Vineyard Terrace Segmentation in the Douro Region Based on Satellite Imagery

Master degree in Data Science

Pedro António Sousa Pedrosa Lopes

Leiria, September of 2023

Vineyard Terrace Segmentation in the Douro Region Based on Satellite Imagery

Master degree in Data Science

Pedro António Sousa Pedrosa Lopes

Dissertation Report under the supervision of Professor Dr. José Martins, and Ana Oliveira

Leiria, September of 2023

Originality and Copyright

This dissertation report is original, made only for this purpose, and all authors whose studies and publications were used to complete it are duly acknowledged.

Partial reproduction of this document is authorized, provided that the Author is explicitly mentioned, as well as the study cycle, *i.e.*, Master degree in Data Science, 2022/2023 academic year, of the School of Technology and Management of the Polytechnic Institute of Leiria, and the date of the public presentation of this work.

Acknowledgments

I would like to express my heartfelt gratitude to my family for their encouragement, and to my friends and girlfriend for their constant support and motivation throughout the countless research and feedback sessions.

Additionally, I extend my deepest appreciation to CEiiA, my supervisors Ana Oliveira and Dr. José Martins, and the remaining CEiiA's Ocean&Space data science team members: Dr. Nelson Pires, Alberto Barbosa, Rute Santos and Pedro Fernandes, for their invaluable contributions, teachings and support throughout the course of this dissertation.

I also could not have achieved this milestone without the guidance and expertise of Miguel Arantes, for his involvement in this dissertation has been instrumental in shaping its direction.

Abstract

The Alto Douro Wine region holds the distinction of being a UNESCO World Heritage Site, known for its traditional vineyard terraces that contribute to its cultural significance. These terraces, engineered to support vine cultivation on the challenging slopes of the Douro valley, were affected by the Phylloxera pest outbreak in the 19th century, resulting in terrace reconstructions for disease control.

Preserving this cultural landscape requires periodic evaluations of the terraces, but current manual field assessments are time-consuming, costly, and prone to errors, leading to infrequent updates. To address these challenges, this dissertation studies alternative approaches using multispectral and SAR satellite imagery, and machine learning to detect and identify vineyards within the terraces, aiming to reduce costs and increase assessment frequency.

The study begins with a review of remote sensing and satellite imaging technologies, followed by a literature review on similar applications and techniques. Data acquisition details are provided, and three segmentation methodologies are explored: band indices, traditional machine learning (support vector machines and random forests) and deep learning (convolutional neural networks).

The deep learning approach, particularly the modified DeepLabV3 model with the ResNet-101 backbone yields the most promising results, despite generalization limitations. Combining the segmented vineyard mask with a slope mask derived from SAR altimetry data increases confidence in identifying vineyards within terraces, offering rough estimations on possible locations of vineyard terraces in the Douro region.

In conclusion, this study presents an alternative and cost-effective approach to preserve the heritage landscape of the Alto Douro Wine region. By leveraging satellite imagery and machine learning, it offers a practical and preliminary means for periodic evaluations, supporting the sustainable conservation of this culturally significant region.

Keywords: Machine learning, Deep learning, Remote sensing, Earth Observation, Alto Douro Vineyard

Resumo

A região do Alto Douro Vinhateiro tem a distinção de ser Património Mundial da UNESCO, conhecida pelos seus socalcos tradicionais de vinha que contribuem para a sua significância cultural. Estes socalcos, concebidos para suportar o cultivo das vinhas nas encostas do Douro, foram afetados pelo surto da praga filoxera no século XIX, resultando na reconstrução dos socalcos para controlo da doença.

A preservação desta paisagem cultural exige avaliações periódicas dos socalcos, contudo as atuais avaliações manuais no terreno são demoradas, dispendiosas e propensas a erros, o que leva a atualizações pouco frequentes. Para responder a estes desafios, esta dissertação estuda abordagens alternativas utilizando imagens de satélite multiespectrais e SAR, e ainda diferentes técnicas de *machine learning* para detetar e identificar vinhas nos socalcos, com o objetivo de reduzir os custos e aumentar a frequência destas avaliações.

O estudo começa com uma revisão das tecnologias de deteção remota e de imagens de satélite, seguida de uma revisão da literatura sobre aplicações e técnicas semelhantes. É explicada a aquisição dos dados e são exploradas três metodologias de segmentação: índices através de combinações de bandas, *machine learning* tradicional (*support vector machines* e *random forests*) e ainda *deep learning* (redes neurais convolucionais).

A abordagem de *deep learning*, em particular o modelo DeepLabV3 modificado com a *backbone* ResNet-101, produz os resultados mais promissores, apesar das limitações de generalização. A combinação da máscara da vinha segmentada com a máscara de declive derivada de dados altimétricos SAR aumenta a confiança na identificação de vinhas em socalcos, oferecendo estimativas sobre possíveis localizações de socalcos de vinha na região.

Em suma, este estudo apresenta uma abordagem alternativa para promover a preservação da paisagem patrimonial da região do Alto Douro Vinhateiro. Ao utilizar imagens de satélite e *machine learning*, é criado um meio prático para realizar avaliações periódicas preliminares, apoiando a conservação sustentável desta região com elevado interesse cultural.

Contents

List of Figures

List of Tables

List of Abbreviations and Acronyms

ADV	*Alto Douro Vinhateiro*
AI	Artificial Intelligence
ANN	Artificial Neural Network
ASPP	Atrous Spatial Pyramid Pooling
ASRS	Airborne Spaceborne Remote Sensing
AUC	Area Under the ROC Curve
BOA	Bottom of Atmosphere
CNN	Convolutional Neural Network
COS	Carta de Uso e Ocupação do Solo
CRS	Coordinate Reference System
DEM	Digital Elevation Model
EDA	Exploratory Data Analysis
EM	Electromagnetic
EO	Earth Observation
ESTG	School of Technology and Management
FCN	Fully Convolutional Network
FN	False Negative
FP	False Positive
FPR	False Positive Rate
GIS	Geographical Information System
HR	High Resolution
IOU	Intersection Over Union
ML	Machine Learning
NIR	Near Infrared
RBF	Radial Basis Function
RF	Random Forests
ROC	Receiver Operating Characteristic Curve
ROI	Region of Interest
SAR	Synthetic Aperture Radar
SMOTE	Synthetic Minority Over-sampling Technique

SNAP	Sentinel Application Platform
SRTM	Shuttle Radar Topography Mission
SVM	Support Vector Machines
TN	True Negative
TOA	Top of Atmosphere
TP	True Positive
TPR	True Positive Rate
UAV	Unmanned Aircraft Vehicle
VHR	Very-High Resolution
XAI	Explainable Artificial Intelligence

1. Introduction

The traditional vineyard terraces present in the landscape of the *Alto Douro* Wine Region (ADV) are a World Heritage Site designated by UNESCO for its cultural landscape significance (UNESCO, 2001).

Although this region around the Douro valley provides beautiful views that attract tourists every year, these man-made structures were not built with the purpose of elegance. To cultivate the vines in the steep, rocky slopes of the Douro valley and its tributaries, the designed solution was to build terraces, using schist walls to support the land (Aguiar, 2002).

With a careful analysis of the organization of vineyards within each terrace, two predominant vineyard terrace configurations emerge. These different configurations are attributed to an infestation by a pest that greatly impacted the global wine industry, particularly in Europe, at the end of the 19th century: the Phylloxera. Phylloxera is an insect pest that attacks the roots of grapevine species, which decimated several vineyards in the late 19th century, with great impact in the Douro vineyard region (Martins, 1991). Consequently, some of the Douro terraces had to be fully reconstructed as regularly shaped, long continuous lines, as to mitigate the propagation of this disease.

The pre-phylloxera terraces are generally horizontal, bearing one or two vine lines, whereas the post-phylloxera terraces were optimized, having a slight slope that provides more sunlight exposure to the vineyards, with wider and continuous terraces supported by taller walls (Carlos, *et al*., 2021).

Although this wine region had its boundaries delineated in 1756, making it one of the first delineated wine regions in the world (Lourenço-Gomes, Pinto, & Rebelo, 2015), wine has been produced in the region for over 2000 years (UNESCO, 2001).

To assure the preservation of this landscape heritage, periodic evaluations must be performed, assessing the areas that may have suffered any alterations, and checking for sites where maintenance of the terraces and its supporting walls is needed.

Currently, these evaluations are performed periodically in the field, where workers manually measure and estimate the preservation of the walls and terraces that support the vineyards. Such processes are prone to errors and deviations from the real values, providing low

confidence results, due to the nature of the respective measurement processes. Not only that, but this type of process requires a significant budget, considering the area covered by the cultural heritage site, the necessary human resources, and the periodicity of the implicit studies. This process also entails a significant time investment before yielding results.

Considering the aforementioned facts and the limited funding provided by UNESCO for such studies, it is reasonable to anticipate substantial intervals between these research initiatives. Prior to the publication of this dissertation, the most recent study was conducted in 2015 (UNESCO) and was performed in limited regions of the whole heritage site. It is also relevant to mention that the periodicity of the conducted studies is not constant, meaning that it is not possible to anticipate when the next study will be released.

In this dissertation, we propose an alternative approach to assess the preservation of this heritage landscape, by using multispectral and SAR satellite images of the site combined with machine learning techniques to detect and identify vineyards located within these terraces, and also measure the perimeter and area occupied by them. If successful, with this process, it is expected that the costs of assessing the preservation of the landscape and the time it takes to perform these assessments will be reduced, and the periodicity of terrain evaluations be performed more regularly.

We start by shortly reviewing the background knowledge, Chapter 2, that enables the reader to comprehend and follow the developed work, where context is provided as to where and why remote sensing satellites originated, and also succinctly explain other relevant topics for this dissertation including remote sensing, Earth Observation (EO) and satellite imaging technologies. In addition, a literature review on similar studies will be conducted in the same chapter, describing developed work with equivalent applications, methods used in those analysis and the state-of-the-art techniques used in the field.

In Chapter 3, the data acquisition process will be outlined, illustrating where and how the data used for this study was obtained, and the platforms used for this purpose.

Moving on to the methodology, we will firstly turn our attention to segmentation using band indices, Chapter 4, where a study of the best combination of bands to enhance vineyard-like spectral signature features will be performed, as to assess the efficiency of such methodology to segment vineyard regions.

In Chapter 5 the data preprocessing techniques and Exploratory Data Analysis (EDA) for the later machine learning tasks will be described, also thoroughly explaining the python package developed during the progress of this dissertation, with the purpose of preprocessing Sentinel-2 imagery for machine learning tasks.

After, in Chapter 6, we define the metrics that will be used to study the performance of the developed models, according to related segmentation articles and also considering the applications of our study.

Another methodology that will be studied is segmentation employing traditional machine learning techniques, Chapter 7, where support vector machines (SVM) and random forest (RF) algorithms will be tested, and conclusions about their predictive capabilities, computational requirements, and overall models training time discussed, based on the obtained results.

On the last chapter of the methodologies studied for segmenting vineyard regions on this dissertation, Chapter 8, the segmentation of vineyard features will be implemented with deep learning techniques, respectively convolutional neural networks. In this chapter, various models based on DeepLabV3 with a ResNet101 backbone will be analysed, and their metrics evaluated and compared. In the end, choosing the model with the best performance.

After generating the segmented mask obtained from the selected model, from the previous methodologies, that outperforms all the others, it will then be overlapped and intersected with a slope mask derived from a digital elevation model (DEM), Chapter 9, in order to filter out vineyards that are not contained in regions with steep geological characteristics, thereby ensuring a strong certainty that the vineyards are contained within terraces.

Finally, in Chapter 10, a discussion about each of the methodologies' results, with a focus on the best performing one will be made, highlighting the possible improvements that could be made to address the flaws observed in them. Within the same chapter, we will present the conclusions reached within the developed study, also acknowledging potential future implementations that could enhance the content's quality and facilitate the advancement of similar research.

2. Background Knowledge & Literature Review

For a better understanding of the contextual framework within this dissertation, some important concepts will be reviewed in the first part of this chapter. From the history of satellite imagery and the basics of remote sensing to GIS software tools and band indices. In short, the most relevant subjects in geospatial analysis that should be of prior knowledge before any work gets implemented will be reviewed and discussed in this chapter. We will then proceed to review the literature on topics related to the ones studied in this dissertation.

2.1. Satellite imagery history

2.1.1. History of Earth Observation satellite imagery

Prior to exploring the European Earth Observation satellite history, which will be a relevant subject of study in this dissertation, there are some global historic facts that should be considered in order to better grasp the origin of EO satellites on a global scale.

EO can be described as "the gathering of information about planet Earth's physical, chemical and biological systems via remote sensing technologies, usually involving satellites carrying imaging devices" (EU Science Hub, 2023).

The first satellite image of Earth was produced on August 14, 1959, by Explorer 6 (Hahn, 2009), a NASA satellite, and is presented on Figure 1.

Figure 1 - First satellite image, adapted from (Internet Archive,

Figure 1 marked a major milestone for human history in satellite imagery. After this satellite, more satellites from the same and different programs were launched with the similar goal of satellite imaging and EO, including the CORONA, LandSat-1 and many others, with the respective quality and volume of the data becoming increasingly higher.

Despite the first image being credited to the United States of America, it was the Soviet Union's Sputnik 1, launched in 1957, that heralded the era of satellite remote sensing (Tatem, Goetz, & Hay, 2008). The dispute between these two nations during the cold war fomented a race for knowledge and technological upper hand in respect to the other nation. Regardless of the obvious atrocious reasons of this war, it unquestionably brought useful improvements and developments in people's lives, with a special emphasis in technology, which also positively impacted the technological advancement of satellites.

2.1.2. History of European Earth Observation satellite imagery

Although the first efforts towards a better observation of Earth with satellites were made by nations such as the United States of America and the former Soviet Union, other nations soon followed with their own research for tools to monitor our planet.

It was on the 17th of July 1991 that the first European satellite for EO was launched into space from an Ariane 4 launch vehicle: the European Remote-Sensing ERS-1 Satellite. The ERS was a program of the European Space Agency (ESA) composed of two missions: ERS-1 and ERS-2, which had a Sun-synchronous polar-orbit, and together observed the Earth for 20 years (ESA, 2022), when they only had an expected 3-year life span.

The main goal of the ERS program was to provide microwave spectrum-based environmental monitoring (ESA, 2022), resorting to several instruments such as a synthetic aperture radar (SAR) for 2D spectra of ocean surface waves, a radar altimeter (RA) to perform altimetry and wind speed tasks on ocean and ice surfaces, and other useful instruments for EO.

During the period when the two satellites were operating simultaneously, achieving the same desired mission objectives, *i.e.*, in a tandem mission, a new radar technique at the time was tested: interferometry. Interferometry is the technique of collecting two or more samples of data (images) with the exact same geolocation and superimposing them, which creates an interferogram. These can be used to measure changes in land surface elevation at a millimetre scale, or just to measure surface elevation, depending on the used configuration.

Envisat was the successor to the ERS, which was launched on the 1st of March 2002, and it was the largest EO satellite until then. Its main goal was to provide continuity for the ERS program, with improvements to the measurements, and enhanced capabilities for remote sensing observation of Earth (ESA, 2022). This satellite was also equipped with a programmable spectrometer, the Medium Resolution Imaging Spectrometer (MERIS), which provided scientists with coloured images of our planet.

As time passed and previous missions life span's getting shorter, with some missions already retired, there came a need to develop a new program of satellites in order to replace the previous ones, so as to provide continuity of data. These series of new-generation EO missions are called Sentinel and are part of the Copernicus programme, with the first satellite (Sentinel-1A), belonging to the Sentinel-1 mission, launched in April 2014 (ESA, 2022).

To this day, the Copernicus programme comprises 7 Sentinel missions: Sentinel-1, Sentinel-2, Sentinel-3, Sentinel-4, Sentinel-5, Sentinel-5P and Sentinel-6. For this study, only imagery from the Sentinel-2 mission was used.

Unlike the common misconception, Sentinel missions are not ordered by technologic evolution, but rather by mission objectives, *i.e.*, the Sentinel-2 mission is not an upgrade of the Sentinel-1 mission, but rather a mission with a different scope, with satellites built with different technologies to study different planetary and atmospheric properties.

2.2.Remote Sensing

A good and general definition of remote sensing in the context of EO studies is provided by Oxford Languages (2022), stating that remote sensing is "the scanning of the Earth by satellite or high-flying aircraft in order to obtain information about it". This means that both satellites and high-flying aircraft such as unmanned aircraft vehicles (UAV's) can be used for remote sensing applications.

For the purpose of detecting vineyard terraces, remote sensing techniques will be utilized, meaning that we will use various types of bands from passive sensor imagery, and through the difference between spectral signatures of the different materials present in the surface of the Earth, we hope to better assess the regions where vineyard plantations are located, which then will be intersected with active sensor imagery to extract the ones contained within terraces.

In Chapter 3, different types of remote sensing satellites used for EO studies will be discussed, and also which ones we will be using for this dissertation.

2.3.Satellite imaging technologies

There are different types of technologies in satellites for imaging purposes, that largely depend on the scope of the study each satellite is going to perform throughout its lifespan. There are two main types of satellite instruments for remote sensing imagery:

- **Passive Sensors**, which are microwave instruments that were designed for receiving and measuring natural emissions produced and reflected by Earth and its atmosphere. These types of sensors can employ multispectral and hyperspectral instruments that create images useful in identifying different features of our planet, by using bands of the electromagnetic (EM) spectrum individually or combining them with the purpose of emphasizing desired features.

  There are different types of sensors to capture a different number of spectral bands with shorter or larger bandwidths in the EM spectrum. There is no clearly defined number of bands that separates a multispectral from a hyperspectral image, but typically multispectral images contain 2 to dozens of bands with relatively large bandwidths and hyperspectral images can contain hundreds and even thousands of bands with narrower bandwidths. What can be used to differentiate these two sensors is the type of technology used in each, which is very different from multispectral to hyperspectral imaging systems.

  There are different types of instruments that are classified as passive sensors, such as spectrometers, radiometers, spectroradiometers and accelerometers.

- **Active Sensors** emit microwave signals which can be reflected, refracted or backscattered by Earth's surface or its atmosphere, and captured again by the satellite. This type of instrument is useful to study Earth's surface regardless of meteorological conditions by using longer microwaves, providing the ability to penetrate through cloud cover, haze, dust and other conditions without significant interference in the signal. Another advantage of active sensors is that they can be used during day and night, providing consistent results since they don't rely on an external EM radiation source to produce data.

There are different types of instruments that are classified as active sensors, such as radars, lidars, lasers and scatterometers.

An illustration that visually represents the main differences between these two sensor technologies can be observed in Figure 2:

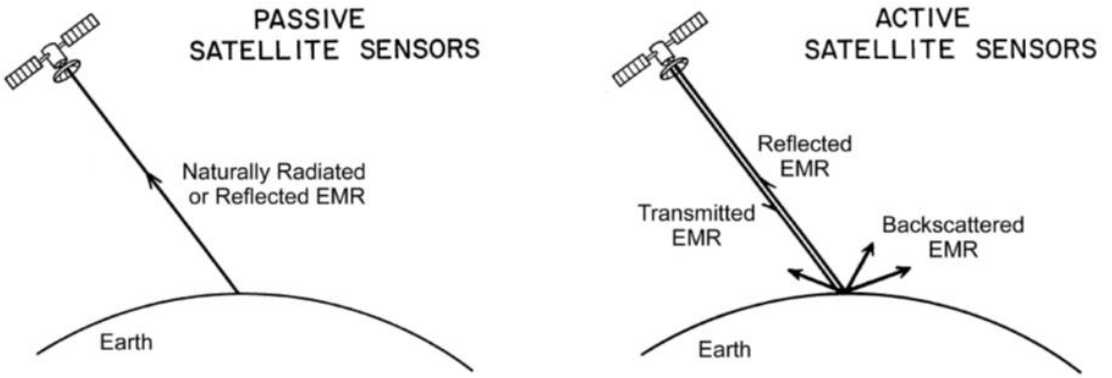

Figure 2 - Passive and active satellite sensors, adapted from (NOAA CoastWatch, 2021)

2.4.GIS

Geographical data, also described as spatial data, is typically stored in a computer system designated Geographic Information System (GIS). This computer system also manages the capture, check and display of data that has geographical characteristics.

2.4.1. GIS tools

To handle and manage geospatial data, GIS software tools are the most commonly used solution. Two of the main available tools are SNAP and QGIS, and, on the present study, both software were used to manage georeferenced data.

- **SNAP** is an acronym for ESA's Sentinel Application Platform and is a Geographical Information System tool used to access, process and analyse Sentinel and other ESA international mission's imagery. Despite its limited data access, SNAP is very user-friendly and is robustly built to work with Sentinel data, and for these reasons it was used to analyse and process ESA mission's data on this work.
- **QGIS** is an open-source Geographical Information System application that allows studies of geographical data from many sources. Unlike SNAP, QGIS is not limited

to a specific data provider, and has diverse functionalities and even third-party plugins that facilitate handling data. For this reason, it is widely used in mostly all geospatial data. On the present study, QGIS is used to manage processed Sentinel data, and to process and analyse data from multiple platforms.

2.4.2. GIS file formats

There are two main types of GIS data: vector and raster data.

- **Vector data** corresponds to data that is stored as points, lines, or polygons. In practice, these may correspond, for instance, to the representation of roads, lakes, and city boundaries in a map. As an example, Google uses vector tiles for its base map in Google maps, when using the '*Map view*' (Google, 2022), as observed in Figure 3:

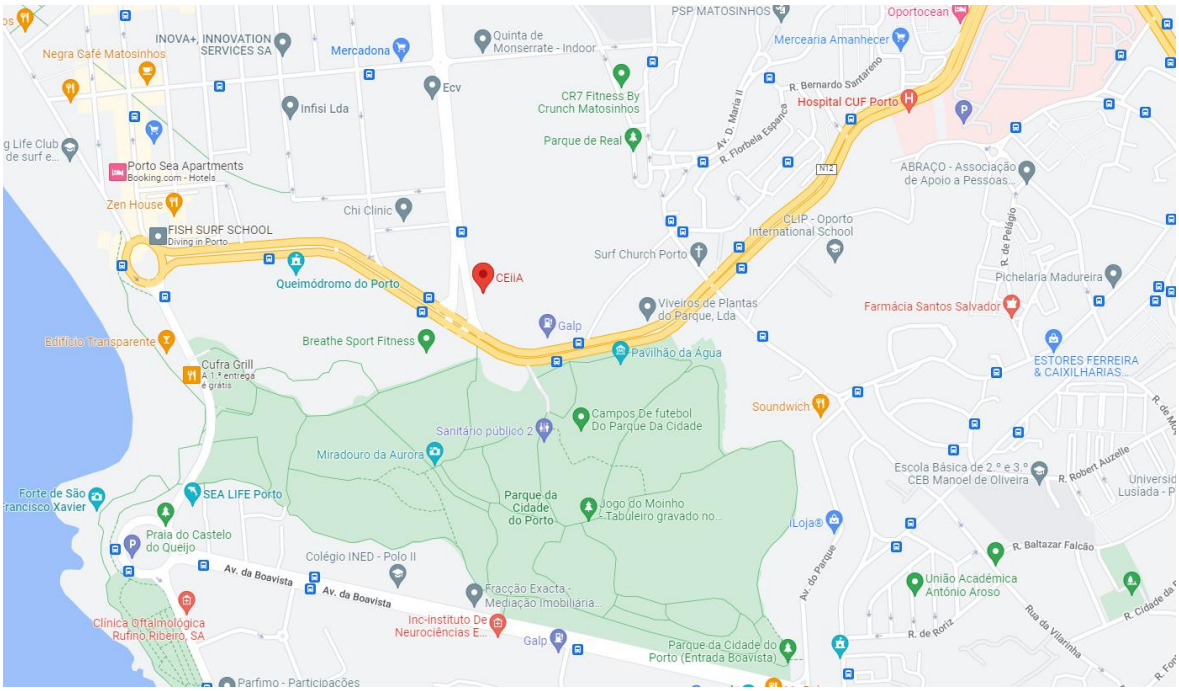

Figure 3 - Example vector data, adapted from (Google, 2022)

- **Raster data** corresponds to gridded data, in which each pixel of an image is associated with a geographical location and is stored in matrices. Data stored in this format represents physical phenomena, and each pixel can store continuous or discrete data. Continuous data can represent elevation or aspect of a location, whereas discrete data often represents objects that are present in that location. Raster data is used to represent digital data. One example are the image tiles present in Google maps when using the '*satellite view*', that takes the format of raster files (Google, 2022), as depicted in Figure 4:

Figure 4 - Example raster data, adapted from (Google, 2022)

2.5.Band indices

Band indices are methods used to combine different bands of the EM spectrum to provide useful representation of desired features in an image.

