type (radar and optical).

The absolute values resulting from normalization by the number of observations represent the magnitude (contribution of each band) in solving the classification problem. This is because the vectors represented by the coefficients indicate the weight of each feature in defining the hyperplanes that separate each binary problem in the multi-class classification scenario.

The impact graphs on the model's performance in this section represent the average variation in model accuracy whenever the features of each band were included or excluded from the training of the SVM models.

Finally, as with the feature selection process using RF, the number of times the features of each band were selected represents a balance between the individual informative capacity of this band, its relationship with other bands, and its impact on classification. This generates a robust feature selection process.

Radar Features

The univariate contribution of radar features, as illustrated in Figure 5.7, revealed that radar features with vertical polarization (VV) exhibited a slightly higher univariate contribution to the classification of agricultural crops compared to radar features with horizontal polarization (VH).

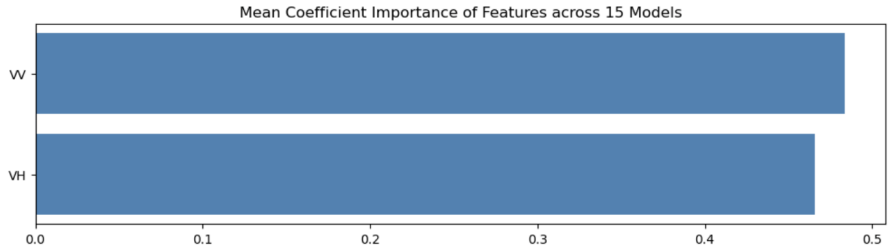

Figure 5.7: Univariate Contribution of Radar Features in SVM Models Based on the Mean Value of the Support Vector Coefficients (Magnitude)

This result is contrary to the result obtained by the RF algorithm. The discrepancy can be attributed to the differing approaches employed by each algorithm in data exploration.

However, the impact on model accuracy (Figure 5.8) aligns with the findings of the RF algorithm, with features with VH polarization (0.0418) demonstrating a greater influence on model performance than features with VV polarization (0.0074) when integrated into the SVM model.

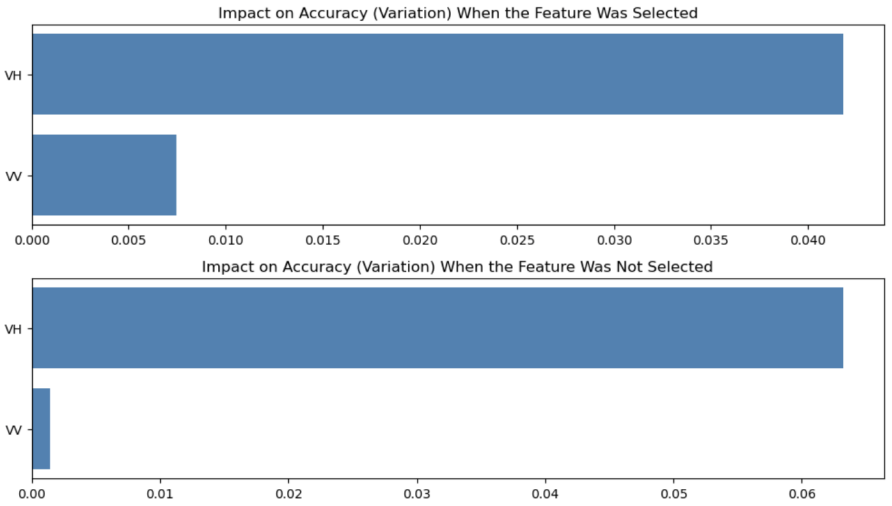

Figure 5.8: Impact of Radar Features on SVM Models

Consequently, the efficacy results (Figure 5.9) of the radar features between the two algorithms also converge, highlighting cross-polarized radar features as the best radar features, with a selection frequency for the optimal feature set equal to 116, while VV-polarized radar features had a selection frequency equal to 112.

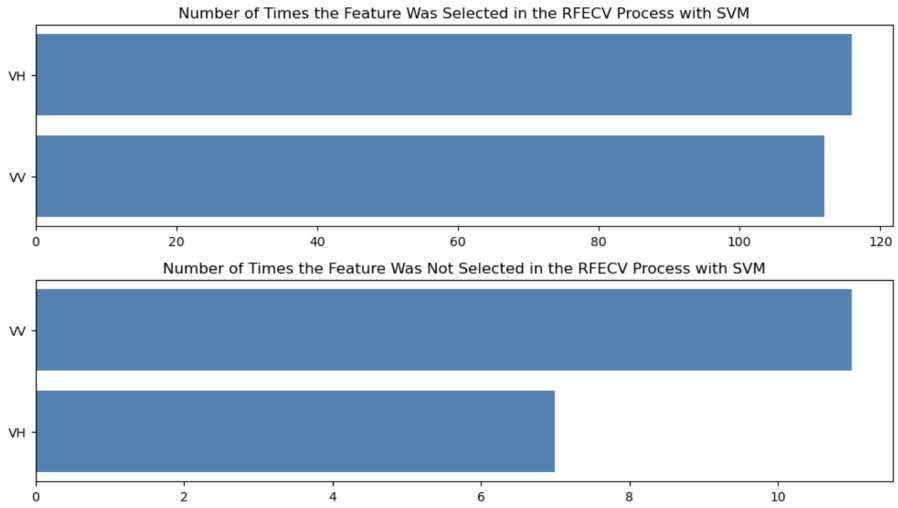

Figure 5.9: Ranking of Radar Feature Effectiveness for SVM Models Based on the Number of Times Each Band's Features are Selected or Not Selected in the RFECV Process

Optical Features

The univariate analysis of optical features (Figure 5.10) revealed that the univariate contribution ranking of features follows this sequence: B8, B11, B2, NDWI, B1, B3, B12, B4, and NDVI exhibited the highest scores, with values of 0.3117, 0.2616, 0.2142, 0.1725, 0.1445, 0.1419, 0.1295, and 0.0737, respectively.

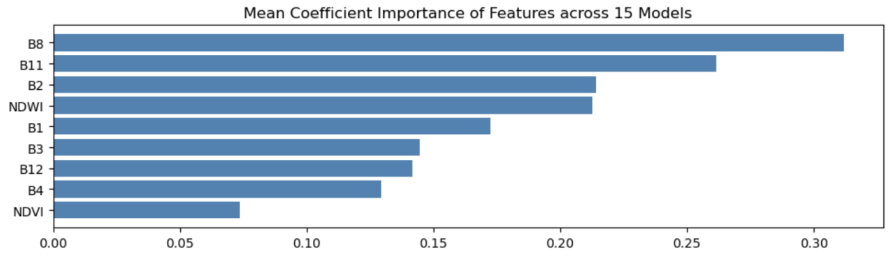

Figure 5.10: Univariate Contribution of Optical Features On SVM Models Based on the Mean Value of the Support Vector Coefficients (Magnitude)

The results of the impact on model performance (Figure 5.11) revealed contributions of 0.0610, 0.0457, 0.0140, 0.0042, 0.0038, 0.0030, and 0.0000. The optimal feature set was selected based on the following values: 0.0005 for B11, 0.005 for B12, 0.0140 for B1, 0.0042 for B2, 0.0038 for B3, 0.0030 for B4, 0.00 for B8, 7.7960e-05 for NDVI, and 7.7960e-05 for NDWI.

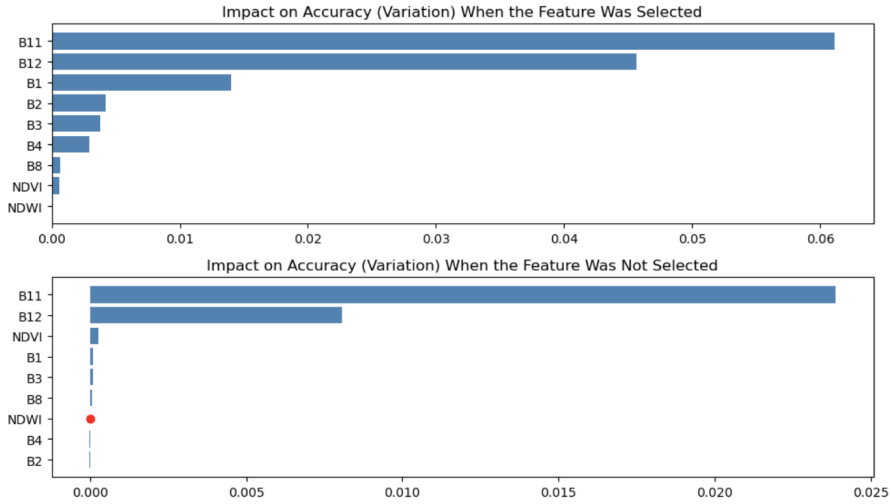

Figure 5.11: Impact of Optical Features on SVM Models Performance (Red Dot Represents a Value Equal to or Below 0)

These results align with those obtained in the RFECV with the RF algorithm, which demonstrated a greater impact of spectral band features, particularly bands B11 and B12, and a minimal impact of spectral indices.

Lastly, with regard to the efficacy of features in relation to other features, the results (Figure 5.12) align with the RF algorithm selection for bands B8 and B11, which were ranked in the top positions with a selection frequency of 57 and 43 times, respectively, for the optimal feature set. However, the results were equally noteworthy with the scores of the bands from the visible region of the electromagnetic spectrum exceeding those of the spectral indices.

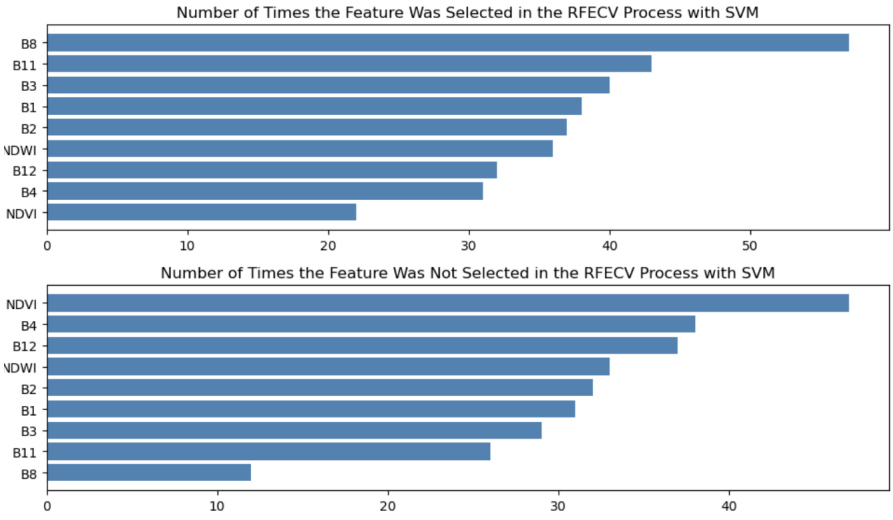

Figure 5.12: Ranking of Radar Feature Effectiveness for SVM Models Based on the Number of Times Each Band's Features are Selected or Not Selected in the RFECV Process

5.1.2 Model Performance Over Time

This section presents an analysis of the metrics (accuracy and F1-score) employed in the validation of RF and SVM models throughout the spring-summer agricultural season. Each metric is evaluated in accordance with the two approaches delineated in Section 4.1.

5.1.2.1 Performance Metrics in RF Models

Accuracy

The accuracy results of each model over time are presented in Table 5.1. The results are organized according to the approaches described in Section 4.1. This table also includes the respective dates associated with each model. These dates relate to the cumulative data intervals used for training each model.

Table 5.1: Accuracy of RF Models

Model ID	Data Time Range	Accuracy in A (%)	Accuracy in B (%)	A-B (%)
RF Model 1	May 01st - May 10th	74.49	70.13	4.36
RF Model 2	May 01st - May 20th	84.81	82.57	2.24
RF Model 3	May 01st - May 30th	86.21	82.39	3.82
RF Model 4	May 01st - Jun 10th	88.36	82.56	5.80
RF Model 5	May 01st - Jun 20th	91.45	85.17	6.28
RF Model 6	May 01st - Jun 30th	92.71	86.69	6.02
RF Model 7	May 01st - Jul 10th	95.47	89.11	6.36
RF Model 8	May 01st - Jul 20th	95.81	93.89	1.92
RF Model 9	May 01st - Jul 30th	96.49	96.11	0.38
RF Model 10	May 01st - Aug 10th	97.25	97.02	0.23
RF Model 11	May 01st - Aug 20th	98.12	98.01	0.11
RF Model 12	May 01st - Aug 30th	98.08	97.78	0.30
RF Model 13	May 01st - Sep 10th	97.61	96.54	1.07
RF Model 14	May 01st - Sep 20th	97.15	96.92	0.23
RF Model 15	May 01st - Sep 30th	96.64	95.21	1.43

The results indicate that the total number of correct predictions by the RF models increases from spring to summer. This is due to the maturation process of agricultural crops and the clarity of the spectral response from the targets.

Moreover, it was demonstrated in Figure 5.13 that models belonging to Approach A exhibited higher accuracy values at the outset of the season, with the number of correct predictions by models belonging to Approach B becoming comparable to those belonging to Approach A as of July 30.

According to this metric, the number of days required for classifications to reach a correct classification rate of 90% or higher is 30 days. This information can be verified in 5.1, where models 5 and 6, highlighted in gray, demonstrate that for Approach A, an accuracy of 91.45% is achieved by June 29, while Approach B

only reaches an accuracy level above 90% by July 20, at which point the accuracy values tend to converge.

However, it is important to note that the test set was not balanced and represents a scenario of unbalanced classes. Therefore, it is essential to exercise caution when using this metric for early mapping identification, as it may provide a false perception of accuracy in minority classes and bias the model toward majority classes. Nevertheless, the use of this metric is highly recommended in scenarios where reality reflects a balance in the occurrence of agricultural classes.

For these reasons, subsequent analyses will focus more on the F-score metric, which provides a balance between precision and recall, given that the reality of the Baixo Mondego presents two predominant classes.

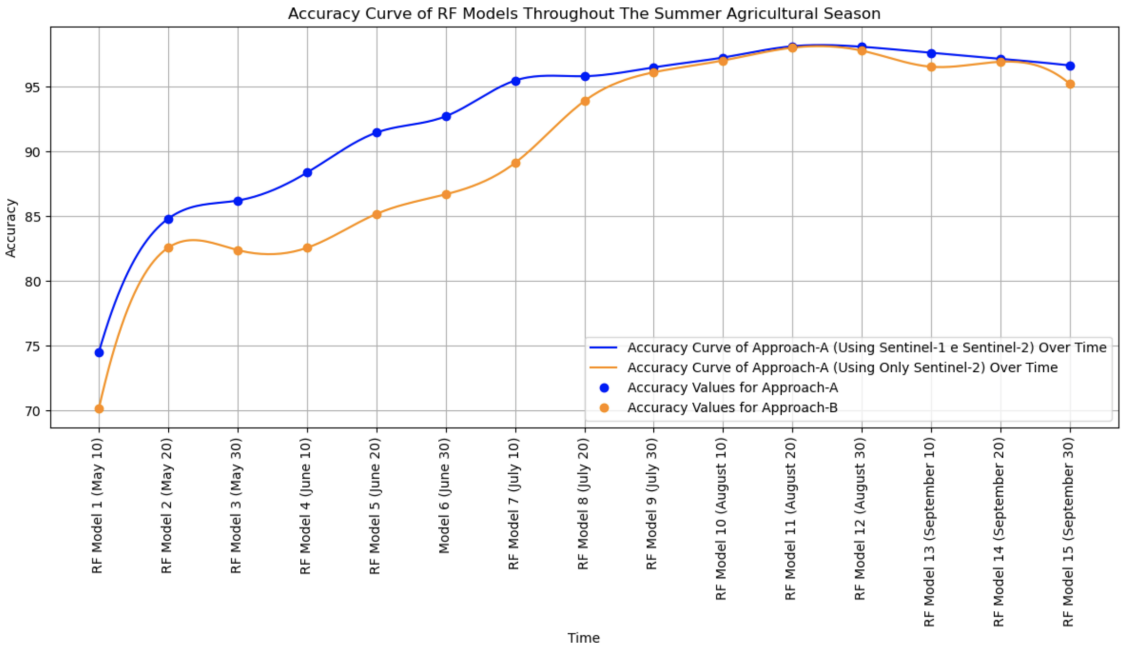

Figure 5.13: Evolution of RF Models' Accuracy Over Time

F1-Score

The results of this metric are described in Table 5.2.

According to these results, the greatest differences in the F1-score values of the models are observed at the beginning of the season, and there is a trend of convergence from spring to summer (Figure 5.14).

The fusion of radar and optical data was also found to result in an improvement in the balance between precision and sensitivity (recall).

This balance represents the model's ability to produce a classification that minimizes both false positives and false negatives simultaneously.

