NOVA IMS
Information Management School

MDSAA

Master's Degree Program in
Data Science and Advanced Analytics

Dynamic Wildfire Prediction in Portugal:

Leveraging Machine Learning and Remote Sensing for Enhanced Risk Mitigation

Afonso Bajanca Lavado

Master Thesis

presented as partial requirement for obtaining a Master's Degree in Data Science and Advanced Analytics

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

Dynamic Wildfire Prediction in Portugal:

Leveraging Machine Learning and Remote Sensing for Enhanced Risk Mitigation

by

Afonso Bajanca Lavado

Master Thesis presented as partial requirement for obtaining the Master's Degree in Data Science and Advanced Analytics, with a specialization in Data Science

Supervised by

Marco Painho, PhD, Nova Information Management School

December, 2024

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledged the Rules of Conduct and Code of Honor from the NOVA Information Management School.

[Lisbon, 02/12/2024]

ACKNOWLEDGEMENTS

This work would not have been possible without the incredible availability and support of my supervisor, Marco Painho, who guided me in this new area of GIS. Coming from a Data Science master’s program, I was warmly welcomed and equipped with all the tools needed to embark on this journey.

I also want to acknowledge the important help of my friends and work colleagues, who pushed me to think differently and tackle the many challenges that arose. They provided great insights, and for that, I am forever grateful.

Lastly, to my family for supporting me every step of the way, from beginning to end. To my older sisters and my father, who brought their wiser experience; to my younger brother, who kept me distracted during stressful times; and to my PhD mother, who served as my consultant - without her, this work would be nowhere near as polished.

I am deeply grateful to all of you.

ABSTRACT

Portugal has the highest annual average of burned areas among Mediterranean countries, facing a significant challenge in managing wildfires, particularly in its southern regions. Rising annual temperatures aggravate this crisis, underscoring the need for innovative solutions.

Although existing studies have successfully applied machine learning models using remote sensing, topographic, weather, and land classification data, they often prioritize accuracy over recall, which limits their effectiveness in predicting wildfire events. Furthermore, these models frequently use low spatial resolutions (e.g., 10 km), which lack the granularity needed for precise predictions.

This study adopts a novel approach by leveraging Sentinel-2 satellite imagery with a high spatial resolution of 20 m, along with NDVI and topographic data. Weather data was excluded due to its unavailability at this resolution, and land classification data was omitted to avoid introducing bias. For each wildfire event with a burned area>=100 ha, we utilized the closest pre-event satellite imagery (occasionally hours before the fire) and converted each 20 m-by-20 m pixel into a geo-localized point. Points within burned areas were labelled as 1, while points surrounding vegetation within a 5 km buffer zone were labelled as 0. To address class imbalance, we maintained a 1:2 ratio between burned and unburned points by randomly removing unburned points, resulting in a training dataset of over 4 million points from 32 wildfire events (2015-2022). The dataset was divided into two subsets based on temporal proximity to wildfire events (≤ 2 days prior and > 2 days prior).

We evaluated Random Forest (RF), XGBoost, and Multi-Layer Perceptron (MLP) models across various feature configurations and fine-tuned them using a time-series split for training. Integrating the entire study area with the unseen data from 2023, rather than just the buffer zone, dramatically increased the imbalance burned-to-unburned ratio from 1:2 to 1:5 000. For subset 1, MLP achieved the highest recall with 8.21%, while for subset 2, it was RF with a recall of 55.42%. However, the models overestimated burned areas - by factors of 5 500 and 93 for subsets 1 and 2, respectively - due to prioritizing maximum recall.

While balancing high recall with accuracy is challenging, dynamic fire risk maps generated by these models could significantly aid in wildfire risk mitigation. Future work should incorporate additional data sources, such as weather, and explore advanced techniques like Convolutional or Recurrent Neural Networks to address these challenges and improve performance to tackle Portugal's wildfire crisis effectively.

KEYWORDS

Remote Sensing; Sentinel-2; Machine Learning; Portugal; Wildfire Prediction

Sustainable Development Goals (SDG):

13 CLIMATE ACTION

15 LIFE ON LAND

TABLE OF CONTENTS

LIST OF FIGURES

Appendix C:

Appendix D:

LIST OF TABLES

Appendix A:

Appendix B:

Appendix E:

Appendix F:

LIST OF ABBREVIATIONS AND ACRONYMS

ANN	Artificial Neural Network.
ARI	Anthocyanin Reflectance Index
BSI	Bare Soil Index
CAOP	Portuguese: Carta de Uso e Ocupação do Solo; English: Official Administrative Charter of Portugal
COS	Portuguese: Carta de Uso e Ocupação do Solo; English: Land Use and Occupation Charter
CSV	Comma-separated values
CVI	Chlorophyll Vegetation Index
DEM	Digital Elevation Model
EVI	Enhanced Vegetation Index
FFWI	Fosberg Fire Weather Index
FN	False Negatives
FP	False Positives
FWI	Fire Weather Index
GBM	Gradient Boosting Machine.
GCI	Green Chlorophyll Index
ICNF	Portuguese: Instituto da Conservação e Natureza Florestal; English: Institute of Nature and Forest Conservation
IPMA	Portuguese: Instituto Português do Mar e da Atmosfera, English: Portuguese Institute of the Sea and Atmosphere
IoT	Internet of Things
LCI	Leaf Chlorophyll Vegetation Index
LFMC	Live Fuel Moisture Content
ML	Machine Learning
MLP	Multi-Layer Perceptron
MSAVI2	Modified Soil-Adjusted Vegetation Index 2
NBR	Normalized Burn Ratio

NDMI	Normalized Difference Moisture Index
NDRE	Normalized Difference Red Edge
NDSI	Normalized Difference Soil Index
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NIR	Near-Infrared
NUTS	Portuguese: Nomenclatura das Unidades Territoriais para Fins Estatísticos; English: Nomenclature of Territorial Units for Statistical Purposes
PCA	Principal Component Analysis
RBF	Radial Basis Function
RF	Random Forest.
SMOTE	Synthetic Minority Over-sampling Technique
SVM	Support Vector Machine.
SWIR1	Short-Wave Infrared 1
SWIR2	Short-Wave Infrared 2
TN	True Negatives
TP	True Positives
t-SNE	t-distributed Stochastic Neighbor Embedding
UMAP	Uniform Manifold Approximation and Projection
XGBoost	eXtreme Gradient Boosting

1. INTRODUCTION

Portugal grapples with a wildfire crisis that ranks among the most severe on a global scale. From 2006 to 2022, the country has consistently reported alarmingly high yearly averages of square metres consumed by wildfires (Statista, 2023). The gravity of this issue is compounded by Portugal's relatively small size compared to other Mediterranean nations. The consequences extend beyond immediate threats to human life and property, casting a long shadow of adverse environmental and socio-economic impacts.

In retrospect, the year 2017 emerged as an anomaly, serving as a stark reminder that the preceding years might have been characterized by fortuitous circumstances. The underlying reason for this impending crisis lies in the failure of government authorities to proactively address and rectify potential issues. The post-2017 period witnessed a gradual deceleration in such crises, yet it's crucial not to be succumb to a false sense of security, considering that a recently burned land takes up to 7 years to grow combustible vegetation (Tedim et al., 2013).

To address the ongoing crisis, some scholars have developed machine learning models that leverage remote sensing to detect wildfires and smoke (Rostami et al., 2022; Y. Sun et al., 2023), or even to predict its spreading behaviour (Adhikari et al., 2023). However, a significant drawback of this approach is that by the time a wildfire is detected by a satellite, authorities are likely already aware of it, and substantial resources are being mobilized for its suppression. This is due to the limited availability of some satellite images, such as Sentinel-2 (every 5 days (ESA, 2023)) and Landsat (every 8 days (*NASA*, 2023)) constellations.

While some researchers have employed remote sensing coupled with artificial intelligence to identify burned areas and assess burn severity (De Luca, Silva, & Modica, 2021; De Luca, Silva, Oom, et al., 2021; Dos Santos et al., 2023; Gómez & Martín, 2011; S. Liu et al., 2020; Tonbul et al., 2023), this method may not be a direct solution to the ongoing wildfire crisis. Nonetheless, its utility extends to diverse applications, encompassing the prediction of losses and the delineation of affected areas through polygon mapping.