Different combinations of EM bands can provide different results in an image. Specific band indices are useful to enhance features like water, vegetation and vegetation health and others.

As an example, in Figure 5[1] a Normalized Differential Vegetation Index (NDVI) index was created using arithmetic operations between the near infrared (NIR) and red bands of a Sentinel-2 image of the Matosinhos region. This index is mostly used for vegetation analysis, since it quantifies and provides information about the presence, density, and vitality of vegetation features. In this example, we can clearly make a distinction between vegetation zones (in blueish tones) like the city park of Porto in the centre-left region of the image, and non-vegetation zones (yellow-greener zones) like the industrial zone of Porto at the central part of the image. Also, the sea is present in the left part of the image, represented in orange.

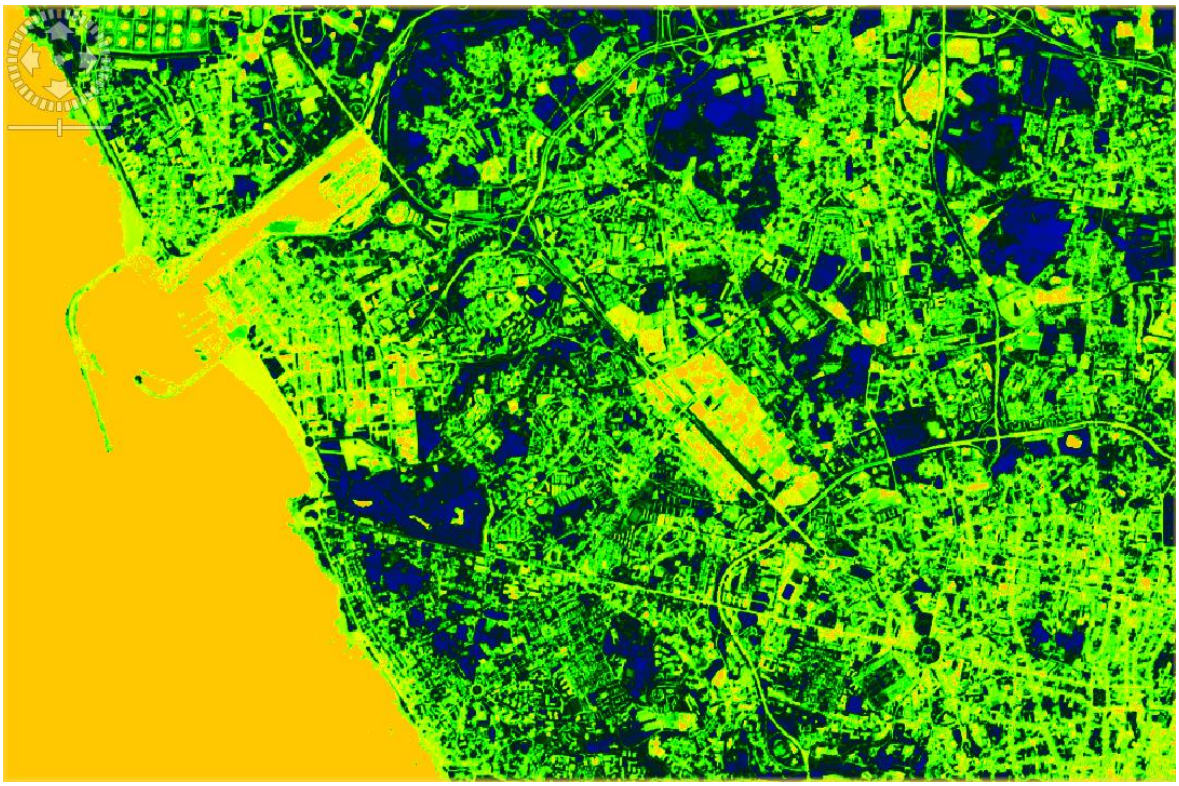

Figure 5 - Application of the NDVI over the Matosinhos region

2.6.Literature Review

Remote sensing techniques have been employed and improved over several decades. Although remote sensing was being used before with photogrammetric measurements, cameras and films, this term was only broadened and recognized in the early 1960's by American entities, respectively the National Aeronautics and Space Administration

[1] Derived from Copernicus Sentinel data 2023, processed by ESA.

(NASA), U.S. Department of Agriculture, U.S. Geological Survey (USGS) and the National Oceanic and Atmospheric Administration (NOAA) (Bauer, 2020).

Since the beginning of the century, airborne and spaceborne remote sensing (ASRS) have witnesses several novel applications (Lei Luo, 2019), which were also possible due to the technological advancements of the time. In this study, we will restrain our focus in spaceborne remote sensing, respectively using passive satellite data for vineyard segmentation and active satellite data for altimetry data.

Satellite remote sensing not only plays a vital role in the study of cultural heritage, but also in several other applications, some of the being environmental monitoring (Li, et al., 2020), resource exploration (Sabins, 1999) (Calvin, Littlefield, & Kratt, 2015), agricultural (Weiss, Jacob, & Duveiller, 2020) and land use (Rawat & Kumar, 2015). There are many other applications in both land and water regions.

Remote sensing satellite imaging techniques may further be divided into 3 categories: multi/hyperspectral imaging (Pettorelli, *et al*., 2005) (Brando & Dekker, 2003) (van der Meer, *et al*., 2012), LiDAR (Bhardwaj, Sam, Bhardwaj, & Martín-Torres, 2016), and SAR imaging (Chen, Lasaponara, & Masini, 2017). The first category relates to passive satellite sensors, and the last 2 categories relate to active satellite sensors.

Several articles have been developed in remote sensing and precision agriculture applications using band vegetation indices, carried out in plantations, including vineyards. According to Giovos *et al*., (2021), most of the articles focus on the application of vegetation indices to study water stress (Soubry, Patias, & Tsioukas, 2017), yield (Kandylakis & Karantzalos, 2016) and vine disease (Albetis, et al., 2017) on vineyards. In this study, we will assess their performance on detecting vineyard terraces using satellite imagery.

Throughout the literature research that was performed, most of the articles related to the identification of vineyard terraces using satellite images were limited, and its publications were quite recent, as such, with this dissertation we hope to provide our contribution to future studies related to this subject.

Although the study of vineyard terraces using satellite images combined with machine learning techniques is still an emerging topic, these techniques have been widely used in several other field applications. Many articles have been developed and published in the detection of different land features using said techniques, e.g., Ozlem and Güngör (2012)

studied the classification of different multispectral satellite image scenes using random forest algorithm. McAllister *et al*., (2022) used machine learning techniques for the extraction of shoreline indicators using multispectral satellite imagery, and Wang *et al*., (2022) studied snow coverage mapping via machine learning algorithms using Sentinel-2 data.

According to Lary *et al*., (2016), some of the most common machine learning algorithms used for classification in remote sensing applications are, among others, support vector machines (SVM), ensemble techniques such as random forests (RF) and deep learning techniques, in particular artificial neural networks (ANN).

One of the few articles found that covered the detection of vineyard regions using deep learning frameworks on satellite imagery, developed by Nuno Figueiredo *et al*. (2022), studied the detection of vineyards in the *Alto Douro* region, the same region of study as this dissertation, using Google Earth satellite images. This article demonstrated a deep-learning approach using a segmentation technique through the ResNetV2 model to detect vineyard plantations of the region, and the results appear to be promising. Another article by Eriita G. Jones *et al*. (2020) studied the impact of pan-sharpening and spectral resolution on vineyard segmentation. Although the segmentation approach is similar in these two studies, it shows that both spectral and spatial resolution play a significant role in the model's performance, when detecting vineyard regions.

In this dissertation we will develop research taking one step further and try to identify individual vineyard terraces, testing both traditional algorithms and state-of-the-art deep learning algorithms to segment vineyards in the studied region.

The resulting outcome of the model will then be combined with altimetry data obtained from SAR sensors, used in this dissertation to study the slope of the region where the segmented vineyards are located, ensuring that the vineyards are located in highland regions, validating the need of a terrace-like arrangement. Other studies have used altimetry data for many other applications, obtaining promising results, including surface water measurements (Alsdorf, Rodriguez, & Lettenmaier, 2007), topography (Berry, 2000) and river monitoring (Michailovsky, McEnnis, Berry, Smith, & Bauer-Gottwein, 2012).

By intersecting the segmented vineyard mask with the slope mask, a higher confidence that vineyards are located within terraces and not on plains or other lowland configurations can be extracted.

3. Data Sources

In this chapter, an exploration of satellite imagery and satellite imagery providers as sources for the raster data used in the methodology, as well as the selected vector data that will provide the label dataset, will be established.

3.1. Earth Observation satellites with spectral imaging capabilities – potential data sources

A list of some of the most relevant and currently operational EO satellites was created, in which we enumerate several satellites that possess multispectral and hyperspectral imaging capabilities, since the scope of this study lies in these types of products.

Some of the next satellites offer open access to their data since they are publicly funded, while others are managed by private entities, hence the availability of their products is subject to purchase costs.

Table 1 - Operational satellites with spectral imaging capabilities

Operational Satellites with Spectral Imaging Capabilities				
Satellite	**Entity**	**Year launched**	**Data Access**	**Description**
Amazônia-1	INPE	2021	Open	Multispectral
Aqua	NASA	2002	Open	Multispectral
Cartosat-1,2,3	ISRO	2005-2019	Open	Multispectral
CBERS-4	CNSA & INPE	2014	Open	Multispectral
Gaofen-2	CNSA	2014	Purchasable	Multispectral
Geosat-1,2	GEOSAT	2009, 2014	Purchasable	Multispectral
Landsat-8,9	NASA & USGS	2013, 2021	Open	Multispectral
Oceansat-2	ISRO	2009	Open	Multispectral
Pleiades-1, Neo	Airbus & CNES	2011-2021	Purchasable	Multispectral
PRISMA	ASI	2019	Open*	Hyperspectral
Sentinel-2	ESA	2015	Open	Multispectral
Suomi NPP	NASA	2011	Open	Multispectral
Terra	NASA	1999	Open	Multispectral
WorldView-1,2,3	MAXAR	2007-2014	Purchasable	Multispectral

*For academic purposes

There also exists data from non-operational satellites that are stored in archives, such as:

Table 2 - Non-operational satellites with spectral imaging capabilities

Non-Operational Satellites with Spectral Imaging Capabilities				
Satellite	**Entity**	**Year launched**	**Data Access**	**Description**
Envisat	ESA	2002	Open	Multispectral
LandSat-1-7	NASA	1972-1999	Open	Multispectral
PROBA-V	ESA	2013	Open	Multispectral
QuickBird	DigitalGlobe	2001	Purchasable	Multispectral
WorldView-4	MAXAR	2019	Purchasable	Multispectral

3.2.Raster data sources – passive sensor

For this study, we need data from land coverage satellites, so coastal and oceanic monitoring satellites are disregarded. We also need good spatial resolution since we want to identify vineyards in images that could occupy a small area (few meters), meaning that we can only use data from high or very-high resolution (VHR) satellite images. This study does not require restrictions on temporal resolution. Given these requirements and the accessibility we have to the following platform's data, we decided to use satellite images from the Sentinel-2 mission and the GEOSAT -2 satellite.

These two platforms are able to provide medium-high (up to 10 meters) and very-high (up to 0.4 meters) spatial resolution images respectively, with multiple bands in the visible, near infrared (NIR) (GEOSAT-2) and shortwave infrared spectral ranges (Sentinel-2A and Sentinel-2B) and, although it was not a necessity, they provide a reasonable revisit period, characteristics that will be clearly explained in the next two subsections.

3.2.1. Sentinel-2

Sentinel-2, illustrated in Figure 6, is an EO multispectral imaging mission with a core focus in land classification. Like the Sentinel-1 mission, this mission also comprises a constellation composed by 2 satellites: Sentinel-2A and Sentinel-2B, phased 180 degrees from each other in a sun-synchronous polar orbit providing high availability (Copernicus, 2022), with a third satellite, Sentinel-2C undergoing testing and preparation for launch in 2024 (ESA, 2022).

Figure 6 - Sentinel-2 mission satellite, adapted from (Copernicus, 2023)

The Sentinel-2 Multispectral Instrument (MSI) samples 13 spectral bands in the visible, NIR and shortwave infrared part of the spectrum. 4 of these bands at 10 meters, 6 bands at 20 meters and 3 bands at 60 meters of spatial resolution. This is considered a medium-high-resolution satellite sensor since it is able to provide images with up to 10 meters of spatial resolution. It also has a high temporal resolution (5 days of revisit time for the constellation, 10 days for each of the Sentinel-2 satellites), and a swath width of 290 kilometres (Copernicus, 2022).

We began the development phase of this study by using images from the Sentinel-2 mission, since the data provided is open and of medium-high-resolution, and expecting to, at least, be able to segment medium-sized vineyards present in the *Alto Douro* region. After this initial study with Sentinel-2 images, we will replicate the methodology with GEOSAT-2 images and analyse the differences in the segmented regions.

3.2.2. GEOSAT-2

The Geosat-2, illustrated in Figure 7, formerly designated as Deimos-2, is a very-high resolution Earth imaging satellite providing images up to 40 cm of spatial resolution after image processing techniques. It was launched on the 19th of June 2014, and is characterized by a sun-synchronous ascending orbit, a swath width of 12 km, and a revisit period of 2 days. It is the second satellite of the GEOSAT Earth Observation system (GEOSAT, 2023).

Figure 7 - GEOSAT-2 Satellite, adapted from (eoportal, 2023)

Its optical imaging sensor captures 5 different bands: blue, green, red, NIR and panchromatic bands. The bands corresponding to the visible and NIR part of the EM spectrum have a spatial resolution of 4 meters, while the panchromatic band has a spatial resolution of 1 meter on nadir conditions. After pan-sharpening and re-sampling techniques are applied, the RGB and NIR bands may also obtain a spatial resolution of 0.75 meters (GEOSAT, 2023). Pan-sharpening is a technique that fuses a panchromatic image with multispectral up-sampled images, obtaining pan-sharpened multispectral images, with the same spatial resolution as the panchromatic image.

It is also possible to obtain 0.40-meter resolution images, by applying artificial intelligence (AI) processing techniques to the pan-sharpened images, thus providing the ability to better assess the different objects and scenes contained in the images.

Remark: Due to external reasons beyond our reach, the data acquisition from this satellite established in the early stages of the dissertation was not possible to execute, hence the exclusive use of Sentinel-2 imagery for the study conducted in this dissertation.

3.3.Raster data source – active sensor

To ensure a higher confidence that the vineyard features are located within terraces, we will use altimetry data obtained from SAR sensors, which will then be processed in order to obtain the slope of the corresponding pixels in degrees. For this study, we chose to use data from the Shuttle Radar Topography Mission (SRTM) (Farr, *et al.*, 2007), which was flown aboard the space shuttle *Endeavour* on February 2000 (USGS, 2018), as illustrated in Figure 8.

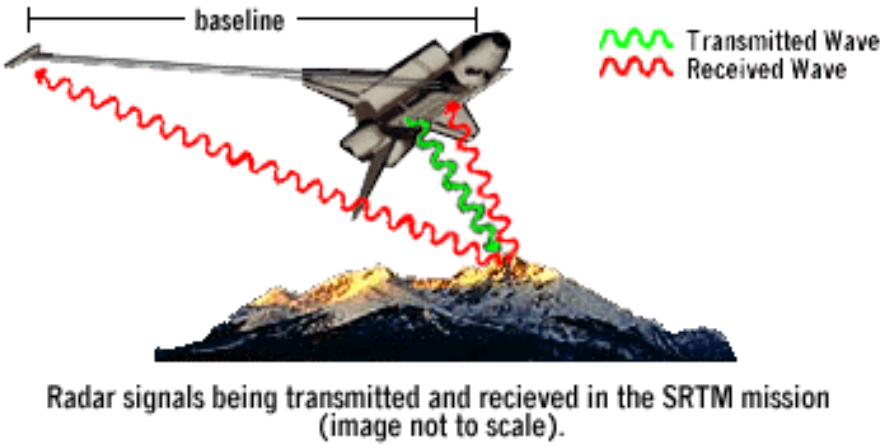

Figure 8 - SRTM illustration, adapted from (NASA, 2023)

The SRTM data is acquired through the creation of a radar beam that interacts with Earth's surface features and is then captured by two SAR sensors, where these sensors have their antennas placed at a fixed distance (baseline). By leveraging the information on the antenna separation and analysing the variations in the reflected radar wave signals of both SAR images *i.e.*, the interferometric pair, an interferogram can be created, where precise measurements of the Earth's surface elevation can be extracted, consequently forming an SRTM Digital Elevation Model.

This decision aligns with the fact that the data has a spatial resolution of 1 arc-second, corresponding to approximately 30 meters, which is suitable for the task we aim to perform, and also the data has open access, contributing to the minimization of costs in this dissertation.

3.4. Vector data source

To evaluate segmentation methods that will be developed during this dissertation, extract performance metrics, and to have labelled data to train, validate and test machine learning models, it is necessary to have a georeferenced dataset with the corresponding studied features prior to developing any methodology. As such, research on already existing datasets with the desired features – vineyards was performed, and a dataset that contains georeferenced labels of several classes, including vineyards, for the Portuguese continental territory was selected: the *Carta de Uso e Ocupação do Solo* (COS) 2018 v2.0[2], which we will refer as COS2018.

The COS2018 is a vector-base land cover classification map of the Portuguese continental territory, which divides the surface occupancy hierarchically by 83 classes, represented in polygons. It has a minimum cartographic unit of one hectare and a minimum distance between lines of 20 meters (Direção Geral do Território, 2022).

Not only does the COS2018 supply information regarding the location of said classes, but it also provides relevant information such as area and perimeter of polygons representing each of the classes.

For the scope of this dissertation, only the vineyard class will be extracted from the COS2018 dataset, and it will be considered as the ground-truth. This class represents the areas of continental Portugal where vineyard plantations are present. By extracting the precisely delineated and georeferenced polygons represented by this class, it enables the development of metrics that can be applied in the evaluation of devised methods, and the proper definition of the dataset for machine learning tasks, contributing to more robust models, which will be studied in depth on Chapters 7 and 8.

[2] Geographical information provided by *Direção-Geral do Território*

4. Segmentation through band indices

In the first proposed method to segment vineyards, we will perform a preliminary analysis as to whether or not it is possible to segment them satisfactorily by using band indices derived from Sentinel-2 imagery. If the results are satisfactory, we will then further develop and elaborate the method and the evaluation techniques applied to the results.

The images provided by the Sentinel-2 mission are composed by 13 bands (12 after BOA processing), hence we will study what is the best band combination to separate vineyards from other features present in the image. In this study, we will try to find the optimal band combination that enhances the reflectance values of vineyards and minimize the reflectance values of other features.

Since the majority of the vines in the Douro region have deciduous characteristics, this method, if efficient, can only be performed during the vine's foliar-active period of the year.

4.1. Index selection

According to Giovos *et al*., (2021), the most common index used in research publications of vineyards is the NDVI (Rouse, Haas, Schell, & Deering, 1974) which is used to study the presence and several characteristics of vegetation features.

The NDVI index is widely applied in remote sensing studies of green vegetation areas and is mostly derived from satellite or aerial imagery. By analysing the reflectance in the NIR and red regions of the electromagnetic spectrum with this index, many green vegetation characteristics such as health and density can be assessed.

The values of an NDVI index range from -1 to 1, with small values indicating regions of lower green vegetation density and high values indicating regions of higher green vegetation density, and are obtain with Equation (4.1):

$$NDVI = \frac{NIR - red}{NIR + red} \tag{4.1}$$

Where NDVI is the normalized difference vegetation index, NIR and red are spectral radiance or reflectance, depending on the image processing level, in this case for Sentinel-2 products, measured respectively in the NIR and red parts of the electromagnetic spectrum (Huang, Tang, Hupy, Wang, & Shao, 2020).