A 90% F1-Score was identified as an acceptable early value. It was found that Approach A minimized the time needed to achieve a value equal to or greater

Table 5.2: F1-Score of RF Models

Model ID	Data Time Range	F1-Score in A (%)	F1-Score in B (%)	A-B (%)
RF Model 1	May 01st - May 10th	72.2	67.28	4.92
RF Model 2	May 01st - May 20th	83.57	80.63	2.94
RF Model 3	May 01st - May 30th	84.28	79.91	4.37
RF Model 4	May 01st - Jun 10th	87.14	81.02	6.12
RF Model 5	May 01st - Jun 20th	89.36	83.1	6.26
RF Model 6	May 01st - Jun 30th	90.88	85.61	5.27
RF Model 7	May 01st - Jul 10th	93.5	87.53	5.97
RF Model 8	May 01st - Jul 20th	94.41	91.46	2.95
RF Model 9	May 01st - Jul 30th	93.91	94.7	-0.79
RF Model 10	May 01st - Aug 10th	95.96	95.44	0.52
RF Model 11	May 01st - Aug 20th	96.97	96.69	0.28
RF Model 12	May 01st - Aug 30th	96.94	95.1	1.84
RF Model 13	May 01st - Sep 10th	95.64	94.54	1.1
RF Model 14	May 01st - Sep 20th	95.53	95.4	0.13
RF Model 15	May 01st - Sep 30th	95.04	93.04	2.00

than 90%. In this approach, the model achieves an F1-score of 90.88% by June 30, whereas for Approach B, it reaches a score on July 20.

Figure 5.14 further illustrates that if the acceptable performance value for the model is reduced to an F1-score below 90%, with 85% as an example, the difference in the time needed to achieve this accuracy level in the two approaches tends to increase, being greater for the approach that uses only optical data.

5.1.2.2 Performance Metrics in SVM Models

Accuracy

The results of the model evaluation using the accuracy metric are presented in Table 5.3.

These results, when compared to those in Table 5.1, reveal that the classification line with SVM models takes longer to achieve an accuracy of 90% or higher.

While the classification line with RF models (Table 5.1) reaches this value for the first time in the fifth classification on June 20th (accuracy equal to 91.45%), the classification line with SVM algorithms only reaches it in the seventh classification on July 10th (94.31%).

Nevertheless, analogous to the classification with RF models, for the SVM models, there is also a 20-day discrepancy in the time required for the approach utilizing solely optical data to achieve an accuracy of 100%. The results are presented in Figure 5.15.

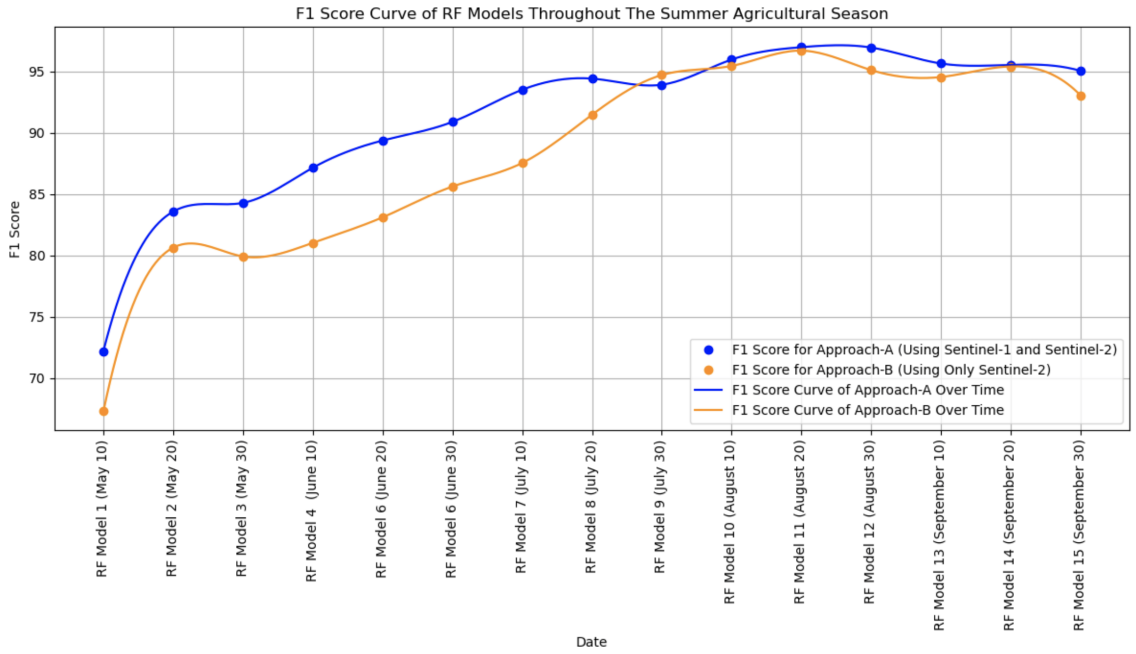

Figure 5.14: Evolution of RF Models' F1-Score Over Time

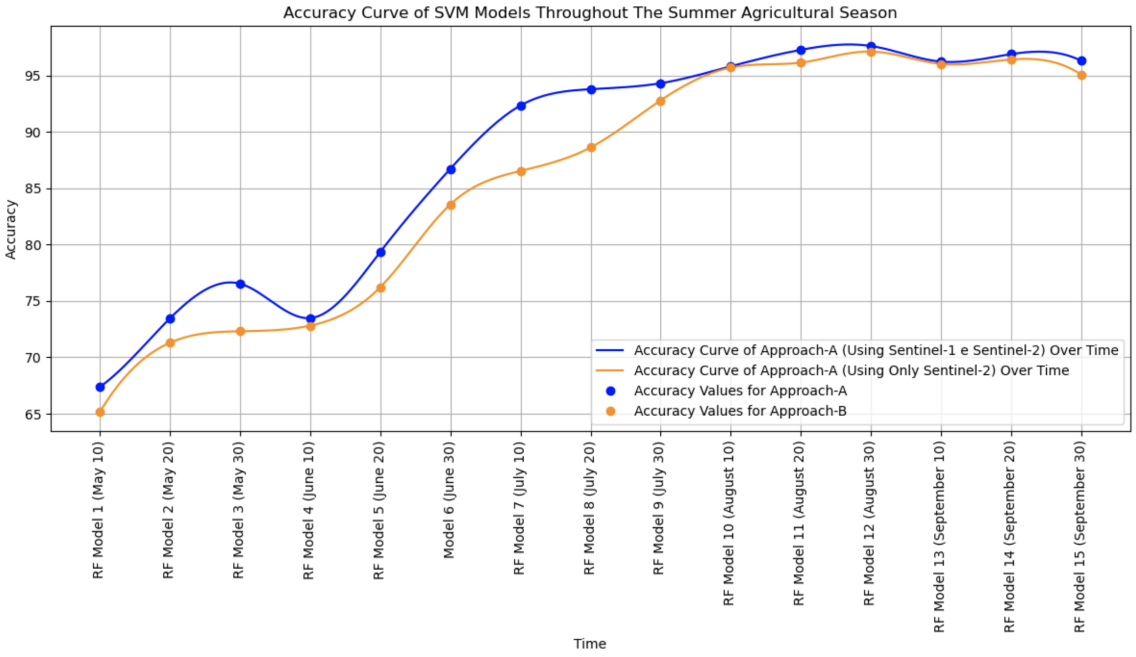

Figure 5.15: Evolution of SVM Models' Accuracy Over Time

F1-Score

The results of the model evaluation using the F1-Score metric are presented in Table 5.4. These results demonstrated that, in comparison to RF models (Table 5.2), the models require a longer time to achieve an F1-Score of 90% or higher.

Table 5.3: Accuracy of SVM Models

Model ID	Data Time Range	Accuracy in A (%)	Accuracy in B (%)	A-B (%)
SVM Model 1	May 01st- May 10th	67.37	65.13	2.24
SVM Model 2	May 01st - May 20th	73.45	71.29	2.16
SVM Model 3	May 01st - May 30th	76.53	72.31	4.22
SVM Model 4	May 01st - Jun 10th	73.48	72.81	0.67
SVM Model 5	May 01st - Jun 20th	79.34	76.21	3.13
SVM Model 6	May 01st - Jun 30th	86.72	83.56	3.16
SVM Model 7	May 01st - Jul 10th	92.33	86.53	5.80
SVM Model 8	May 01st - Jul 20th	93.78	88.61	5.17
SVM Model 9	May 01st - Jul 30th	94.31	92.79	1.52
SVM Model 10	May 01st - Aug 10th	95.81	95.70	0.11
SVM Model 11	May 01st - Aug 20th	97.27	96.14	1.13
SVM Model 12	May 01st - Aug 30th	97.61	97.12	0.49
SVM Model 13	May 01st - Sep 10th	96.23	96.02	0.21
SVM Model 14	May 01st - Sep 20th	96.89	96.42	0.47
SVM Model 15	May 01st - Sep 30th	96.30	95.07	1.23

Table 5.4: F1-Score of SVM Models

Model ID	Data Time Range	F1-Score A (%)	F1-Score B (%)	Difference (%)
SVM Model 1	May 01st - May 10th	64.72	62.96	1.76
SVM Model 2	May 01st - May 20th	72.03	68.36	3.67
SVM Model 3	May 01st - May 30th	74.85	70.91	3.94
SVM Model 4	May 01st - Jun 10th	76.58	70.85	5.73
SVM Model 5	May 01st - Jun 20th	77.67	73.97	3.70
SVM Model 6	May 01st - Jun 30th	85.17	81.23	3.94
SVM Model 7	May 01st - Jul 10th	90.58	85.43	5.15
SVM Model 8	May 01st - Jul 20th	92.47	86.67	5.80
SVM Model 9	May 01st - Jul 30th	91.86	90.79	1.07
SVM Model 10	May 01st - Aug 10th	94.16	92.72	1.44
SVM Model 11	May 01st - Aug 20th	96.04	94.58	1.46
SVM Model 12	May 01st - Aug 30th	95.97	94.89	1.08
SVM Model 13	May 01st - Sep 10th	94.12	93.70	0.42
SVM Model 14	May 01st - Sep 20th	94.55	94.86	-0.31
SVM Model 15	May 01st - Sep 30th	93.64	92.35	1.29

While the RF model classification line records 91.45% for the F1-Score metric on June 20th, the SVM models only register 90.58% on July 10th. With regard to the two approaches, it was also determined that there is a 20-day discrepancy in the classification line when only optical data is utilized to achieve an accuracy of 90.79%. The aforementioned results are depicted in Figure 5.16.

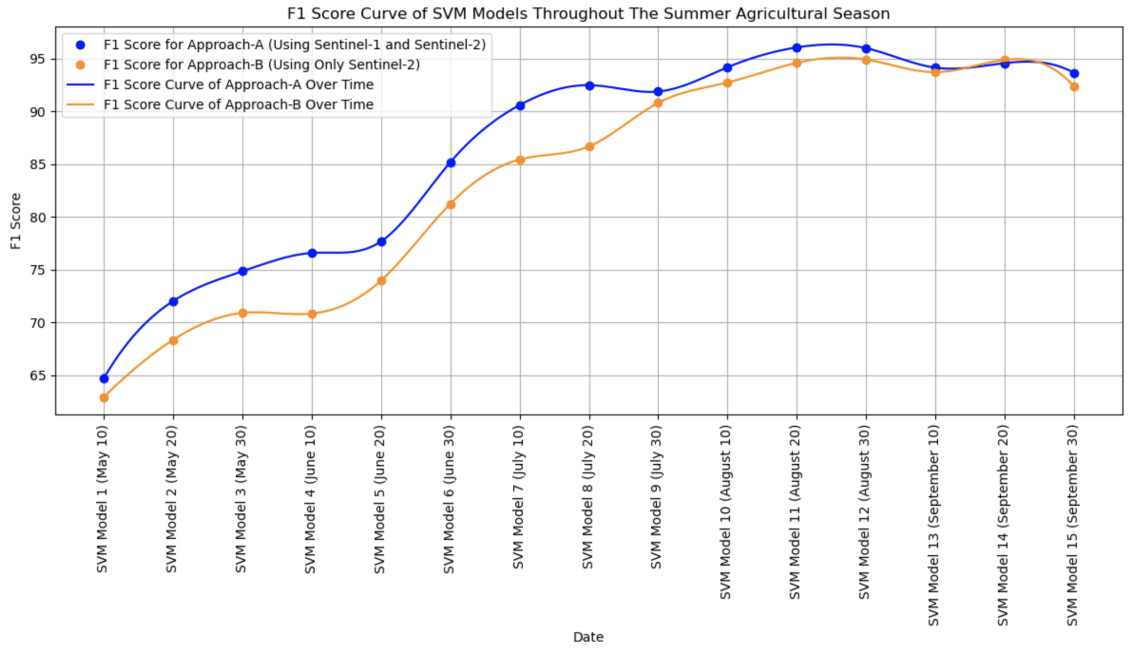

Figure 5.16: Evolution of SVM Models' F1-Score Over Time

5.1.3 Model Tuning

The results of hyperparameter optimization with randomized search and cross-validation for SVM revealed that the SVM's C parameter ranged from 3.85 to 9.60. This adjustment to the permissiveness of correct classifications during training resulted in better generalizability on the test set.

For the RF, the number of trees varied from 317 to 316, the number of variables considered at each split varied between 3 and 6, the minimum number of samples varied between 26 and 30, and the maximum number of nodes varied between 18 and 24. These results highlight the necessity for model tuning as the amount of time series data grows.

5.1.4 Early Crop Mapping Identification

The comparison of the times in which each of the 60 models achieves a performance equal to or greater than 90%, according to the F1 score metric that best matches the reality of the test data in this study, revealed that the model with the ID 'RF Model 6', trained using Approach A (see Figures 5.14 and 5.16), is the one that minimizes the time the most, achieving an F1 score on June 30th.

Figure 5.17 shows the time required for the models to achieve an F1 score performance of 90% or higher during the iterative classification process. For each model, the performance in both approaches is shown.

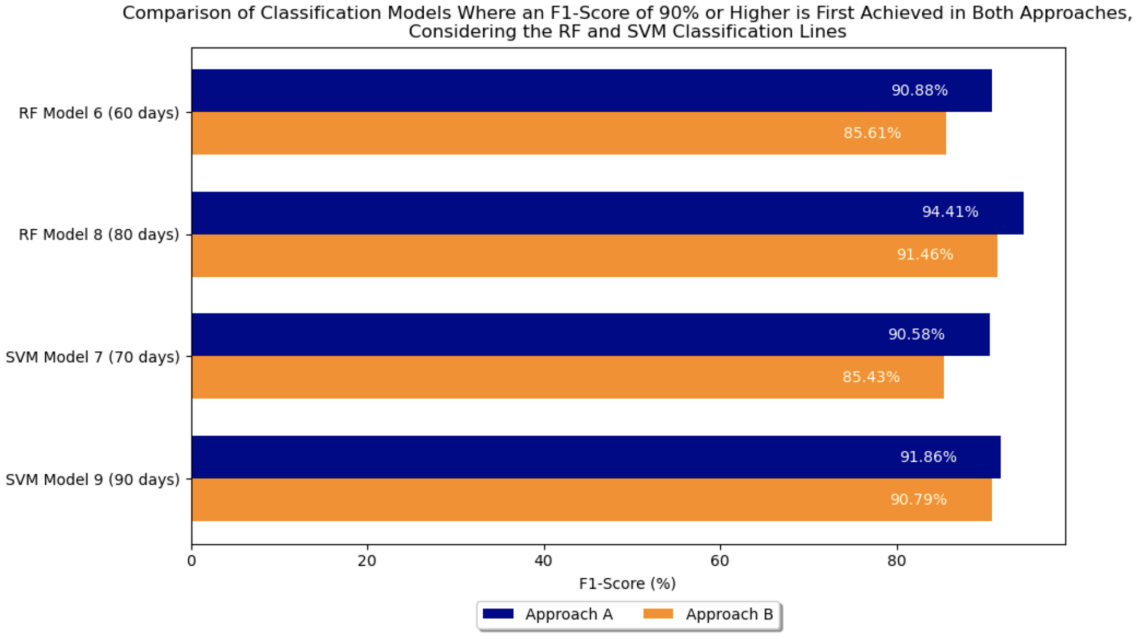

Figure 5.17: Comparison of Classification Models Where an F1-Score of 90% or Higher is First Achieved in Both Approaches, Considering the RF and SVM Classification Lines

The spatial distribution map generated by the application of the early mapping model (RF Model 6) is shown in Figure 5.18.

In regard to the preliminary mapping framework within the phenological cycle of rice and maize crops in Baixo-Mondego, the evolution of classification model performance over time and the onset and phenological stages of the two crops, as documented by the Coimbra Agricultural Innovation Center and the Food and Food and Agriculture Organization (FAO), were considered. It was determined that the early agricultural mapping approach utilizing the RF 6 model enables the attainment of a classification with an F1-Score metric value of 90.88% when rice plants are in the vegetative stage and maize plants are in the initial reproductive phase. This information is illustrated in Figure 5.19.