As outlined by (Mira & Lourenço, 2013), wildfire prevention revolves around mitigating the three components of the "fire triangle":

- Activation energy: e.g., cigarettes, bonfires, abandoned equipment, grills.
- Fuel: including dead leaves, dried bushes, and flammable vegetation like eucalyptus.
- Oxidizing agents: primarily oxygen.

While reducing all three factors would be optimal, the reduction of oxygen is impractical, and addressing activation energy is a target for government awareness campaigns. Therefore, to properly address the wildfire crisis, the primary focus should be on minimizing fuel sources.

With that in mind, our research is focused on developing a suite of machine learning models aimed at predicting areas prone to major forest fires and determining the most effective model through comparative analysis. By focusing on wildfire prediction, this research aims to bridge the gap between reactive and proactive wildfire management strategies. The outcome has the potential to aid resource allocation, prioritize high-risk areas, and ultimately reduce the socio-economic and environmental toll of wildfires.

Our primary choice for a machine learning model capable of discerning the spectral signature of vegetation prone to burning is the Multi-Layer Perceptron (MLP) architecture, a well-established deep learning model commonly used for classification tasks, as supported by recent studies (Castrejon et al., 2023). Additionally, we considered the Random Forest (RF) model, known for its strong results in previous wildfire fuel characterization research (Santos et al., 2023), and eXtreme Gradient Boosting (XGBoost), recognized for its robust handling of complex and non-linear datasets (Guria et al., 2024). Drawing insights from historical fire data provided by the ICNF (Portuguese: Instituto da Conservação e Natureza Florestal; English: Institute of Nature and Forest Conservation), our models aim to enhance predictive accuracy by integrating environmental variables such as terrain data (elevation, slope and aspect). This comprehensive approach strengthens the models' effectiveness in forecasting fire risk.

In the next chapter, we will explore related works, diving into the differences between this study and existing research. This discussion will also address the overarching question: **Is it possible to predict wildfire-prone areas using high spatial resolution data?**

Additionally, we will examine how wildfires are propagated, as well as the measures and technologies used to prevent them. Following the discussion of related works, we will present the study area, datasets used, and the tools developed to predict areas prone to burn. The results of our analysis will then be shared, concluding with a summary of our findings and key takeaways.

2. LITERATURE REVIEW

Wildfires pose a significant threat to Portugal, needing a thorough exploration of the literature to understand the various facets of these destructive events. This chapter provides a comprehensive review of the existing knowledge on wildfires, covering topics ranging from their spreading behaviours to historical occurrences and contemporary preventive measures and technologies. In the end, we will analyse related works and highlights how this thesis distinguishes itself within the field.

2.1. FOREST FIRES

2.1.1. Definition

In a recent survey of over 220 specialists, consensus emerged around terms such as "Natural Hazard," "Disturbance," and "Climate-sensitive" to describe wildfires (Tedim & Leone, 2020).

Interestingly, human-caused fires are often classified as wildfire hazards when they pose threats to lives, property, social and economic structures, or the environment. While wildfires by definition are inherently natural phenomena, they can simultaneously be beneficial and hazardous. For certain ecosystems, it can be beneficial because they play a critical role in maintaining health and biodiversity. However, in fire-sensitive ecosystems, they can be devastating, affecting livelihoods, and leading to injuries or fatalities.

Most wildfires today are linked to human activities, challenging the notion that they are purely natural hazards. It's important to recognize that while not all wildfires result in disasters, certain conditions can escalate them to catastrophic levels, impacting vulnerable ecosystems and communities. Consequently, effective wildfire risk reduction and preparedness strategies are essential to minimize potential damage and save lives.

To conclude, wildfires, often referred to as forest fires or bushfires, are uncontrolled fires that spread rapidly through vegetation, potentially causing extensive damage to ecosystems and human infrastructure. In Portugal, for a fire to be considered a wildfire, it must be equal to or greater than 100 hectares (Tedim et al., 2013).

2.1.2. Spreading behaviour

According to (J. Sun et al., 2023) in "Facing the Wildfire Spread Risk Challenge: Where Are We Now and Where Are We Going?", while wildfires can exhibit discernible patterns, they remain infinitely complex and nonlinear systems, with their spread driven by meteorological, fuel, topographical, and human factors. Additionally, the article "Wildfire Likelihood's Elements: A Literature Review" by (Mhawej et al., 2015) provides a detailed classification of these factors, further enriching our understanding.

- **Meteorological**: The increasing global temperatures have led to drier forests, which facilitate the ignition and spread of wildfires. Precipitation is essential in maintaining moisture levels across various elements; however, during the drier months aggravated by climate change, this is often insufficient, leading to worsening conditions. Additionally, wind speed, direction and patterns significantly influence wildfire direction by spreading embers to unburned areas, creating multiple ignition points. Other relevant climatic factors include air humidity, and illumination time and intensity.

- **Fuel**: Fuel sources are divided into live and dead categories. Live fuel typically includes vegetation, with different types exhibiting varying ignition points and spread rates. Dead fuel encompasses materials such as litter (paper, cans, cigarettes) and dead vegetation (pine needles, dry grass, branches, trunks and fallen leaves), which are highly combustible. Additionally, the density of dead fuel affects oxygen availability; lower density increases the risk due to higher oxygen levels and faster spreading speeds.
- **Topographical**: Variations in elevation and slope can significantly impact both ignition and fire spread. Higher elevations typically feature lower temperatures and moister conditions in both flora and fauna, making it more challenging for fires to start and spread. In contrast, steeper slopes often have drier conditions because they retain less moisture from rainfall, leading to faster burn rates. Additionally, the effect of convection columns, which heat vegetation above the fire, increases the speed of fire propagation in an upward direction. Consequently, fires spread much faster uphill than downhill (de Castro et al., 2006). The aspect of the slope also plays a crucial role; in the Northern Hemisphere, south- or east-facing slopes, which receive more sunlight, tend to be drier, whereas north- or west-facing slopes, which are shadier, retain more moisture.
- **Human**: Human activities are the primary cause of wildfire ignitions, whether through negligence or intentional acts. Areas with higher population densities are more susceptible to a greater number of wildfires. Additional anthropogenic factors include proximity to agricultural land, roads, wildland-urban interface zones, recreation areas, breeding grounds, and exploitation zones.

Other researchers, such as (Z. Liu et al., 2012), have also examined wildfire spread and highlighted that climate change, alongside rising global temperatures, will likely worsen fire occurrences caused by both human activities and natural lightning events.

To build a machine learning (ML) model capable of predicting areas prone to burn, it is essential to incorporate variables that encompass the four primary factors influencing wildfire spread. **Meteorological** factors, such as precipitation, humidity, wind, and temperature, can be monitored using data from meteorological databases. **Fuel** factors can be analysed through satellite-derived multispectral bands to assess live and dead moisture. **Topographical** factors, including elevation, slope and aspect, can be evaluated employing terrain digital models. Finally, **human** factors can be assessed by examining the proximity of wildfires to densely populated areas to understand the influence of human activity.

2.2. WILDFIRES IN PORTUGAL

2.2.1. Comparison with other Mediterranean countries

A comparative analysis of forest fires in Southern Europe highlights significant differences in management and outcomes between Portugal and its neighbouring countries from 1981 to 2010.

While Greece, France, Italy, and Spain have shown success in reducing fire occurrences and burned areas, Portugal has struggled. Despite its smaller size, Portugal has not significantly reduced fire occurrences, and it is the only country with an increasing trend in burned areas. This emphasizes the urgent need for effective fire management programs in Portugal, as other countries have managed to implement more successful strategies (Mira & Lourenço, 2013).

2.2.2. Ignition causes

According to (Nunes et al., 2016), it is unsurprising that ignition numbers are higher in densely populated urban and suburban municipalities compared to rural areas. However, decades of rural-to-urban migration have led to the abandonment of agricultural lands, which are now predominantly covered by shrubs. As a result, rural areas experience a larger extent of burned land than high-density coastal cities. This discrepancy can also be attributed to the inaction and lack of resources in some rural villages compared to urban centres.