The previous equation can be translated to Equation (4.2) using Sentinel-2 bands:

$$NDVI = \frac{B8 - B4}{B8 + B4} \tag{4.2}$$

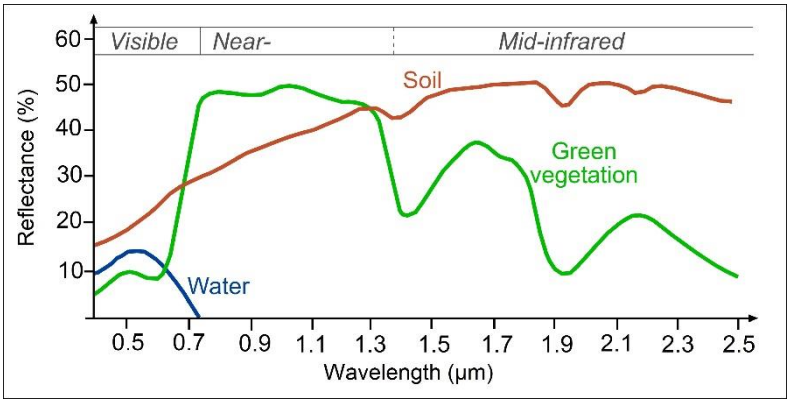

Figure 9 - Reflectance of green vegetation in different wavelengths, adapted from (SEOS, 2023)

In Figure 9 the spectral reflectance of green vegetation is depicted. Comparing with the spectral reflectance of vine's leaves, depicted in Figure 10:

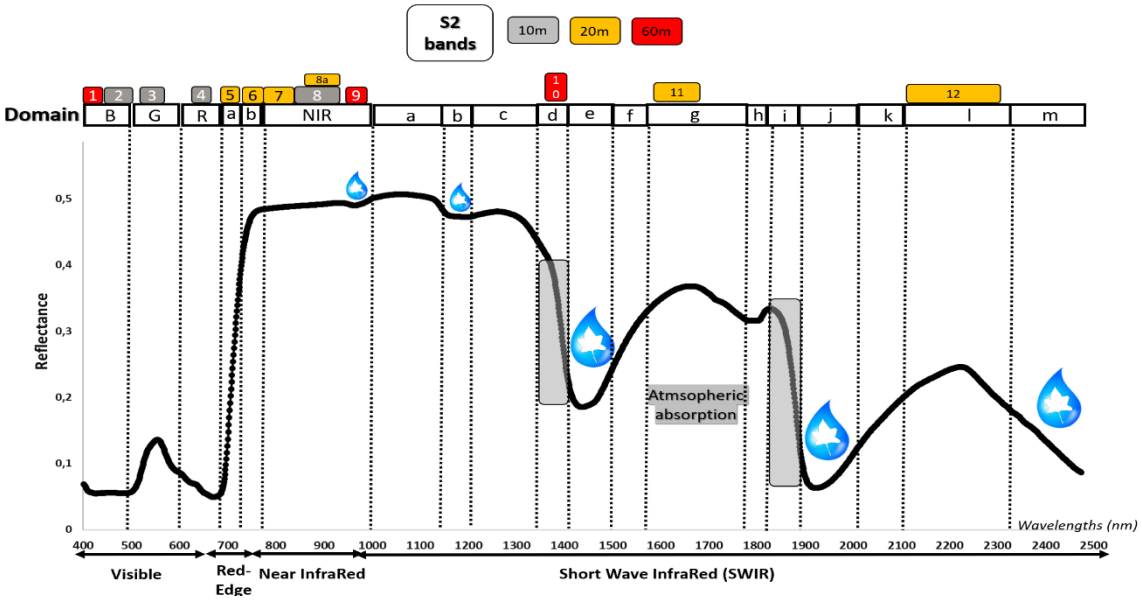

Figure 10 - Reflectance of vine's leaves in different wavelengths over Sentinel-2 channels, adapted from (Laroche-Pinel, et al., 2021)

It can be inferred that both spectral signatures have an identical behaviour, as expected since vine's leaves are characterized with a green colour, thus further corroborating to the use of an NDVI index for this implementation.

According to the graphic corresponding to Figure 10, it is possible to observe how the reflectance of the vine's leaves varies according to the wavelength. We can observe a maximum reflectance value of approximately 50% between wavelengths of 770nm to 1300nm, explained by the spongy mesophyll layer present in the vine's leaves, and a minimum value of less than 10% at 450 nm. The leave's spectral signature is also characterized with a local maxima at around 550nm, resulting from the chlorophyll absorption, and several local maxima and minima ranging from 1,400nm to 2,250nm, which result from water/moisture absorption (Tayade, *et al*., 2022), also represented by the water droplets in the image.

By using the NDVI index, the wavelengths at which the vine's leaves exhibit maximum reflectance will be emphasized, while wavelengths with minimum reflectance are suppressed. This ensures that these vegetation features are highlighted, distinguishing them from other non-relevant features.

One point to consider is that, since vine's leaves have a spectral signature comparable to other green vegetation, features with similar spectral signatures present in the image, such as forests or orchards, may also be highlighted, which can impact the efficiency of this method.

Giovos, *et al*. (2021) also demonstrated that band index classification is heavily dependent on the spatial resolution of the aerial images:

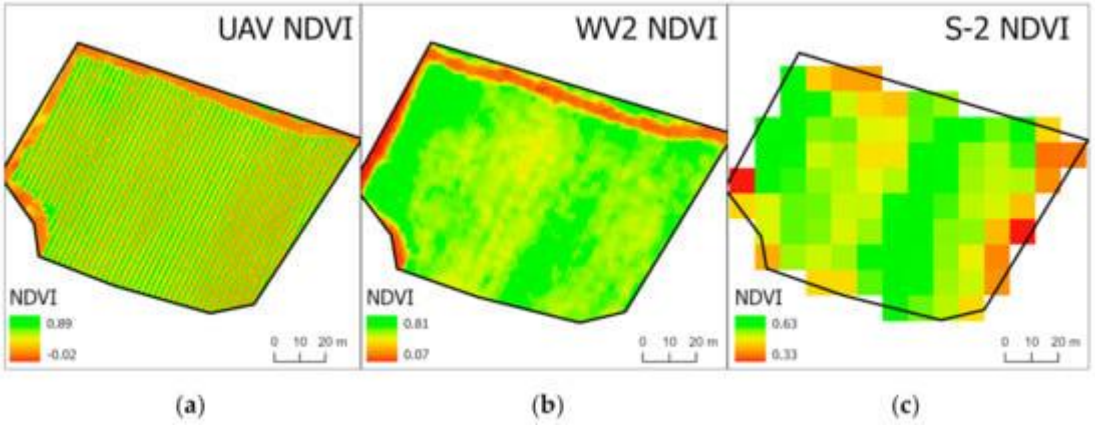

Figure 11 - NDVI map from different resolution imagery (UAV, Worldview-2, Sentinel-2), adapted from (Giovos, Tassopoulos, Kalivas, Lougkos, & Priovolou, 2021)

In Figure 11, it is possible to visualize that, by using very-high resolution imagery such as UAV imagery, it is possible to separate the canopy of the vines from the soil, which leads to more accurate results after applying index classification methods, whereas by using medium-high-resolution imagery, such as Sentinel-2 imagery, this is not possible to observe, due to the lower resolution image having both soil and vine features mixed within the same pixel, which, after applying index classification methods, leads to less accurate results.

Nevertheless, since we only have access to Sentinel-2 imagery, we will study the application of the NDVI to those products, perform the analysis of the resulting index layer, and extract the respective conclusions from these results.

In Figure 12, a study conducted by ADVID (2021) on two common grape varieties in the Douro vineyard region, the "Touriga Francesa" and "Touriga Nacional" shows their vegetative cycle over the year.

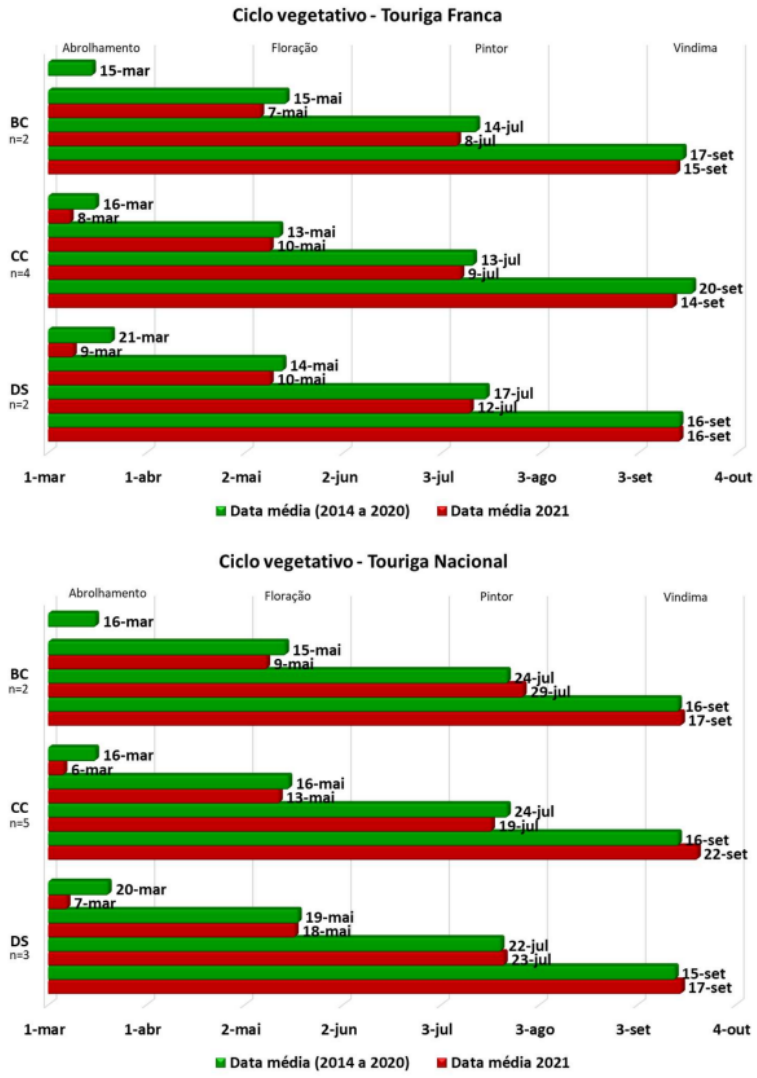

Figure 12 - Vegetative cycle of famous grape varieties present in the Douro vineyard region, adapted from (ADVID, 2021)

It is possible to observe that the cycle starts, on average, with the budburst in mid-March, with the first flowers sprouting two months later, in mid-May, and the grape harvest being performed in mid-September. Taking these observations into consideration, the application of an NDVI index to segment vineyard regions is optimal during the period between early June until mid-September, where the vines have the highest leave density coverage.

4.2.Data acquisition

To develop this methodology, we relied on the Copernicus Scihub platform to select and download the Sentinel-2 image product, and the SNAP application from ESA for all the other processes of the study: from image data processing to the classification by indices using band math.

Firstly, we selected the region of interest (ROI) as observed in Figure 13 which, for this purpose, corresponds to an area around *Peso da Régua*, a village adjacent to the Douro River, and performed a query for Sentinel-2 products for the foliar-active period of the vines (from June until mid-September), with minimal cloud coverage (0% to 10%) for level 1C processing, meaning that the image was already submitted to, among others, radiometric corrections in Top of Atmosphere (TOA) reflectance, and also for level 2A processing, which processes level 1C images and per-pixel radiometric measurements are provided in Bottom of Atmosphere (BOA) reflectance (Copernicus, 2022). We then selected the desired product and downloaded it. A decision to use both levels of processing was made as to evaluate which one performs best to identify the respective vineyard regions in the satellite images.

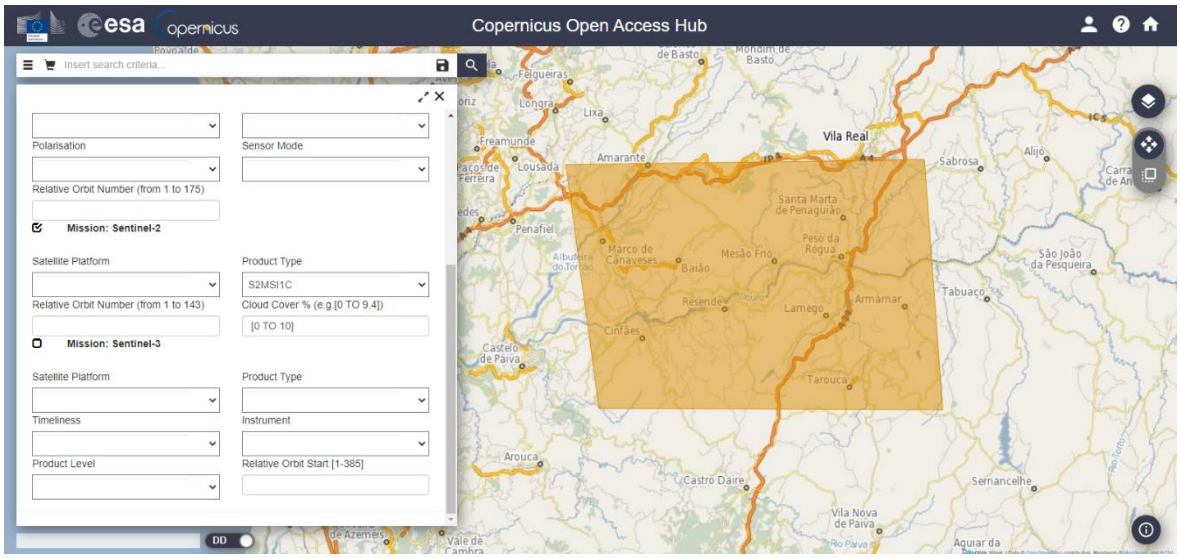

Figure 13 - Copernicus Scihub product query, adapted from (Copernicus, 2023)

After the product is acquired, we then analysed it using SNAP. In an effort to perform arithmetic operations between bands with different spatial resolution, a resampling process of these bands is needed in order to obtain identical spatial resolutions between bands. Since the chosen bands to create the selected index have the same spatial resolution of 10 meters, this process will not have to be performed.

4.3.Index Analysis

Taking the previous observations into consideration, we then computed this index in SNAP, by using the Sentinel-2 image bands 4 and 8, corresponding to the red and NIR part of the EM spectrum respectively, as illustrated in Figure 14:

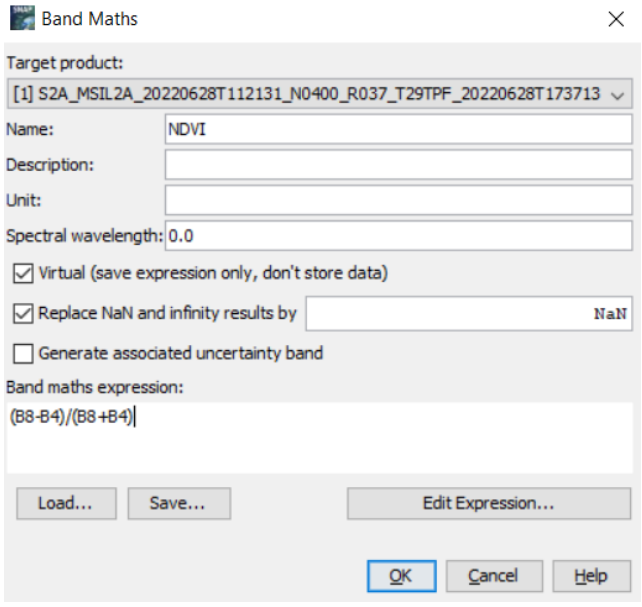

Figure 14 - SNAP computation of band math to create an NDVI map

After the index was created for both images with different processing levels, we compared them on the same area of the image where we know there are vineyard terraces present, and obtained the following results:

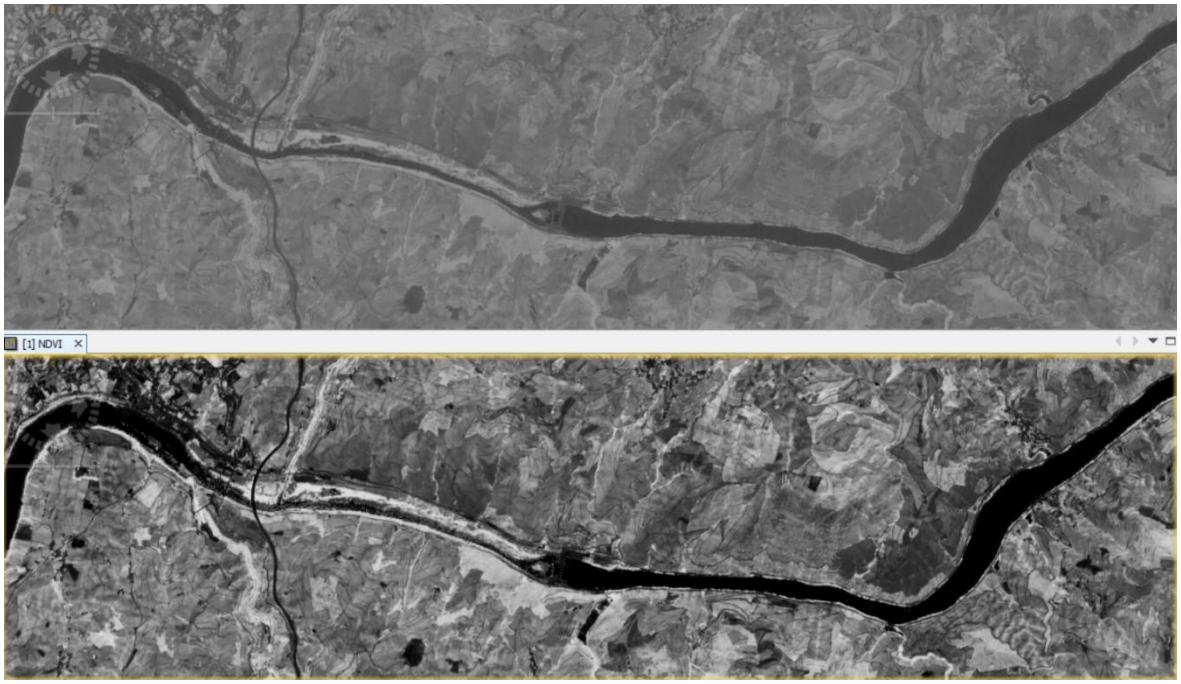

Figure 15 - Application of the NDVI index over Sentinel-2 imagery

We can observe in Figure 15[3] that the TOA (top image) and BOA (bottom image) reflectance bring different results using the NDVI index. For the level 1C image (TOA reflectance), there appears to be more noise in the image, compared to the level 2A image (BOA reflectance) that appears cleaner. This can be explained due to the atmospheric effect overlapped with the reflectance of surface features present in the level 1C processed image.

[3] Derived from Copernicus Sentinel data 2023, processed by ESA

For a better analysis and comparison of both images, a colour map will be used. For this purpose, we chose to use a gradient with red tones for regions of no vegetation, and green tones for regions with vegetation, with an orange/yellow transition colour, as depicted in Figure 16[4]:

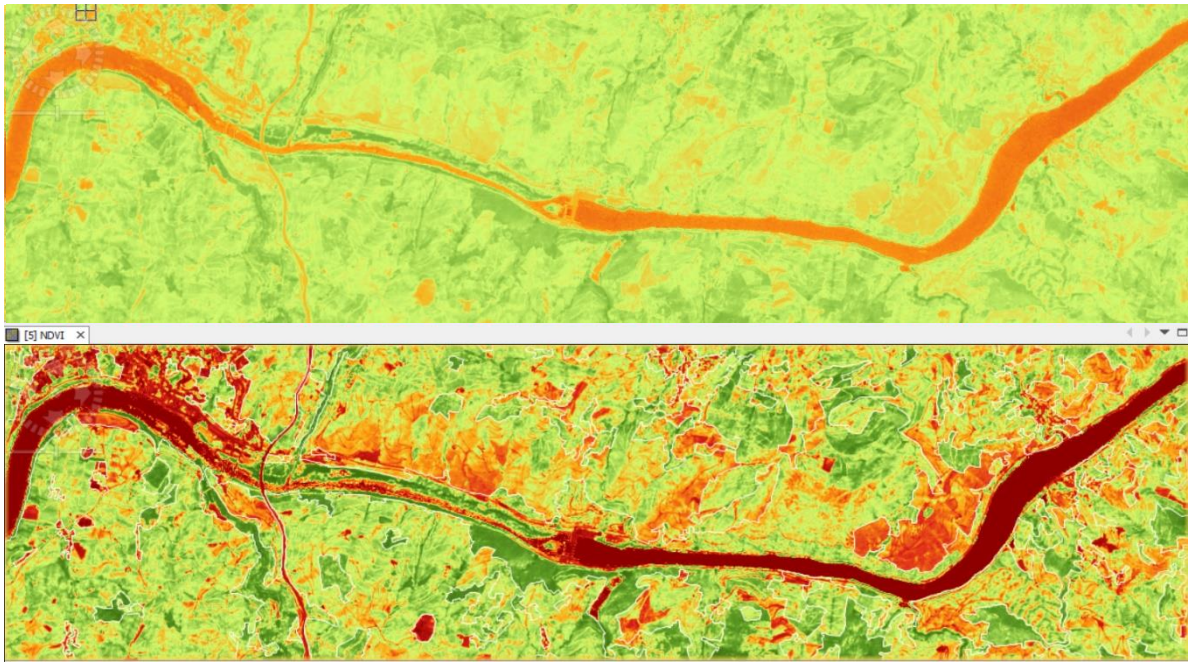

Figure 16 - Application of a color map to the previous NDVI map over Sentinel-2 imagery

Analysing the differences in non-vegetation and vegetation regions of Figure 16, a clear distinction can be made in both these zones. In the level 2A product, the features are more clearly depicted and classified, compared to the level 1C product, where due to atmospheric effects, the different features present in the image are more difficult to discern. Also, the difference in vegetation regions (depicted in different tones of green, where the darker the green the denser the vegetation is) and non-vegetation regions is clearer in the level 2A product.

Considering the points mentioned earlier, level 1C products can be discarded from this study.

[4] Derived from Copernicus Sentinel data 2023, processed by ESA

In the next figure, representing a closeup to the previews level 2A product with identifier S2A_MSIL2A_20220628T112131_N0400_R037_T29TPF_20220628T173713, Figure 17, a comparison between the NDVI map and the overlapped label from the COS2018 with the vineyard regions defined can be evaluated:

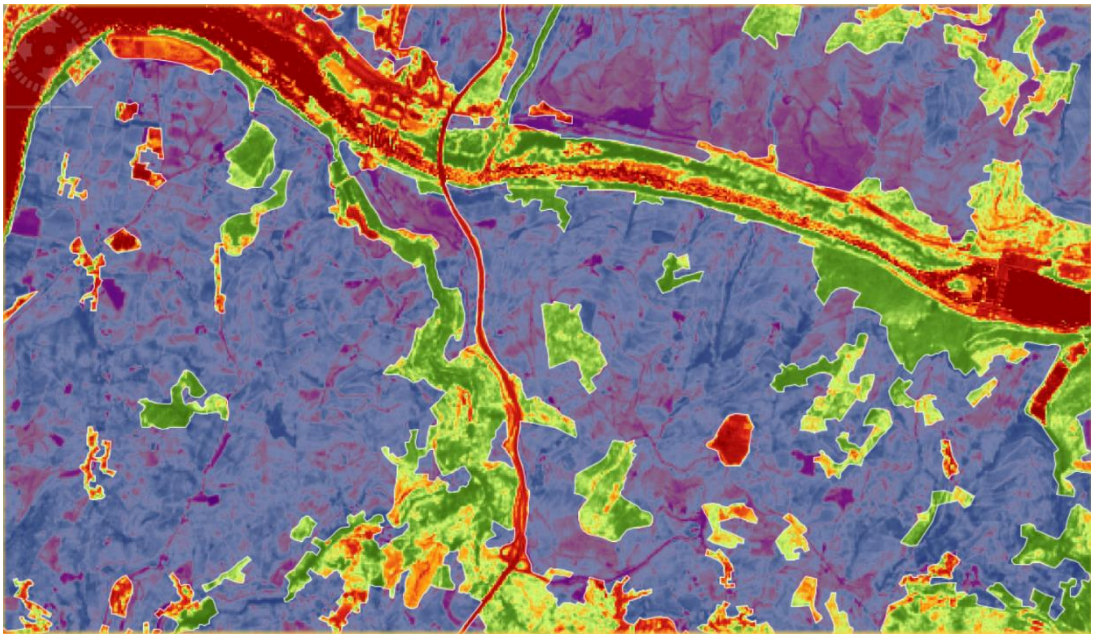

Figure 17 - Overlap of the vineyard polygons from the COS2018 over the NDVI map

In Figure 17, it is possible to visually inspect the quality of classification using the NDVI index. The vineyard regions extracted from the COS2018 and overlapped with the previous NDVI map are represented in blueish tones, with some transparency to observe the features beneath them.