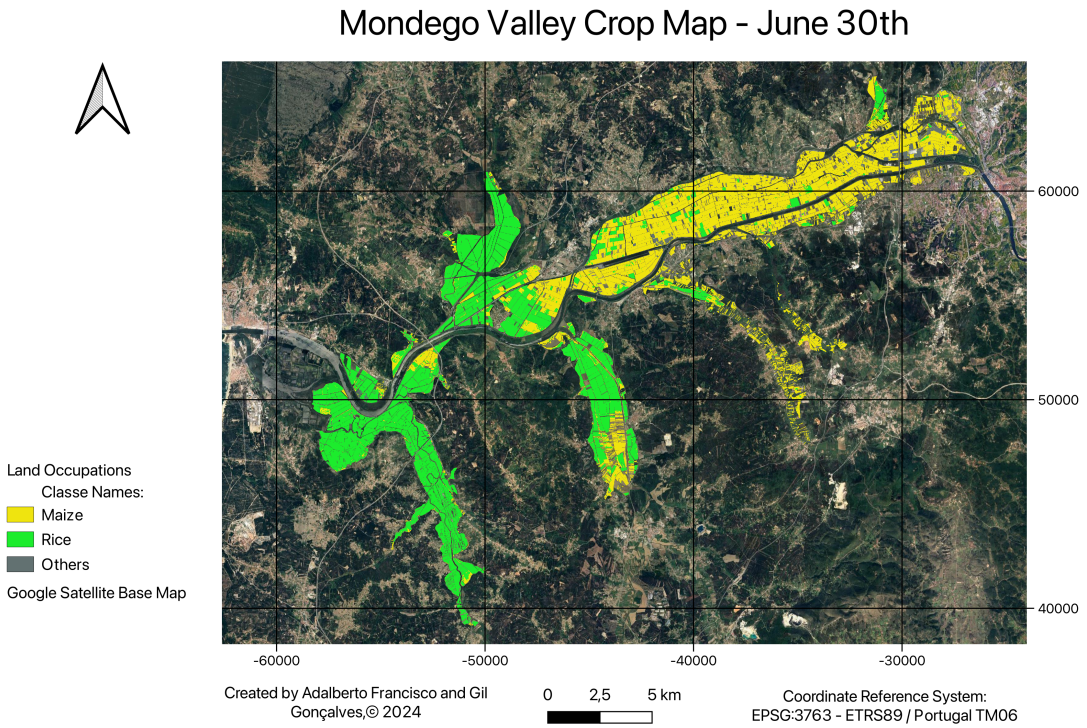

Figure 5.18: Early Crop Mapping of the Mondego Valley as of July 30th

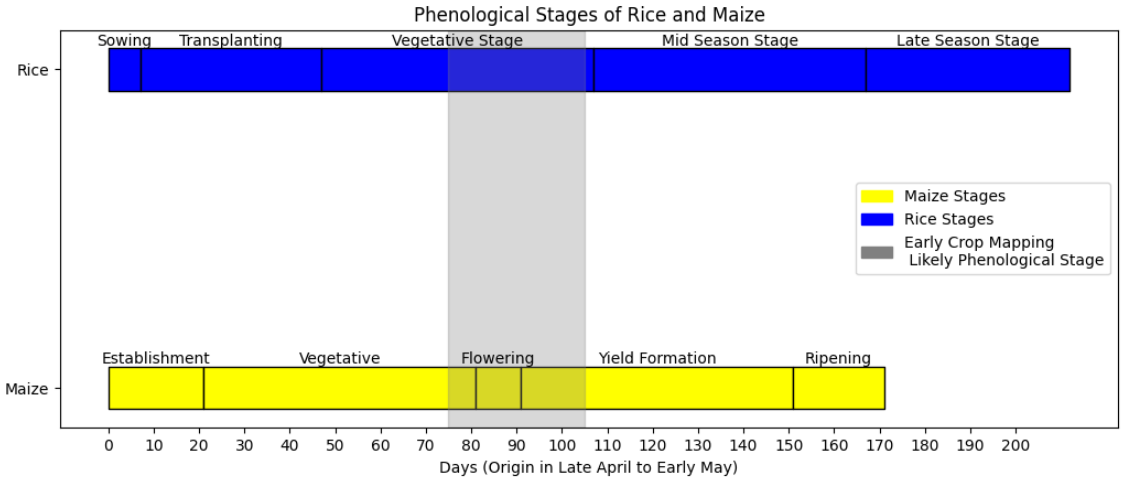

Figure 5.19: Phenological Stages of Rice and Maize

5.1.5 Post-Classification Results

This section presents the results of the post-classification of the agricultural crop spatial distribution map.

5.1.5.1 Sampling Design

The spatial distribution of the 396 points utilized for post-classification of the agricultural crops map is illustrated in Figure 5.20. The methodology employed in delineating the sample is delineated in Section 4.6.1.

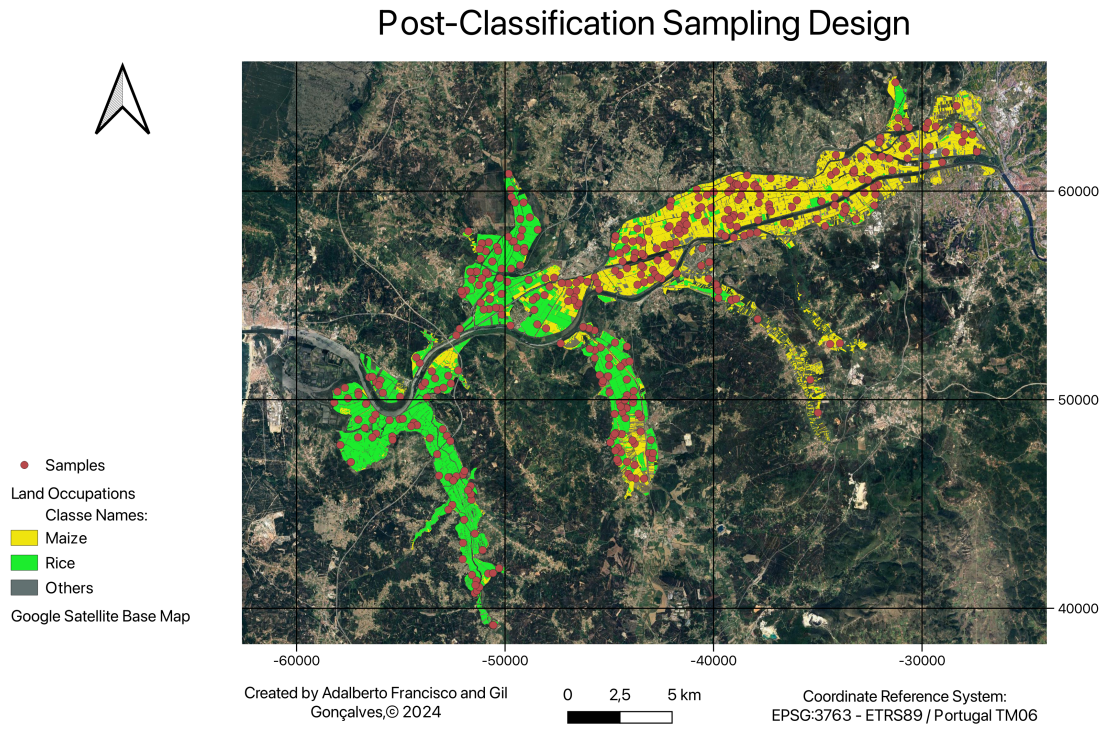

Figure 5.20: Post-Classification Sampling Design

5.1.5.2 Response Design

As described in Section 4.6.2, the response design used the IFAP map, shown in Figure 5.21.

5.1.5.3 Post-Classification Metrics

The confusion matrix presented in Table 5.5 was derived from the validation process of the early mapping of the spatial distribution of agricultural crops in the Baixo Mondego region. This process utilized monitoring information from IFAP.

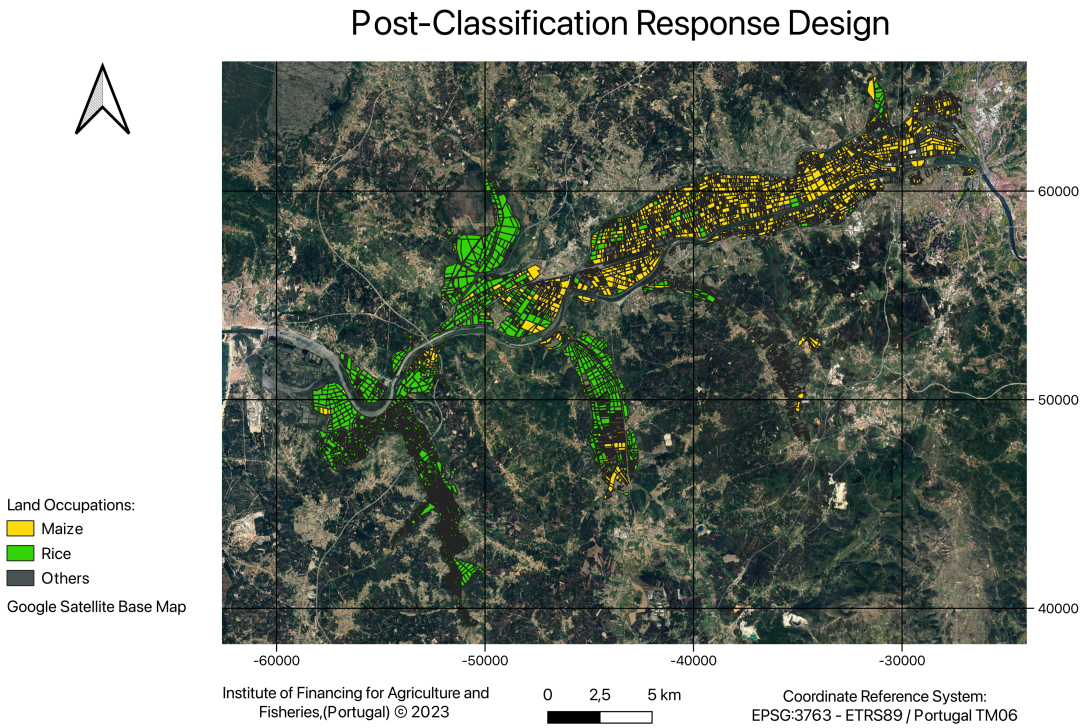

Figure 5.21: Post-Classification Response Design

5.1.5.4 Confusion Matrix

Table 5.5: Confusion Matrix

	Maize (1)	**Rice (2)**	**Others Occupations (3)**	**Total**	**User Accuracy**
1	174	9	10	193	0.90155
2	3	182	3	188	0.96809
3	0	2	13	15	0.86667
Total	177	193	26	396	
Producer Accuracy	0.98299	0.94316	0.50000		

5.1.5.5 Post-Classification Metrics

The post-classification metrics demonstrated that the map exhibited an overall accuracy of 93.19%.

Regarding the robustness of the model from the perspective of the likelihood of field verification of the information on the map, translated by user accuracy (precision), the results show a high reliability above 85%, with emphasis on the classes of interest in this study, showing accuracy values of 90.16% and 96.81% for rice and maize, respectively.

For the producer's accuracy (recall), despite the high values recorded for the maize and rice classes at 98.30% and 94.32%, the results also show that the model

could be improved for the minority classes, where it registers an accuracy of 50.01%.

Table 5.6: Post-Classification Metrics

Category	**Accuracy**	**Standard Deviation**
Overall	0.93187	0.0126
User's Accuracy		
Maize	0.90155	0.0215
Rice	0.96809	0.01285
Others Occupations	0.86667	0.09085
Producer's Accuracy		
maize	0.98299	0.00961
Rice	0.94316	0.01599
Others Occupations	0.5007	0.06295

According to these results, the estimated area of each of the classes and their respective margins of error are shown in Table 5.7.

Table 5.7: Class Area Adjusted

Class	**Area Adjusted (ha)**	**Error (ha)**	**Lower Limit (ha)**	**Upper Limit (ha)**
Maize	4 590,65	116,58	4 362,15	4 819,16
Rice	5 022,69	104,96	4 816,97	5 228,40
Others Occupations	675,81	98,38	482,98	868,64
Total	10 289,15			

Chapter 6

Discussion

The presentation of the results discussion in this chapter is divided into 6.1. Effectiveness of Selected Features; 6.2. Contribution of Optical Radar Fusion; 6.3. Fused Data Exploration Capability and 6.4 Comparison with Previous Studies.

6.1 Effectiveness of Selected Features

With regard to the efficacy of the features, the enhanced performance of the cross-polarization features can be attributed to their heightened sensitivity to the geometry and orientation of the vegetation, coupled with a diminished responsiveness to soil state returns. These returns can exhibit variability across plots with identical crops. Consequently, this radar feature was selected to a slightly greater extent than any other by both RFECV with RF models (Figure 5.3) and RFECV with SVM models (Figure 5.9).

With regard to the optical image features, the results of the RFECV with RF (Figure 5.6) and RFECV with SVM (Figure 5.12) are in agreement with regard to the placement of bands in the infrared, shortwave infrared, and the green channel of the visible, which are identified as the most efficient bands.

These results demonstrate that the vegetation vigor of band 3 and the water content were crucial for distinguishing agricultural crops. This is corroborated by the fact that NDWI occupies an intermediate to high position and NDVI in the selection rankings of the RFECV processes with SVM models (Figure 5.12) and RF (Figure 5.6), respectively.

The observation that NDVI and band 4 (the red channel of the visible region of the electromagnetic spectrum) occupy intermediate to low positions in the selection rankings of RFECV processes with RF models (and SVM models) indicates that biomass quantity was not as crucial as water content in vegetation and soil.

It is also worth noting the utility of the RFECV analysis process. Rather than selecting features based solely on their univariate importance (discriminative power when considered individually), this process weighs this information according to its impact on model performance and its relationship to other features. As a result

of this weighting, the selection outcomes of the same features may vary between models due to the differing architectural characteristics of each. However, cross-analyses such as the one presented above can facilitate the interpretation of the relative value of the information for the specific problem at hand.

6.2 Contribution of Optical and Radar Data Fusion

This subsection presents an analysis of the contribution of data fusion to early mapping.

As evidenced in Tables 5.2 and 5.4, the fusion of radar and optical data yielded a F1-Score metric exceeding 90% with a 30-20 day lead time, surpassing the performance of the approach utilizing solely optical data. This outcome was particularly pronounced when a lower performance threshold was deemed acceptable.

These results demonstrate that the integration of radar and optical data optimizes the process of early mapping, thereby minimizing the time required to achieve the desired level of performance at the outset of the agricultural season.

The primary reason for this is the limited availability of high-quality optical data at the conclusion of spring and the advent of summer. As illustrated in Figure 6.1, there is a period of stagnation in the quantity of optical images with less than 10% cloud coverage from May 1 to June 20, with only three high-quality optical images available during this interval.

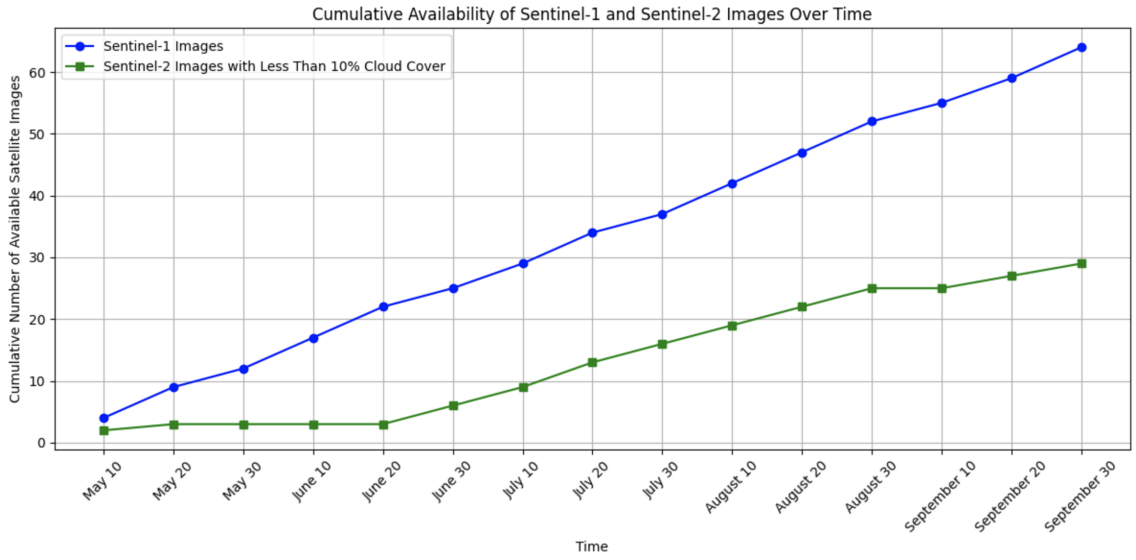

Figure 6.1: Cumulative Availability of Sentinel-1 and Sentinel-2 Images Over Time

However, from June 20 onward, there is an increase in the availability of high-quality optical images. This increase is one of the reasons for the trend of increasing accuracy convergence observed in classification lines with RF and SVM algorithms after June 20. This analysis justifies the fact that the fused data approach

minimizes the time to reach the performance of models in both the classification line with ML algorithms and the classification line with SVM algorithms.

6.3 Capability to Explore the Fused Data

This subsection presents an analysis of the capacity of two algorithms to identify patterns in merged data.

The statistical calculations presented in Table 6.1, which include the maximum, minimum, average, and standard deviation of the difference curves for the F1-Score in each classification line approach using the RF and SVM algorithms, allow us to draw the following results regarding the pattern exploration capacity in merged data:

Table 6.1: Difference Between F1-Score Ratings in Two Approaches (A and B) Mapped According to Classification Lines (RF and SVM)

Model	Mean Difference	Max Difference	Min Difference	Standard Deviation
RF	2.93	6.26	-0.79	2.34
SVM	2.68	5.80	0.11	1.76

The mean values of the two curves (2.93 and 2.68 for the RF and SVM curves, respectively) suggest that the RF algorithm may have a slight advantage in exploring patterns in the merged data, particularly when the time period is shorter (Figure 6.2).