Between 1995 and 2015, 34.5% of forest fires in Portugal were attributed to unknown causes, possibly due to insufficient investigation. However, this percentage has been decreasing since 2000 due to improved investigative methods. During the same period, negligent and intentional causes were responsible for 29.9% and 20.4% of forest fires, respectively. Intentional forest fires can be divided into two categories: malevolent acts of pyromania, and fires caused by agricultural or livestock-related activities. As the proportion of unknown causes has declined, the detection of human-caused forest fires has increased. Reactivations - instances where areas re-burns due to latent heat, sparks, or embers igniting remaining fuel - constituted 14.6% of the total ignitions. In contrast, natural causes accounted for only 0.6% of the investigated forest fire ignitions (Meira Castro et al., 2020).

2.2.3. Current measures and technologies to prevent spreading

This section reviews the existing strategies, legislative frameworks, and technologies in Portugal aimed at mitigating wildfire spread, including land management practices, community engagement, and innovative tech-driven solutions.

In Portugal, entities like ICNF, along with government bodies, have established recommendations and guidelines to reduce fire risks during fire seasons. For instance, they discourage working with flammable materials on hot days, using motorized tools near rocks (to prevent spark creation), and launching fireworks in forested areas, just to name a few. However, as highlighted by (Mira & Lourenço, 2013), one of the most significant issues in Portuguese forests is excessive fuel accumulation. Over time, unmanaged biomass significantly increases the risk of wildfire ignition and spread. Effective fuel management practices can mitigate this risk by making it more difficult for fires to ignite and propagate yet delays in implementing these strategies in key areas remain a concern.

2.2.3.1. Legislative measures

To address fuel management, the Portuguese government enacted Decree-Law No. 156/2004 (Ministério Público Portugal, 2024) To remain within the scope of this thesis, we will not provide the full text of the Decree-Law. However, it is worth noting that Article 16 provides detailed specifications, which are summarized below for the reader:

- **Minimum clearances** of 10 metres on either side of roads, railways, and major electrical transmission lines during critical fire periods.
- **A minimum 50-metre buffer** around rural structures and 100 metres around forest-adjacent settlements, managed by municipal councils or parish authorities.
- **Protective strips of at least 100 metres** surrounding industrial zones and landfills near forested areas.

These fuel break zones significantly reduce fire spread by creating barriers that limit vegetation fuel. Researchers (Afonso et al., 2020; Lourenço et al., 2023; Pereira-Pires, Aubard, Ribeiro, et al., 2020; Pereira-Pires, Aubard, Silva, et al., 2020; Pereira-Pires et al., 2022) took the challenge, and are using satellite imagery combined with ML techniques to identify stretches of land that require a minimum 10-metre fuel break—an area cleared of vegetation—along certain roads to prevent wildfire spread, demonstrating effective ways of integrating AI technologies with remote sensing into wildfire prevention.

2.2.3.2. Early detection systems

Preventing wildfire spread also relies heavily on early detection. A few Portuguese startups, such as Bee2FireDetection (Bee2Fire, 2024) and Leitek (Leitek, 2024), are developing IoT-based solutions to detect wildfires early by placing sensors in forests. Other researchers (Guede-Fernández et al., 2021; Martins et al., 2022) have explored deep learning approaches to identify smoke columns in images, adding another layer of early warning capabilities.

While these detection systems are promising, their implementation faces challenges due to the fragmented nature of Portuguese forest ownership—84.2% privately owned and 13.8% communal (Meira Castro et al., 2020), which complicates large-scale adoption.

2.2.3.3. Fire risk maps

Another preventative approach involves the use of fire risk maps. National entities, such as IPMA leverage weather data to issue fire risk warnings. A commonly accepted guideline is the “30-30-30” rule: temperatures above 30°, wind speeds over 30 km/, and humidity below 30% indicate high fire risk, triggering alerts in vulnerable areas.

Some researchers (Dimuccio et al., 2011; Ermitão et al., 2023; Sivrikayaq et al., 2024) have built upon this by developing higher-resolution fire risk maps that incorporate remote sensing, spectral, vegetation, terrain, land classification maps and other relevant data. These advanced maps offer a finer spatial resolution compared to IPMA’s district-level data, helping authorities pinpoint specific high-risk zones more accurately. However, these maps often remain static, which limits their usefulness over time as vegetation and environmental conditions change. Ideally, dynamic maps with spatial resolutions of no less than 250 metres per pixel would provide more timely and actionable insights.

While this thesis also aims to contribute to wildfire prevention, it takes a distinct approach. Given the extensive research on fuel break detection and fire risk mapping, this study focuses on predicting areas that are particularly prone to burning. In the next section, we will examine the most relevant and related works in more detail, distinguishing the specific contributions of this thesis to the field of wildfire prevention.

2.3.RELATED WORKS

This thesis focuses on predicting wildfire-prone areas rather than solely on prevention. We identified four studies that have applied remote sensing and artificial intelligence to wildfire prediction. In this section, we review each of these studies in terms of their purpose, methodologies, data sources, and key contributions, and we highlight how our research distinguishes itself through unique methodologies.

(Kaur et al., 2023).in "A Data Integration Framework with Multi-Source Big Data for Enhanced Forest Fire Prediction" presents a data integration framework combining remote sensing, weather, and other data to create an 18-years database for wildfire prediction in Alberta, Canada. Using ML models like Random Forest (RF), eXtreme Gradient Boosting (XGBoost), and Multi-Layer Perceptron (MLP), the authors addressed data imbalance by adjusting spatial resolution and using spatial subsampling and downsampling techniques, ultimately selecting a 10 km x 10 km grid as the most suitable. While testing a higher resolution (1 km x 1 km) was unsuitable due to imbalance, the integration of ERA5-Land a Copernicus Climate dataset and Alberta DEM (Digital Elevation Model) data allowed for comprehensive weather and topographical modelling. Marking each cell as a binary target (ignition or not), XGBoost emerged as the top performer, achieving an AUC-ROC of 87% and a recall of 75%. The study contributes a proactive risk assessment framework, offering a robust approach for wildfire mitigation and ecosystem preservation.

(Ntinopoulos et al., 2023).in "Fusion of Remotely-Sensed Fire-Related Indices for Wildfire Prediction through the Contribution of Artificial Intelligence" developed a model to predict wildfire ignitions by combining fire-related indices, such as the Fire Weather Index (FWI), Fosberg Fire Weather Index (FFWI), NDVI, and NDMI, with artificial intelligence techniques, specifically Artificial Neural Networks and Radial Basis Function (RBF). Using data from CHIRPS precipitation records, MODIS satellite reflectance, and Climate Forecast System Reanalysis, the study leveraged multi-source data across various spatial resolutions to train models on 13 attributes per point, including climatic, vegetation, and fire indices. A new index, FWIveg, which combines FWI with the Vegetation Health Index, showed high predictive accuracy, especially for gradual fire risk assessment.

(Cheruku et al., 2022) in "Predicting Wildfire using Live Fuel Moisture Content with Machine Learning" investigated the predictive value of Live Fuel Moisture Content (LFMC) for near-term wildfire occurrence across an 8 km x 8 km grid in the western US. By combining LFMC data with historical wildfire records, the study applied ML models—including K-Nearest Neighbors, Support Vector Machine (SVM), RF, and Neural Networks—to evaluate fire risk. The authors used LFMC maps, updated every 15 days at a 250-metre resolution, and analysed the data at multiple time intervals for each wildfire: the date closest to the wildfire event, one month prior, and two months prior. Results showed that incorporating LFMC trends over time improved recall, though models like SVM and ResNet-18 underperformed compared to others.