It is possible to observe that some non-vegetation regions are misclassified as being vegetation (possible false-positive), and some regions of dense and less dense vegetation are also classified under the same class, taking into consideration that the likelihood of denser vegetation being forest or other similar non-vineyard feature is extremely high. The opposite also happens, where regions of less dense vegetation, in which the likelihood of being vineyards is higher are not classified as such (possible false-negative).

Overall, the results of the classification of vineyards with band indices were inconclusive in terms of vineyard classification, and a further investigation of this method and respective evaluation metrics for the task of this dissertation is not efficient, since there are other potential methods that have a higher likelihood of providing better results and can be more suitable for the task. These methods will be studied in detail in Chapters 7 and 8.

5. Data preprocessing

The data used for this study is not in a format suitable to train machine learning algorithms, *i.e.*, the level-2A processed data provided by the Copernicus platform is delivered as a zip file, containing raster data in .jp2 format, organized in different folders according to the spatial resolution (10, 20 and 60m), containing not only information regarding each band of the EM spectrum, but also additional information, including Aerosol Optical Thickness (AOT) map, Water Vapour (WVP) map and a Scene Classification (SCL) map. Also, each of the bands (after the resampling process) has an approximate size of 10,980x10,980px, which is not efficient for machine learning tasks computationally and memory-wise.

In order to preprocess the data into a more suitable format for machine learning tasks, several python scripts were built for this purpose, which, for a cleaner and more reliable and maintainable approach for future work, were then used to develop a python package which we designated as SentProd.

5.1. Sentprod

The SentProd package takes the directory where the zip file is stored as an argument, and performs several tasks, including:

- (Optional) **Unzip** or not unzip the file (default): since this package is built on top of esa-snappy, a python interface for the SNAP library, it is not required to unzip the product to access and manipulate its data, conserving a considerable amount of disk space.
- (Optional) **Spectral subsetting**: the user can choose which bands to use, improving this way the time it takes to perform the required computations, simultaneously conserving space in the disk. The default subset bands are B1~B12, discarding all the additional derived bands, saving up to 25GB of space per product and processing the task approximately 6x faster.
- **Resampling the selected bands**: since the bands have different spatial resolutions, matrix operations between different sized bands is not possible. To solve this problem, all selected bands are resampled to the same size (default is 10m), enabling this way to perform mathematical operations between them.

- **Merge the resampled bands**: To facilitate the handling of the data subsequently, the bands are translated from .jp2 format to TIFF format due to the flexibility, versatility compatibility with legacy systems and overall wide support of the later format compared to the first. These translated bands are then ordered according to central wavelength and concatenated into a single TIFF file, which makes manipulating the data easier for machine learning tasks on later stages.

With the first stage of raster data preprocess completed, the attention is now turned to the vector data. Given an image that was submitted to the previous preprocessing event, we now want to extract the vineyard mask delineated by boundaries of that image, *i.e.*, clip the vector data by the raster extent, within the same geolocation. Given that passive sensor satellite images are vulnerable to atmospheric conditions e.g., clouds, a simple algorithm that segments clouds by using their spectral signature was developed, and, given the generated could mask, extracted from the previously processed image, and the vineyard mask obtained from the COS2018, a cloudless vineyard mask resulting from the subtraction of the intersection between cloud polygons and vineyard polygons over the original vineyard mask is created. Assuring in this way that images where vineyards covered by clouds are not being wrongly classified as vineyards, which could lead to false information fed to the future machine learning model, affecting its performance.

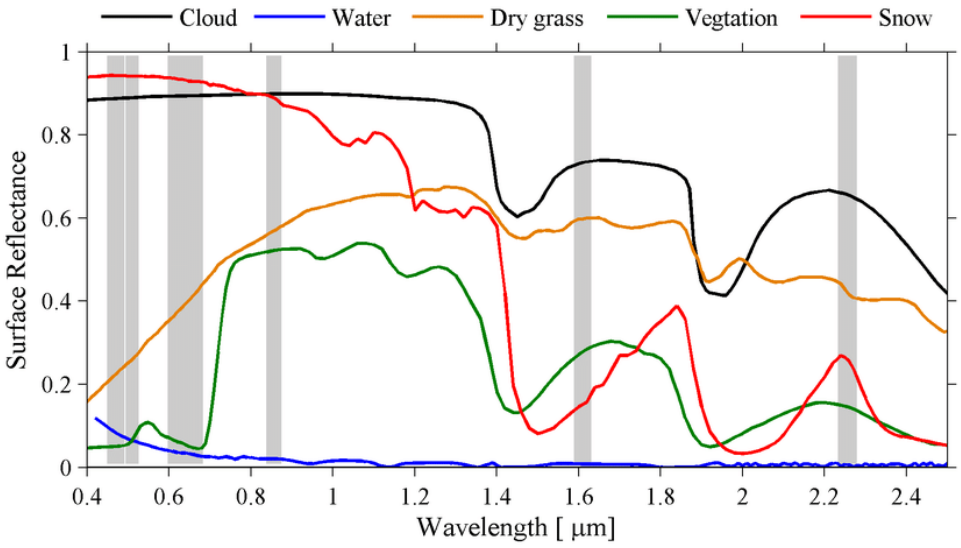

Figure 18 - Reflectance of cloud surfaces in different wavelengths, adapted from (Zhuge, Zou, & Wang, 2017)

Considering Figure 18, it is possible to analyse the spectral signature of several surfaces. Focusing on the cloud spectral signature, which is the one relevant for this subsection's study, it is possible to infer that it has maximum reflectance in the interval between 0.4 to 1.4μm, with local maxima values around approximately 1.6μm and 2.2μm. In Sentinel bands, this corresponds to the spectral range of all bands, in accordance with Table 3:

Table 3 - Wavelengths and bandwidths of the three spatial resolutions of the MSI instruments, adapted from (Copernicus, 2023)

Spatial Resolution (m)	Band Number	S2A		S2B	
		Central Wavelength (nm)	Bandwidth (nm)	Central Wavelength (nm)	Bandwidth (nm)
10	2	492.4	66	492.1	66
	3	559.8	36	559.0	36
	4	664.6	31	664.9	31
	8	832.8	106	832.9	106
20	5	704.1	15	703.8	16
	6	740.5	15	739.1	15
	7	782.8	20	779.7	20
	8a	864.7	21	864.0	22
	11	1,613.7	91	1,610.4	94
	12	2,202.4	175	2,185.7	185
60	1	442.7	21	442.2	21
	9	945.1	20	943.2	21
	10	1,373.5	31	1,376.9	30

Taking the previous points into consideration, the cloud removal algorithm checks if all elements along the depth axis of a spectral subset of the original image array, corresponding to band 1 to band 12 of the image, evaluate to True. This condition is met when all 12 pixel values are higher than the provided reflectance threshold, in this case the computed threshold is equal to 2,500, obtained from pixel analysis of a cloud in the Sentinel-2 product. If this condition is met, then the respective segmented mask pixel takes the value of 1, if the condition isn't met, then it takes the value 0. Examples of the output of this algorithm can be observed in Figure 19[5]:

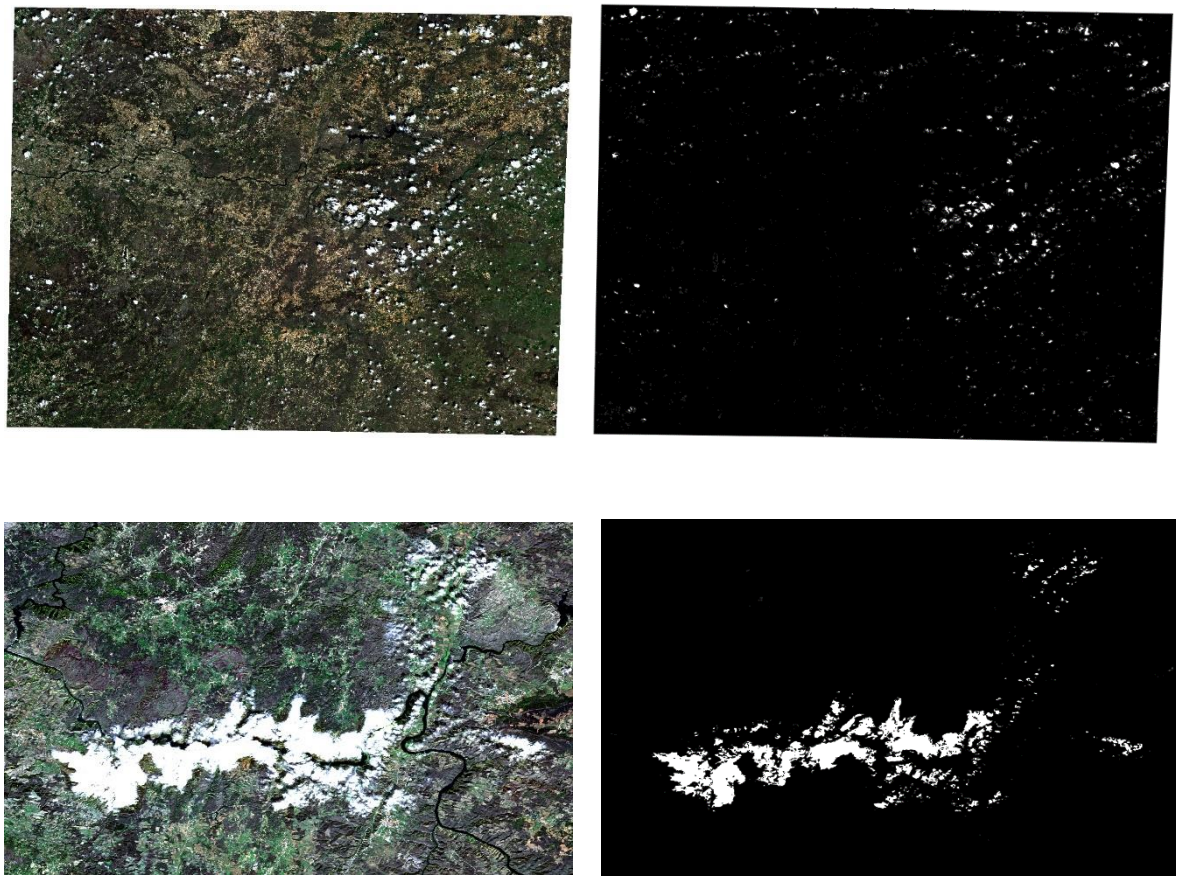

Figure 19 - Cloud mask segmentation

This process of removing clouds from the labels is particularly relevant in images taken during the winter season, where it is difficult to obtain products with minimal cloud coverage.

To do this, the optional argument "--cloud_removal" can be provided to the main.py script, followed by the directory path where the shapefile corresponding to the COS2018 is located. By providing this argument, the pipeline to remove clouds from the vector data, and the

[5] Derived from Copernicus Sentinel data 2023, processed by ESA.

vector file creation given the image boundaries *i.e.*, the entire vector data preprocessing pipeline is defined, ensuring accurate representation and integration of the spatial information.

To successfully execute the pipeline, another argument is required, namely "--class_name" followed by the class name that corresponds to the COS2018n4_L attribute list, which, for this study, are the "*Vinhas*". If the raster data has already been preprocessed earlier, we can call the argument "--no_merge" to only execute the vector data preprocessing pipeline.

The resulting layer is illustrated in Figure 20[6]:

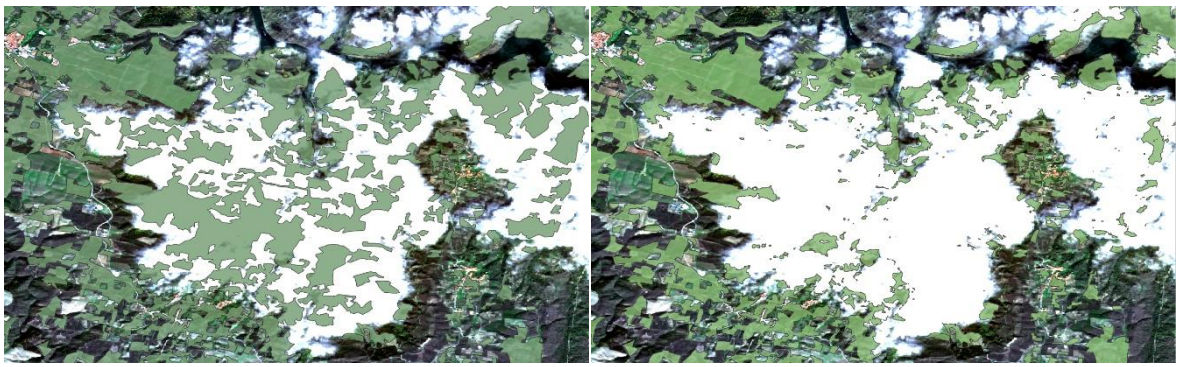

Figure 20 - Removal of vineyard labels overlapped with clouds

Given the previously supplied arguments, the main.py script can be executed, and the label file can be successfully created.

One last feature of this package is its ability to divide both the image and corresponding label into smaller image chips with the argument "--tile", making it easier to feed and train a machine learning model using batches of smaller subsets of the dataset, instead of providing the original over-sized dataset with a unique image and label.

[6] Derived from Copernicus Sentinel data 2023, processed by ESA.

5.2.Exploratory Data Analysis

Before supplying the data to the machine learning models, both to train and to infer, an Exploratory Data Analysis (EDA) will be performed, to examine the data and better understand its characteristics and patterns, and a subsequent preprocess of the data given the conclusions retrieved from the EDA process will be developed.

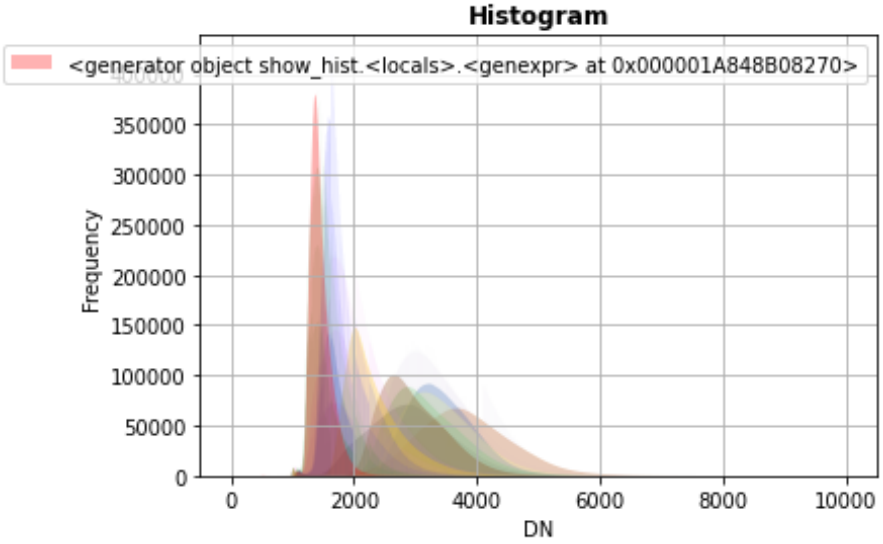

Figure 21 - Sentinel-2 image histogram

Previewing the histogram of the image where the image chips were extracted, *i.e.*, the original Sentinel-2 image, Figure 21, it is possible to observe that the reflectance values of all 12 bands, originally between [0:10,000] in the supplied level 2A product, are positively skewed, indicating that the distribution of pixels is skewed towards lower values.

Studying in more depth the histogram of each of the Sentinel-2 bands, Figure 22, it is possible to observe that the positive skewness behaviour is common to every band. This may visually represent darker scenes present in the image, meaning scenes with low reflectance values in the respective bands.

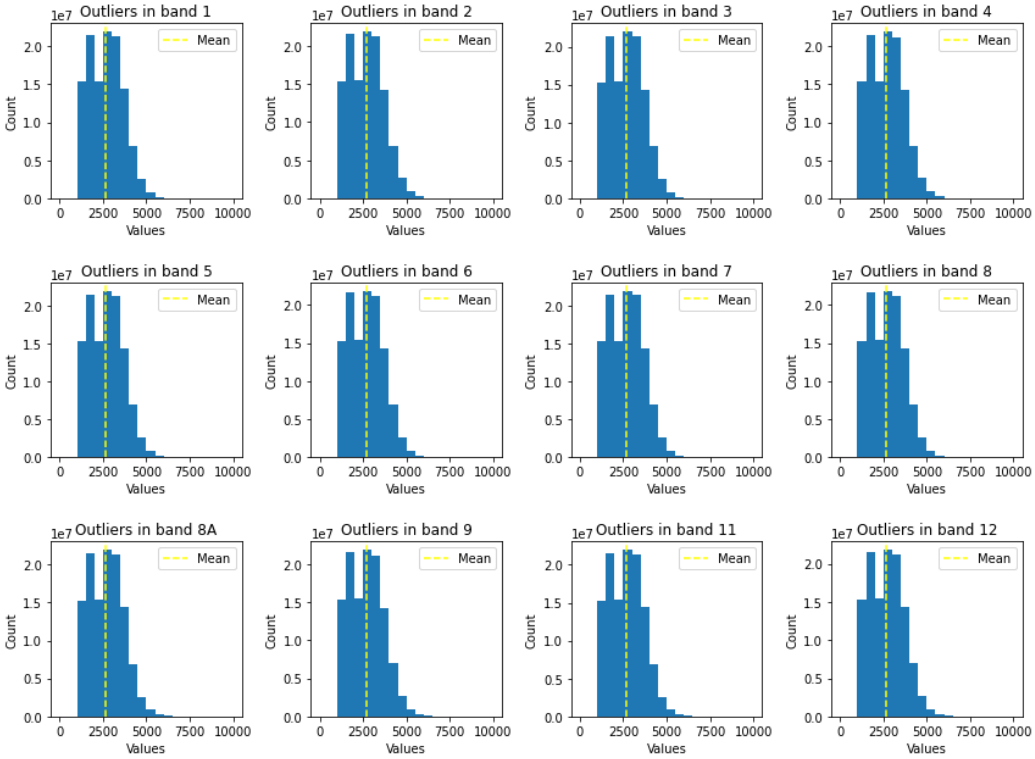

Figure 22 - Sentinel-2 image histogram decomposed by band

Regarding outlier detection, the previous histograms for each of the bands were recomputed using boundary-lines for the lower and upper outlier boundary regions, Figure 23. For this study we chose to classify a value as an outlier if $|x - \mu| > 3\sigma$.

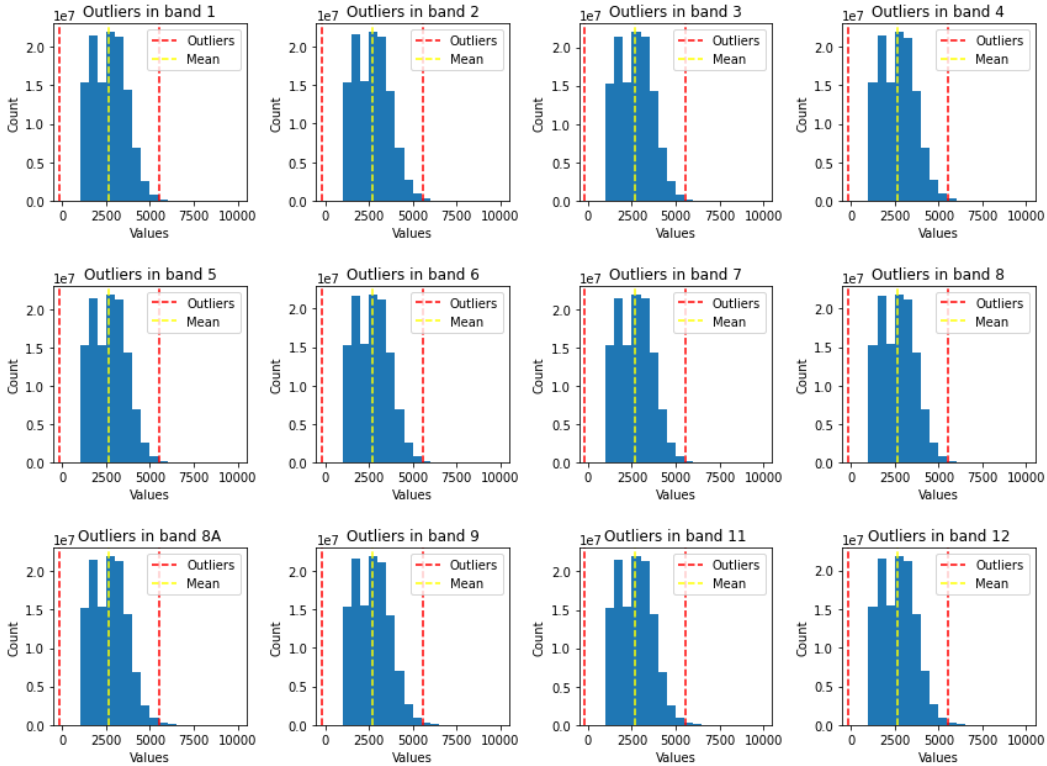

Figure 23 - Sentinel-2 image histogram decomposed by band with outlier boundaries

The previous histograms, Figure 23, reveal the presence of outliers, specifically over the upper outlier boundary. On average, the outliers account for 1,371,893 pixel values of each band, which have a total of 120,560,400 pixels values each. Overall, the outliers correspond to an averaged total of 1.14% of the bands pixel values.

5.3.Data Preprocessing for subsequent machine learning tasks

If the data were to be normalized using maximum and minimum pixel values, as described in the normalization equation, Equation (5.1):

$$X_{normalized} = \frac{x - x_{minimum}}{x_{maximum} - x_{minimum}} \quad (5.1)$$

the outliers that have a high value would bias the result of the denominator $x_{maximum} - x_{minimum}$, increasing it, which would then lead to a higher division value, resulting in a lower normalized pixel value. This would further substantiate the previous situation where it was inferred that the image may already have dark scenes.

To solve this situation, instead of using maximum and minimum pixel values, we will use percentiles to perform the normalization. After analysing the data and performing several tests with different percentile values, the best results were obtained with the percentiles 2 and 98, this way most of the original information was conserved. This was done using the following modified normalization equation, Equation (5.2):

$$X_{normalized} = \frac{x - x_{percentile(2)}}{x_{percentile(98)} - x_{percentile(2)}} \quad (5.2)$$

After the normalization process, the pixel values then passed through a clipping process, in which normalized pixels with values smaller than 0 were replaced with 0 and pixels with values larger than 1 were replaced with 1, assuring pixel values between [0;1].