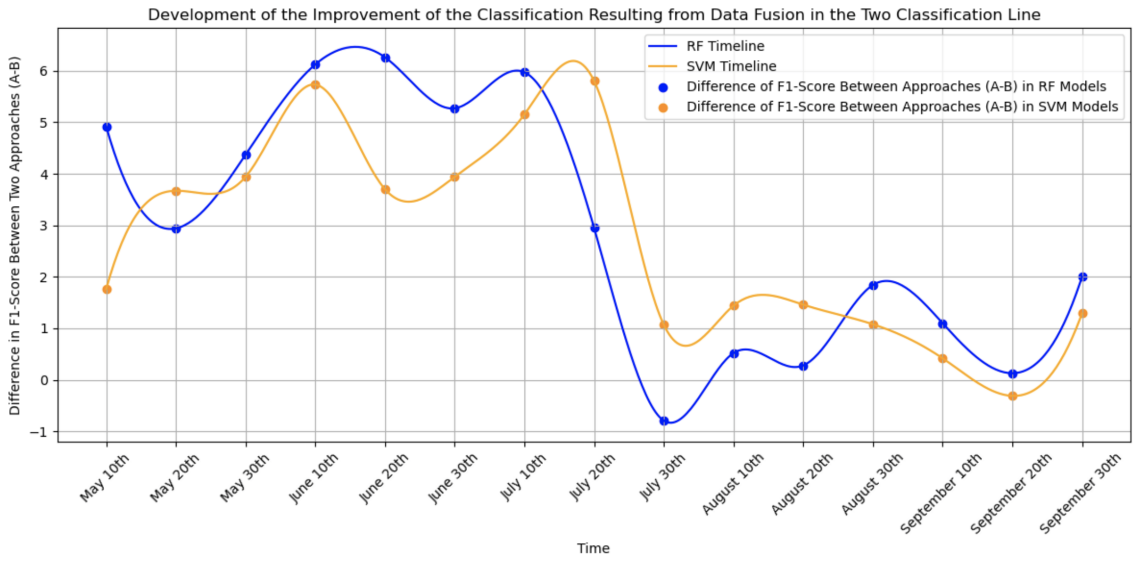

Figure 6.2: Development of the Improvement of the Classification Resulting from Data Fusion in the Two Classification Lines

The maximum values (6.26 and 5.80) for RF and SVM curves, respectively, and the tendency of the RF curve to overlap the SVM curve at the start of the season

(Figure 6.2) reveal a superior capacity of the RF algorithm to explore the merged data early in the season, when radar data still had limited spectral clarity, caused by the low detail in the three-dimensional structure of crops that were still in an early phase of development.

The standard deviation values of each curve indicate that the exploration of the merged data by the SVM algorithm was more stable throughout the season, although slightly lower. This explains why the difference curve in the F1-Score value of models trained by the SVM algorithm did not register any negative values.

These analyses assist in elucidating the rationale behind the observation that models trained with the RF algorithms exhibited a minimal requirement for performance equal to or greater than 90% over the course of the study, given that SVM algorithms were also trained with an approach utilizing merged data.

6.4 Comparison with Previous Studies

6.4.1 Methodological Perspective

As in prior studies conducted by Inglada et al. (2016), Rahmati et al. (2022), Hao et al. (2018), and Skakun et al. (2017), this study employed an incremental classification approach, enabling the tracking of the classification's evolution throughout the spring-summer agricultural season.

This study demonstrated that the methodology employed enables the mapping of the evolution of the classification throughout the agricultural season and the identification of the early temporal window in which classification models reach acceptable performance. This approach not only saves time but also reduces the computational resources required for classifying images from the entire agricultural season.

Table 6.2 presents a comparison of the value of the increment of the time series of satellite data (in days) used in this study with those of other studies on the early mapping of agricultural crops from satellites.

Table 6.2: Time Series Increment Across Various Studies

Study	Time Series Increment	Overall Accuracy (%)
This Study	10 days	90.88
Inglada et al. (2016)	14 days	75.00
Rahmati et al. (2022)	Whenever a Quality Image Was Available	80.12
Hao et al. (2018)	8, 16 and 32 days	91.89
Skakun et al. (2017)	8 days	95.00
Navarro et al. (2021)	5 days	86

6.4.2 Data Fusion Improvements Perspective

As in the studies by Santos (2022), Blickensdörfer et al. (2022), Asam et al. (2022), and Qadir & Mondal (2020), this study used the fusion of radar data with optical data and found that it could improve classification by up to 6.26%.

A mapping of the maximum improvement percentages resulting from the fusion of optical and radar data is presented in Table 6.3.

Table 6.3: Enhanced classification through the fusion of optical and radar images

Study	Crops Studied	Improvement (%)
This Study	Maize, Rice	6.26
Santos (2021)	Oat, Ryegrass, Barley, Rapeseed, Pea, Broad Bean, Birdsfoot Trefoil, Wheat, Triticale	2.2
Blickensdörfer et al. (2022)	Grassland, Winter Wheat, Winter Rye, Winter Barley, Spring Oat, Spring Cereals, Other Cereals, Rapeseed, Silage Maize, Grain Maize, Potato, Sugar Beet, Legume, Sunflower, Strawberry, Asparagus, Onion, Other Vegetables, Carrot, Hops, Vineyard, Orchard, Small Woody Features	10
Asam et al. (2022)	Winter Wheat, Maize, Sugar beet, Rapeseed	9
Qadir and Mondal (2020)	Soybean, Rice	3

6.4.3 Algorithm Performance Perspective

Similar to the studies by Zhou et al. (2017), Qadir & Mondal (2020), Xue et al. (2023), and Erdanaev et al. (2022), this study compared the performance of the RF and SVM algorithms and found that the RF algorithm demonstrated a greater ability to exploit the time series of the merged data, outperforming SVM with an average and maximum improvement of 0.25% and 0.98%, respectively.

A comparative mapping between this study and other studies in which both algorithms have been tested in the classification of agricultural crops from satellite imagery is shown in Table 6.4.

6.4.4 Enhancements to the Surface Surveillance System in Portugal

The implementation of the Common Agricultural Policy (CAP) of the European Union in Portugal saw the introduction of the Surface Surveillance System (SVS).

Table 6.4: Best Machine Learning Algorithm in Different Studies

Study	Algorithms	Best Machine Learning Algorithm
This Study	RF and SVM	RF
Zhou et al. (2017)	RF and SVM	SVM
Qadir and Mondal (2020)	RF and SVM	RF
Xue et al. (2023)	RF and SVM	RF
Erdanaev et al. (2022)	RF and SVM	RF

As outlined by Navarro et al. (2021) on the IFAP portal [8], the SVS is based on the classification of series of satellite images interpolated every five days, high-resolution images, field views, and georeferenced photographs sent by farmers themselves.

In this context, the fusion of data from Sentinel-2 satellites with radar data from Sentinel-1 would provide two contributions to the Surface Surveillance System:

1. An enhanced ability to discriminate agricultural crops, as radar data includes new information based on the structure and moisture content of vegetation;

2. The fusion of data would make the system less dependent on optical data, thereby reducing the need for field visits, particularly during critical periods when cloud coverage is higher. As stated by Navarro et al. (2021), the existing system addresses the deficiency in quality optical data by employing linear interpolation.

[8]Surface Surveillance System

Chapter 7

Conclusions and Future Work

This chapter presents the principal findings of the study, the implications of the results in the agricultural context, and an outlook for future work.

7.1 Summary of Key Findings

The principal findings of this study are presented in the following list:

- Over the past two decades, the most effective methodologies for agricultural mapping have involved the integration of time-series satellite data;
- In the field of applying artificial intelligence for mapping automation, there has been a slight increase in the application of deep learning algorithms. However, the majority of applications continue to be machine learning algorithms, notably SVM, RF, ML, k-NN, and DT;
- The application of incremental classification methodology has the advantage of identifying the early temporal window in which classification models reach acceptable performance, saving both time and computational resources that would be used to classify images from the entire agricultural season;
- The feature selection process using the RFECV algorithm provided three benefits:
    1. Robust reduction based on balancing the univariate relevance of each feature, the multivariate relevance of the features, and the impact on model performance;
    2. Contextual reduction for each model, taking into account the architecture of the model. This factor is critical because different models may assign slightly different relevance weights to the same information;
    3. More intuitive interpretability of feature selection than other techniques such as PCA.

- The cross-polarization features were the most efficient among the radar data for distinguishing corn and rice, as these features allow the mapping of the three-dimensional structure of the crop and are minimally influenced by soil return, which can differ between plots with the same crop and add noisy information.;
- Among the optical features in the "raw" bands, the most pertinent were those associated with the reflectance of the near-infrared and short-wave infrared, which are sensitive to the moisture content in the soil and vegetation. For spectral indices, NDWI demonstrated greater efficacy than NDVI. These findings underscore the pivotal role of optical information pertaining to soil and vegetation water content in differentiating between corn and rice. Additionally, it is noteworthy that these results indicate that while NDVI and NDWI indices are derived from the "raw" bands, they may not supplant the necessity for combining them with bands in the classification process;
- Data fusion improved classification due to the addition of complementary information related to the geometry and orientation of the crops, and penetrating power over clouds. For these reasons, for the distinction of corn and rice, the improvement is greater the lower the availability of quality optical images, which in this case study occurs in the spring and early summer;
- The RF algorithm was more effective in identifying the pattern in both approaches (radar data fused with optical data and only optical data);
- Google Earth Engine facilitates access to and processing of various satellite data impressively and is practical for classification tasks of these data. However, from a processing standpoint, it sacrifices some flexibility in favor of computational power. For this reason, integration with Colab via Google Drive was beneficial for implementing Oversampling, Parameter Optimization, and Recursive Feature Elimination functions through cross-validation.

7.2 Implications for Agricultural Monitoring

The findings of this study, which focused on the early mapping of rice and corn crops using optimized models of Random Forest (RF) and Support Vector Machine (SVM) with multispectral and radar satellite data, have significant implications for agricultural monitoring. These implications extend to various aspects related to the optimization of the agricultural management process, economic impacts, and advancements in agricultural technology.

7.2.1 Agricultural Management Optimization

The improved accuracy and reliability of early classification allows farmers and agricultural stakeholders to make timely decisions. Each case describes the contributions of accurate early mapping:

- From a production point of view, findings demonstrate that proposed approach allows early accurate classification of agricultural crops during the late vegetative and early reproductive phases;
- From the perspective of support from governmental and non-governmental organizations, findings also demonstrate that proposed approach allows development of resilient monitoring systems for farmers' declarations based on satellite image classifications.

Regarding the construction of these monitoring systems, in the context of Portugal, where a system based on the classification of interpolated optical image time series is used every 5 days, the proposed approach would not only increase the system's capacity to discriminate crops, but also reduce the system's dependence on optical data.

7.2.2 Economic and Environmental Impacts

In terms of economic and environmental impact, the approach is cost-effective as it permits the processing of substantial quantities of data at no additional expense and on remote servers. This circumvents the potential for increased production costs while also contributing to the reduction of environmental impact.

7.2.3 Technological Advancements in Agriculture

The successful integration of GEE with Colab for automating agricultural crop mapping through machine learning demonstrates the potential of advanced computational tools, setting a precedent for the adoption of sophisticated, cost-effective, flexible technologies capable of handling large volumes of information in agriculture. The scalability of this methodology to other crops and the accessibility of satellite data for different geographic regions ensure that even small farmers can benefit from these technologies, promoting inclusive agricultural development.

7.3 Recommendations for Future Research

Future work involves the following tasks:

1. Improve the model's capability for classifying minority classes, which was a limitation of the early mapping model;
2. Explore the application of the early mapping model using approaches such as transfer learning, fine-tuning, and integration of auxiliary data (such as meteorological information);
3. Experiment with different time series increments as done in the studies by Hao et al. (2018);

4. Explore the capabilities of deep learning algorithms;
5. Create an independent ground truth data for evaluating the 2024 crop season.

References

Asam, S., Gessner, U., González, R. A., Wenzl, M., Kriese, J., & Kuenzer, C. (2022). Mapping crop types of germany by combining temporal statistical metrics of sentinel-1 and sentinel-2 time series with lpis data. *Remote Sensing*, *14*. https://doi.org/10.3390/rs14132981

Blickensdörfer, L., Schwieder, M., Pflugmacher, D., Nendel, C., Erasmi, S., & Hostert, P. (2022). Mapping of crop types and crop sequences with combined time series of sentinel-1, sentinel-2 and landsat 8 data for germany. *Remote Sensing of Environment*, *269*. https://doi.org/10.1016/j.rse.2021.112831

Chabalala, Y., Adam, E., & Ali, K. A. (2022). Machine learning classification of fused sentinel-1 and sentinel-2 image data towards mapping fruit plantations in highly heterogenous landscapes. *Remote Sensing*, *14*. https://doi.org/10.3390/rs14112621

Chen, J., Jönsson, P., Tamura, M., Gu, Z., Matsushita, B., & Eklundh, L. (2004). A simple method for reconstructing a high-quality ndvi time-series data set based on the savitzky-golay filter. *Remote Sensing of Environment*, *91*, 332–344. https://doi.org/10.1016/j.rse.2004.03.014

Erdanaev, E., Kappas, M., & Wyss, D. (2022). Irrigated crop types mapping in tashkent province of uzbekistan with remote sensing-based classification methods. *Sensors*, *22*(15). https://doi.org/10.3390/s22155683

Fonte, C. C., Duarte, D., Jesus, I., Costa, H., Benevides, P., Moreira, F., & Caetano, M. (2024). Accuracy assessment and comparison of national, european and global land use land cover maps at the national scale—case study: Portugal. *Remote Sensing*, *16*(9). https://doi.org/10.3390/rs16091504

Hao, P., Wu, M., Niu, Z., Wang, L., & Zhan, Y. (2018). Estimation of different data compositions for early-season crop type classification. *PeerJ*, *2018*. https://doi.org/10.7717/peerj.4834

He, S., Peng, P., Chen, Y., & Wang, X. (2022). Multi-crop classification using feature selection-coupled machine learning classifiers based on spectral, textural and environmental features. *Remote Sensing*, *14*. https://doi.org/10.3390/rs14133153

Heinl, M., Walde, J., Tappeiner, G., & Tappeiner, U. (2009). Classifiers vs. input variables-the drivers in image classification for land cover mapping. *International Journal of Applied Earth Observation and Geoinformation*, *11*, 423–430. https://doi.org/10.1016/j.jag.2009.08.002

HU, Q., bin WU, W., SONG, Q., LU, M., CHEN, D., yi YU, Q., & jun TANG, H. (2017). How do temporal and spectral features matter in crop classification in heilongjiang province, china? *Journal of Integrative Agriculture*, *16*, 324–336. https://doi.org/10.1016/S2095-3119(15)61321-1

Immitzer, M., Vuolo, F., & Atzberger, C. (2016). First experience with sentinel-2 data for crop and tree species classifications in central europe. *Remote Sensing*, *8*. https://doi.org/10.3390/rs8030166

Inglada, J., Vincent, A., Arias, M., & Marais-Sicre, C. (2016). Improved early crop type identification by joint use of high temporal resolution sar and optical image time series. *Remote Sensing*, *8*. https://doi.org/10.3390/rs8050362

Jayakumari, R., Nidamanuri, R. R., & Ramiya, A. M. (2021). Object-level classification of vegetable crops in 3d lidar point cloud using deep learning convolutional neural networks. *Precision Agriculture*, *22*, 1617–1633. https://doi.org/10.1007/s11119-021-09803-0

Joshi, A., Pradhan, B., Gite, S., & Chakraborty, S. (2023). Remote-sensing data and deep-learning techniques in crop mapping and yield prediction: A systematic review. *Remote Sensing*, *15*(8). https://doi.org/10.3390/rs15082014

Kussul, N., Lemoine, G., Gallego, F. J., Skakun, S. V., Lavreniuk, M., & Shelestov, A. Y. (2016). Parcel-based crop classification in ukraine using landsat-8 data and sentinel-1a data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *9*, 2500–2508. https://doi.org/10.1109/JSTARS.2016.2560141

Langley, S. K., Cheshire, H. M., & Humes, K. S. (2001). A comparison of single date and multitemporal satellite image classifications in a semi-arid grassland. *Journal of Arid Environments*, *49*, 401–411. https://doi.org/10.1006/jare.2000.0771

Machichi, M. A., loubna El mansouri, yasmina imani, Bourja, O., Lahlou, O., Zennayi, Y., Bourzeix, F., Houmma, I. H., & Hadria, R. (2023). Crop mapping using supervised machine learning and deep learning: A systematic literature review. https://doi.org/10.1080/01431161.2023.2205984

Navarro, A., Silva, I., Catalão, J., & Falcão, J. (2021). An operational sentinel-2 based monitoring system for the management and control of direct aids to the farmers in the context of the common agricultural policy (cap): A case study in mainland portugal. *International Journal of Applied Earth Observation and Geoinformation*, *103*, 102469. https://doi.org/https://doi.org/10.1016/j.jag.2021.102469

Niel, T. G. V., & McVicar, T. R. (2004). Determining temporal windows for crop discrimination with remote sensing: A case study in south-eastern australia. *Computers and Electronics in Agriculture*, *45*, 91–108. https://doi.org/10.1016/j.compag.2004.06.003