(Malik et al., 2021) in “Wildfire Risk Prediction and Detection using Machine Learning in San Diego, California” developed a ML model to predict and detect wildfire risk in San Diego, California, by integrating satellite, weather, and historical fire data. Leveraging fire records from agencies like CAL Fire and Bureau of Land Management, weather data from the NCEO, and remote sensing data from Landsat 8 (via Google Earth Engine), the authors addressed data imbalance using Synthetic Minority Over-sampling Technique (SMOTE). They also applied atmospheric corrections to the Red and Near-Infrared bands of the satellite imagery and used a 14 km x 14 km grid for spatial analysis. By using an ensemble model that combined separate models for fire and weather data (SVM, XGBoost, RF) and fire and remote sensing data (MLP, CNN, LSTM) and aggregating predictions through majority voting, the study achieved an astonishing 100% predictive accuracy, therefore, 100% recall.

2.3.1. Research gap

Unlike previous studies that employ low spatial resolutions, such as 1 km or even coarser scales like 10 km, this thesis adopts a distinctive approach by utilizing a spatial resolution of 20 metres, offering 2 500 times greater detail than 1 km and 250 000 times greater than 10 km. This high resolution enables a more granular differentiation between burned and unburned areas. By leveraging remote sensing data from the closest available date to each wildfire event, this study aims to train a model capable of identifying the intrinsic factors that distinguish vegetation prone to burning from vegetation that remained unaffected.

We could not find any related work in Portugal that utilizes remote sensing and ML to predict areas specifically prone to burning. This thesis intends to predict areas prone to burn without relying on Land Classification Maps data from GIS-sources, which classifies terrain types (e.g., vegetation, buildings, water bodies). Instead, we focus solely on spectral data from satellite imagery, terrain information from a DEM, and weather data from meteorological stations. Ideally, our model will learn to differentiate between vegetation types, such as eucalyptus (high fire risk) and cork oak (low fire risk), solely from spectral signatures - without additional land cover classifications. Once the desired granularity is achieved, the model’s predictions could be used to generate up-to-date fire risk maps, identifying high-risk hotspots with each new satellite image and weather report. While such maps represent a practical outcome, achieving accurate and granular fire susceptibility predictions remains the primary objective. To address these gaps, this research poses the following key questions:

1. Can areas prone to burning be effectively predicted using a 20-metre spatial resolution, without relying on GIS-based land cover classifications?
2. To what extent can the selected models differentiate between vegetation-dense areas prone to burning and other less susceptible areas, such as urban regions or wetlands?
3. Which features are most strongly correlated with fire susceptibility, and how do they contribute to the models’ predictions?

3. METHODOLOGY

Building on the identified gaps and distinctive approach outlined in the previous section, this chapter details the methodology employed to achieve the objectives of this thesis. We provide an in-depth overview of the study area, outline the data sources and technologies used, and describe the methods applied to train and test our Machine Learning (ML) models for distinguishing vegetation prone to burning from areas more resistant to fire.

3.1.STUDY AREA

This thesis focuses on the southern region of Portugal (Figure 3-1), specifically the districts of Faro and the southern part of Beja (covering the municipalities of Odemira, Ourique, Aljustrel, Castro Verde, Mértola, and Almodôvar). Geographically, the study area corresponds to the NUTS III regions of Algarve and the southern part of both Alentejo Litoral and Baixo Alentejo (INE, 2024). These districts feature distinct landscapes: Faro is characterized by the Monchique mountains and a relatively humid atmosphere due to its proximity to the Mediterranean, while Beja is drier, with extensive plains (Cunha, 2021; Moreira, 2018).

Figure 3-1 Study area delimited by yellow-highlighted municipalities.

According to the Strategy and Studies Office of the Portuguese Republic (E. do D. de F. Gabinete de Estratégia e Estudos da República Portuguesa, 2023), Faro spans an area of 4 997 km², which represents 5.4% of Portugal’s total area, and has a population of 484 122, accounting for 4.6% of the national population. By contrast, the Beja district covers a larger area of 10 263 km² but has a lower population of 148 881, representing just 1.4% of Portugal’s total (E. do D. de B. Gabinete de Estratégia e Estudos da República Portuguesa, 2023). Because this study focuses only on the six southern municipalities of Beja, the area of interest in Beja is approximately 5 341 km². This brings our total study area to 10 338 km² across both districts.

Despite Beja’s larger size, its population density is lower, and the vegetation density is also lower than in Faro, which may explain why Beja has a smaller burned area compared to Faro. Between 2015 and 2023, the total area burned in Beja was approximately 172.04 km², while in Faro, it reached 506.62 km², according to ICNF data. Over these nine years span, 10.14% of Faro’s land was affected by fires, compared to only 1.68% in Beja. This discrepancy reflects Faro’s denser vegetation, making it more prone to fires.

One key reason for selecting Faro and the southern part of Beja as the study area is the significant wildfire risk highlighted by (Tedim et al., 2013). While the increase in wildfire incidents in the Algarve was modest (6.5% between 1980 and 2003), the burned area rose dramatically by 508%—more than double the national rate of around 200%. This trend suggests that fires in southern Portugal, particularly in the Algarve, are more challenging to control and extinguish, potentially due to limited resources, which further emphasizes the importance of developing a fire prediction and prevention system for this critical region.

Forests in southern Portugal, particularly in the Algarve, are predominantly made up of oaks, in contrast to the pine and eucalyptus forests that dominate the northern regions (Costa et al., 2007). These southern forests are economically important, supplying raw materials for industries such as cork, paper, and furniture. Notably, private forest properties in the south are larger on average, covering approximately 100 hectares, which is significantly more than in other parts of the country (Meira Castro et al., 2020).

The study area extends from the northernmost point at 37° 52’ 15” N, 8° 26’ 5” W to the southernmost point at 36° 57’ 42” N, 7° 53’ 5” W, and from the westernmost point at 37° 1’ 20” N, 8° 59’ 51” W to the easternmost point at 37° 9’ 58” N, 7° 23’ 36” W.

3.2.DATA COLLECTION

This section is essential for understanding the complex factors influencing wildfire risk in the study area. It provides a comprehensive description of the key data sources that serve as the foundation of our predictive analysis. This detailed overview of data sources and processing can be valuable for researchers looking to build on this work or explore alternative methodologies in similar studies.

3.2.1. Fire events data

ICNF (Portuguese: Instituto da Conservação e Natureza Florestal; English: Institute of Nature and Forest Conservation) provides GIS shapefiles with historical fire events data, including detailed records on fire locations, types, causes, sizes, and durations (ICNF, 2024). While the ICNF dataset covers fires nationwide from 1975 to 2023, this study focuses specifically on wildfires within the study area and after the launch of Sentinel-2 on June 23, 2015. Following the definition of wildfires established on 2.1.1 Definition, we exclude wildfires smaller than 100 hectares (ha). This refined dataset, comprising 35 significant wildfire events (Figure 3-2 and Table 0-1 in Appendix A) serves as a vital foundation for analysing wildfire behaviour.