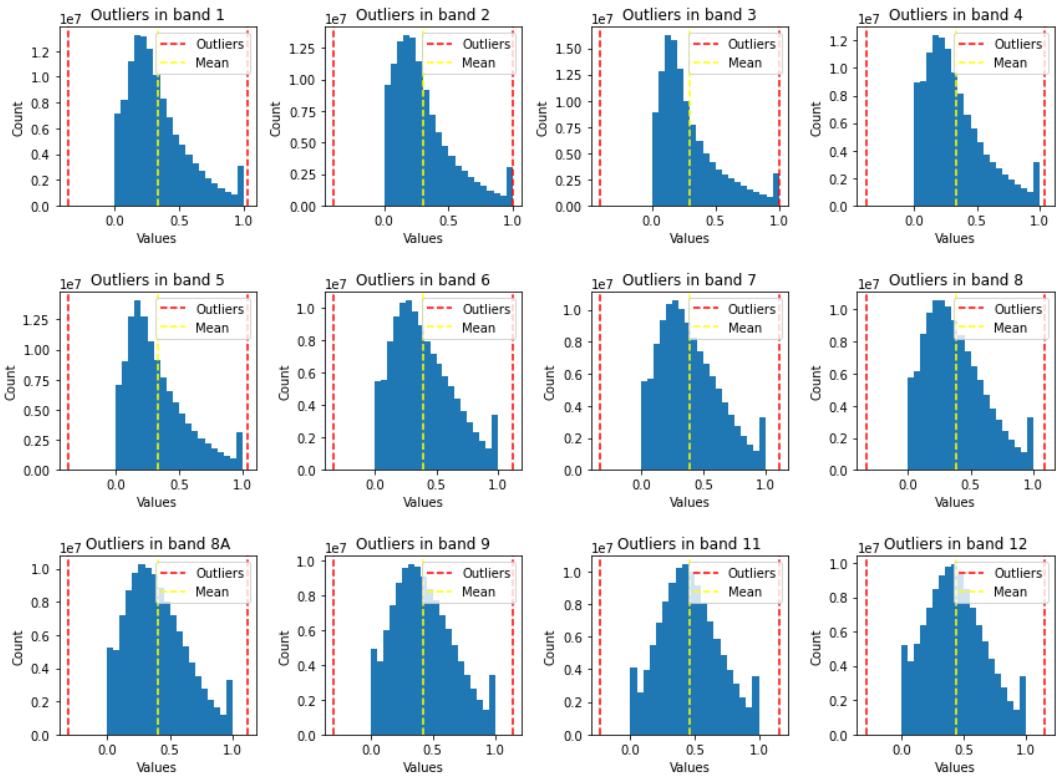

Figure 24 - Sentinel-2 normalized image histogram decomposed by band with outlier boundaries

In Figure 24, the histogram of each band after the normalization and clipping processes can be analysed. It is possible to determine that the skewness of the histograms remains similar to the original skewness, but with data points more concentrated towards the mean values.

The presence of outliers also reduced considerably, corresponding now, on average, to 405,361 pixel values of each band, with, as before, a total of 120,560,400 pixels values each. This corresponds to an averaged total of 0.33% of the bands' pixel values.

The outcomes of the previous techniques can be seen in Figure 25[7], where a comparison of a true colour composition of the image with minimum maximum normalization vs percentile normalization with clipped values is performed:

[7] Copernicus Sentinel data 2023, processed by ESA.

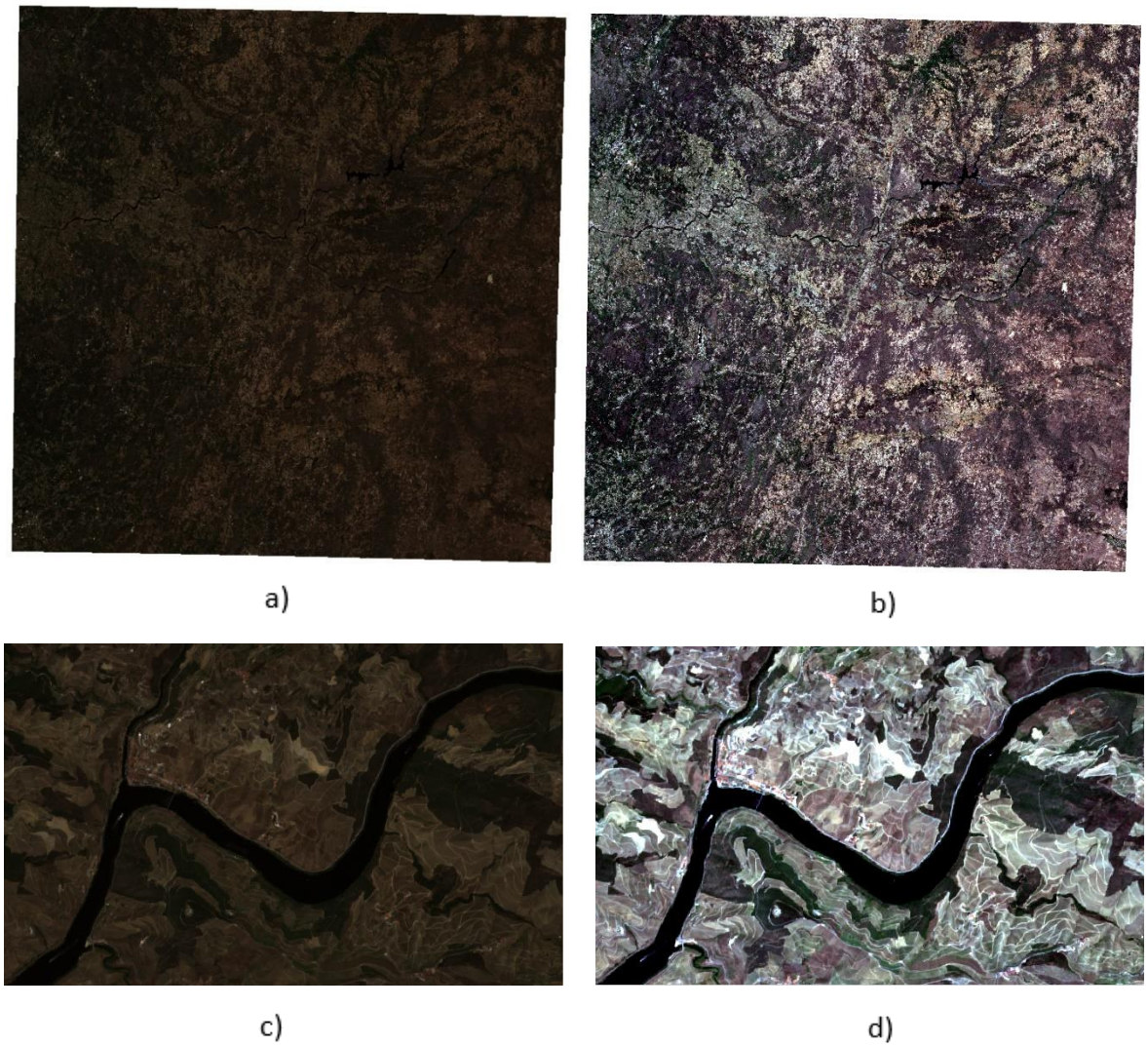

Figure 25 - Comparison between conventional normalization, a) and c), vs percentile normalization with clipped values, b) and d)

The visualization of b) and d) images of Figure 25 is clearer and more appealing, but more importantly the differences between different features are sharper, compared to images a) and c), contributing to the forthcoming models' capability to differenciate them.

Given all the preprocessing tasks performed to the Sentinel-2 image and corresponding labels, the whole preprocessed dataset used to train the models in Chapters 7 and 8 is depicted in Figure 26[8]:

[8] Derived from Copernicus Sentinel data 2023, processed by ESA.

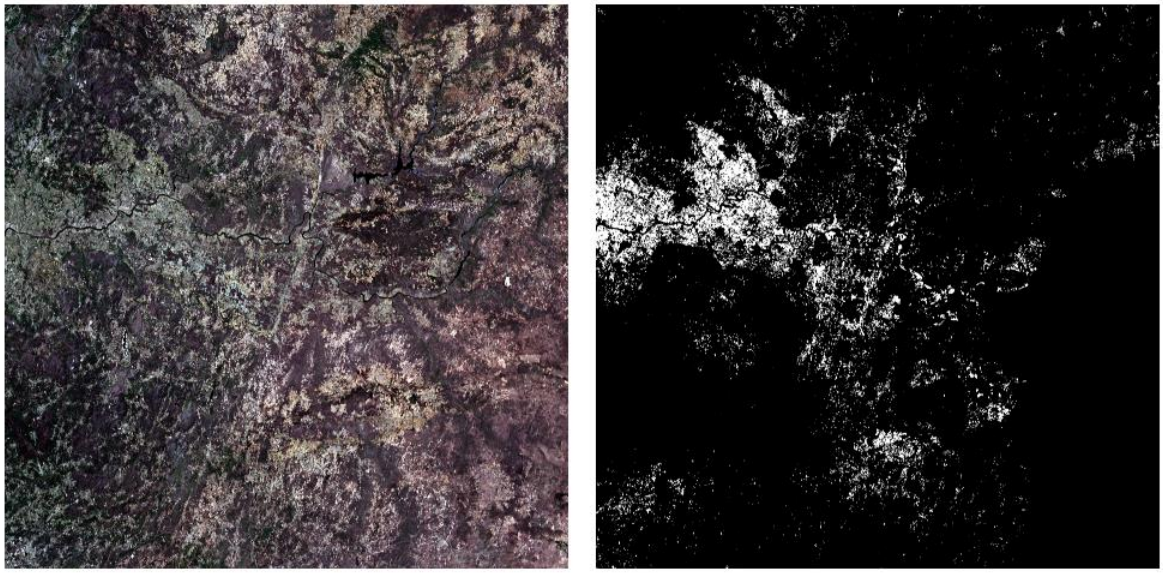

Figure 26 - Train validation and test dataset, with image data depicted on the left and corresponding vineyard labels on the right

This dataset will be posteriorly divided into train test and validation sets.

5.4. Class distribution

Concerning vector data and the class distribution of the binary classes vineyard and background, our dataset exhibits an inherent class imbalance, with a significantly higher number of samples belonging to the class background, as depicted in Figure 27. One should take into account that pixels in the segmented mask with value "True" or 1 correspond to the vineyard class and pixels with value "False" or 0 correspond to the background class.

```
Counter({False: 113227044, True: 7333356})
```

Figure 27 - Total number of samples for background class (False) and vineyard class (True)

Thus, the class background represents aproximately 94% of the dataset, and only 6% represents the class vineyard. This imbalance could lead to poor model performance, including biased predictions and poor generalization.

To mitigate the impact of an imbalanced dataset, a Synthetic Minority Over-sampling Technique (SMOTE) will be employed to the data used in Chapter 7, over-sampling the minority class using synthetic examples, created by connecting the original samples to its *k*

nearest neighbors (Chawla, Bowyer, Hall, & Kegelmeyer, 2002). This approach helps to balance the class distribution and improve class representation.

Although this analysis was previously performed to an entire Sentinel-2 image, the selected techniques will be employed individually to each image chip extracted from the original image, totalling 36 chips, ensuring that the values used for the respective transformations coincide to each of the raster chips and corresponding labels. Since 4 of the 36 chips do not have any vineyard features in them, they will be excluded, hence the dataset will be composed with a total of 32 image chips.

6. Evaluation metrics

To evaluate the models' performance developed in the next two chapters, a set of different metrics were chosen. The bibliography demonstrates that the most common evaluation metrics used in machine learning algorithms for image segmentation are, among others, accuracy, precision, recall, F1-score, Jaccard index and the Area Under the ROC Curve (AUC) (Melki, *et al*., 2022) (Moccia, De Momi, El Hadji, & Mattos, 2018) (Müller, Soto-Rey, & Kramer, 2022). Since these metrics have values ranging between 0 and 1, the higher the metric value, the better the performance of the model is according to that metric. Not only will the mean of these metrics be evaluated, but the confusion matrix of the predictions made by the model will also be analysed. The confusion matrix is a tabular representation that summarizes and depicts the performance of a classification algorithm, depicting both true and false negative and positive instances.

- Accuracy measures the number of correct predictions in relation to the total number of predictions, Equation (6.1):

$$Accuracy = \frac{1}{N}\sum_{i}^{N} 1(y_i = \hat{y}_i) \tag{6.1}$$

- Precision measures the ratio of correctly classified positives over all classified positives, Equation (6.2):

$$Precision = \frac{TP}{TP + FP} \tag{6.2}$$

- Recall measures ratio of correctly classified positives over all real positives, Equation (6.3):

$$Recall = \frac{TP}{TP + FN} \tag{6.3}$$

- The F1-score is a harmonic mean of precision and recall metrics, meaning that it gives more weight to lower values, penalizing models that have a large difference between precision and recall, Equation (6.4):

$$F1_{score} = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (6.4)$$

- Jaccard index, also known as Intersection over Union (IOU), is a metric used to measure the similarity or overlap between two sets or objects, represented by A and B in the next equation. In practice, it measures the similarities between the ground-truth and the model's predicted labels, Equation (6.5):

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \quad (6.5)$$

- The AUC metric represents the area under the ROC curve. The ROC curve represents the true positive rate (TPR), also known as sensitivity, against the false positive rate (FPR), which is equal to 1 minus specificity, at various classification thresholds. An example can be observed in Figure 28:

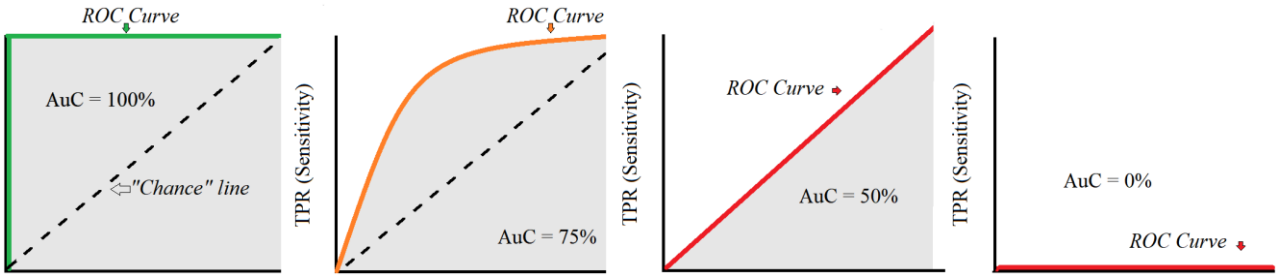

Figure 28 - AUC example, adapted from (Paperspace, 2020)

The AUC can be interpreted as follows:

- AUC = 0.5: The model's predictive capabilities demonstrate no advantage over random guessing between positive and negative instances.
- AUC > 0.5: The model demonstrates a certain degree of predictive ability, with higher values indicating a higher predictive performance distinguishing between positive and negative instances.
- AUC = 1: The model has perfect predictive capabilities, precisely differentiating between positive and negative instances.

We will focus on maximizing the F1-score of our models by testing several classification thresholds and selecting the one that produces the best F1-score. This was the chosen metric to maximize since it balances two important metrics: precision and recall, this way accounting for the class imbalance of our dataset, while simultaneously considering false classifications.

7. Segmentation through the application of traditional machine learning algorithms

An alternative method to segment vineyard regions on satellite imagery involves the application of traditional machine learning techniques, also referred to as Explainable AI (XAI) techniques.

From the literature review we inferred that two main supervised learning algorithms are used for the purpose of image segmentation: support vector machines (SVM) and random forests (RF). To develop this task, we will experiment these algorithms for the purpose of performing a binary classification on individual pixels as to if they belong or not to a specific class: vineyard. These algorithms will receive a Sentinel-2 image previously chipped using the SentProd library, preprocessed and evaluated using the techniques and metrics seen in the previous chapters.

Both models were trained with the parameter "probability" set to True. Although computationally more expensive, this parameter allows, during inference, the creation of a probability map, providing the probability of a pixel to belong to one class or the other.

The confidence map will be used instead of the classic classification output because it allows for a greater flexibility and versatility for post-processing of the outcome. This way, during the model's test stage, the F1-score will be computed for a range of thresholds = [1;100] with unitary increments, returning the confidence threshold that maximizes the F1-score for each image. This threshold will then be summed up and divided by the total number of images used in test, providing the mean threshold score, which will then be set as the default threshold for all inferences with the corresponding model.

7.1. Support Vector Machines

SVMs are a supervised machine learning algorithm that employs a binary classifier based on an optimal separation hyperplane for these two classes (Suykens & Vandewalle, 1998), with a maximized margin between opposite parallel planes to the optimal hyperplane, created by the support vectors of each class in the feature space. These classes are mapped from the input space to the feature space through the nonlinear function $\varphi(.)$: $\mathbb{R}^{\mathrm{n}} \rightarrow \mathbb{R}^{\mathrm{n_b}}$, given a training set $\{x_k;\ y_k\}_{k=1}^{N}$ with independent variables $x_k \in \mathbb{R}^{\mathrm{n}}$ and the corresponding binary class labels $y_k \in \{-1; 1\}$ (Suykens, 2001).

After training SVM models with different hyperparameter configurations, using the balanced dataset, the models were then tested.

Two different kernels will be experimented with this algorithm:

- **Linear kernel**: this kernel is used when the data is linearly separable through the algorithm's separation hyperplane, calculating the dot product between data points to measure their proximity. The linear kernel is defined in Equation (7.1) as:

$$K(x, y) = x * y \tag{7.1}$$

  Given the complexity of the data, and that the data points corresponding to each feature present in the image can have similar reflectance values in the same band, it is unlikely that the linear separation of the data points will provide good results.

- **RBF kernel**: the Radial Basis Function (RBF) kernel is a non-linear kernel that relies on a Gaussian similarity measure to capture the non-linear relationships between data points, and consequently find the optimal non-linear decision boundaries. The RBF kernel is defined in Equation (7.2) as:

$$K(x, y) = e^{-1*\frac{||x-y||^2}{2*\sigma^2}} (=) e^{-\gamma*||x-y||^2}, where\ \gamma = \frac{1}{2 * \sigma^2} \tag{7.2}$$

  It is expected that, given the more complex decision boundaries, these boundaries will fit the data more accurately, hence we predict to output better results than the previous kernel, when training the model with the optimal hyperparameters.

One other hyperparameter that will be altered in different models is the penalty parameter C. This parameter trades of correct classification of training samples with a maximization of the decision's function margin.

Another hyperparameter that was programmed in all models was the gamma γ parameter, which controls the influence that each training example makes on the decision boundary, affecting the smoothness of the decision boundary. Too large values of γ will cause the

model to overfit, too small values on the other hand constraint the model, reducing its ability to capture the complexity of the data.

Given the complexity and volume of the data, training a non-linear kernel in a Central Processing Unit (CPU) would need tremendous computational power. As such, instead of training this algorithm relying on CPU using the Scikit-learn package, it was trained on a Graphics Processing Unit (GPU) using the RAPIDS cuML library, converting the Numpy arrays corresponding to each of the image bands and the label binary mask to Cupy arrays suitable for GPU handling. For this study we will use a NVIDIA RTX A6000 GPU, with 48GB GDDR6 of memory.

In Table 4 the models' evaluation metrics can be analysed:

Table 4 – Support Vector Machine metrics results using the models trained with the balanced dataset

Balanced dataset										
Kernel	**n_iter**	**C**	**γ**	**mF1**	**mAccuracy**	**mPrecision**	**mRecall**	**mJaccard**	**mAUC**	**Threshold @maxF1**
Linear	1000	10^0	-	0.3127	0.8825	0.2308	0.4762	0.2073	0.6778	0.68
RBF	500	10^0	5^{-1}	**0.325**	**0.8693**	**0.2306**	**0.5502**	**0.2095**	**0.7084**	0.55
	1000	10^0	5^{-1}	0.2863	0.8502	0.2062	0.477	0.1793	0.6655	0.62
	1000	10^1	5^{-1}	0.2236	0.8112	0.1739	0.4973	0.1586	0.6482	0.53

Due the size and volume of the dataset and possibly also due the recent implementation of the previous algorithm in the used package, several unexpected errors related to CUDA memory allocation were encountered both in training and inference of the models. As such, given the training and inference time, even when relying on GPU for all heavy computing tasks, we limited the number of outer iterations per training epoch to 1000, and didn't vary the γ parameter in an effort to minimize these constraints. Even after applying these measures, it was only possible to train and infer a small number of models using this algorithm in due time.

Nevertheless, it is possible to observe that the results extracted from the metrics of Table 4 are not very satisfactory. The model with the best performance was limited to 500 outer iterations, with $C = 1$ and $\gamma = \frac{1}{n_{features}} = \frac{1}{2} = 0.5$.

Although it was expected that the RBF model would provide better metrics as the number of outer iterations increased, this was not what happened. Instead, the best metrics were obtained in a more constrained model. This could be due to a number of reasons, one of them being that the model is converging faster than what was previously anticipated, which could lead to overfit when trained for a larger number of iterations over the dataset.

Increasing the penalty parameter also results in worst metric results, which indicates that by increasing the decision function margin, trading correctly classified samples, has a downgrading effect in the output resulting metrics. In this case, it seems that the decision function margin should be narrower, allowing for a higher number of correctly classified samples, improving the output results.

The confusion matrix obtained from the inference of the testing dataset is depicted in Figure 29:

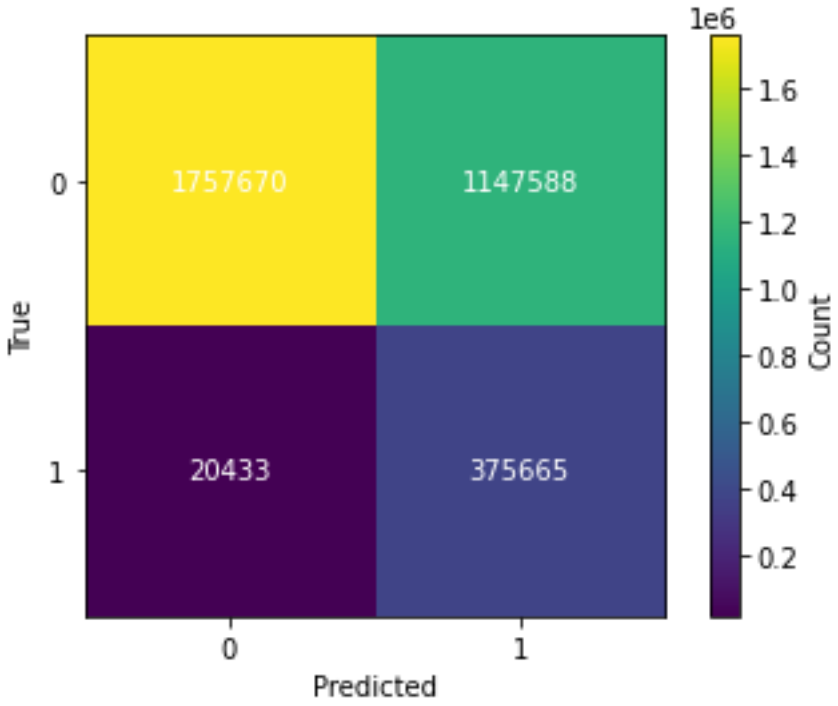

Figure 29 - Confusion matrix resultant from the test of the best SVM model

From the analysis of the confusion matrix, it is possible to conclude that there exists a total of 1,757,670 TN, 1,147,588 FP, 20,433 FN and 375,665 TP in the output segmented mask, where several metrics discussed next were derived.