Potgieter, A. B., Zhao, Y., Zarco-Tejada, P. J., Chenu, K., Zhang, Y., Porker, K., Biddulph, B., Dang, Y. P., Neale, T., Roosta, F., & Chapman, S. (2021). Evolution and application of digital technologies to predict crop type and crop phenology in agriculture. https://doi.org/10.1093/insilicoplants/diab017

Qadeer, M. U., Saeed, S., Taj, M., & Muhammad, A. (2021). Spatio-temporal crop classification on volumetric data. *Proceedings - International Conference on Image Processing, ICIP*, *2021-September*, 3812–3816. https://doi.org/10.1109/ICIP42928.2021.9506046

Qadir, A., & Mondal, P. (2020). Synergistic use of radar and optical satellite data for improved monsoon cropland mapping in india. *Remote Sensing*, *12*. https://doi.org/10.3390/rs12030522

Rahmati, A., Zoej, M. J. V., & Dehkordi, A. T. (2022). Early identification of crop types using sentinel-2 satellite images and an incremental multi-feature ensemble method (case study: Shahriar, iran). *Advances in Space Research*, *70*, 907–922. https://doi.org/10.1016/j.asr.2022.05.038

Santos, M. J. G. (2021). Classificação de culturas agrícolas de inverno com recurso à plataforma google earth engine e imagens dos satélites sentinel-1 e sentinel-2. http://hdl.handle.net/10451/48443

Skakun, S., Franch, B., Vermote, E., Roger, J.-C., Becker-Reshef, I., Justice, C., & Kussul, N. (2017). Early season large-area winter crop mapping using modis ndvi data, growing degree days information and a gaussian mixture model. *Remote Sensing of Environment*, *195*, 244–258. https://doi.org/10.1016/j.rse.2017.04.026

Verma, A., Kumar, A., & Lal, K. (2019). Kharif crop characterization using combination of sar and msi optical sentinel satellite datasets. *Journal of Earth System Science*, *128*. https://doi.org/10.1007/s12040-019-1260-0

Xue, H., Xu, X., Zhu, Q., Yang, G., Long, H., Li, H., Yang, X., Zhang, J., Yang, Y., Xu, S., Yang, M., & Li, Y. (2023). Object-oriented crop classification using time series sentinel images from google earth engine. *Remote Sensing*, *15*. https://doi.org/10.3390/rs15051353

Yi, Z., Jia, L., Chen, Q., Jiang, M., Zhou, D., & Zeng, Y. (2022). Early-season crop identification in the shiyang river basin using a deep learning algorithm and time-series sentinel-2 data. *Remote Sensing*, *14*. https://doi.org/10.3390/rs14215625

yu HAO, P., jun TANG, H., xin CHEN, Z., yan MENG, Q., & peng KANG, Y. (2020). Early-season crop type mapping using 30-m reference time series. *Journal of Integrative Agriculture*, *19*, 1897–1911. https://doi.org/10.1016/S2095-3119(19)62812-1

Zhou, T., Pan, J., Zhang, P., Wei, S., & Han, T. (2017). Mapping winter wheat with multi-temporal sar and optical images in an urban agricultural region. *Sensors*, *17*(6). https://doi.org/10.3390/s17061210

Appendices

Appendix A

Glossary

- RF Classification Line - Continuous classification throughout the season using the RF algorithm;
- SVM Classification Line - Continuous classification throughout the season using the SVM algorithm;
- Approach A - Classification using merged data (radar + optical);
- Approach B - Classifications that use only optical data;
- Machine Learning Algorithms - A set of mathematical functions used by the computer to learn patterns in data.
- Model - Algorithms + Features + Parameters (Internal Model Parameters) + Hyperparameters (External Model Parameters);

Appendix B

Flowchart of the Methodological Framework in Landscape Orientation

Incremental Process with Geospatial Intelligence and Supervised Learning Throughout the 2023 Summer Agricultural Season (i=1,2,3...15)
Non-Iterative Process
1.Feature Extraction
2.Building the Reference Database
3.Feature Selection
4.Training Models
5.Test Models
6.Models' Evaluation
7.Post-Classification
Approach A
Approach B
i10-Day Time Series with Sentinel-1 Data (i=1,2,3...15)
i10-Day Time Series with Sentinel-2 Data (i=1,2,3...15)
i10-Day Time Series with Sentinel-2 Data (i=1,2,3...15)
SOMTE Oversampling
SOMTE Oversampling
70%
30%
Samples Ground Truth
Recursive Feature Elimination with CV
Recursive Feature Elimination with CV
Best Features of A
Best Features of B
Train RF and SVM Models
Train RF and SVM Models
Apply Models to Plots
Apply Models to Plots
Accuracy Assment
Accuracy Assment
Model Accuary for A
Model Accuary for B
Early Crop Map
tmin = min{t|P(t) ≥ 0.9}
Overlay the Accuracy Curves Values
a(A) - a(B)
Sentinel-1 Contribution
IFAP 2023 Monitoring Map
Accuracy Assessment
Classification Accuracy
input 10-Day Incremental Time Series(i=1,2,3...15)
Input Map
Input Database
Process
Outputs
Best Features
Notations:
P - Performance;
i - Iteration;
t- Time;
A - Approach Using S-1 and S-2 Data;
B - Approach Using Only S-2 Data.

Appendix C

Coding

C.1 Coding in GEE

C.1.1 Prepare Samples to Oversampling

```
var imbalaced_points = rojects/ee-adalbertodissertation/assets/imbalaced_points

var ROI = projects/ee-adalbertodissertation/assets/ROI
var initial_time = 'yyyy-mm-dd';
var final_time = 'yyyy-mm-dd';
var cloud_cover_thrshold = 10;
var sentinel_1_bands = ['VV', 'VH'];
var sentinel_2_bands = ['B1', 'B2', 'B3', 'B4', 'B8', 'B11', 'B12'];


function calculate_ndvi(image) {
  var ndvi = image.normalizedDifference(['B8', 'B4']).rename('NDVI');
  return image.addBands(ndvi);
}

function reproject_to_10m(image) {
  return image
  .reproject({crs: 'EPSG:32629', scale: 10})
  .reduceResolution({
    reducer: ee.Reducer.mean(),
    maxPixels: 1024
  });
}

function calculate_ndwi(image) {
  var ndwi = image.normalizedDifference(['B3', 'B8']).rename('NDWI');
  return image.addBands(ndwi);
}

function reduce_to_10_days_mean(collection_name, start_date, end_date,
  interval_days) {
  var ee_start_date = ee.Date(start_date);
  var ee_end_date = ee.Date(end_date);
  var total_days = ee_end_date.difference(ee_start_date, 'day');
  var sequence = ee.List.sequence(0, total_days, interval_days);
```

```
35
36    var medians_of_collection = sequence.map(function(i) {
37      var start_of_stage = ee_start_date.advance(i, 'day');
38      var end_of_stage = start_of_stage.advance(interval_days, 'day');
39      var median_of_stage = collection_name.filterDate(start_of_stage,
      end_of_stage)
40        .median()
41        .set('system:time_start', start_of_stage.millis());
42      return reproject_to_10m(median_of_stage);
43    });
44    return medians_of_collection;
45  }
46
47  function add_features_date_to_name(image, medians_of_collection) {
48    var new_band_names = medians_of_collection.map(function(img) {
49      var date = ee.Image(img).date().format('YYYY-MM-dd');
50      return ee.Image(img).bandNames().map(function(bandName) {
51        return ee.String(bandName).cat('_').cat(date);
52      });
53    }).flatten();
54
55    return image.rename(new_band_names);
56  }
57
58  var filtered_sentinel_2_imgs_resampled = filtered_sentinel_2_imgs_and_SI.map(
      reprojectTo10m);
59  var filtered_sentinel_1_imgs_resampled = filtered_sentinel_1_imgs.map(
      reprojectTo10m);
60  var s1_images = filtered_sentinel_1_imgs_resampled;
61  var s2_images = filtered_sentinel_2_imgs_resampled;
62  var reduced_s1_images = reduce_to_10_days_mean(
      filtered_sentinel_1_imgs_resampled, initial_time, final_time, 10);
63  var reduced_s2_images = reduce_to_10_days_mean(
      filtered_sentinel_2_imgs_resampled, initial_time, final_time, 10);
64  var s1_bands = add_features_date_to_name(ee.ImageCollection.fromImages(
      reduced_s1_images).toBands(), reduced_s1_images);
65  var s2_bands = add_features_date_to_name(ee.ImageCollection.fromImages(
      reduced_s2_images).toBands(), reduced_s2_images);
66  var fused_satellite_data = s1_bands.addBands(s2_bands);
67  var s1_s2_data = fused_satellite_data.clip(ROI)
68
69
70  var sampleData = s1_s2_data.sampleRegions({
71    collection: sample_points,
72    properties: ['class_ID'],
73    scale: 10,
74    geometries: true
75  });
76  var sampleData_ = sampleData.randomColumn('random')
77  var split_threshold = 0.7
78  var training = sampleData_.filter(ee.Filter.lt('random', split_threshold));
79  var testing = sampleData_.filter(ee.Filter.gte('random', split_threshold));
80
81  var number_of_sentinel_1_imgs = filtered_sentinel_1_imgs.size();
82  var S1_list = filtered_sentinel_1_imgs.toList(filtered_sentinel_1_imgs.size());
83  var S1_dates = S1_list.map(function(image) {
84    return ee.Image(image).date().format();
85  });
```

```
86
87  var number_of_sentinel_2_imgs = filtered_sentinel_2_imgs_and_SI.size();
88  var S2_list = filtered_sentinel_2_imgs.toList(filtered_sentinel_2_imgs.size());
89  var S2_dates = S2_list.map(function(image) {
90    return ee.Image(image).date().format();
91  });
92
93  Export.table.toDrive({
94    collection: training,
95    description: 'imbalanced_training_data',
96    folder: 'GEE_Exports',
97    fileNamePrefix: 'training_data',
98    fileFormat: 'CSV'
99  });
```

C.1.2 Classification

```
1   var balanced = projects/ee-adalbertodissertation/assets/balanced_points
2   var ROI = projects/ee-adalbertodissertation/assets/ROI
3   var initial_time = 'yyyy-mm-dd';
4   var final_time = 'yyyy-mm-dd';
5   var cloud_cover_thrshold = 10;
6   var sentinel_1_bands = ['VV', 'VH'];
7   var sentinel_2_bands = ['B1', 'B2', 'B3', 'B4', 'B8', 'B11', 'B12'];
8
9
10  function calculate_ndvi(image) {
11    var ndvi = image.normalizedDifference(['B8', 'B4']).rename('NDVI');
12    return image.addBands(ndvi);
13  }
14
15  function reprojectTo10m(image) {
16    return image
17    .reproject({crs: 'EPSG:32629', scale: 10})
18    .reduceResolution({
19      reducer: ee.Reducer.mean(),
20      maxPixels: 1024
21    });
22  }
23
24  function calculate_ndwi(image) {
25    var ndwi = image.normalizedDifference(['B3', 'B8']).rename('NDWI');
26    return image.addBands(ndwi);
27  }
28
29  function reduce_to_10_days_mean(collection_name, start_date, end_date,
      interval_days) {
30    var ee_start_date = ee.Date(start_date);
31    var ee_end_date = ee.Date(end_date);
32    var total_days = ee_end_date.difference(ee_start_date, 'day');
33    var sequence = ee.List.sequence(0, total_days, interval_days);
34
35    var medians_of_collection = sequence.map(function(i) {
36      var start_of_stage = ee_start_date.advance(i, 'day');
37      var end_of_stage = start_of_stage.advance(interval_days, 'day');
```

```
      var median_of_stage = collection_name.filterDate(start_of_stage,
    end_of_stage)
      .median()
      .set('system:time_start', start_of_stage.millis());
      return reprojectTo10m(median_of_stage);
    });
    return medians_of_collection;
  }

  function add_features_date_to_name(image, medians_of_collection) {
    var new_band_names = medians_of_collection.map(function(img) {
      var date = ee.Image(img).date().format('YYYY-MM-dd');
      return ee.Image(img).bandNames().map(function(bandName) {
        return ee.String(bandName).cat('_').cat(date);
      });
    }).flatten();

    return image.rename(new_band_names);
  }

  function Z_Score_to_Image(image, bands, region) {
    var mean = image.select(bands).reduceRegion({
      reducer: ee.Reducer.mean(),
      geometry: region,
      scale: 10,
      maxPixels: 1e9
    });
    var stdDev = image.select(bands).reduceRegion({
      reducer: ee.Reducer.stdDev(),
      geometry: region,
      scale: 10,
      maxPixels: 1e9
    });
    return image.select(bands).subtract(mean.toImage(bands)).divide(stdDev.
    toImage(bands));
  }

  function calculate_stats(collection, bands) {
    var means = {};
    var stdDevs = {};

    bands.forEach(function(band) {
      var mean = collection.reduceColumns(ee.Reducer.mean(), [band]).get('mean');
      var stdDev = collection.reduceColumns(ee.Reducer.stdDev(), [band]).get('
    stdDev');
      means[band] = mean;
      stdDevs[band] = stdDev;
    });

    return {means: means, stdDevs: stdDevs};
  }

  function Z_Score_to_Bands(feature, bands, stats) {
    var normDict = {};
    bands.forEach(function(band) {
      var value = ee.Number(feature.get(band));
      var mean = ee.Number(stats.means[band]);
      var stdDev = ee.Number(stats.stdDevs[band]);
```

```
93        var normalizedValue = value.subtract(mean).divide(stdDev);
94        normDict[band] = normalizedValue;
95      });
96      return feature.set(normDict);
97    }
98
99    function Z_Score_to_Collections(feature) {
100     feature = Z_Score_to_Bands(feature, sentinel1Bands, S_1_Stats); //Cooment
        this line in case of approach B
101     feature = Z_Score_to_Bands(feature, sentinel2Bands, S_2_Stats);
102     return feature;
103   }
104
105   var filtered_sentinel_2_imgs = ee.ImageCollection('COPERNICUS/S2_SR_HARMONIZED'
        )
106   .filterBounds(ROI)
107   .filterDate(ee.Date(initial_time), ee.Date(final_time))
108   .filter(ee.Filter.lt('CLOUDY_PIXEL_PERCENTAGE', cloud_cover_thrshold))
109   .select(sentinel_2_bands);
110
111   var filtered_sentinel_2_imgs_and_SI = filtered_sentinel_2_imgs.map(
        calculate_ndvi).map(calculate_ndwi);
112
113   var filtered_sentinel_1_imgs = ee.ImageCollection('COPERNICUS/S1_GRD_FLOAT')
114   .filterBounds(ROI)
115   .filterDate(ee.Date(initial_time), ee.Date(final_time))
116   .filter(ee.Filter.eq('instrumentMode', 'IW'))
117   .select(sentinel_1_bands);
118
119   var filtered_sentinel_2_imgs_resampled = filtered_sentinel_2_imgs_and_SI.map(
        reprojectTo10m);
120   var filtered_sentinel_1_imgs_resampled = filtered_sentinel_1_imgs.map(
        reprojectTo10m);
121   var s1_images = filtered_sentinel_1_imgs_resampled;
122   var s2_images = filtered_sentinel_2_imgs_resampled;
123   var reduced_s1_images = reduce_to_10_days_mean(
        filtered_sentinel_1_imgs_resampled, initial_time, final_time, 10);
124   var reduced_s2_images = reduce_to_10_days_mean(
        filtered_sentinel_2_imgs_resampled, initial_time, final_time, 10);
125   var s1_bands = add_features_date_to_name(ee.ImageCollection.fromImages(
        reduced_s1_images).toBands(), reduced_s1_images);
126   var s2_bands = add_features_date_to_name(ee.ImageCollection.fromImages(
        reduced_s2_images).toBands(), reduced_s2_images);
127   var fused_satellite_data = s1_bands.addBands(s2_bands);
128   var s1_s2_data = fused_satellite_data.clip(ROI);
129
130   var multiespectral_band_indexis = ee.List.sequence('int_number','int_number');
131   var multiespctral_set_bands = s1_s2_data.select(multiespectral_band_indexis);
132
133   //RANDOM FOREST CASE
134   var sampleData = s1_s2_data.sampleRegions({ //Change s1_s2_data to
        multiespctral_set_band in case of Approach B
135     collection: sample_points,
136     properties: ['class_ID'],
137     scale: 10,
138     geometries: true
139   });
140
```

```
var sampleData_ = sampleData.randomColumn('random')
var split_threshold = 0.7
var training = sampleData_.filter(ee.Filter.lt('random', split_threshold));
var testing = sampleData_.filter(ee.Filter.gte('random', split_threshold));
var balanced_training_set = balanced;
var best_features_for_RF = s1_s2_data.select(['Best Feature Names'])
var nomes_s1_s2_data = s1_s2_data.bandNames()


var trainedClassifier = ee.Classifier.smileRandomForest({
  numberOfTrees: 307,
  variablesPerSplit: 3,
  minLeafPopulation: 26,
  bagFraction: 1.0,
  maxNodes: 18,
  seed: 0
}).train({
  features: balanced,
  classProperty: 'class_ID',
  inputProperties: best_features_for_RF.bandNames()
});
var crop_map =  s1_s2_data.classify(trainedClassifier); //Change s1_s2_data to
  multiespctral_set_band in case of Approach B


//SUPPORT VECTOR MACHINE CASE
var s1_s2_data_without_normalization = s1_s2_data
var s1_normalized_data = Z_Score_to_Image(s1_s2_data_without_normalization, ['S
  -1 Best Features'], ROI);
var s2_normalized_data = Z_Score_to_Image(s1_s2_data_without_normalization, ['S
  -2 Best Features'], ROI);
var s1_s2_normalized_data = s1_normalized_data.addBands(s2_normalized_data);
var spectral_index = s1_s2_data_without_normalization.select(['SIs']) //In
  order to maintain the index scales
var s1_s2_IS_normalized = s1_s2_normalized_data.addBands(spectral_index);

var only_s2_and_IS = s2_normalized_data.addBands(spectral_index); //In Case of
  Aproach B

var sampleData = s1_s2_data_without_normalization.sampleRegions({ //to avoide
  data leakage...,
  //...sampling is not performed on the normalized multiband image
  collection: sample_points,
  properties: ['class_ID'],
  scale: 10,
  geometries: true
});
var sampleData_ = sampleData.randomColumn('random')
var split_threshold = 0.7
var training = sampleData_.filter(ee.Filter.lt('random', split_threshold));
var testing = sampleData_.filter(ee.Filter.gte('random', split_threshold));
var balanced_training_set = balanced;

var S_1_Bands = ['S-1 Best Features'];
var S_2_Bands = ['S-2 Best Features',];
var indices = ['SI Best Features'];
var S_1_Stats = calculate_stats(balanced_training_set, S_1_Bands);
var S_2_Stats = calculate_stats(balanced_training_set, S_2_Bands);
```

```
var normalized_training = balanced_training_set.map(Z_Score_to_Collections);


var trainedClassifier = ee.Classifier.libsvm({
  kernelType: 'linear',
  cost: 9.6
}).train({
  features: normalized_training,
  classProperty: 'class_ID',
  inputProperties: s1_s2_IS_normalized.bandNames() //change to only_s2_and_IS.bandNames() in case of approach B
});

var crop_map =  s1_s2_IS_normalized.classify(trainedClassifier); //change to only_s2_and_IS in case of approach B


var classes = [1, 2, 3, 4, 5, 6];
var testingSample = crop_map.sampleRegions({
  collection: testing,
  properties: ['class_ID'],
  scale: 10
});

var error_matrix = testingSample.errorMatrix('class_ID', 'classification', classes);
print('Confusion Matrix', error_matrix);
print('Accuracy', error_matrix.accuracy());
var producerAccuracy = error_matrix.producersAccuracy();
var consumerAccuracy = error_matrix.consumersAccuracy();
print('Producer Accuracy (Recall)', producerAccuracy);
print('Consumer Accuracy (Precision)', consumerAccuracy);
var F1_score = error_matrix.fscore()
print('F1 Score', F1_score)
var f1ScoreGeneral = ee.Array(F1_score).reduce(ee.Reducer.mean(), [0]).get([0]);
print('General F1 Score', f1ScoreGeneral);


// var rgb_view = s1_s2_data.visualize({
  //   bands: ['B4', 'B3', 'B2'],
  //   min: 0,
  //   max: 3000,
  //   gamma: [0.95, 0.95, 0.95]
  // });
// Map.addLayer(rgb_view, {}, 'RGB Data View')

//Plot Classified Map

Map.addLayer(crop_map, {
  min: 1,
  max: 6,
  palette: ['yellow', 'green', 'blue', 'grey','red', 'cyan','green']
}, 'Crop Map');


Export.image.toDrive({
  image: crop_map,
```

```
    description: 'crop_map',
    scale: 10,
    region: ROI,
    maxPixels: 1e13,
    crs:'EPSG:3763'
  });