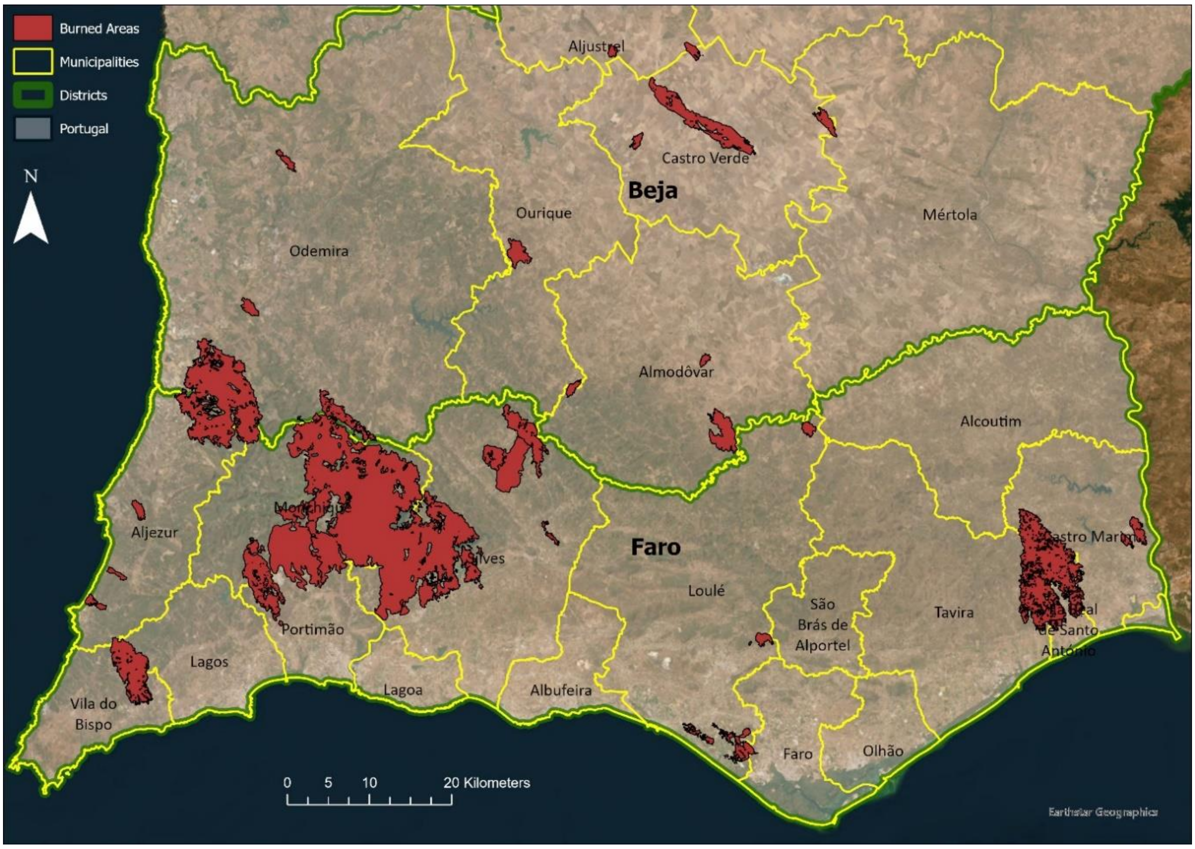


Figure 3-2 Burned areas (≥ 100 ha) in the study area (2015-2023), highlighted in red.

Wildfires on 2016-09-03 and 2018-08-03 took several days to be fully extinguished, and both experienced reignition during that period. While ICNF recorded them as a single wildfire each, due to their prolonged duration, we were able to capture satellite imagery during the events. This allowed us to split each wildfire event into two separate instances for analysis. By doing this, our model could still train on the vegetation that was about to burn and learn that already burned land cannot burn again in the following years. As a result, we ended up with 37 distinct wildfires records in our whole dataset.

3.2.2. Satellite imagery

For Fire Events data, we had only one source, but for remote sensing data, we considered both NASA's Landsat 8 and the European Space Agency's (ESA) Sentinel-2 satellites. Both are public and accessible. Landsat 8, launched on February 11th, 2013, offers a spatial resolution of 15 metres for true colour images and 30 metres for multispectral data, with a revisit time of 16 days (i.e., the time required to capture the same area again). In contrast, Sentinel-2, part of the Copernicus program and launched on July 23rd, 2015, consists of two synchronized satellites (Sentinel-2A and Sentinel-2B) that together achieve a 5-day revisit time. It also provides better spatial resolution—10 metres for true colour and 20 metres for multispectral data (ESA, 2023; NASA, 2023).

For this thesis, we selected Sentinel-2 for its faster revisit time and higher spatial resolution. Sentinel-2 captures a wide range of spectral bands, including visible light (red, green and blue) and near- and mid-infrared bands (Table 0-2). Figure 3-3 and Figure 3-4 (Source: Copernicus,

2023) provide a visual representation of the electromagnetic spectrum of Sentinel-2 bands with a spatial resolution of 10 and 20 meters, respectively.

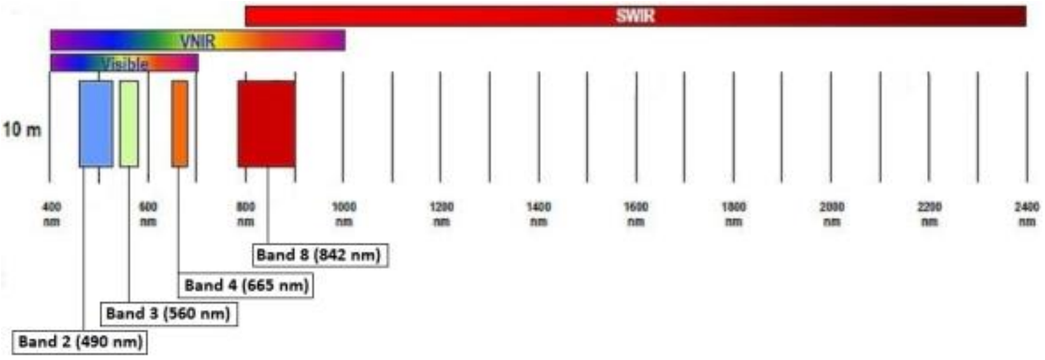


Figure 3-3 Sentinel-2 bands with 10-meter spatial resolution and their corresponding wavelengths in the electromagnetic spectrum.

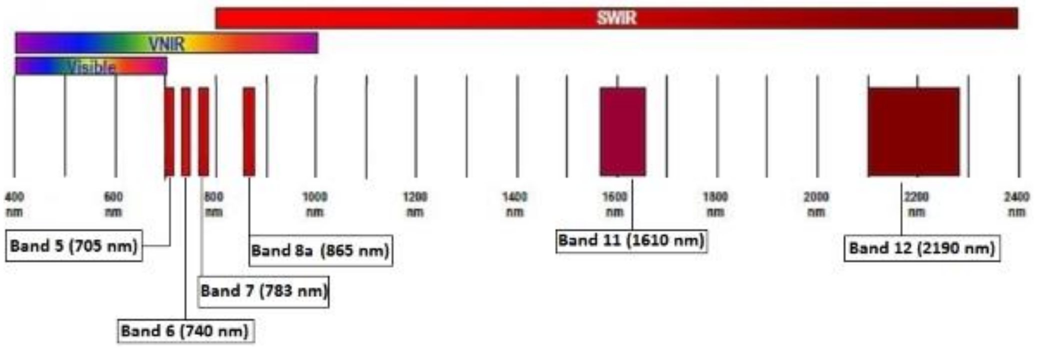


Figure 3-4 Sentinel-2 bands with 20-meter spatial resolution and their corresponding wavelengths in the electromagnetic spectrum.

Although Sentinel-2 offers additional bands at a 60-metre resolution, we focused solely on bands that can have a spatial resolution of 20 metres: B2, B3, B4, B5, B6, B7, B8a, B11, and B12. This spatial resolution was selected to avoid quadrupling the data size, which would have exceeded our computational capacity if using 10-metre spatial resolution. Consequently, each pixel represents a 20 m-by-20 m area, covering 400 square metres (equal to 0.04 hectares).

To cover all wildfire events in our dataset, we downloaded satellite imagery TIF files from the Copernicus Browser website (Copernicus Browser, 2024), by selecting “Sentinel-2 - MSI - L2A”. These files capture that captured the study area prior to each wildfire event. Table 0-3 in Appendix A provides a detailed list of all the satellite imagery files downloaded and used in ArcGIS Pro for this analysis.

As observed, a total of 32 satellite images were utilized in this thesis. Although the dataset includes 37 wildfires records, it may seem that some satellite images are missing, however, this is not the case. In several instances, wildfires occurred in close temporal proximity to one another, allowing a single satellite image to capture multiple wildfire events.

As shown in Table 0-3 in Appendix A, the “Zone” column contains two distinct values: West side and East side. This distinction was necessary because the satellite could not capture the entire study area in a single image. The dividing line is located just to the right of the city of Faro. Wildfire events were subsequently processed based on the side where they occurred. Figure 3-5 illustrates the West side (left) and East side (right), respectively.