Regarding the F1-score, the best mean score obtained has a value of 0.3250, suggesting that the model's ability to correctly classify and distinguish vineyards from other features is

relatively low. Taking into consideration the definition of the F1-score, Equation ((6.4), this result demonstrates that the model has difficulties to achieve both high precision, meaning that it has a low ability to correctly predict positive instances, and high recall, meaning that it has a low ability to correctly identify all positive instances; given the mean precision and mean recall metric results, with best results of 0.2306 and 0.5502 respectively, the model fails to achieve high results in either.

Taking into consideration the accuracy score, the best mean result has a value of 0.8693, indicating that the model correctly predicted approximately 86.93% of the instances (TP + TN) in the dataset. After a more detailed study on this metric, it was observed that the results are skewed by the majority class: the more imbalanced de testing scenario is, the smaller the accuracy is in the minority class, but the results are biased due to the majority class. By studying this metric alone, *i.e.*, without considering other metrics that, for example, study the proportion of type 1 (FP) and type 2 (FN) errors during inference, one could assume that the model has an almost excellent performance, which certainly is not the case. By studying a combination of different metrics, a true assessment of the models can be developed.

Observing the Jaccard index scores, the best model returns a mean value of 0.2095. This score indicates that the similarities between the true labels and the predicted labels is low.

Finally, the best mean AUC score is 0.7084, demonstrating that the model has an acceptable discrimination ability, implying that the model performs reasonably well in terms of sensitivity and specificity.

7.2.Random Forests

Random forests, RF, are also a supervised machine learning algorithm but have a completely different approach to the problem, using an ensemble of decision trees. According to Breiman (2001) “A random forest is a classifier consisting of a collection of tree-structured classifiers $\{h(x, \Theta_k), k = 1, \dots\}$ where the $\{\Theta_k\}$ are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input x.”

Although, for the previously mentioned reasons, the SVM models were only trained with the balanced dataset, for assessment purposes, a small set of models with different hyperparameter configurations will be trained using the imbalanced dataset for the random forest algorithm, prior to training models with the balanced dataset.

Using the same dataset split identically as in the previous algorithm, the random forest model was also trained using different combinations of hyperparameters in both the imbalanced and balanced dataset to subsequently find the optimal hyperparameters that maximize the model's metrics, specifically the F1-score.

In this algorithm, the number of estimators, *i.e.*, the number of decision trees contained in the random forest will be altered as to understand the changes in the metrics, and consequently the behavioural changes within the model. Also, a comparison between training the models using the original imbalanced dataset and the balanced dataset will be performed. The models were also trained with a maximum depth until all leaves contain less than 2 samples, or until all leaves are pure, *i.e.*, until every leaf node contains instances from one class.

The RF algorithm will be trained on a CPU since the computational power required is not sufficient to justify training on the previous GPU used to train the SVM models.

The results of the models test results can be observed in Table 5 and Table 6:

Table 5 - Random Forest metrics results using the models trained with the imbalanced dataset

Imbalanced dataset							
n_estimators	**mF1**	**mAccuracy**	**mPrecision**	**mRecall**	**mJaccard**	**mAUC**	**Threshold @maxF1**
5	0.2529	**0.9008**	0.2283	0.2819	0.16866	0.6006	0.35
10	0.2724	0.9002	0.2311	0.3217	0.1789	0.6189	0.37
20	0.2748	0.8986	0.2343	0.3257	0.1834	0.6206	0.37
50	**0.2795**	0.898	**0.2357**	**0.3379**	**0.187**	**0.6252**	0.38

Table 6 - Random Forest metrics results using the models trained with the balanced dataset

Balanced dataset							
n_estimators	**mF1**	**mAccuracy**	**mPrecision**	**mRecall**	**mJaccard**	**mAUC**	**Threshold @maxF1**
5	0.265	0.8981	0.2346	0.3017	0.1754	0.61	0.5
10	0.275	**0.9044**	0.2486	0.3049	0.1822	0.6171	0.53
20	0.283	0.9001	0.2445	0.3315	0.188	0.6257	0.52
50	0.288	0.9004	0.2444	0.3389	0.1904	0.6284	0.52
100	**0.29**	0.901	**0.2514**	**0.3415**	**0.1912**	**0.6288**	0.52

Comparing Table 5 with Table 6, it is possible to conclude that the balanced dataset generally provides slightly better results than the imbalanced dataset. This could be due to the algorithm's inherent ability to handle imbalanced datasets. Given its implementation, different techniques including bagging, feature randomness and voting mechanism contribute to the algorithm's ability in handling imbalanced datasets.

Focusing on the algorithm's results trained with the balanced dataset, Table 6, generally the results improve when increasing the number of estimators, with the overall best metric results obtained when using 100 estimators. Given the increasing complexity and consequent increase in computational power and training time, the maximum number of estimators used was limited to 100. The best mean F1-score obtained corresponds to 0.2895, suggesting that the model's ability to correctly classify and distinguish vineyards from other features is relatively low, with low mean precision (0.2514) and low mean recall (0.3415) scores.

Regarding accuracy, the mean results are high, with values approximate to 0.9, meaning that the model has a good correct classification rate. Given that the dataset used in inference, *i.e.*, the original test dataset is imbalanced, this high result could be explained due to the model correctly predicting the majority class (background), even if it has a low performance in classifying the underrepresented class (vineyard).

Taking into consideration the Jaccard index, the similarities between the predicted and actual classes is also low, with a best mean score of 0.1912, indicating a low overlap between predicted and actual classes.

Finally, focusing on the mean AUC score, the values for all different models are greater than 0.6, with the best mean result of 0.6288, suggesting a satisfactory predictive ability to distinguish between vineyards and background.

The probability thresholds that maximize the F1-score are very similar to each other within the same training dataset but have a large discrepancy between the two training datasets. This can be explained due to the differences in class representation between datasets. In the imbalanced dataset, due to differences between the number of samples of each class that the model was train with, the F1-score is maximized with a lower probability that the pixel corresponds to vineyard. On the other, on the balanced dataset, since the classes are equally represented during training, the F1-score is maximized with a probability of approximately 52% that the pixel corresponds to vineyard.

7.3. Further conclusions

One major factor that can be contributing to the low performance metrics in both algorithms is the spatial resolution of the image, which can be leading to the model's decrease in discrimination power due to the loss of detail of the present features.

Some factors could potentially impact in a positive way the resulting metrics, including using different techniques to solve the imbalanced dataset, since in this study we only focused on the SMOTE technique, and increasing the number of estimators in the RF algorithm, since we can observe an increasing trend in metrics performance with the increase in the number of estimators of the random forest, escalating this parameter could potentially lead to better results.

One relevant observation made in both algorithms is the large decrease in performance, according to the previous metrics, when the class distribution is more noticeable. The best model reaches F1-scores of approximately 0.7~0.8 and similarly high metrics when tested using image chips that have a more similar class distributions, and much smaller metrics, including F1-scores, when using image chips with a more noticeable disparity between class distribution. This behaviour is in accordance with the fact that testing a model in imbalanced scenarios often leads to imbalanced metric performance, generally with high accuracy but low F1-scores, which is precisely what is happening in our case. It is possible that, given that the model has more information regarding the majority class of the testing image, compared to the minority class, it can be biased towards the majority class, struggling to correctly classify instances from the minority class, leading to a decrease in metrics that evaluate false predictions.

To visually evaluate the performance of the best model, meaning the model using the SVM algorithm with a limit of 500 outer iterations trained with the balanced dataset, a segmentation mask was generated with it in a balanced and unbalanced scenario, and can be observed in Figure 30[9] and Figure 31[9]:

[9] Adapted from Copernicus Sentinel data 2023, processed by ESA.

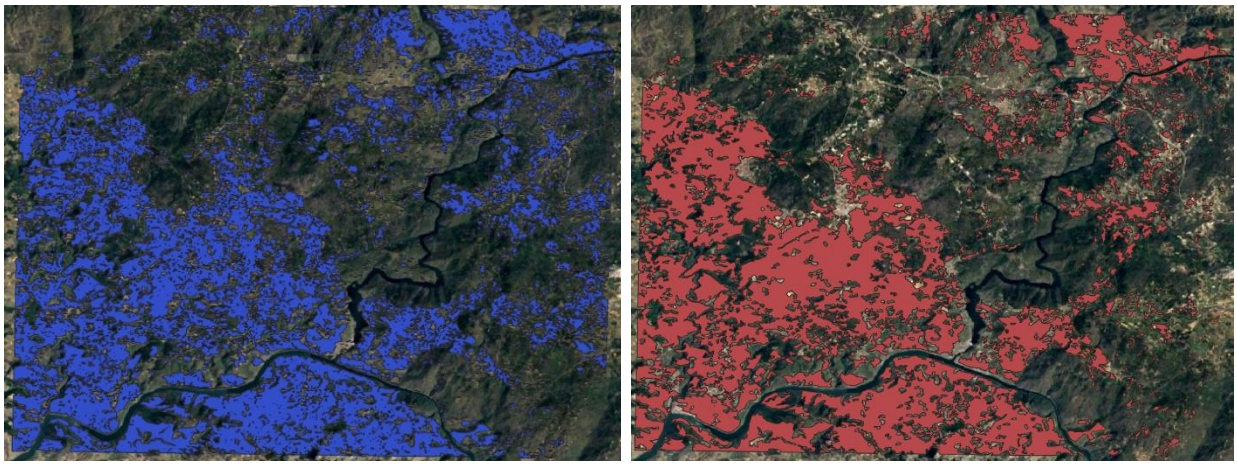

Figure 30 - Prediction of the model (left) vs original label (right) on a balanced scenario

In the previous figure, it is possible to observe that the model behaviour is, in fact, very acceptable given that the class distribution in the image is balanced. The vineyard polygons are coloured while the background polygons are transparent and not observable. The previous vector data is then overlapped with a satellite image. There are some misclassification regions in the image, but overall, the results are good.

When evaluating the results of the inference in an imbalanced scenario, the situation exhibits a significant difference, as depicted in Figure 31:

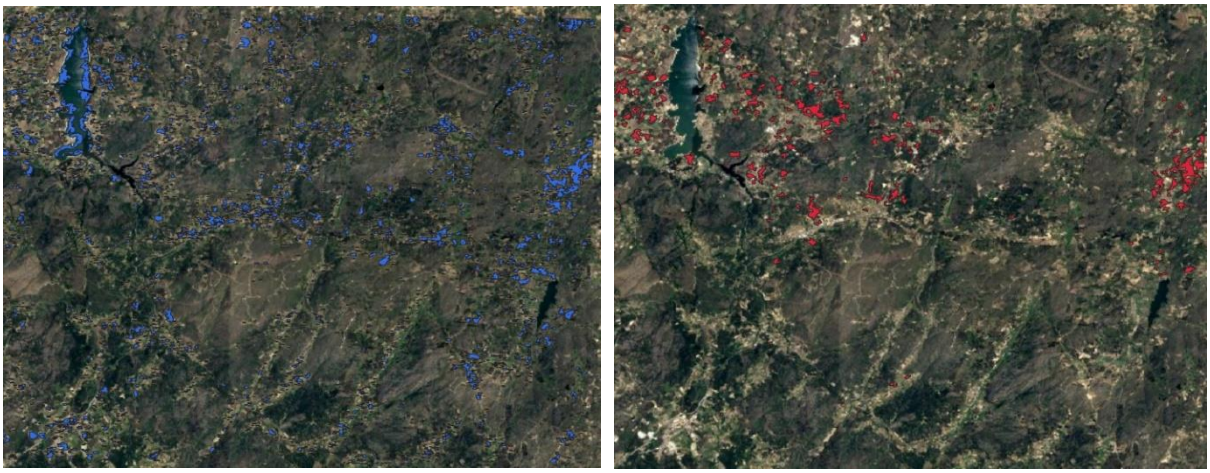

Figure 31 - Prediction of the model (left) vs original label (right) on an imbalanced scenario

Seeing that the segmentation output of these two algorithms were below the expectations, in Chapter 8, we will focus on developing deep learning models in an effort to obtain better results.

8. Deep Learning approach to vineyard segmentation

We will start this chapter by developing a comparative study on different convolutional neural network architectures and backbones previously developed for semantic segmentation and choose the best architecture and backbone to apply in this study, *i.e.*, to segment the vineyards present in the same Sentinel-2 image used in the previous chapter.

We are going to develop the corresponding python code using the PyTorch framework, since it allows for a greater flexibility in model adaptations, has a strong community which is helpful to debug possible issues, and overall is widely used by the research community.

8.1. Model architecture

At the time this subsection was developed, there exist three different DL models implemented within the PyTorch package for semantic segmentation, according to PyTorch's documentation (PyTorch, 2023):

- **LRASPP** – A semantic segmentation MobileNet model based on the MobileNetV3, tuned to run in mobile phone CPUs. For this reason, this model can be excluded for this study.
- **FCN** – The Fully Convolutional Network is a deep learning model, which concept is to convert conventional feedforward neural networks, that use fully connected layers, into a fully convolutional architecture, with fully connected layers converted into convolutional layers, that takes as input an image of arbitrary size and outputs a segmented mask with identical size (Long, Shelhamer, & Darrell, 2015).
- **DeepLabV3** – The DeepLabV3 model is characterized by atrous convolutions, which are a type of convolution that introduces gaps within the convolutional kernel, applying the kernel at selected positions of the feature map rather than applying the kernel at all positions of the feature map like in a traditional convolution. This allows for a larger effective receptive field, without necessarily increasing the number of parameters, obtaining a model with increased multi-scale contextual information without a relevant increase in computational complexity. The spacing between positions of the kernel is defined by the dilation rate, hence this technique is also designated as dilated convolutions.

The technique used in this model is designated as Atrous Spatial Pyramid Pooling (ASPP), consisting in applying multiple parallel atrous convolutions with different dilated rates. This allows the model to capture information at different scales and to model contextual information effectively (Chen, Papandreou, Schroff, & Adam, 2017).

Given these two possible models, we opted to use the DeepLabV3 model over FCN since it demonstrates good results in benchmark studies on known datasets and is effective with fine-grained details and objects of varying scales, such as the vineyard regions in our study. These regions may be sparse and distant from each other or occupy large areas separated by small distances. The DeepLabV3 model also preserves fine details in the segmentation mask, thanks to the use of the ASPP technique.

According to the research performed, the DeepLabV3 model is preferred in situations where more complex contextual modelling is needed, and when dealing with fine-grained objects that may also have different scales, at the expense of computational cost.

8.2.Backbone

There are two different suitable backbones in the previously mentioned PyTorch's documentation to use with this model for our study: the ResNet50 and the ResNet101. The Residual Network (ResNet) architecture was introduced by He *et al*., (2016) to address the problem of vanishing gradients on very deep neural networks, by introducing skip or residual connections between layers, enabling the network to learn more efficiently.

The ResNet50 differs from the ResNet101 mainly due to the depth of these neural networks: while the ResNet101 has 101 layers, consisting of convolutional layers, residual blocks and the last fully connected layer, the ResNet50 has 50 layers. These differences result in contrasting complexity and performance. The ResNet101 generally has the ability to learn more complex representations, but may also be computationally heavier, taking more time to train, and needing more data to avoid overfit. On the other hand, the ResNet50 may not learn such complex representations, but uses less resources and has a smaller training time. Since in our study the vineyards can have different class distributions between each image chip, having different reflectance values in each of the 12 bands used to train and infer the models, a potentially more suitable backbone architecture is the ResNet101, hence we will use it as the backbone for the DeepLabV3 model architecture.

Each residual block in the ResNet101 consists of multiple convolutional layers and a residual connection, as depicted in Figure 32. This connection involves element-wise addition between the output of its convolutions and the input of the block. The skip connection performs identity mapping, providing a direct path for gradient flow during training, bypassing one or more layers.

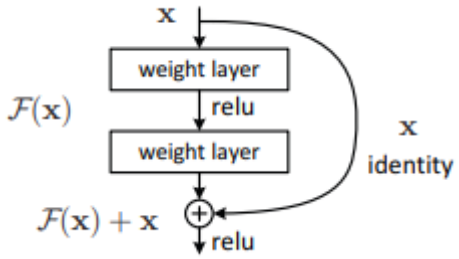

Figure 32 - Residual block, adapted from (He, Zhang, Ren, & Sun, 2016)

The ResNet101 backbone will not be used with pre-trained weights, and the first convolutional layer will be modified to accommodate the difference in the number of features of the images fed to the model. The modified layer will receive 12 features for the 12 Sentinel-2 level 2A image bands. As such, the model is defined in python the following way, Figure 33:

```
model = deeplabv3_resnet101(pretrained=False, weights = None, num_classes=1)

backbone = model.backbone

new_conv_layer = nn.Conv2d(12, 64, kernel_size=7, stride=2, padding=3, bias=False)
backbone.conv1 = new_conv_layer
```

Figure 33 - DeepLabV3 with ResNet101 backbone and modified convolutional layer

8.3. Train, validation and test dataset

To train and test this model, we will be using the same data as in the previous chapter, although separated differently due to image size constrains: the previous dataset was trained and tested with the original Sentinel-2 image divided into 26 and 6 image chips respectively, while in this chapter it was necessary to cut the original Sentinel-2 image and corresponding labels from the original size of 10,980x10,980 into 10,752x10,752, as to ensure that the dataset is divisible by multiples of 2^n, in order to create the dataset chips. Because of this, the resultant dataset has 1,412, 176 and 176 training, validating and testing image chips respectively when the dataset images have a size of 256x256, and 353, 44 and 44 training,

validating and testing image chips respectively when the dataset images have a size of 512x512.

Due to these differences, the dataset split process in this chapter diverges from the previous split process. This way, although the data used is the same, the training and testing sets are not identical. To address this issue, a possible solution would be to preprocess the dataset in the previous chapter by cutting the original Sentinel-2 image to the previous size before creating identical train and test image chips. This approach would allow us to use the same sets with different image sizes in both chapters, which cannot be carried out due to limited resources at this stage. Nonetheless, we will compare the testing results, taking these constraints into consideration.

8.4.Test results

After successfully creating the modified DeepLabV3 model and having the dataset defined, we trained the model with the corresponding data and different hyperparameter/technique configurations, obtaining the following results:

Table 7 - DeepLabV3 test results

Hyperparameters & Techniques						Metrics						Classification threshold
Batch size	Img size	Data Augmentation	Number of epochs	Learning Rate	1_conv layer Bias	mF1	mAccuracy	mPrecision	mRecall	mJaccard	mAUC	Threshold@ maxF1
4	256x256	No	50	0.001	False	0.4123	0.9632	0.4347	0.4346	0.3043	0.6712	0.45
16	256x256	No	50	0.001	False	0.6494	0.9683	0.7067	0.6255	0.5008	0.805	0.44
16	256x256	Yes	50	0.001	False	0.6294	0.9705	0.7393	0.5689	0.4868	0.779	0.4
16	512x512	No	50	0.001	False	0.6187	**0.9719**	0.8085	0.544	0.4822	0.7681	0.46
16	512x512	Yes	50	0.001	False	0.5793	0.9638	0.6333	0.5441	0.4327	0.7641	0.6
32	256x256	No	50	0.001	False	0.6364	0.9665	0.7228	0.5814	0.4766	0.7838	0.71
32	256x256	Yes	50	0.001	False	0.6843	0.9702	0.7476	0.639	0.5293	0.8131	0.21
32	256x256	Yes	50	0.01	False	0.6162	0.9663	0.691	0.5638	0.4599	0.7752	0.82
32	**256x256**	**Yes**	**100**	**0.001**	**False**	**0.705**	0.9686	0.708	**0.7129**	**0.552**	**0.847**	0.32
32	256x256	Yes	100	0.001	True	0.6864	0.9697	0.73	0.6505	0.53	0.8186	0.47
64	256x256	No	50	0.001	False	0.6416	0.9701	**0.8794**	0.5319	0.5249	0.7598	0.45
64	256x256	Yes	50	0.001	False	0.5606	0.9566	0.5773	0.5469	0.4013	0.7626	0.47

Observing Table 7, the results of the inference with the previously trained models on the test dataset displays far better results than the traditional machine learning models studied in the previous chapter.

The DeepLabV3 model was trained with different hyperparameters and data augmentation techniques, respectively:

- **Batch size**: the model was trained and tested with batch sizes ranging from 4 to 64, where it is possible to observe that overall, the behaviour of the metrics improves as the batch size increases, until it reaches a batch size of 32, where the metrics are maximized. When using a batch size greater than 32, *i.e.*, a batch size of 64, the metric results start to deteriorate.
- **Image size**: The model was also trained and tested with different image sizes. After training and testing the model with the 512x512 image chips, we concluded that the behaviour of the metrics when the model is trained with the 256x256 images is better. For that reason, the remaining training and testing instances were performed exclusively on 256x256 image chips.
- **Data augmentation**: The dataset was interchangeably trained with data augmentation techniques to increase the length of the training dataset, introducing samples that may improve the model's performance.

  To implement this method, we used different data augmentation techniques, including ShiftScaleRotate, which is a technique that applies affine transformations, respectively shifting, scaling and rotating the image according to user's defined values and RandomCrop, which randomly crops the images according to the user's desired values, in our case, we performed random crops with a size of 32x32.
- **Number of training epochs**: Due to computational limitations, the models where almost exclusively trained using 50 epochs, except the models that had the most promising results when trained with 50 epochs and simultaneously hadn't yet converged, where the training loss was still decreasing.
- **Learning rate**: We trained the different models with a default learning rate of 10^{-3}, and tested a different learning rate of 10^{-2} at the best performing model at the time, obtaining worst metric results. For that reason, we continued to train the remaining models with the first learning rate value.
- **1st conv layer Bias**: The modified convolutional layer of the ResNet101 backbone has a programable bias parameter, which could be set to True of False, *i.e.*, to introduce a bias parameter or not in this layer. We tested all the first models without a bias parameter in the first backbone convolutional layer, experimenting introducing a bias parameter on the best performing model, obtaining slightly worst results. Due to this, the bias parameter was set to false in all the other model's training and testing.

The best metric scores, including the F1-score, were obtained in the model using 256x256 image chips, with 32 image chips per batch, with a learning rate of 10^{-3}, without a bias term in the first convolutional layer, trained with data augmentation for 100 epochs. The F1-score obtained was 0.705, which indicates that the model has a reasonable harmonic balance between precision and recall and a reasonable ability to identify positive instances simultaneously minimizing misclassifications. Both precision and recall scores were also reasonably high with values 0.708 and 0.7129 respectively, suggesting a moderate capability of correctly predicting and identifying true positive instances.

The accuracy score for this model is 0.9686, indicating that approximately 96.86% of the predictions made by the model are correct.

The obtained Jaccard score was 0.552, suggesting a moderate overlap between the ground truth labels and the predicted labels.