```

C.2 Coding in Colab

C.2.1 Oversampling

```
import pandas as pd
from imblearn.over_sampling import SMOTE
import geopandas as gpd
from shapely.geometry import Point
import json

umbalanced_training_set = pd.read_csv('training_data.csv')

umbalanced_training_set['geometry'] = umbalanced_training_set['.geo'].apply(
  lambda x: Point(json.loads(x)['coordinates']))
umbalanced_training_set['LAT'] = umbalanced_training_set['geometry'].apply(
  lambda g: g.y)
umbalanced_training_set['LON'] = umbalanced_training_set['geometry'].apply(
  lambda g: g.x)

X = umbalanced_training_set.drop(['class_ID', '.geo', 'geometry', 'random', '
  system:index'], axis=1)
y = umbalanced_training_set['class_ID']

smote = SMOTE(random_state=42)
X_resampled, y_resampled = smote.fit_resample(X, y)

balanced_training_set = pd.DataFrame(X_resampled, columns=X.columns)
balanced_training_set['class_ID'] = y_resampled
balanced_training_set['lat'] = umbalanced_training_set.iloc[:len(y_resampled)][
  'LAT']
balanced_training_set['lon'] = umbalanced_training_set.iloc[:len(y_resampled)][
  'LON']

balanced_training_set['lat'] = pd.to_numeric(balanced_training_set['LAT'],
  errors='coerce')
balanced_training_set['lon'] = pd.to_numeric(balanced_training_set['LON'],
  errors='coerce')

balanced_training_set.to_csv('balanced_data.csv', index=False)
```

C.2.2 Hyperparameters Optimization and Feature Selection

C.2.2.1 In Random Forest Models

```
from sklearn.ensemble import RandomForestClassifier
import pandas as pd
from sklearn.model_selection import RandomizedSearchCV
from scipy.stats import randint, uniform
from sklearn.svm import SVC
from sklearn.model_selection import train_test_split
from sklearn.feature_selection import RFECV
import numpy as np

```

```
df = pd.read_csv('/content/drive/MyDrive/Dados/balanced_data_.csv')
clean_data = df.drop(columns= ['LAT','LON','lat','lon'])

#Feature Selection
#S1
sentinel_1_df = clean_data.loc[:, clean_data.columns.str.startswith('V')]
sentinel_1_df['class_ID'] = clean_data['class_ID']
X_s1 = sentinel_1_df.drop('class_ID', axis=1)
y_s1 = sentinel_1_df['class_ID']
Xs1_train, Xs1_test, ys1_train, ys1_test = train_test_split(X_s1, y_s1, test_size
    =0.30, random_state=42)

RF_base_model_s1 = RandomForestClassifier(random_state=42)
rfecv_RF_1 = RFECV(estimator=RF_base_model_s1, step=1, cv=5, scoring='accuracy')
rfecv_RF_1.fit(Xs1_train, ys1_train)

s1_feature_names = Xs1_train.columns
cv_scores_RF = rfecv_RF_1.cv_results_['mean_test_score']
s1_accuracy_changes = np.ediff1d(cv_scores_RF, to_begin=cv_scores_RF[0] -
    cv_scores_RF[0])

s1_feature_names = Xs1_train.columns
s1_feature_importances = rfecv_RF_1.estimator_.feature_importances_
feature_data = pd.DataFrame({
  'Feature': s1_feature_names,
  'Importance': np.concatenate(([0] * (len(s1_feature_names) - len(
    s1_feature_importances)), s1_feature_importances)),
  'Selected': ['Yes' if x else 'No' for x in rfecv_RF_1.support_],
  'Accuracy Change': s1_accuracy_changes
})
s1_feature_data = feature_data[['Feature', 'Selected', 'Importance', 'Accuracy
    Change']]


#S2
sentinel_2_df = clean_data.loc[:, ~clean_data.columns.str.startswith('V')]
sentinel_2_df['class_ID'] = clean_data['class_ID']
sentinel_2_df['class_ID'] = clean_data['class_ID']
X_s2 = sentinel_2_df.drop('class_ID', axis=1)
y_s2 = sentinel_2_df['class_ID']

RF_base_model_2 = RandomForestClassifier(random_state=42)
rfecv_RF_2 = RFECV(estimator=RF_base_model, step=1, cv=5, scoring='accuracy')
rfecv_RF_2.fit(Xs2_train, ys2_train)


cv_scores_RF_2 = rfecv_RF_2.cv_results_['mean_test_score']
s2_accuracy_changes = np.ediff1d(cv_scores_RF_2, to_begin=cv_scores_RF_2[0] -
    cv_scores_RF_2[0])
s2_feature_names = Xs2_train.columns
s2_feature_importances = rfecv_RF_2.estimator_.feature_importances_
feature_data_2 = pd.DataFrame({
  'Feature': s2_feature_names,
  'Importance': np.concatenate(([0] * (len(s2_feature_names) - len(
    s2_feature_importances)), s2_feature_importances)),
  'Selected': ['Yes' if x else 'No' for x in rfecv_RF_2.support_],
  'Accuracy Change': s2_accuracy_changes
})
```

```
s2_feature_data_2 = feature_data_2[['Feature', 'Selected', 'Importance', '
    Accuracy Change']]


s2_feature_data_2.to_csv('/content/drive/MyDrive/Dados/S1_feature_selection.csv')
s1_feature_data.to_csv('/content/drive/MyDrive/Dados/sS2_feature_selection.csv')


#RF HYPERPARAMTER OPTIMIZATION
X = clean_data.drop('class_ID', axis=1)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.30,
    random_state=42)

p_distributions_RF = {
  'n_estimators': randint(100, 500),
  'max_features': uniform(0.1, 0.9),
  'max_depth': randint(5, 50),
  'min_samples_split': uniform(0.01, 0.1),
  'min_samples_leaf': uniform(0.01, 0.1),
  'bootstrap': [True, False]
}

RF_classifier = RandomForestClassifier()

random_search_RF = RandomizedSearchCV(
estimator=RF_classifier,
param_distributions=p_distributions_RF,
n_iter=10,
cv=5,
random_state=42,
n_jobs=-1
)

random_search_RF.fit(X_train, y_train)

best_RF_model = random_search_RF.best_estimator_
best_RF_model_score = best_RF_model.score(X_test, y_test)
print("Performance of the best RF model on the test set:", best_RF_model_score)
best_RF_parameters = random_search_RF.best_params_
print('Best_Hyperparameter:', best_RF_parameters)

#Standardization Min Leaf Population and Variables Per Split for Entry into GEE
total_samples = len(X_train)
min_leaf_fraction = best_RF_parameters['min_samples_leaf']
min_leaf_population = int(round(total_samples * min_leaf_fraction))
print(min_leaf_population)
total_features = X_train.shape[1]
fration = best_RF_parameters['max_features']
variablesPerSplit = int(round(total_features * fration))
variablesPerSplit
```

C.2.2.2 In Random Support Vector Machine Models

```
import numpy as np
import pandas as pd
from sklearn.model_selection import RandomizedSearchCV
```

```
from scipy.stats import randint, uniform
from sklearn.svm import SVC
from sklearn.model_selection import train_test_split
from sklearn.feature_selection import RFECV
from sklearn.svm import SVC
from sklearn.preprocessing import StandardScaler
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import StandardScaler

df = pd.read_csv('/content/drive/MyDrive/Dados/balanced_data.csv')
clean_data = df.drop(columns= ['LAT','LON','lat','lon'])

#FEATURE SELECTION
sentinel_1_df = clean_data.loc[:, clean_data.columns.str.startswith('V')] #
    Because of all Sentinel-1 Features Starts with 'V'.
sentinel_1_df['class_ID'] = clean_data['class_ID']
X_s1 = sentinel_1_df.drop('class_ID', axis=1)
y_s1 = sentinel_1_df['class_ID']
Xs1_train, Xs1_test, ys1_train, ys1_test = train_test_split(X_s1, y_s1, test_size
    =0.30, random_state=42)

pipe_s1 = Pipeline([
('scaler', StandardScaler()),
('svm', SVC(kernel='linear', random_state=42))
])
rfecv_SVM_for_s1 = RFECV(
estimator=pipe_s1,
step=1,
cv=5,
scoring='accuracy',
importance_getter='named_steps.svm.coef_'
)
rfecv_SVM_for_s1.fit(Xs1_train, ys1_train)


s1_feature_names = Xs1_train.columns
cv_scores_SVM_1 = rfecv_SVM_for_s1.cv_results_['mean_test_score']

kernel = pipe_s1.named_steps['svm'].kernel
if kernel == 'linear':
svm_model = rfecv_SVM_for_s1.estimator_.named_steps['svm']
coef_matrix_1 = svm_model.coef_
importance_s1 = np.mean(np.abs(coef_matrix_1), axis=0)
selected_features = Xs1_train.columns[rfecv_SVM_for_s1.support_]
feature_importance = dict(zip(selected_features, importance_s1))
feature_data_1 = pd.DataFrame({
  'Feature': Xs1_train.columns,
  'Importance': [feature_importance.get(x, 0) for x in Xs1_train.columns],
  'Selected': ['Yes' if x else 'No' for x in rfecv_SVM_for_s1.support_],
  'Accuracy Change': np.ediff1d(rfecv_SVM_for_s1.cv_results_['mean_test_score'],
    to_begin=0)
})
else:
feature_data_1 = pd.DataFrame({
  'Feature': Xs1_train.columns,
  'Importance': [0] * len(Xs1_train.columns),
  'Selected': ['Yes' if x else 'No' for x in rfecv_SVM_for_s1.support_],
```

```
  'Accuracy Change': np.ediff1d(rfecv_SVM_for_s1.cv_results_['mean_test_score'], to_begin=0)
})


sentinel_2_df = clean_data.loc[:, ~clean_data.columns.str.startswith('V')]
sentinel_2_df['class_ID'] = clean_data['class_ID']
sentinel_2_df['class_ID'] = clean_data['class_ID']
X_s2 = sentinel_2_df.drop('class_ID', axis=1)
y_s2 = sentinel_2_df['class_ID']
Xs2_train, Xs2_test, ys2_train, ys2_test = train_test_split(X_s2, y_s2, test_size=0.30, random_state=42)

pipe_s2 = Pipeline([
('scaler', StandardScaler()),
('svm', SVC(kernel='linear', random_state=42))
])

rfecv_SVM_for_s2 = RFECV(
estimator=pipe_s2,
step=1,
cv=5,
scoring='accuracy',
importance_getter='named_steps.svm.coef_'
)
rfecv_SVM_for_s2.fit(Xs2_train, ys2_train)

s2_feature_names = Xs2_train.columns
cv_scores_SVM_2 = rfecv_SVM_for_s2.cv_results_['mean_test_score']

kernel_2 = pipe_s2.named_steps['svm'].kernel

if kernel == 'linear':
svm_model_2 = rfecv_SVM_for_s2.estimator_.named_steps['svm']
coef_matrix_2 = svm_model_2.coef_
importance_s2 = np.mean(np.abs(coef_matrix_2), axis=0)
selected_features = Xs2_train.columns[rfecv_SVM_for_s2.support_]
feature_importance = dict(zip(selected_features, importance_s2))
feature_data_2 = pd.DataFrame({
  'Feature': Xs2_train.columns,
  'Importance': [feature_importance.get(x, 0) for x in Xs2_train.columns],
  'Selected': ['Yes' if x else 'No' for x in rfecv_SVM_for_s2.support_],
  'Accuracy Change': np.ediff1d(rfecv_SVM_for_s2.cv_results_['mean_test_score'], to_begin=0)
})
else:
feature_data_2 = pd.DataFrame({
  'Feature': Xs2_train.columns,
  'Importance': [0] * len(Xs2_train.columns),
  'Selected': ['Yes' if x else 'No' for x in rfecv_SVM_for_s2.support_],
  'Accuracy Change': np.ediff1d(rfecv_SVM_for_s2.cv_results_['mean_test_score'], to_begin=0)
})


feature_data_1.to_csv('/content/drive/MyDrive/Dados_SVM/s1_fetaure_selection.csv')
```

```
feature_data_2.to_csv('/content/drive/MyDrive/Dados_SVM/s2_feature_selection.csv')

#HYPERPARAMTERS OPTIMIZATION
scaler = StandardScaler()
X_s1_scalested = scaler.fit_transform(X_s1)
X_s1_scalested = pd.DataFrame(X_s1_scalested, columns=X_s1.columns)
X_s2_scalested = scaler.fit_transform(X_s2)
X_s2_scalested = pd.DataFrame(X_s2_scalested, columns=X_s2.columns)
s1_s2_df = pd.concat([X_s1_scalested, X_s2_scalested], axis=1)
s1_s2_df['class_ID'] = y_s1

X_train_scalested, X_test_scalested, y_train_scalested, y_test_scalested = train_test_split(
s1_s2_df.drop('class_ID', axis=1), s1_s2_df['class_ID'], test_size=0.30, random_state=42
)

svm_classifier = SVC(kernel='linear')
param_distributions = {
  'C': uniform(0.1, 10)
}
random_search = RandomizedSearchCV(
estimator=svm_classifier,
param_distributions=param_distributions,
n_iter=10,
cv=5,
random_state=42,
n_jobs=-1
)
random_search.fit(X_train_scalested, y_train_scalested)
best_parameters = random_search.best_params_
print('Best SVM Hyperparameters:', best_parameters)
```