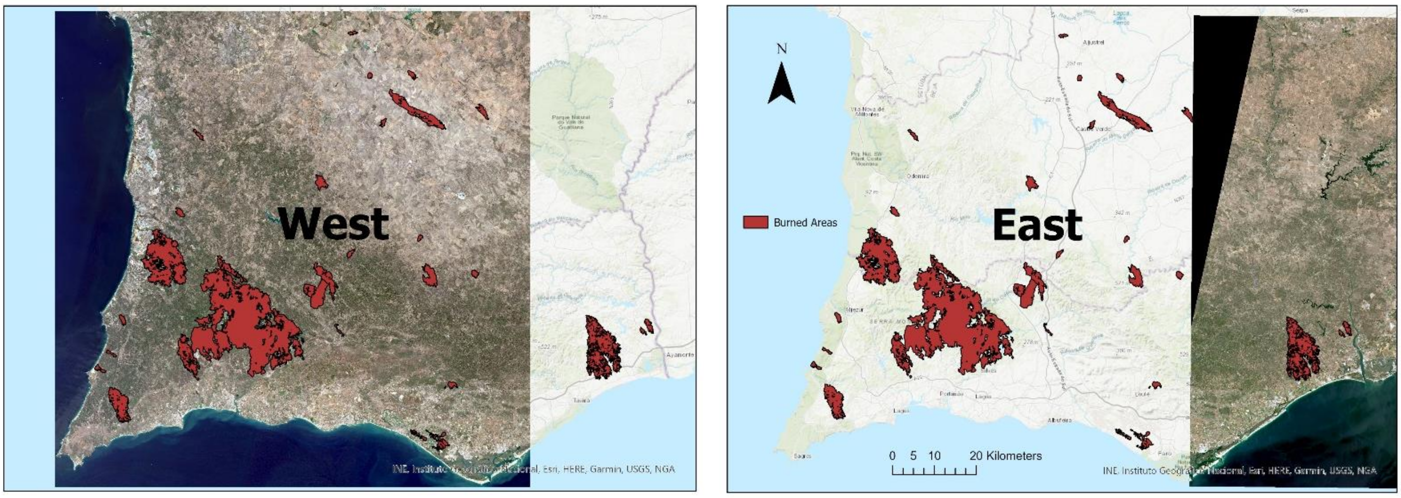


Figure 3-5 Satellite imagery of the study area. Left: West side; Right: East side.

In the following section, we will explore the application of specific vegetation indices, including the Normalized Difference Vegetation Index (NDVI), among others, to enhance the detection and analysis of wildfire-prone areas.

3.2.2.1. Multispectral indices

Multispectral indices are derived from mathematical operations—such as addition, subtraction, multiplication, and division—applied to specific spectral bands in satellite imagery. These operations generate various indices and ratios, typically ranging from -1 to 1, which are crucial for interpreting different land cover characteristics. Originally proposed by (Rouse & Haas, 1973) and later developed by (Lyon, 1998), multispectral indices have become invaluable tools for extracting meaningful insights from satellite imagery in wildfire research and related fields, such as vegetation fuel analysis, burned area mapping, drought monitoring, and burn severity assessment (Das et al., 2023; Oliveira et al., 2023).

There are potentially infinite multispectral indices that can be created, as researchers continue to develop new combinations based on specific study objectives. For this thesis, however, we focused on a selected set of indices that are widely recognized in the literature for their ability to analyse various environmental conditions related to wildfires. As mentioned by several authors (Lazaj, 2016; Novo et al., 2024), these indices cover a broad range of areas, including healthy vegetation, dry vegetation, water bodies, soil, and vegetation change. This approach ensures that we capture diverse factors contributing to wildfire risk. The indices and equations used in this thesis are widely recognized in wildfire research and are also available through resources such as the Copernicus browser (Copernicus Browser, 2024), which provides tools and data for Earth observation analysis.

Our primary goal in using these multispectral indices was not to explore or enhance any particular index but rather to improve the model's ability to predict vegetation areas prone to burning. By leveraging these indices, we aimed to capture more comprehensive environmental data that could help the model better detect wildfire-prone zones. In section 3.5.2 Data preprocessing, we will explain how the final set of indices was narrowed down by selecting only those that showed a significant correlation with the target variable.

To improve clarity, we have standardized the naming of the spectral bands used in the formulas. Instead of referring to the specific band numbers from Sentinel-2 (e.g., B02, B03), we will use the more commonly known names for these spectral bands: B02 as Blue, B03 as Green, B04 as Red, B05 as Red Edge, B08 as NIR (Near-Infrared), B11 as SWIR1 (Short-Wave Infrared 1), and B12 as SWIR2 (Short-Wave Infrared 2). This decision was made to align with the broader literature (Pacheco et al., 2023), as different satellite systems (such as Landsat) assign different numbers to similar bands. Furthermore, using standard names rather than spectral magnitudes (e.g., NIR is often referred to as 865 nm and Red as 665 nm) makes the formulas more universally applicable and easier to understand in future works.

In this thesis, thirteen indices were initially chosen: NDVI, NDRE, NDWI, EVI, LCI, CVI, GCI, ARI, MSAVI2, BSI, NDMI, NBR, and NDSI. Each of these indices was selected based on its proven utility in related studies (Avetisyan et al., 2023).

Normalized Difference Vegetation Index (NDVI):

Used to measure vegetation health by comparing near-infrared and red-light absorption.

$$NDVI = \frac{NIR-Red}{NIR+Red} \tag{3-1}$$

Normalized Difference Red Edge (NDRE):

Focuses on detecting vegetation stress using the near-infrared and red-edge bands.

$$NDRE = \frac{NIR-Red\ Edge}{NIR+Red\ Edge} \tag{3-2}$$

Normalized Difference Water Index (NDWI):

Designed to monitor water content in vegetation and surface water bodies.

$$NDWI = \frac{Green- NIR}{Green+\mathrm{NIR}} \tag{3-3}$$

Enhanced Vegetation Index (EVI):

Improves sensitivity to dense vegetation and minimizes atmospheric influences.

$$EVI = 2.5 * \frac{NIR-Red}{(NIR+6*Red)-(7.5*Blue+1)} \tag{3-4}$$

Leaf Chlorophyll Vegetation Index (LCI):

Estimates leaf chlorophyll content by comparing the near-infrared and red-edge bands.

$$LCI = \frac{NIR - Red\ Edge}{NIR + Red} \tag{3-5}$$

Chlorophyll Vegetation Index (CVI):

Assesses chlorophyll concentration using the near-infrared and green bands.

$$CVI = \frac{NIR * Red}{Green^2} \tag{3-6}$$

Green Chlorophyll Index (GCI):

Measures the amount of chlorophyll in vegetation using near-infrared and green light.

$$GCI = \frac{NIR}{Green} - 1 \tag{3-7}$$

Anthocyanin Reflectance Index (ARI):

Detects anthocyanin pigments in vegetation to assess stress or senescence.

$$ARI = \frac{1}{Green} - \frac{1}{Red\ Edge} \tag{3-8}$$

Modified Soil-Adjusted Vegetation Index 2 (MSAVI2):

Optimized for areas with bare soil to improve vegetation monitoring.

$$MSAVI2 = \frac{2 * NIR + 1 - \sqrt{(2 * NIR + 1)^2 - 8 * (NIR - Red)}}{2} \tag{3-9}$$

Bare Soil Index (BSI):

Identifies bare soil by contrasting the visible red, blue, and infrared bands.

$$BSI = \frac{(Red + SWIR1) - (NIR + Blue)}{(Red + SWIR1) + (NIR + Blue)} \tag{3-10}$$

Normalized Difference Moisture Index (NDMI):

Monitors vegetation moisture by comparing near-infrared and shortwave infrared bands.

$$NDMI = \frac{NIR - SWIR1}{NIR + SWIR1} \tag{3-11}$$

Normalized Burn Ratio (NBR):

Used to assess burn severity by comparing near-infrared and shortwave infrared bands.

$$NBR = \frac{NIR - SWIR2}{NIR + SWIR2} \tag{3-12}$$

Normalized Difference Soil Index (NDSI):

Differentiates between soil and other land cover types using green and shortwave infrared.

$$NDSI = \frac{Green - \mathrm{SWIR1}}{Green + SWIR1} \tag{3-13}$$

These multispectral indices played a critical role in the preprocessing of our ML models, enhancing the dataset with vital information on vegetation, moisture, and soil conditions. By incorporating both Sentinel-2 spectral bands and additional environmental features, we ensured that the models had a comprehensive view of the landscape, making them more robust and accurate in predicting wildfire-prone areas. This holistic approach forms the foundation for integrating other data sources, such as weather and terrain, which further enrich the predictive models for wildfire risk assessment.