Finally, the AUC score obtained from this model is 0.847, which suggests that the model's ROC curve has a good discrimination capability between vineyards and background instances.

The train and validation loss of the model were also measured where, although the training loss is still decreasing after 100 epochs, the validation loss plateaus after approximately 20 epochs, with some occasional higher loss values during some validation epochs, which could indicate potential overfitting, as depicted in Figure 34:

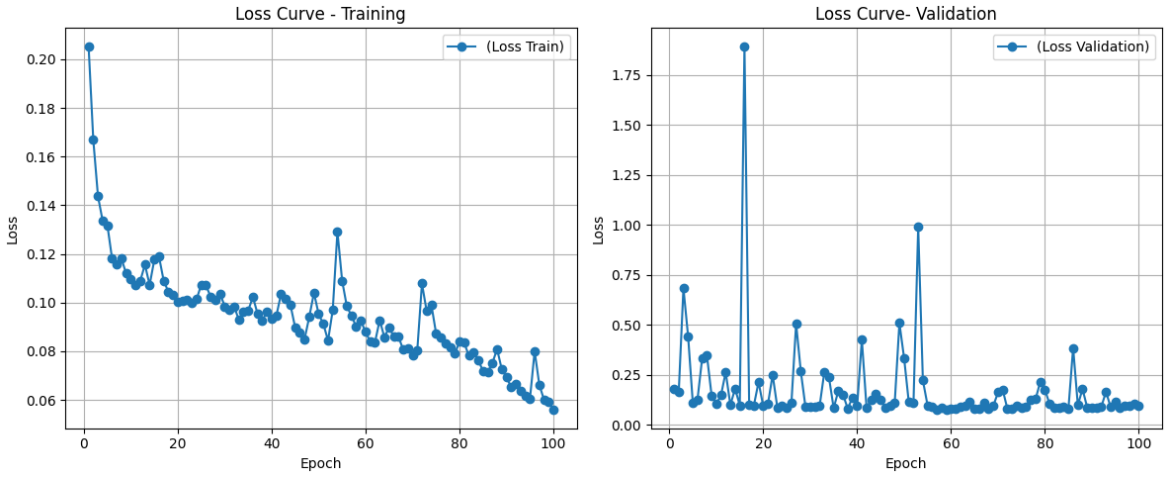

Figure 34 - Train and Validation loss curve for the best model

The best classification threshold that maximizes the F1-score for this model has a value of 0.32, meaning that the instances in the probability map that have a probability of being vineyard larger or equal to 32% are classified as vineyard, while the rest is classified as background. This threshold is relatively low, which might indicate that a) the best classification threshold is actually smaller than what would be expected and the model is sensitive to lower probabilities of a pixel being vineyard, b) due to the model complexity, it could be overfitting to the dominant class (background), as a result of the class imbalance issue in the dataset, or c) the original dataset split process, which randomly allocated image chips into the train validation and test datasets, allocated a large number of image chips with high vineyard representation to the test dataset, resulting in a low representation of background instances in the testing scenario compared to a real-world scenario.

To ensure the previously obtained results are valid, we tested the model using a new Sentinel-2 image of a different region that also contains vineyards in it, with identifier S2A_MSIL2A_20230713T112121_N0509_R037_T29TNF_20230713T190857, and retrieved the resultant metrics, as observed in Figure 35:

```
Mean Test Accuracy = 0.9377754820244653
Mean Test Precision = 0.15315009935043886
Mean Test Recall = 0.46339485528213636
Mean Test IOU = 0.1371704436501854
Mean Test AUC = 0.7032451225178582
Mean Test F1-score = 0.21950256709325394
```

Figure 35 - Test results in a different Sentinel-2 image

We also tested the best classification threshold that maximizes the F1-score for this case, obtaining a result equal to 0.89.

As we can infer, the new testing results are very disparate from the previous results, with a significant difference in the classification threshold. This indicates that, most likely, a mixture of points a) and b) mentioned earlier could be the primary contributing factors to this issue. Given that the best classification threshold is 0.89 for this new testing scenario and 0.32 for the first, it suggests that indeed the former test dataset may have an overrepresentation of vineyard features in the corresponding image chips.

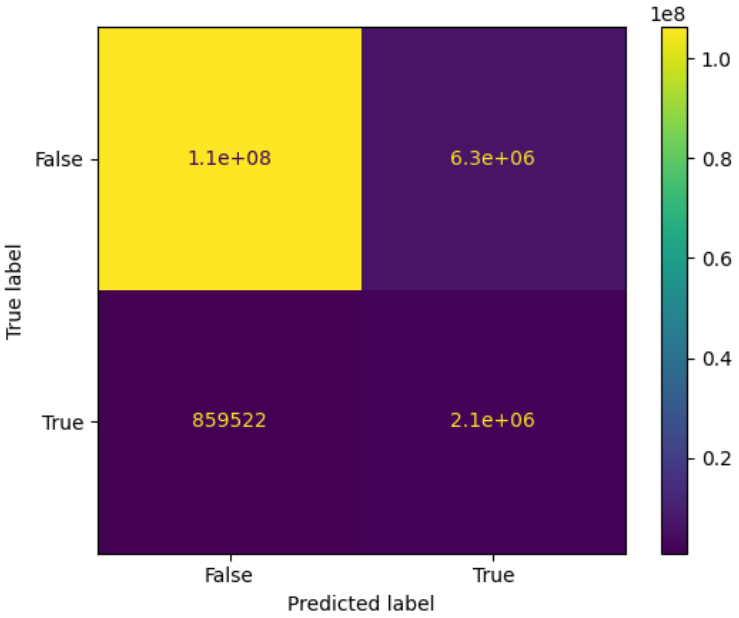

Figure 36 - Confusion matrix resultant from the model test in the 2nd dataset

The resultant confusion matrix, depicted in Figure 36, provides strong evidence of an imbalance of the testing scenario, with far more classified background instances than vineyard instances. Regardless, it is possible to infer that there are 2,086,431 TP, 6,342,416 FP, 106,317,135 TN and 859,522 FN in the output segmented mask, where most of the previous metrics were derived from.

While in the first test dataset the results demonstrate that the segmented mask is similar to the original mask, with a small proportion of misclassifications, as depicted in the following example, Figure 37[10]:

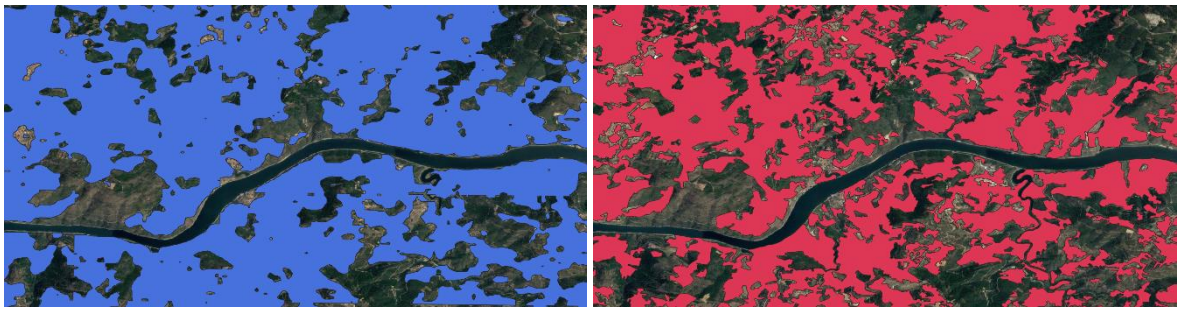

Figure 37 – Sampled segmented mask (left) vs ground truth (right) 1st test dataset

[10] Adapted from Copernicus Sentinel data 2023, processed by ESA.

In the second test scenario, the results exhibit suboptimal performance, with a large number of misclassifications, respectively type I errors, where a larger number of background instances are mistakenly classified as vineyards (FP), as depicted in the example below, Figure 38[11]:

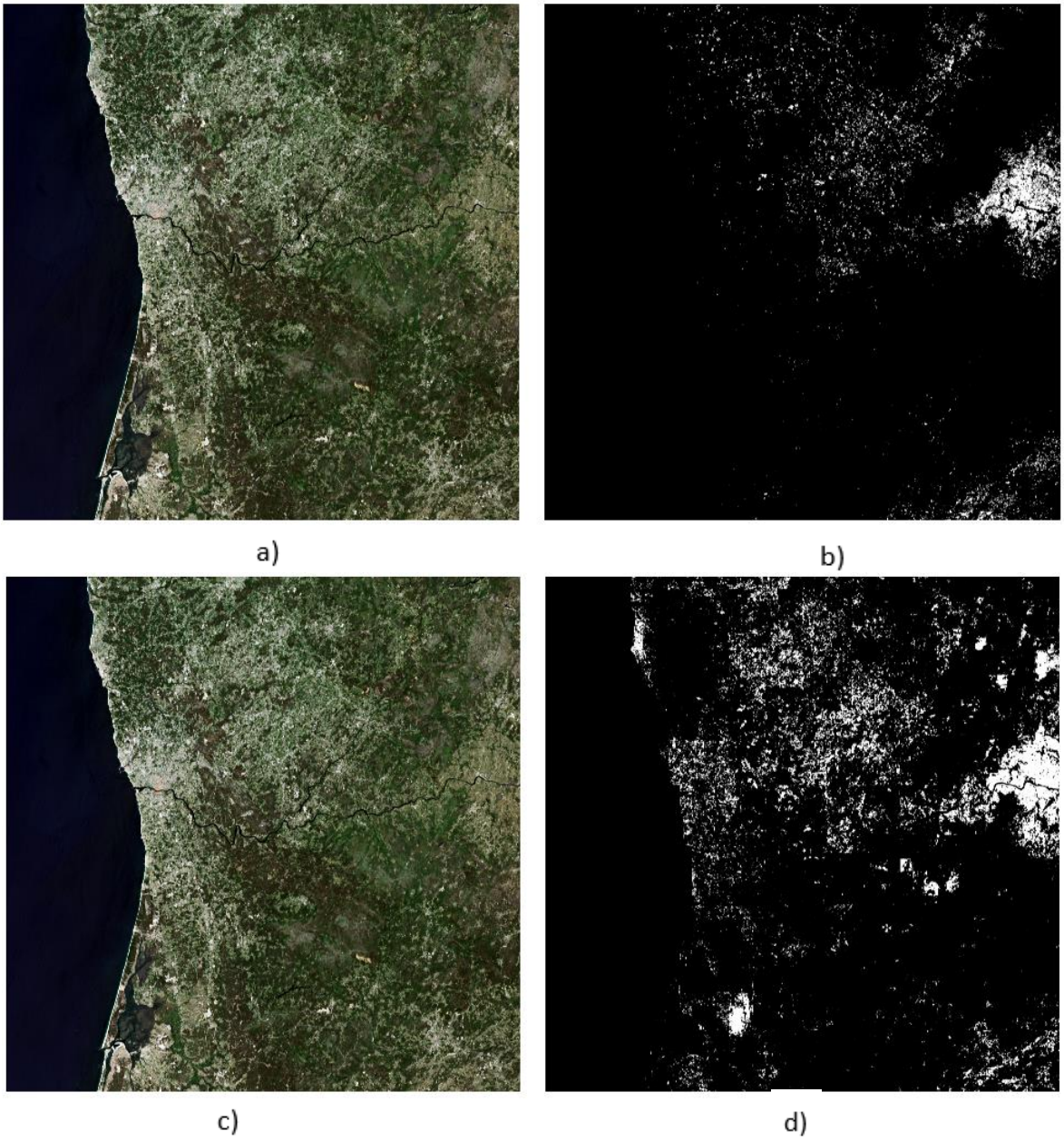

Figure 38 -natural color image, a) and c), with ground truth labels, b) vs segmented mask, d), for the 2nd test dataset

[11] Adapted from Copernicus Sentinel data 2023, processed by ESA.

Where images a) and b) correspond to the natural colour image, followed by the ground truth labels for the vineyards, and images c) and d) correspond also the natural colour image, but followed by the resultant segmented mask.

While regions with a higher vineyard density, particularly those located at the central right and lower right sections of the image appear to be satisfactorily classified, the remaining image has a high number of false positive instances, as mentioned previously.

Given the differences in metrics between the first and second testing scenarios, the original dataset should again be split randomly, and the models retrained and retested given the new data separation. Due to computational resource access limitations, it was not possible to perform this process again in a newly split dataset.

9. Altimetry Data

To have a higher confidence that the vineyard features previously segmented are located within terraces, in this topic we will study the creation of slope masks, given a desired slope threshold, which will then be intersected with the segmented vineyards, ensuring this way that the vineyards are contained in regions with positive or negative slopes, circumstances that may necessitate the construction of terraces, and not on flatlands where there are no geological constraints that justify building terraces.

The SRTM NASA data was obtained from the DWTKNS srtm30m interface, a 30 meter SRTM tile visualization and download interface developed by Derek Watkins (2023). For our application, we selected a tile that covers the most part of the *Alto Douro* region, as depicted in Figure 39:

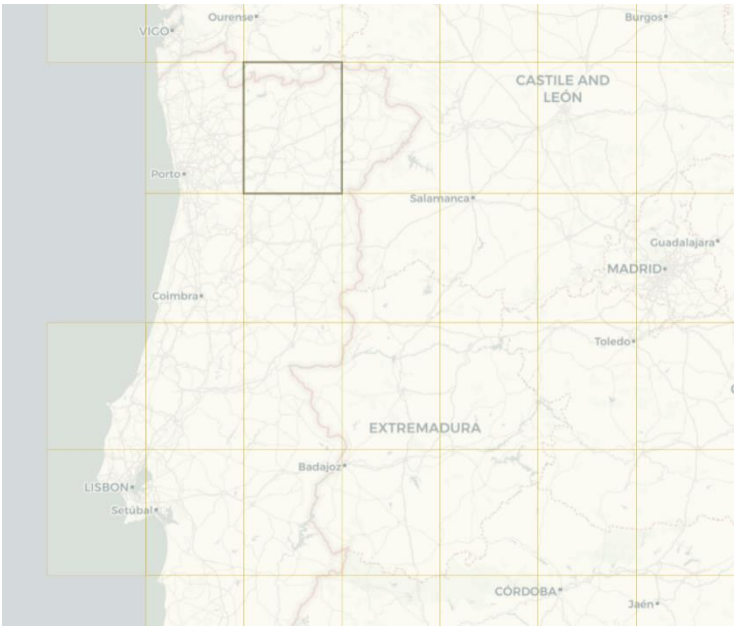

Figure 39 - Altimetry tile for the Alto Douro region, adapted from (Watkins, 2023)

With the corresponding altimetry raster data, as depicted in Figure 40[12]:

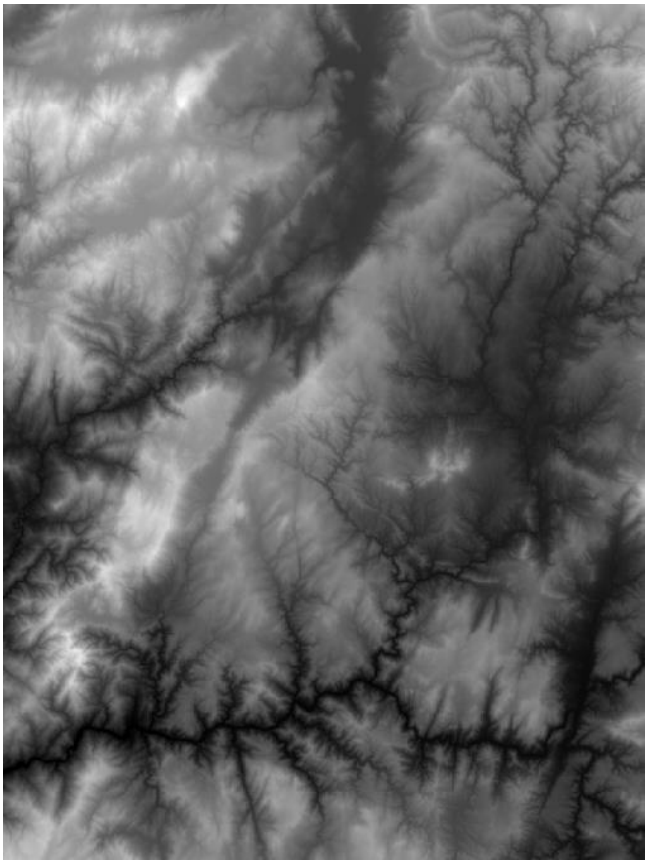

Figure 40 - Altimetry data from the previous tile

Darker pixels corresponds to lower regions, while brighter pixels correspond to higher regions. After a short EDA analysis on this data, it is possible to determine that the lowest point in this tile has a height of 36 meters and the highest point a height of 1,524 meters. The mean height and mean absolute slope of the image are 627 meters and 10.6º respectively, indicating that this region is moderately composed by highlands.

According to Fonseca (2023), a renowned wine producer from the region and one of the first wine producer houses of the vintage Porto wine, approximately two thirds of the region's vineyard planted area is located on rocky hillsides with a slope higher than 30%, equivalent to 16.7º. Given that we want to maximize the segmented vineyard area, we will use an empirical conservation factor of 0.25, this way defining the slope threshold to 13º, increasing the size of the intersection between both vector data.

[12] NASA Shuttle Radar Topography Mission (SRTM) (2013). Shuttle Radar Topography Mission (SRTM) Global

To extract the slope from the elevation values, we first calculate the gradient of the selected region, which, as demonstrated in Equation (9.1), is mathematically defined in $\mathbb{R}^2$as:

$$\nabla f = \frac{\partial f}{\partial x}\hat{\imath} + \frac{\partial f}{\partial y}\hat{\jmath} \tag{9.1}$$

Performing a variation to fit our study, we get Equation (9.2):

$$\nabla z = (\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}) \tag{9.2}$$

In this case, we define $\frac{dz}{dx}$ and $\frac{dz}{dy}$ with a central finite difference approximation approach as depicted in Equations (9.3) and (9.4), obtaining the relative slopes in x and y directions:

$$\frac{\partial z}{\partial x} = \frac{z[i, j+1] - z[i, j-1]}{2 * \Delta x} \tag{9.3}$$

$$\frac{\partial z}{\partial y} = \frac{z[i+1, j] - z[i-1, j]}{2 * \Delta y} \tag{9.4}$$

Where i and j correspond to the indices of the pixels located in x and y coordinates of image, computationally representing the indices of the pixels in the 2D NumPy array, and Δx and Δy the pixel spacing, in this case 30 meters.

We then calculate the magnitude of the gradient by performing the square root over the previously obtained partial derivatives squared, multiplied by the arctangent to obtain the slope of the terrain in radians at each point of the image, Equation (9.5):

$$slope_{rad} = \arctan(\sqrt{\left(\frac{dz}{dx}\right)^2 + \left(\frac{dz}{dy}\right)^2}) \tag{9.5}$$

Which then is converted to degrees with Equation (9.6):

$$slope_{deg} = slope_{rad} * \frac{180º}{\pi} \quad (9.6)$$

Despite the mathematical approach previously explained to derive the slope in degrees from altitude, most of the methods used for our calculations are already implemented in python functions.

From here, we provided the previously defined absolute slope threshold, which is 13º, and created a mask where all values larger than that threshold obtain the value of 1, and 0 for all other absolute slope values, Equation (9.7):

$$slope_{mask}(x, y) = \begin{cases} 1, if\ |\nabla z(x, y)| > 13 \\ 0, otherwise \end{cases} \quad (9.7)$$

After performing all necessary computations, we obtained the slope mask, depicted in Figure 41:

Figure 41 - slope mask with slope values > 13º

Having obtained the raster slope mask, it then went through a polygonization process, *i.e.*, it was converted from raster to vector data. This was performed to ensure that it would be possible to perform the intersection between the slope mask polygons and the segmented vineyard polygons, obtaining this way exclusively vineyard polygons contained within terrain with a slope higher than 13º. The result of the polygonization, overlapped with a subset of a segmented vineyard mask can be observed in Figure 42:

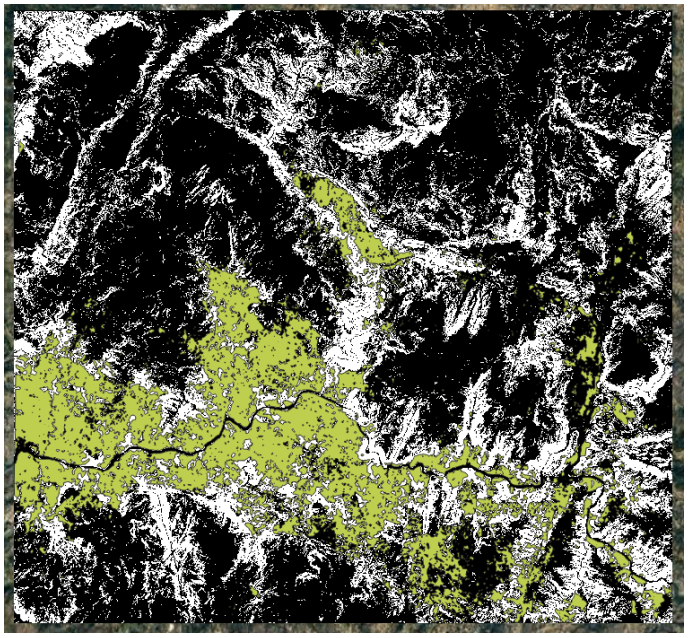

Figure 42 - Slope mask overlapped with the segmented vineyard mask

Finally, the intersection process of both vector masks was performed, obtaining the final vineyard terrace mask, depicted in Figure 43:

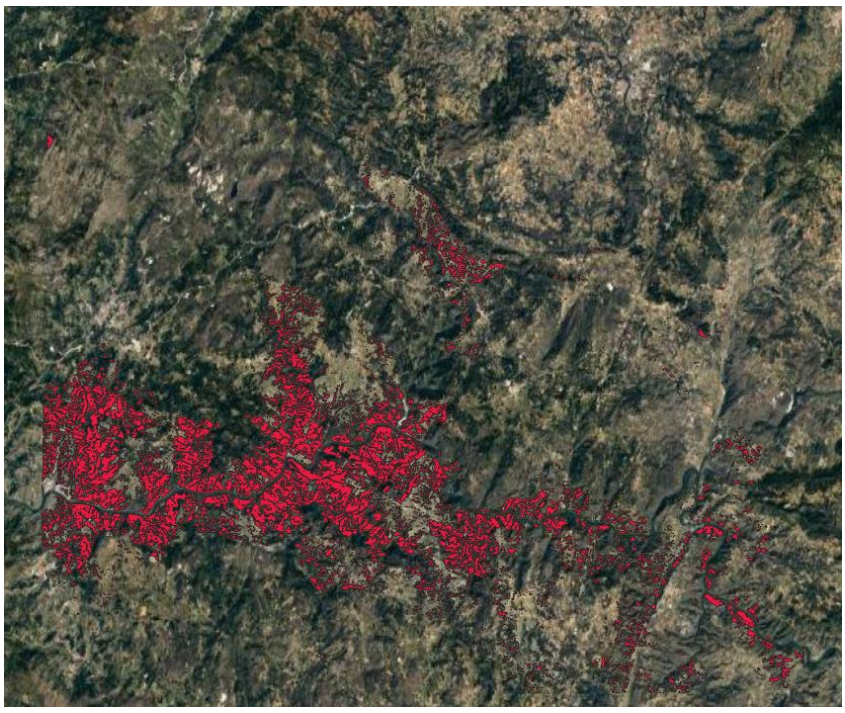

Figure 43 - Vineyard terrace mask

Observing the vineyard terrace mask, it is possible to verify that most of the vineyard terraces of this region are adjacent to the Douro River, built within the rocky terrain over the Douro valley. Performing a closer inspection to a subset of the previous figure, Figure 44, one can infer that, overall, most of the vineyard terraces of that subset region are accurately represented over the segmented terrace mask.