Appendix D

Reference Data Acquisition

Figure D.1: Reference Data Acquisition Equipment

Appendix E

Satellite Image Acquisition

```
Satellite Image Time Series

    Iteration 1
        Interval of Days (May 1st to May 10th)
            Times Series

                Sentinel-1
                    Number of Images:
                        4
                    Acquisition Date:
                        0: 2023-05-01T18:27:42
                        1: 2023-05-01T18:28:07
                        2: 2023-05-06T18:36:09
                        3: 2023-05-07T06:34:54

                Sentinel-2:
                    Number of Images
                        2

                    Acquision Date:
                        0: 2023-05-04T11:30:15
                        1: 2023-05-07T11:40:10


    Iteration 2
        Interval of Days (May 1st to May 20th)
            Times Series
                Sentinel-1
                    Number of Images:
                        9
                    Acquisition Date:
                        0: 2023-05-01T18:27:42
                        1: 2023-05-01T18:28:07
                        2: 2023-05-06T18:36:09
                        3: 2023-05-07T06:34:54
                        4: 2023-05-12T06:43:08
                        5: 2023-05-13T18:27:43
                        6: 2023-05-13T18:28:08
                        7: 2023-05-18T18:36:10
                        8: 2023-05-19T06:34:55
                Sentinel-2:
```

```
                    Number of Images
                        3
                    Acquision Date:
                        0: 2023-05-12T11:40:13
                        1: 2023-05-17T11:40:12
                        2: 2023-05-19T11:30:19

    Iteration 3
        Interval of Days (May 1st - May 30th)
            Times Series
                Sentinel-1
                    Number of Images:
                        12
                    Acquisition Date:
                        0: 2023-05-01T18:27:42
                        1: 2023-05-01T18:28:07
                        2: 2023-05-06T18:36:09
                        3: 2023-05-07T06:34:54
                        4: 2023-05-12T06:43:08
                        5: 2023-05-13T18:27:43
                        6: 2023-05-13T18:28:08
                        7: 2023-05-18T18:36:10
                        8: 2023-05-19T06:34:55
                        9: 2023-05-24T06:43:09
                        10: 2023-05-25T18:27:43
                        11: 2023-05-25T18:28:08
                Sentinel-2:
                    Number of Images
                        3
                    Acquision Date:
                        0: 2023-05-12T11:40:13
                        1: 2023-05-17T11:40:12
                        2: 2023-05-19T11:30:19

    Iteration 4
        Interval of Days (May 1st to June 10th)
            Times Series
                Sentinel-1
                    Number of Images:
                        17
                    Acquisition Date:
                        0: 2023-05-01T18:27:42
                        1: 2023-05-01T18:28:07
                        2: 2023-05-06T18:36:09
                        3: 2023-05-07T06:34:54
                        4: 2023-05-12T06:43:08
                        5: 2023-05-13T18:27:43
                        6: 2023-05-13T18:28:08
                        7: 2023-05-18T18:36:10
                        8: 2023-05-19T06:34:55
                        9: 2023-05-24T06:43:09
                        10: 2023-05-25T18:27:43
                        11: 2023-05-25T18:28:08
                        12: 2023-05-30T18:36:10
                        13: 2023-05-31T06:34:56
                        14: 2023-06-05T06:43:09
                        15: 2023-06-06T18:27:44
                        16: 2023-06-06T18:28:09
```

```
                Sentinel-2:
                    Number of Images:
                        3
                    Acquision Date:
                        0. 2023-05-12T11:40:13
                        1. 2023-05-17T11:40:12
                        2. 023-05-19T11:30:19

    Iteration 5
        Interval of Days (May 1st - June 20th)
            Times Series
                Sentinel-1
                    Number of Images:
                        22
                    Acquisition Date:
                        0: 2023-05-01T18:27:42
                        1: 2023-05-01T18:28:07
                        2: 2023-05-06T18:36:09
                        3: 2023-05-07T06:34:54
                        4: 2023-05-12T06:43:08
                        5: 2023-05-13T18:27:43
                        6: 2023-05-13T18:28:08
                        7: 2023-05-18T18:36:10
                        8: 2023-05-19T06:34:55
                        9: 2023-05-24T06:43:09
                        10: 2023-05-25T18:27:43
                        11: 2023-05-25T18:28:08
                        12: 2023-05-30T18:36:10
                        13: 2023-05-31T06:34:56
                        14: 2023-06-05T06:43:09
                        15: 2023-06-06T18:27:44
                        16: 2023-06-06T18:28:09
                        17: 2023-06-11T18:36:11
                        18: 2023-06-12T06:34:56
                        19: 2023-06-17T06:43:10
                        20: 2023-06-18T18:27:45
                        21: 2023-06-18T18:28:10
                Sentinel-2:
                    Number of Images
                        3
                    Acquision Date:
                        0: 2023-05-12T11:40:13
                        1: 2023-05-17T11:40:12
                        2: 2023-05-19T11:30:19

    Iteration 6
        Interval of Days (May 1st - June 30th)
            Times Series
                Sentinel-1
                    Number of Images:
                        22
                    Acquisition Date:
                        0: 2023-05-01T18:27:42
                        1: 2023-05-01T18:28:07
                        2: 2023-05-06T18:36:09
                        3: 2023-05-07T06:34:54
                        4: 2023-05-12T06:43:08
                        5: 2023-05-13T18:27:43
```

```
                        6: 2023-05-13T18:28:08
                        7: 2023-05-18T18:36:10
                        8: 2023-05-19T06:34:55
                        9: 2023-05-24T06:43:09
                        10: 2023-05-25T18:27:43
                        11: 2023-05-25T18:28:08
                        12: 2023-05-30T18:36:10
                        13: 2023-05-31T06:34:56
                        14: 2023-06-05T06:43:09
                        15: 2023-06-06T18:27:44
                        16: 2023-06-06T18:28:09
                        17: 2023-06-11T18:36:11
                        18: 2023-06-12T06:34:56
                        19: 2023-06-17T06:43:10
                        20: 2023-06-18T18:27:45
                        21: 2023-06-18T18:28:10
                        22: 2023-06-23T18:36:12
                        23: 2023-06-24T06:34:57
                        24: 2023-06-29T06:43:10
                Sentinel-2:
                    Number of Images
                        6
                    Acquision Date:
                        0: 2023-05-12T11:40:13
                        1: 2023-05-17T11:40:12
                        2: 2023-05-19T11:30:19
                        3: 2023-06-23T11:30:18
                        4: 2023-06-26T11:40:14
                        5: 2023-06-28T11:30:19

    Iteration 7
        Interval of Days (May 1st - July 10th)
            Times Series
                Sentinel-1
                    Number of Images:
                        29
                    Acquisition Date:
                    0:  2023-05-01T18:27:42
                    1: 2023-05-01T18:28:07
                    2: 2023-05-06T18:36:09
                    3: 2023-05-07T06:34:54
                    4: 2023-05-12T06:43:08
                    5: 2023-05-13T18:27:43
                    6: 2023-05-13T18:28:08
                    7: 2023-05-18T18:36:10
                    8: 2023-05-19T06:34:55
                    9: 2023-05-24T06:43:09
                    10: 2023-05-25T18:27:43
                    11: 2023-05-25T18:28:08
                    12: 2023-05-30T18:36:10
                    13: 2023-05-31T06:34:56
                    14: 2023-06-05T06:43:09
                    15: 2023-06-06T18:27:44
                    16: 2023-06-06T18:28:09
                    17: 2023-06-11T18:36:11
                    18: 2023-06-12T06:34:56
                    19: 2023-06-17T06:43:10
                    20: 2023-06-18T18:27:45
```

```
                    21: 2023-06-18T18:28:10
                    22: 2023-06-23T18:36:12
                    23: 2023-06-24T06:34:57
                    24: 2023-06-29T06:43:10
                    25: 2023-06-30T18:27:45
                    26: 2023-06-30T18:28:10
                    27: 2023-07-05T18:36:12
                    28: 2023-07-06T06:34:58

                Sentinel-2:
                    Number of Images
                        9
                    Acquision Date:
                        0: 2023-05-12T11:40:13
                        1: 2023-05-17T11:40:12
                        2: 2023-05-19T11:30:19
                        3: 2023-06-23T11:30:18
                        4: 2023-06-26T11:40:14
                        5: 2023-06-28T11:30:19
                        6: 2023-07-01T11:40:15
                        7: 2023-07-03T11:30:19
                        8: 2023-07-06T11:40:15


    Iteration 8
        Interval of Days (May 1st to July 20th)
            Times Series
                Sentinel-1
                    Number of Images:
                        34
                    Acquisition Date:
                        0: 2023-05-01T18:27:42
                        1: 2023-05-01T18:28:07
                        2: 2023-05-06T18:36:09
                        3: 2023-05-07T06:34:54
                        4: 2023-05-12T06:43:08
                        5: 2023-05-13T18:27:43
                        6: 2023-05-13T18:28:08
                        7: 2023-05-18T18:36:10
                        8: 2023-05-19T06:34:55
                        9: 2023-05-24T06:43:09
                        10: 2023-05-25T18:27:43
                        11: 2023-05-25T18:28:08
                        12: 2023-05-30T18:36:10
                        13: 2023-05-31T06:34:56
                        14: 2023-06-05T06:43:09
                        15: 2023-06-06T18:27:44
                        16: 2023-06-06T18:28:09
                        17: 2023-06-11T18:36:11
                        18: 2023-06-12T06:34:56
                        19: 2023-06-17T06:43:10
                        20: 2023-06-18T18:27:45
                        21: 2023-06-18T18:28:10
                        22: 2023-06-23T18:36:12
                        23: 2023-06-24T06:34:57
                        24: 2023-06-29T06:43:10
                        25: 2023-06-30T18:27:45
                        26: 2023-06-30T18:28:10
```

```
                        27: 2023-07-05T18:36:12
                        28: 2023-07-06T06:34:58
                        29: 2023-07-11T06:43:12
                        30: 2023-07-12T18:27:42
                        31: 2023-07-12T18:28:07
                        32: 2023-07-17T18:36:13
                        33: 2023-07-18T06:34:58
                Sentinel-2:
                    Number of Images:
                        13
                    Acquision Date:
                        0: 2023-05-12T11:40:13
                        1: 2023-05-17T11:40:12
                        2: 2023-05-19T11:30:19
                        3: 2023-06-23T11:30:18
                        4: 2023-06-26T11:40:14
                        5: 2023-06-28T11:30:19
                        6: 2023-07-01T11:40:15
                        7: 2023-07-03T11:30:19
                        8: 2023-07-06T11:40:15
                        9: 2023-07-11T11:40:15
                        10: 2023-07-13T11:30:19
                        11: 2023-07-16T11:40:15
                        12: 2023-07-18T11:30:20

    Iteration 9
        Interval of Days (May 1st to July 30th)
            Times Series
                Sentinel-1
                    Number of Images:
                        37
                    Acquisition Date:
                        0: 2023-05-01T18:27:42
                        1: 2023-05-01T18:28:07
                        2: 2023-05-06T18:36:09
                        3: 2023-05-07T06:34:54
                        4: 2023-05-12T06:43:08
                        5: 2023-05-13T18:27:43
                        6: 2023-05-13T18:28:08
                        7: 2023-05-18T18:36:10
                        8: 2023-05-19T06:34:55
                        9: 2023-05-24T06:43:09
                        10: 2023-05-25T18:27:43
                        11: 2023-05-25T18:28:08
                        12: 2023-05-30T18:36:10
                        13: 2023-05-31T06:34:56
                        14: 2023-06-05T06:43:09
                        15: 2023-06-06T18:27:44
                        16: 2023-06-06T18:28:09
                        17: 2023-06-11T18:36:11
                        18: 2023-06-12T06:34:56
                        19: 2023-06-17T06:43:10
                        20: 2023-06-18T18:27:45
                        21: 2023-06-18T18:28:10
                        22: 2023-06-23T18:36:12
                        23: 2023-06-24T06:34:57
                        24: 2023-06-29T06:43:10
                        25: 2023-06-30T18:27:45
```

```
                        26: 2023-06-30T18:28:10
                        27: 2023-07-05T18:36:12
                        28: 2023-07-06T06:34:58
                        29: 2023-07-11T06:43:12
                        30: 2023-07-12T18:27:42
                        31: 2023-07-12T18:28:07
                        32: 2023-07-17T18:36:13
                        33: 2023-07-18T06:34:58
                        34: 2023-07-23T06:43:12
                        35: 2023-07-24T18:27:47
                        36: 2023-07-24T18:28:12
                        37: 2023-07-29T18:36:14

                Sentinel-2:
                    Number of Images:
                        16
                    Acquision Date:
                        0: 2023-05-12T11:40:13
                        1: 2023-05-17T11:40:12
                        2: 2023-05-19T11:30:19
                        3: 2023-06-23T11:30:18
                        4: 2023-06-26T11:40:14
                        5: 2023-06-28T11:30:19
                        6: 2023-07-01T11:40:15
                        7: 2023-07-03T11:30:19
                        8: 2023-07-06T11:40:15
                        9: 2023-07-11T11:40:15
                        10: 2023-07-13T11:30:19
                        11: 2023-07-16T11:40:15
                        12: 2023-07-18T11:30:20
                        13: 2023-07-23T11:30:19
                        14: 2023-07-26T11:40:15
                        15: 2023-07-28T11:30:20

    Iteration 10
        Interval of Days (May 1st to  August 10th)
            Times Series
                Sentinel-1
                    Number of Images:
                        42
                    Acquisition Date:
                        0: 2023-05-01T18:27:42
                        1: 2023-05-01T18:28:07
                        2: 2023-05-06T18:36:09
                        3: 2023-05-07T06:34:54
                        4: 2023-05-12T06:43:08
                        5: 2023-05-13T18:27:43
                        6: 2023-05-13T18:28:08
                        7: 2023-05-18T18:36:10
                        8: 2023-05-19T06:34:55
                        9: 2023-05-24T06:43:09
                        10: 2023-05-25T18:27:43
                        11: 2023-05-25T18:28:08
                        12: 2023-05-30T18:36:10
                        13: 2023-05-31T06:34:56
                        14: 2023-06-05T06:43:09
                        15: 2023-06-06T18:27:44
                        16: 2023-06-06T18:28:09
```

```
                        17: 2023-06-11T18:36:11
                        18: 2023-06-12T06:34:56
                        19: 2023-06-17T06:43:10
                        20: 2023-06-18T18:27:45
                        21: 2023-06-18T18:28:10
                        22: 2023-06-23T18:36:12
                        23: 2023-06-24T06:34:57
                        24: 2023-06-29T06:43:10
                        25: 2023-06-30T18:27:45
                        26: 2023-06-30T18:28:10
                        27: 2023-07-05T18:36:12
                        28: 2023-07-06T06:34:58
                        29: 2023-07-11T06:43:12
                        30: 2023-07-12T18:27:42
                        31: 2023-07-12T18:28:07
                        32: 2023-07-17T18:36:13
                        33: 2023-07-18T06:34:58
                        34: 2023-07-23T06:43:12
                        35: 2023-07-24T18:27:47
                        36: 2023-07-24T18:28:12
                        37: 2023-07-29T18:36:14
                        38: 2023-07-30T06:34:59
                        39: 2023-08-04T06:43:13
                        40: 2023-08-05T18:27:48
                        41: 2023-08-05T18:28:13
                Sentinel-2:
                    Number of Images
                        19

                    Acquision Date:
                        0: 2023-05-12T11:40:13
                        1: 2023-05-17T11:40:12
                        2: 2023-05-19T11:30:19
                        3: 2023-06-23T11:30:18
                        4: 2023-06-26T11:40:14
                        5: 2023-06-28T11:30:19
                        6: 2023-07-01T11:40:15
                        7: 2023-07-03T11:30:19
                        8: 2023-07-06T11:40:15
                        9: 2023-07-11T11:40:15
                        10: 2023-07-13T11:30:19
                        11: 2023-07-16T11:40:15
                        12: 2023-07-18T11:30:20
                        13: 2023-07-23T11:30:19
                        14: 2023-07-26T11:40:15
                        15: 2023-07-28T11:30:20
                        16: 2023-07-31T11:40:15
                        17: 2023-08-05T11:40:15
                        18: 2023-08-07T11:30:20

    Iteration 11
        Interval of Days (May 1st to August 20th)
            Times Series
                Sentinel-1
                    Number of Images:
                        47
                    Acquisition Date:
                        0: 2023-05-01T18:27:42
```

```
                        1: 2023-05-01T18:28:07
                        2: 2023-05-06T18:36:09
                        3: 2023-05-07T06:34:54
                        4: 2023-05-12T06:43:08
                        5: 2023-05-13T18:27:43
                        6: 2023-05-13T18:28:08
                        7: 2023-05-18T18:36:10
                        8: 2023-05-19T06:34:55
                        9: 2023-05-24T06:43:09
                        10: 2023-05-25T18:27:43
                        11: 2023-05-25T18:28:08
                        12: 2023-05-30T18:36:10
                        13: 2023-05-31T06:34:56
                        14: 2023-06-05T06:43:09
                        15: 2023-06-06T18:27:44
                        16: 2023-06-06T18:28:09
                        17: 2023-06-11T18:36:11
                        18: 2023-06-12T06:34:56
                        19: 2023-06-17T06:43:10
                        20: 2023-06-18T18:27:45
                        21: 2023-06-18T18:28:10
                        22: 2023-06-23T18:36:12
                        23: 2023-06-24T06:34:57
                        24: 2023-06-29T06:43:10
                        25: 2023-06-30T18:27:45