3.2.3. Terrain data

To better understand the dynamics of wildfire spread within our study area, we recognized the significance of terrain (elevation, slope and aspect) as crucial factor, as discussed in section 2.1.2 Spreading behaviour. Given Portugal's diverse landscape—including mountains, rivers, beaches, cities, and plains—it was essential to analyse the impact of wildfires in relation to these terrain features. To facilitate this analysis, we obtained a Digital Terrain Model (DTM) for mainland Portugal in TIF format from ESRI Portugal (ArcGIS, 2023), ensuring compatibility with the coordinate systems used in this thesis, ETRS89 (dgTerritório, 2024). Using ArcGIS Pro, we calculated both slope and aspect from the elevation data to enhance our understanding of how terrain influences wildfire behaviour.

Elevation data represents the altitude of the terrain, measure from the sea-level, where higher values indicate greater altitude. Slope measures the steepness of the terrain, which is a critical factor in fire spread; steeper slopes tend to accelerate fire movement. Additionally, aspect refers to the direction a slope faces, which affects moisture levels and, consequently, fire behaviour. As mentioned in section 2.1.2 Spreading behaviour, slopes facing north retain more moisture, while south-facing slopes are typically drier and more prone to fire.

While there were no significant challenges with the terrain data, some missing data points in ocean areas required attention, and these were handled appropriately (explained in more detail in section 3.5.2 Data preprocessing). Overall, the inclusion of terrain data was a key factor in enriching our model and improving its ability to predict wildfire-prone areas.

3.2.4. Land classification and delimitation data

Land classification and delimitation data are essential for various applications in land management, urban planning, and land use research, among others. Datasets such as COS (Portuguese: Carta de Uso e Ocupação do Solo, English: Land Use and Occupation Charter) for land cover classification (*Dados.Gov*, 2023a), and CAOP (Portuguese: Carta Administrativa Oficial de Portugal, English: Official Administrative Charter of Portugal) for administrative boundaries (*Dados.Gov*, 2023b), are available as open data from government platforms.

The CAOP dataset, providing official administrative boundaries of Portugal, is central to this thesis as it allows precise delimitation of the study area using tools within ArcGIS Pro.

Specifically, we used district and municipality polygon datasets to exclude wildfire events outside the designated study area.

Although COS holds potential for classification models, we used it only for model evaluation, not as a predictive feature. As highlighted in section 2.3.1 Research gap, our goal is to identify fire-prone areas using only spectral and terrain data, avoiding reliance on land cover labels that could introduce bias. Moreover, land cover maps are not frequently updated; the latest COS dataset is from 2018, while our satellite imagery extends to 2023, risking discrepancies between actual spectral data and outdated classifications. Despite this, the COS dataset remains valuable for evaluation, offering context on land cover patterns that may affect model performance.

3.2.5. Weather data

Our initial goal was to incorporate weather data into the model to improve its predictive capabilities, by including key environmental variables such as wind speed, precipitation, temperature, and humidity, as explained in section 2.1.2 Spreading behaviour. We expected to find detailed, granular data that would vary across different zones within the study area, such as highly dense forests versus dried and bare land, which would enhance the model's accuracy.

Aligning weather data with the satellite imagery's temporal and spatial resolution was crucial, however, unfortunately, the available weather data from IPMA did not meet these requirements. IPMA's API consisted of monthly averages for each district or municipality, lacking the resolution necessary to capture localized weather variations. Using such broad averages would introduce noise into the model and provide no meaningful correlation with the target variable.

While we explored other sources, open and public data had similar limitations, and private services were not pursued due to time and resource constraints. As a result, we proceeded without incorporating weather data into the model.

3.3.MACHINE LEARNING MODELS

Machine learning (ML) models are computational algorithms that can be trained to recognize specific patterns within a dataset. In this thesis, we used supervised ML models, which operate by learning from a labelled dataset containing features and a target variable, 'Burned', which has a binary output: 0 for unburned areas and 1 for burned areas.

There are many types of supervised ML models available but testing all of them to determine the best option, was impractical for this study. Instead, by exploring related studies (previously mentioned in section 2.3 Related works) and deepening our understanding of supervised ML models (Khan et al., 2024; Zekić-Sušac et al., 2014), we narrowed our selection to those capable of efficiently handling extensive amounts of data while uncovering nuanced patterns. This was essential given that our specific dataset is large, high-dimensional and complex.

Three models were selected for this study for their simplicity and effectiveness: **Random Forest (RF)**, a bagging ensemble method; **eXtreme Gradient Boosting (XGBoost)**, an optimized version of Gradient Boosting Machine (GBM) known for its superior performance and speed with large datasets; and **Multi-Layer Perceptron (MLP)**, a neural network model.

We deliberately excluded more complex deep learning models, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). While these advanced neural networks excel in pattern recognition and time-series forecasting, their substantial complexity and the need for advanced fine-tuning hyperparameters techniques (e.g., filters, kernel sizes, padding, cell types), posed challenges that were beyond the scope of this study.

Each model is discussed in detail in the following sections (3.3.1 - 3.3.3).

3.3.1. Random Forest

To understand RF, it is helpful to first examine Decision Trees (DT), initially developed by J.R. Quinlan (Quinlan, 1986). A decision tree is structured like a tree, where each branch represents a decision or split based on a feature, and each leaf represents a final classification. While a decision tree can be a simple yet powerful classifier, it often overfits because the model captures specific details in the training data rather than generalizing well to unseen data.

RF, developed by Leo Breiman (Breiman, 2001), addresses this issue by creating an ensemble, or ‘forest’, of decision trees. It does so by generating trees based on random subsets of data (a method known as bootstrapping) and random subsets of features at each split, thereby creating a diverse set of trees that generalize better to new data. Once the trees are constructed, RF aggregates their predictions for a given data point, typically using a majority vote (in classification tasks) to determine the final outcome. This combined process—bootstrapping data and aggregating results—is known as ‘bagging’, making RF a bagging technique.

For this thesis we used RF Classifier from Scikit-Learn packages (*Scikit Learn*, 2023b) prioritizing the following input hyperparameters:

- **n_estimators**: Specifies the number of trees in the forest. Higher values typically enhance accuracy but also increase computational time.
- **criterion**: Determines the function used to evaluate the quality of a split.
- **max_depth**: Limits the maximum depth that each tree can grow. Greater depth can improve accuracy but may also increase the risk of overfitting.
- **min_samples_split**: The minimum number of samples required to split a leaf node, preventing splits on smaller nodes and helping to control overfitting.
- **min_samples_leaf**: The minimum number of samples a leaf node must have after a split, preventing overly specific leaves and reducing the risk of overfitting.
- **max_features**: The maximum number of features considered for each split. Limiting this number encourages tree diversity and can reduce overfitting.
- **class_weight**: Adjusts classes weight in the target variable, treating them equally if set to ‘None’ or assigning weights to diminished classes when set to ‘balanced’.

Overall, RF is a robust supervised ML model that handles large and complex data well, without overfitting, making it one of the selected models.

3.3.2. XGBoost

XGBoost is an optimized version of the traditional GBM classifier model developed by Jerome H. Friedman (Friedman, 2001). As explained earlier, both RF and GBM are supervised ensemble methods, but RF uses bagging, while GBM employs boosting.

In RF, trees are built in parallel—several trees are constructed simultaneously and then aggregated to form the final model. GBM, on the other hand, builds trees sequentially (a process known as boosting). It begins with a weak learner (a simple, naive model) and iteratively adds new trees, each one reducing the residual errors of the previous model based on a specified loss function. Without being too technical, this process is called gradient boosting because it uses gradient descent to minimize the loss with each new learner.