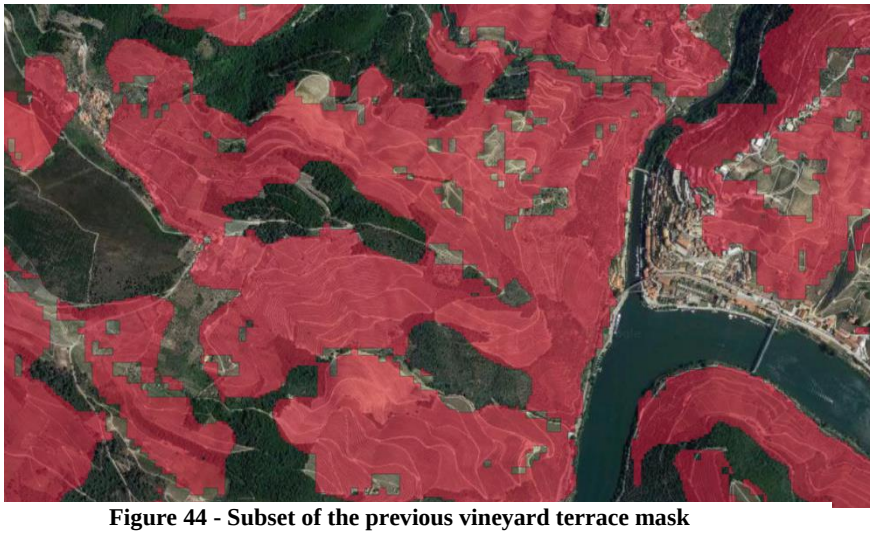

Figure 44 - Subset of the previous vineyard terrace mask

From these vector masks, since they use the EPSG:3857 coordinate reference system (CRS), also designated as WGS 84 / Pseudo-Mercator, not only is it possible to visually and geographically identify vineyard terrace regions, but also calculate useful metrics from them. Taking Figure 44 as an example, we can calculate the respective geometry attributes information of the represented polygons, as depicted in Figure 45:

Added geom info — Features Total: 6, Filtered: 6, Selected: 0

	fid	DN	id	class	area	perimeter
1	53759	1	1335	vineyard	5227705.5...	60729.35...
2	53759	1	1936	vineyard	391352.46...	6328.972...
3	53759	1	1134	vineyard	90649.754...	1649.063...
4	53759	1	1110	vineyard	83939.887...	2045.873...
5	35103	1	1335	vineyard	1292.5064...	143.8057...
6	34911	1	1335	vineyard	1292.5064...	143.8057...

Figure 45 - Geometry attributes information

From the former figure, it is possible to infer that the total number of segmented vineyard terrace polygons in Figure 44 is 6, with the larger polygon having a perimeter of approximately 60,729 m and an area of approximately 5,227,706 m^2 and the smallest polygon a perimeter of approximately 144 m and an area of approximately 1293 m^2.

The total occupied perimeter by vineyard terraces identified with the modified DeepLabV3 model in Figure 44 amounts to approximately to 71,041 m with an approximate area of 5,796,233 m^2, thus demonstrating valuable information for an efficient management of this World Heritage Site.

10. Discussion & Conclusions

In this chapter we will discuss the overall progress, results and challenges faced during the making of this dissertation, and provide with the conclusions obtained from the developed work.

10.1. Discussion

In this study we performed three different techniques in an effort to identify vineyard features within the Sentinel-2 images.

Given the spatial resolution limitations to identify such small features in very large regions, the overall behaviour of all three techniques delivered the best achievable results, although not meeting the desired expectations.

The first technique, segmentation using band indices, discerns the worst results out of the three, with vineyard and non-vineyard features mixture within the same region both in the vegetation as well as in the non-vegetation region.

The second technique, segmentation using traditional machine learning techniques, provided better results, and a more complete metric assessment was viable to perform. Unfortunately, the outcome of the metrics of all models trained were below a satisfactory level, with complex features such as the vineyards not being correctly predicted within the expected range. For this reason, we continued the search for a more suitable technique to identify these features.

Of the three techniques and several models trained within the last two, the model that yielded the best test performance was attained using the third and last technique explored in this study: a deep learning approach for vineyard segmentation, respectively, a convolutional neural network, with its model relying on the modified DeepLabV3 architecture with the ResNet101 backbone, trained with a batch size of 32 with 256x256 image chips with data augmentation for 100 epochs and without a bias parameter in the first convolutional layer, failed to generalize the predictions on a different testing scenario.

Although the model failed to meet the first test scenario metric results, in a scenario with a smaller concentration of vineyard plantations, it performs satisfactorily in regions with a

higher vineyard density. If we apply this model on regions where we know there exists vineyards *a priori*, the results improve greatly.

Better results could be extracted using the DL methodology, studying the behaviour of other semantic segmentation architectures, including the ones excluded in the beginning of Chapter 8, and also the U-Net architecture, which provided good results in other semantic segmentation articles.

Other variables that could impact the metric results of the model are the class distribution, which, as we have seen, this dataset has an inherent high class imbalance. Given the availability of more time and computational resources, the training and testing of more data augmentation variations and combinations could be performed, increasing the probability that the model would discern better results. The experiment of other data preprocessing techniques, such as feature selection of the Sentinel-2 band may also improve metric results and training time by only selecting bands relevant to the identification of vineyard features. Another variable that could improve metric results is the training and testing of other randomly split sets. As we have seen, the model's performance on the test dataset didn't correspond to the actual metrics on a more accurate representation of a real test scenario.

Regardless of these results, the previously obtained segmented vineyard mask was combined with a slope mask derived from SAR altimetry data, in order to obtain a higher confidence that the vineyards segmented are located within terraces. To do this, an intersection between both vector data was performed, and a final vineyard terrace mask was created. Although it shouldn't be taken as a high confidence or true vineyard terrace segmented mask, it could provide not only UNESCO and state site managers or conservation officers, but also farmers, agricultural engineers, wine producers and other interested parts a rough estimation of whether the terrain is used for vineyard plantations and whether or not it is inserted in a terrain or in a lowland field configuration, in any time of the year.

A more precise estimation could be performed if we combine the altimetry data with the vineyard polygons directly extracted from the COS2018. However, it's worth noting that the data is already 6 years old, which could arguably be considered legacy data in order to perform a more recent land study. New vineyard regions may have emerged, while others might have disappeared during such a long period of time. Hence, there is a need for a product that could remain available and updated all-year round.

10.2. Conclusions

In conclusion, this study aimed to identify vineyard features within Sentinel-2 images using different techniques, with a primary focus on the DeepLabV3 model, as an alternative method to already labour intensive existing ones, to assess vineyard terrace conservation and preservation.

Despite significant efforts, the spatial resolution limitations of Sentinel-2 imagery posed challenges in achieving optimal results for the identification of such small features in vast regions. Given that the comparative study between high and very-high resolution satellite images in order to identify vineyard terraces in the Douro region, planned in the beginning of the dissertation, was not able to be performed due to data acquisition limitations, an analysis of the differences in metrics given by the different spatial resolutions could not be developed. Regardless, it is expected that better spatial resolutions will lead to the models ability to better differentiate between the vineyard features and the other background features present in the images.

While the segmentation using band indices and the segmentation through traditional machine learning techniques displayed less satisfactory outcomes, the segmentation using a deep learning approach showed some improvement, yet did not meet the desired metrics for identifying complex vineyard features.

The best test performance was achieved with a convolutional neural network, specifically the modified DeepLabV3 model with the ResNet101 backbone. However, its generalization on a different testing scenario proved less effective, indicating the need for further refinement to suit both dense and non-dense vineyard scenarios.

Future work should explore the use of other semantic segmentation architectures, such as the U-Net and FCN models. By investigating these different architectures, we can assess their potential to yield better results than the DeepLabV3 architecture used in this study. Studying the application of the ResNet50 backbone architecture could also be advantageous to analyse the differences in the metrics from the ResNet101 and determine if the results yield by the Resnet101 outweighs its computational cost. Additionally, addressing class distribution imbalances from the dataset, i.e., vineyard and background features, and employing more extensive data augmentation techniques with increased computational resources could enhance the model's performance.

Despite the challenges inherent to the segmentation processes developed in this dissertation, the resultant segmented vineyard mask, when combined with slope data derived from SAR altimetry, offers a valuable tool for obtaining a preliminary estimation of vineyards within terrace terrain configurations. This information could benefit various stakeholders, both private and public.

As part of future improvements, efforts should be directed towards enhancing the cloud removal algorithm which could rely on a machine learning model on itself to segment the desired cloud features we want to intersect and remove from the vineyard mask, further developing the SentProd library for production-level deployment, and consider the seasonality factors to ensure year-round vineyard identification, leading to a higher availability of the vineyard terrace segmentation products of the region.

Furthermore, comparing the results using very-high resolution imagery, such as GEOSAT-2, could yield valuable insights into refining the analysis and identification of vineyard terraces.

Ultimately, while the study acknowledges the limitations and areas for improvement encountered in the different methods used, the obtained vineyard mask, when combined with altimetry data, serves as a valuable resource for preliminary assessments of the presence and area occupied by vineyard terraces.

11. **Bibliography**

ADVID. (2021). *Boletim Balanco ano viticola 2021*. Retrieved from https://www.advid.pt/uploads/DOCUMENTOS/Subcategorias/ano-vitico/Boletim_Balanco_ano_viticola_2021_VFact.pdf

Aguiar, F. B. (2002). O Alto Douro Vinhateiro, uma paisagem cultural, evolutiva e viva. *DOURO - Estudos & Documentos*, 143-152.

Akar, Ö., & Güngör, O. (2012). Classification of multispectral images using Random Forest algorithm. *Journal of Geodesy and Geoinformation*, 105-112.

Albetis, J., Duthoit, S., Guttler, F., Jacquin, A., Goulard, M., Poilvé, H., . . . Dedieu, G. (2017). Detection of Flavescence dorée Grapevine Disease Using Unmanned Aerial Vehicle (UAV) Multispectral Imagery. *remote sensing*, 308.

Alsdorf, D. E., Rodriguez, E., & Lettenmaier, D. P. (2007, May 09). Measuring surface water from space. *Reviews of Geophysics*.

Bauer, M. E. (2020). Remote Sensing of Environment: History, Philosophy, Approach and Contributions, 1969 –2019. *Remote Sensing of Environment*, 111522.

Berry, P. A. (2000). Topography from land radar altimeter data: Possibilities and restrictions. *Physics and Chemistry of the Earth, Part A: Solid Earth and Geodesy*, 81-88.

Bhardwaj, A., Sam, L., Bhardwaj, A., & Martín-Torres, F. J. (2016). LiDAR remote sensing of the cryosphere: Present applications and future prospects. *Remote Sensing of Environment*, 125-143.

Brando, V. E., & Dekker, A. G. (2003). Satellite hyperspectral remote sensing for estimating estuarine and coastal water quality. *IEEE Transactions on Geoscience and Remote Sensing*, 1378-1387.

Breiman, L. (2001). Random Forests. *Machine Learning*, 5–32.

Calvin, W. M., Littlefield, E. F., & Kratt, C. (2015). Remote sensing of geothermal-related minerals for resource exploration in Nevada. *Geothermics*, 517-526.

Carlos, C., Gonçalves, C., Costa, J., Costa, H., Alcazar, R., Sousa, J., . . . Gomes, E. (2021). *Guia de Construção de Muros de Pedra Seca.* ADVID - Associação para o Desenvolvimento da Viticultura Duriense.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 321-357.

Chen, F., Lasaponara, R., & Masini, N. (2017). An overview of satellite synthetic aperture radar remote sensing in archaeology: From site detection to monitoring. *Journal of Cultural Heritage*, 5-11.

Chen, L.-C., Papandreou, G., Schroff, F., & Adam, H. (2017). Rethinking Atrous Convolution for Semantic Image Segmentation. *Computer Vision and Pattern Recognition (cs.CV)*.

Copernicus. (2022, 11 25). *sentinel-2 satellite description orbit*. Retrieved from https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2/satellite-description/orbit

Copernicus. (2022, 11 25). *technical guides sentinel-2 msi*. Retrieved from https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-2-msi

Copernicus. (2023, 04 12). *scihub copernicus.* Retrieved from https://scihub.copernicus.eu/dhus/#/home

Copernicus. (2023, 05 19). *sentinel2 instrument payload resolution and swath.* Retrieved from https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2/instrument-payload/resolution-and-swath

Copernicus. (2023, 01 25). *Sentinel-2-MSI overview*. Retrieved from https://sentinels.copernicus.eu/documents/247904/266422/Sentinel-2-MSI_overview.jpg

Direção Geral do Território. (2022). *Carta de Uso e Ocupação do Solo - 2018 metadata*. Retrieved from dgterritorio: https://snig.dgterritorio.gov.pt/rndg/srv/por/catalog.search#/metadata/b498e89c-1093-4793-ad22-63516062891b

eoportal. (2023, 01 25). *deimos 2 geosat 2.* Retrieved from https://www.eoportal.org/satellite-missions/deimos-2-geosat-2

ESA. (2022, 11 25). *Gearing up for third Sentinel-2 satellite*. Retrieved from https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-2/Gearing_up_for_third_Sentinel-2_satellite

ESA. (2022, 11 25). *missions envisat*. Retrieved from https://earth.esa.int/eogateway/missions/envisat

ESA. (2022, 11 24). *missions ers*. Retrieved from https://earth.esa.int/eogateway/missions/ers

ESA. (2022, 11 25). *sentinel missions*. Retrieved from https://sentinel.esa.int/web/sentinel/missions

EU Science Hub. (2023, 01 16). *earth observation*. Retrieved from https://joint-research-centre.ec.europa.eu/scientific-activities-z/earth-observation_en

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., . . . Alsdorf, D. (2007). The Shuttle Radar Topography Mission. *Reviews of Geophysics*.

Figueiredo, N., Neto, A., Cunha, A., Sousa, J. J., & Sousa, A. (2022). Deep Learning Approach for Terrace Vineyards Detection from Google Earth Satellite Imagery. *International Symposium on Geoscience and Remote Sensing* (pp. 5824-5827). Kuala Lumpur, Malaysia: IEEE.

Fonseca. (2023, 07 28). *vinhas douro vale do douro terroir*. Retrieved from https://www.fonseca.pt/pt/vinhas-douro/vale-do-douro/terroir

GEOSAT. (2023, 01 02). *GEOSAT 2 Imagery User Guide*. Retrieved from https://geosat.space/wp-content/uploads/2022/04/GEOSAT-2-Imagery-User-Guide-v3.2.pdf

Giovos, R., Tassopoulos, D., Kalivas, D., Lougkos, N., & Priovolou, A. (2021). Remote Sensing Vegetation Indices in Viticulture: A Critical Review. *agriculture*, 457.

Google. (2022). *maps*. Retrieved from https://www.google.com/maps

Google. (2022, 12 29). *vector map*. Retrieved from https://developers.google.com/maps/documentation/javascript/vector-map

Hahn, M. (2009, 08 20). *GRIN*. Retrieved from https://web.archive.org/web/20091130171224/http://grin.hq.nasa.gov/ABSTRACTS/GPN-2002-000200.html

He, K., Zhang, X., Ren, S., & Sun, J. (2016, June). Deep Residual Learning for Image Recognition. *Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770 - 778). Las Vegas, NV, USA: IEEE.

Huang, S., Tang, L., Hupy, J. P., Wang, Y., & Shao, G. (2020). A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. *Journal of Forestry Research*, 1–6.

Internet Archive. (2003). *GPN-2002-000200*. Retrieved from https://web.archive.org/web/20091222041935/http://grin.hq.nasa.gov/IMAGES/SMALL/GPN-2002-000200.jpg

Jones, G. E., Wong, S., Milton, A., Sclauzero, J., Whittenbury, H., & McDonnell, M. D. (2020). The Impact of Pan-Sharpening and Spectral Resolution on Vineyard Segmentation through Machine Learning. *remote sensing*, 934.

Kandylakis, Z., & Karantzalos, K. (2016). Precision Viticulture from Multitemporal, Multispectral Very High Resolution Satellite Data. *ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 919.

Laroche-Pinel, E., Albughdadi, M., Duthoit, S., Chéret, V., Rousseau, J., & Clenet, H. (2021). Understanding Vine Hyperspectral Signature through Different Irrigation Plans: A First Step to Monitor Vineyard Water Status. *remote sensing*, 536.

Lary, D. J., Alavi, A. H., Gandomi, A. H., & Walker, A. L. (2016). Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, 3-10.

Li, J., Pei, Y., Zhao, S., Xiao, R., Xiao, S., & Zhang, C. (2020). A Review of Remote Sensing for Environmental Monitoring in China. *remote sensing*, 1130.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully Convolutional Networks for Semantic Segmentation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Lourenço-Gomes, L., Pinto, L. M., & Rebelo, J. (2015). Wine and cultural heritage. The experience of the Alto Douro Wine Region. *Wine Economics and Policy*, 78-87.

Luo, L., Wang, X., Guo, H., Lasaponara, R., Zong, X., Masini, N., . . . Yao, Y. (2019). Airborne and spaceborne remote sensing for archaeological and cultural heritage applications: A review of the century (1907–2017). *Remote Sensing of Environment*, 111280.

Martins, C. A. (1991). A filoxera na viticultura nacional. *Análise Social*, 653-688.

McAllister, E., Payo, A., Novellino, A., Dolphin, T., & Medina-Lopez, E. (2022). Multispectral satellite imagery and machine learning for the extraction of shoreline indicators. *Coastal Engineering*, 104102.

Melki, P., Bombrun, L., Millet, E., Diallo, B., ElGhor, H. E., & Da Costa, J.-P. (2022). Exploratory Analysis on Pixelwise Image Segmentation Metrics with an Application in Proximal Sensing. *remote sensing*, 996.

Michailovsky, C. I., McEnnis, S., Berry, P. A., Smith, R., & Bauer-Gottwein, P. (2012). River monitoring from satellite radar altimetry in the Zambezi River basin. *Hydrology and Earth System Sciences*, pp. 2181–2192.

Moccia, S., De Momi, E., El Hadji, S., & Mattos, L. S. (2018). Blood vessel segmentation algorithms — Review of methods, datasets and evaluation metrics. *Computer Methods and Programs in Biomedicine*, 71-91.

Müller, D., Soto-Rey, I., & Kramer, F. (2022). Towards a guideline for evaluation metrics in medical image segmentation. *Research Notes*, 210.

NASA. (2023, July 29). *srtm instrument interferometry*. Retrieved from https://www2.jpl.nasa.gov/srtm/instrumentinterferometry.html

NOAA CoastWatch. (2021). *remote sensing basics*. Retrieved from https://coastwatch.gitbook.io/satellite-course/lectures/remote-sensing-basics

Oxford Languages. (2022, 12 16). *oxford english dictionary*. Retrieved from https://languages.oup.com/research/oxford-english-dictionary/

Paperspace. (2020). *auc area under the roc curve*. Retrieved from machine learning paperspace: https://machine-learning.paperspace.com/wiki/auc-area-under-the-roc-curve

Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J.-M., J.Tucker, C., & Stenseth, N. C. (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in Ecology & Evolution*, 503-510.

PyTorch. (2023, July 01). *vision stable models*. Retrieved from https://pytorch.org/vision/stable/models.html

Rawat, J. S., & Kumar, M. (2015). Monitoring land use/cover change using remote sensing and GIS techniques: A case study of Hawalbagh block, district Almora, Uttarakhand, India. *The Egyptian Journal of Remote Sensing and Space Science*, 77-84.

Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. *Goddard Space Flight Center 3d ERTS-1 Symp., Vol. 1, Sect. A* (pp. 309-317). Remote Sensing Center, Texas A&M University, College Station, Texas: NASA.

Sabins, F. F. (1999). Remote sensing for mineral exploration. *Ore Geology Reviews*, 157-183.

SEOS. (2023, 02 12). *classification.* Retrieved from https://seos-project.eu/classification/classification-c01-p05.html

Soubry, I., Patias, P., & Tsioukas, V. (2017). Monitoring vineyards with UAV and multi-sensors for the assessment of water stress and grape maturity. *Journal of Unmanned Vehicle Systems*, 37-50.

Suykens, J. A. (2001). Support Vector Machines: A Nonlinear Modelling and Control Perspective. *European Journal of Control*, 311-327.

Suykens, J. A., & Vandewalle, J. P. (1998). *Advanced Black-Box Techniques, Nonlinear Modeling.*

Tatem, A. J., Goetz, S. J., & Hay, S. I. (2008). Fifty Years of Earth Observation Satellites: Views from above have lead to countless advances on the ground in both scientific knowledge and daily life. *American Scientist*, 390–398.

Tayade, R., Yoon, J., Lay, L., Khan, A. L., Yoon, Y., & Kim, Y. (2022). Utilization of Spectral Indices for High-Throughput Phenotyping. *plants*, 1712.

UNESCO. (2001). *list 046*. Retrieved from https://whc.unesco.org/en/list/1046/

UNESCO. (2015). *State of Conservation Report, Alto Douro Wine Region World Heritage Property (1046).*

USGS. (2018, 07 30). *usgs eros archive digital elevation shuttle radar topography mission srtm 1*. Retrieved from https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-shuttle-radar-topography-mission-srtm-1

van der Meer, F. D., van der Werff, H. M., van Ruitenbeek, F. J., Hecker, C. A., Bakker, W. H., Noomen, M. F., . . . Woldai, T. (2012). Multi- and hyperspectral geologic remote sensing: A review. *International Journal of Applied Earth Observation and Geoinformation*, 112-128.

Wang, Y., Su, J., Zhai, X., Meng, F., & Liu, C. (2022, February 8). Snow Coverage Mapping by Learning from Sentinel-2 Satellite Multispectral Images via Machine Learning Algorithms. *remote sensing*, 782.

Watkins, D. (2023, July 15). *srtm30m.* Retrieved from dwtkns: https://dwtkns.com/srtm30m/

Weiss, M., Jacob, F., & Duveiller, G. (2020). Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment*, 111402.

Zhuge, X.-y., Zou, X., & Wang, Y. (2017, 11). A Fast Cloud Detection Algorithm Applicable to Monitoring and Nowcasting of Daytime Cloud Systems. *Transactions on Geoscience and Remote Sensing*, 6111 - 6119.