                        26: 2023-06-30T18:28:10
                        27: 2023-07-05T18:36:12
                        28: 2023-07-06T06:34:58
                        29: 2023-07-11T06:43:12
                        30: 2023-07-12T18:27:42
                        31: 2023-07-12T18:28:07
                        32: 2023-07-17T18:36:13
                        33: 2023-07-18T06:34:58
                        34: 2023-07-23T06:43:12
                        35: 2023-07-24T18:27:47
                        36: 2023-07-24T18:28:12
                        37: 2023-07-29T18:36:14
                        38: 2023-07-30T06:34:59
                        39: 2023-08-04T06:43:13
                        40: 2023-08-05T18:27:48
                        41: 2023-08-05T18:28:13
                        42: 2023-08-10T18:36:14
                        43: 2023-08-11T06:34:59
                        44: 2023-08-16T06:43:13
                        45: 2023-08-17T18:27:48
                        46: 2023-08-17T18:28:13
                Sentinel-2:
                    Number of Images:
                        22
                    Acquision Date:
                        0: 2023-05-12T11:40:13
                        1: 2023-05-17T11:40:12
                        2: 2023-05-19T11:30:19
                        3: 2023-06-23T11:30:18
                        4: 2023-06-26T11:40:14
                        5: 2023-06-28T11:30:19
                        6: 2023-07-01T11:40:15
                        7: 2023-07-03T11:30:19
```

```
                        8: 2023-07-06T11:40:15
                        9: 2023-07-11T11:40:15
                        10: 2023-07-13T11:30:19
                        11: 2023-07-16T11:40:15
                        12: 2023-07-18T11:30:20
                        13: 2023-07-23T11:30:19
                        14: 2023-07-26T11:40:15
                        15: 2023-07-28T11:30:20
                        16: 2023-07-31T11:40:15
                        17: 2023-08-05T11:40:15
                        18: 2023-08-07T11:30:20
                        19: 2023-08-10T11:40:16
                        20: 2023-08-12T11:30:19
                        21: 2023-08-17T11:30:20
    Iteration 12
        Interval of Days (May 1st to August 30th)
            Times Series
                Sentinel-1
                    Number of Images:
                        52
                    Acquisition Date:
                        0: 2023-05-01T18:27:42
                        1: 2023-05-01T18:28:07
                        2: 2023-05-06T18:36:09
                        3: 2023-05-07T06:34:54
                        4: 2023-05-12T06:43:08
                        5: 2023-05-13T18:27:43
                        6: 2023-05-13T18:28:08
                        7: 2023-05-18T18:36:10
                        8: 2023-05-19T06:34:55
                        9: 2023-05-24T06:43:09
                        10: 2023-05-25T18:27:43
                        11: 2023-05-25T18:28:08
                        12: 2023-05-30T18:36:10
                        13: 2023-05-31T06:34:56
                        14: 2023-06-05T06:43:09
                        15: 2023-06-06T18:27:44
                        16: 2023-06-06T18:28:09
                        17: 2023-06-11T18:36:11
                        18: 2023-06-12T06:34:56
                        19: 2023-06-17T06:43:10
                        20: 2023-06-18T18:27:45
                        21: 2023-06-18T18:28:10
                        22: 2023-06-23T18:36:12
                        23: 2023-06-24T06:34:57
                        24: 2023-06-29T06:43:10
                        25: 2023-06-30T18:27:45
                        26: 2023-06-30T18:28:10
                        27: 2023-07-05T18:36:12
                        28: 2023-07-06T06:34:58
                        29: 2023-07-11T06:43:12
                        30: 2023-07-12T18:27:42
                        31: 2023-07-12T18:28:07
                        32: 2023-07-17T18:36:13
                        33: 2023-07-18T06:34:58
                        34: 2023-07-23T06:43:12
                        35: 2023-07-24T18:27:47
                        36: 2023-07-24T18:28:12
```

```
                        37: 2023-07-29T18:36:14
                        38: 2023-07-30T06:34:59
                        39: 2023-08-04T06:43:13
                        40: 2023-08-05T18:27:48
                        41: 2023-08-05T18:28:13
                        42: 2023-08-10T18:36:14
                        43: 2023-08-11T06:34:59
                        44: 2023-08-16T06:43:13
                        45: 2023-08-17T18:27:48
                        46: 2023-08-17T18:28:13
                        47: 2023-08-22T18:36:15
                        48: 2023-08-23T06:35:00
                        49: 2023-08-28T06:43:14
                        50: 2023-08-29T18:27:49
                        51: 2023-08-29T18:28:14
                Sentinel-2:
                    Number of Images
                        2
                    Acquision Date:
                        0: 2023-05-12T11:40:13
                        1: 2023-05-17T11:40:12
                        2: 2023-05-19T11:30:19
                        3: 2023-06-23T11:30:18
                        4: 2023-06-26T11:40:14
                        5: 2023-06-28T11:30:19
                        6: 2023-07-01T11:40:15
                        7: 2023-07-03T11:30:19
                        8: 2023-07-06T11:40:15
                        9: 2023-07-11T11:40:15
                        10: 2023-07-13T11:30:19
                        11: 2023-07-16T11:40:15
                        12: 2023-07-18T11:30:20
                        13: 2023-07-23T11:30:19
                        14: 2023-07-26T11:40:15
                        15: 2023-07-28T11:30:20
                        16: 2023-07-31T11:40:15
                        17: 2023-08-05T11:40:15
                        18: 2023-08-07T11:30:20
                        19: 2023-08-10T11:40:16
                        20: 2023-08-12T11:30:19
                        21: 2023-08-17T11:30:20
                        22: 2023-08-20T11:40:15
                        23: 2023-08-22T11:30:21
                        24: 2023-08-27T11:30:19
    Iteration 13
        Interval of Days (May 1st to September 10th)
            Times Series
                Sentinel-1
                    Number of Images:
                        55
                    Acquisition Date:
                        0: 2023-05-01T18:27:42
                        1: 2023-05-01T18:28:07
                        2: 2023-05-06T18:36:09
                        3: 2023-05-07T06:34:54
                        4: 2023-05-12T06:43:08
                        5: 2023-05-13T18:27:43
                        6: 2023-05-13T18:28:08
```

```
                        7: 2023-05-18T18:36:10
                        8: 2023-05-19T06:34:55
                        9: 2023-05-24T06:43:09
                        10: 2023-05-25T18:27:43
                        11: 2023-05-25T18:28:08
                        12: 2023-05-30T18:36:10
                        13: 2023-05-31T06:34:56
                        14: 2023-06-05T06:43:09
                        15: 2023-06-06T18:27:44
                        16: 2023-06-06T18:28:09
                        17: 2023-06-11T18:36:11
                        18: 2023-06-12T06:34:56
                        19: 2023-06-17T06:43:10
                        20: 2023-06-18T18:27:45
                        21: 2023-06-18T18:28:10
                        22: 2023-06-23T18:36:12
                        23: 2023-06-24T06:34:57
                        24: 2023-06-29T06:43:10
                        25: 2023-06-30T18:27:45
                        26: 2023-06-30T18:28:10
                        27: 2023-07-05T18:36:12
                        28: 2023-07-06T06:34:58
                        29: 2023-07-11T06:43:12
                        30: 2023-07-12T18:27:42
                        31: 2023-07-12T18:28:07
                        32: 2023-07-17T18:36:13
                        33: 2023-07-18T06:34:58
                        34: 2023-07-23T06:43:12
                        35: 2023-07-24T18:27:47
                        36: 2023-07-24T18:28:12
                        37: 2023-07-29T18:36:14
                        38: 2023-07-30T06:34:59
                        39: 2023-08-04T06:43:13
                        40: 2023-08-05T18:27:48
                        41: 2023-08-05T18:28:13
                        42: 2023-08-10T18:36:14
                        43: 2023-08-11T06:34:59
                        44: 2023-08-16T06:43:13
                        45: 2023-08-17T18:27:48
                        46: 2023-08-17T18:28:13
                        47: 2023-08-22T18:36:15
                        48: 2023-08-23T06:35:00
                        49: 2023-08-28T06:43:14
                        50: 2023-08-29T18:27:49
                        51: 2023-08-29T18:28:14
                        52: 2023-09-03T18:36:16
                        53: 2023-09-04T06:35:01
                        54: 2023-09-09T06:43:15
                Sentinel-2:
                    Number of Images
                        25
                    Acquision Date:
                        0: 2023-05-12T11:40:13
                        1: 2023-05-17T11:40:12
                        2: 2023-05-19T11:30:19
                        3: 2023-06-23T11:30:18
                        4: 2023-06-26T11:40:14
                        5: 2023-06-28T11:30:19
```

```
                        6: 2023-07-01T11:40:15
                        7: 2023-07-03T11:30:19
                        8: 2023-07-06T11:40:15
                        9: 2023-07-11T11:40:15
                        10: 2023-07-13T11:30:19
                        11: 2023-07-16T11:40:15
                        12: 2023-07-18T11:30:20
                        13: 2023-07-23T11:30:19
                        14: 2023-07-26T11:40:15
                        15: 2023-07-28T11:30:20
                        16: 2023-07-31T11:40:15
                        17: 2023-08-05T11:40:15
                        18: 2023-08-07T11:30:20
                        19: 2023-08-10T11:40:16
                        20: 2023-08-12T11:30:19
                        21: 2023-08-17T11:30:20
                        22: 2023-08-20T11:40:15
                        23: 2023-08-22T11:30:21
                        24: 2023-08-27T11:30:19

    Iteration 14
        Interval of Days (May 1st to September 20th)
            Times Series
                Sentinel-1
                    Number of Images:
                        59
                    Acquisition Date:
                        0: 2023-05-01T18:27:42
                        1: 2023-05-01T18:28:07
                        2: 2023-05-06T18:36:09
                        3: 2023-05-07T06:34:54
                        4: 2023-05-12T06:43:08
                        5: 2023-05-13T18:27:43
                        6: 2023-05-13T18:28:08
                        7: 2023-05-18T18:36:10
                        8: 2023-05-19T06:34:55
                        9: 2023-05-24T06:43:09
                        10: 2023-05-25T18:27:43
                        11: 2023-05-25T18:28:08
                        12: 2023-05-30T18:36:10
                        13: 2023-05-31T06:34:56
                        14: 2023-06-05T06:43:09
                        15: 2023-06-06T18:27:44
                        16: 2023-06-06T18:28:09
                        17: 2023-06-11T18:36:11
                        18: 2023-06-12T06:34:56
                        19: 2023-06-17T06:43:10
                        20: 2023-06-18T18:27:45
                        21: 2023-06-18T18:28:10
                        22: 2023-06-23T18:36:12
                        23: 2023-06-24T06:34:57
                        24: 2023-06-29T06:43:10
                        25: 2023-06-30T18:27:45
                        26: 2023-06-30T18:28:10
                        27: 2023-07-05T18:36:12
                        28: 2023-07-06T06:34:58
                        29: 2023-07-11T06:43:12
                        30: 2023-07-12T18:27:42
```

```
                        31: 2023-07-12T18:28:07
                        32: 2023-07-17T18:36:13
                        33: 2023-07-18T06:34:58
                        34: 2023-07-23T06:43:12
                        35: 2023-07-24T18:27:47
                        36: 2023-07-24T18:28:12
                        37: 2023-07-29T18:36:14
                        38: 2023-07-30T06:34:59
                        39: 2023-08-04T06:43:13
                        40: 2023-08-05T18:27:48
                        41: 2023-08-05T18:28:13
                        42: 2023-08-10T18:36:14
                        43: 2023-08-11T06:34:59
                        44: 2023-08-16T06:43:13
                        45: 2023-08-17T18:27:48
                        46: 2023-08-17T18:28:13
                        47: 2023-08-22T18:36:15
                        48: 2023-08-23T06:35:00
                        49: 2023-08-28T06:43:14
                        50: 2023-08-29T18:27:49
                        51: 2023-08-29T18:28:14
                        52: 2023-09-03T18:36:16
                        53: 2023-09-04T06:35:01
                        54: 2023-09-09T06:43:15
                        55: 2023-09-10T18:27:49
                        56: 2023-09-10T18:28:14
                        57: 2023-09-15T18:36:16
                        58: 2023-09-16T06:35:02
                Sentinel-2:
                    Number of Images
                        2
                    Acquision Date:
                        0: 2023-05-12T11:40:13
                        1: 2023-05-17T11:40:12
                        2: 2023-05-19T11:30:19
                        3: 2023-06-23T11:30:18
                        4: 2023-06-26T11:40:14
                        5: 2023-06-28T11:30:19
                        6: 2023-07-01T11:40:15
                        7: 2023-07-03T11:30:19
                        8: 2023-07-06T11:40:15
                        9: 2023-07-11T11:40:15
                        10: 2023-07-13T11:30:19
                        11: 2023-07-16T11:40:15
                        12: 2023-07-18T11:30:20
                        13: 2023-07-23T11:30:19
                        14: 2023-07-26T11:40:15
                        15: 2023-07-28T11:30:20
                        16: 2023-07-31T11:40:15
                        17: 2023-08-05T11:40:15
                        18: 2023-08-07T11:30:20
                        19: 2023-08-10T11:40:16
                        20: 2023-08-12T11:30:19
                        21: 2023-08-17T11:30:20
                        22: 2023-08-20T11:40:15
                        23: 2023-08-22T11:30:21
                        24: 2023-08-27T11:30:19
                        25: 2023-09-14T11:40:14
```

```
                        26: 2023-09-19T11:40:14
    Iteration 15
        Interval of Days (May 1st to September 30th)
            Times Series
                Sentinel-1
                    Number of Images:
                        64
                    Acquisition Date:
                        0: 2023-05-01T18:27:42
                        1: 2023-05-01T18:28:07
                        2: 2023-05-06T18:36:09
                        3: 2023-05-07T06:34:54
                        4: 2023-05-12T06:43:08
                        5: 2023-05-13T18:27:43
                        6: 2023-05-13T18:28:08
                        7: 2023-05-18T18:36:10
                        8: 2023-05-19T06:34:55
                        9: 2023-05-24T06:43:09
                        10: 2023-05-25T18:27:43
                        11: 2023-05-25T18:28:08
                        12: 2023-05-30T18:36:10
                        13: 2023-05-31T06:34:56
                        14: 2023-06-05T06:43:09
                        15: 2023-06-06T18:27:44
                        16: 2023-06-06T18:28:09
                        17: 2023-06-11T18:36:11
                        18: 2023-06-12T06:34:56
                        19: 2023-06-17T06:43:10
                        20: 2023-06-18T18:27:45
                        21: 2023-06-18T18:28:10
                        22: 2023-06-23T18:36:12
                        23: 2023-06-24T06:34:57
                        24: 2023-06-29T06:43:10
                        25: 2023-06-30T18:27:45
                        26: 2023-06-30T18:28:10
                        27: 2023-07-05T18:36:12
                        28: 2023-07-06T06:34:58
                        29: 2023-07-11T06:43:12
                        30: 2023-07-12T18:27:42
                        31: 2023-07-12T18:28:07
                        32: 2023-07-17T18:36:13
                        33: 2023-07-18T06:34:58
                        34: 2023-07-23T06:43:12
                        35: 2023-07-24T18:27:47
                        36: 2023-07-24T18:28:12
                        37: 2023-07-29T18:36:14
                        38: 2023-07-30T06:34:59
                        39: 2023-08-04T06:43:13
                        40: 2023-08-05T18:27:48
                        41: 2023-08-05T18:28:13
                        42: 2023-08-10T18:36:14
                        43: 2023-08-11T06:34:59
                        44: 2023-08-16T06:43:13
                        45: 2023-08-17T18:27:48
                        46: 2023-08-17T18:28:13
                        47: 2023-08-22T18:36:15
                        48: 2023-08-23T06:35:00
                        49: 2023-08-28T06:43:14
```

```
                        50: 2023-08-29T18:27:49
                        51: 2023-08-29T18:28:14
                        52: 2023-09-03T18:36:16
                        53: 2023-09-04T06:35:01
                        54: 2023-09-09T06:43:15
                        55: 2023-09-10T18:27:49
                        56: 2023-09-10T18:28:14
                        57: 2023-09-15T18:36:16
                        58: 2023-09-16T06:35:02
                        59: 2023-09-21T06:43:15
                        60: 2023-09-22T18:27:50
                        61: 2023-09-22T18:28:15
                        62: 2023-09-27T18:36:16
                        63: 2023-09-28T06:35:02
                Sentinel-2:
                    Number of Images
                        29
                    Acquision Date:
                        0: 2023-05-12T11:40:13
                        1: 2023-05-17T11:40:12
                        2: 2023-05-19T11:30:19
                        3: 2023-06-23T11:30:18
                        4: 2023-06-26T11:40:14
                        5: 2023-06-28T11:30:19
                        6: 2023-07-01T11:40:15
                        7: 2023-07-03T11:30:19
                        8: 2023-07-06T11:40:15
                        9: 2023-07-11T11:40:15
                        10: 2023-07-13T11:30:19
                        11: 2023-07-16T11:40:15
                        12: 2023-07-18T11:30:20
                        13: 2023-07-23T11:30:19
                        14: 2023-07-26T11:40:15
                        15: 2023-07-28T11:30:20
                        16: 2023-07-31T11:40:15
                        17: 2023-08-05T11:40:15
                        18: 2023-08-07T11:30:20
                        19: 2023-08-10T11:40:16
                        20: 2023-08-12T11:30:19
                        21: 2023-08-17T11:30:20
                        22: 2023-08-20T11:40:15
                        23: 2023-08-22T11:30:21
                        24: 2023-08-27T11:30:19
                        25: 2023-09-14T11:40:14
                        26: 2023-09-19T11:40:14
                        27: 2023-09-26T11:30:18
                        28: 2023-09-29T11:40:13
```