While GBM has certain advantages over RF in terms of accuracy and flexibility, its sequential nature can result in slower training times. XGBoost, developed by Tianqi Chen and Carlos Guestrin (Chen & Guestrin, 2016), addresses this by introducing multiple optimizations, such as regularization, column and row subsampling, and GPU support. These improvements make XGBoost faster and more efficient while retaining the core principles of GBM.

We implemented XGBoost using the official version from its GitHub repository (dmlc, 2021), ensuring access to the latest updates and optimizations provided by the original developers. Below is a comprehensive list of the hyperparameters fine-tuned in this thesis for optimizing the XGBoost model:

- **n_estimators**: The number of boosting rounds. Similar to RF, increasing this value generally improves accuracy but it also increases training time.
- **learning_rate (eta)**: Controls the rate at which new trees adjust the model, with lower values slowing learning to improve accuracy but requiring more training iterations.
- **max_depth**: Similar to the RF, it limits the depth to control overfitting.
- **min_split_loss (gamma)**: The minimum loss reduction required to further split a leaf node. Higher values reduce overfitting by preventing unnecessary splits.
- **min_child_weight**: Similar to the min_samples_leaf in RF, it sets the minimum sum of instance weights required in a leaf node to allow further splitting.
- **subsample**: The fraction of the training data randomly sampled to grow each tree. For example, setting this to 0.5 means that only half of the training data is used in each boosting round, adding data variability to help prevent overfitting.
- **sampling_method**: Specifies the method for sampling training instances.
- **colsample_bytree**: Defines the fraction of features randomly sampled to grow each tree, similar to subsample but applied to features rather than instances.
- **tree_method**: Specifies the tree construction algorithm to be used.

XGBoost, the most advanced model used, was chosen for its performance, speed, and ability to handle complex, high-dimensional datasets.

3.3.3. Multi-Layer Perceptron

MLP is a Deep Learning (DL) model inspired by the structure of the human brain, placing it within the category of Artificial Neural Networks (ANN). An ANN is a network of interconnected nodes; belonging to the broader family of Neural Networks, which in the scope of this thesis is helpful for also being a supervised ML technique. While neural networks and perceptrons have a long history dating back to 1943 (Mcculloch & Pitts, 1943) with significant advances in the 1980s (Fukushima, 1980; Minsky & Papert, 1969; Rosenblatt, 1958) this thesis will focus specifically on how MLP functions as a model.

To understand MLP, let’s begin with its basic unit: the perceptron. A perceptron is a simplified model that mimics a biological neuron, where each neuron receives signals, processes them, and passes them on to other neurons Similarly, a perceptron has inputs (signals from other perceptrons), an activation function (which processes information), and outputs (signals to other perceptrons). A single perceptron alone has limited functionality, but when arranged in layers, perceptrons form a powerful network capable of recognizing complex patterns, mimicking the functioning of the human brain. In an MLP, the first layer, called the input layer, receives the raw data (e.g., satellite imagery spectral bands). Following the input layer, there are hidden layers that add complexity to the model; these layers allow the network to capture intricate relationships within the data. We are able to determine the number of hidden layers as well as the configuration of the perceptrons; where if we have a single hidden layer, the MLP will not fall under into the category of Deep Learning (but it is still considered an ANN). The final layer, known as the output layer, produces the prediction. In our case, the MLP is trained to classify areas as either prone to burn (1) or not (0), so the output layer will contain a single neuron representing this binary prediction.

MLP learns by using a technique called backpropagation (Popescu et al., 2009). After an initial prediction, the model calculates its error and propagates it backward, from the output layer, to the hidden, and then to the input layer, to adjust each perceptron’s weights. These weights are responsible of how each perceptron processes the data. This iterative process, called an epoch, is repeated over several times until the model converges, meaning that further adjustments yield minimal improvements. The challenge lies in balancing the number of epochs: too many may lead to overfitting, while too few can result in underfitting.

For this thesis, we used the MLP Classifier from Scikit-Learn (*Scikit Learn*, 2023a) and focused on fine-tuning the following key hyperparameters:

- **hidden_layer_sizes**: Configuration of the hidden layers, specifying both the number of layers and the size of each.
- **activation**: Activation function that will be applied to the hidden layers.
- **solver**: Optimization algorithm for updating weights.
- **learning_rate**: The learning rate method for weight updates.
- **learning_rate_init**: The initial learning rate applied at the start of training.
- **max_iter**: The maximum number of iterations (or epochs). If the model converges before reaching this number, training stops early.

- **shuffle**: Boolean to specify whether to shuffle samples in each iteration/epoch.

SMOTE was used to balance classes, compensating for MLP's lack of a 'class_weight' hyperparameter to improve sensitivity to minority classes.

MLP offers some possible configurations that will be explored in the Results section. In the following subsection, we will review the software tools used throughout this thesis.

3.4. TOOLS USED

This section presents a detailed overview of the technologies employed in this research, highlighting their roles in addressing the complexities of wildfire prediction. The technologies range from advanced machine learning models to sophisticated software applications, illustrating the multi-faceted approach of this study.

To support the various stages of this research, the following tools were used during this thesis:

- **Scopus** and **ResearchGate**: Used to retrieve academic and research papers with the aid of queries and filtering functionalities
- **ChatGPT** - Assisted with code-related questions, improved efficiency, and facilitated brainstorming and English writing suggestions
- **Zotero** - Served as a centralized platform for managing references and automatically generating the bibliography
- **PowerPoint** - Used to create some images and diagrams (remaining figures were produced using ArcGIS Pro®, Python's Matplotlib and Seaborn libraries)

3.4.1. ArcGIS Pro

ArcGIS Pro was a central tool throughout this thesis. It facilitated the import and integration of satellite, burned area, and elevation data, transforming these datasets into tabular form for ML models. Additionally, ArcGIS Pro was used to generate fire risk raster maps based on the predicted results of the models.

Tools: In ArcGIS Pro, we utilized tools such as 'Merge', 'Dissolve', 'Intersect', 'Calculate Metrics', and query filtering to process burned areas from ICNF. To create the tabular dataset for Python from satellite imagery, we applied tools like 'Composite Bands', 'Extract by Mask', 'Buffer', 'Clip', 'Raster to Point', 'Create Random Points', 'Extract Multi Values to Points', 'Add XY Coordinates', 'Calculate Field', 'Merge', and 'Intersect'. For exporting the datasets, we used 'Export Table', and after processing in Python, the data was re-imported to ArcGIS Pro using 'Create Points From Table'. Finally, dynamic fire risk maps were generated with the 'Point to Raster' tool, leveraging predictions from the best ML model and applying various symbology techniques.

3.4.2. Jupyter Notebook (Google Colab)

Leveraging the advantages of Google Colab, a Jupyter Notebook was used extensively for data cleaning, analysis, feature selection, processing, and model evaluation. It provided the computational environment where the RF, XGBoost and MLP models were trained and tested.

Libraries: In this research, a variety of programming libraries were utilized to handle different aspects of data processing, visualization, analysis, and model building. Here is an overview of the key libraries and their specific applications in our study:

Standard Libraries:

- **time**: Used to track the duration of code execution, particularly useful during the training of ML models.
- **re**: Used for handling regular expressions, facilitating text processing tasks.
- **copy (from copy)**: Used for creating shallow or deep copies of objects, ensuring original data remains unaffected during processing
- **defaultdict (from collections)**: Provides a dictionary that defaults to a specified type, allowing for efficient handling of missing values in data processing tasks.

Data Processing and Manipulation:

- **pandas (pd)**: Essential for data manipulation and analysis, extensively used to organize data into readable and editable formats like DataFrames, handle missing values, and process large datasets.
- **numpy (np)**: Supports large, multi-dimensional arrays and matrices, and provides a large collection of mathematical functions for array operations, making it ideal for efficient numerical computations.
- **stats (from scipy)**: Contains statistical functions and tests, which are useful for detecting outliers, performing hypothesis testing, and analysing data distributions.

Imputation and Scaling:

- **KNNImputer (from sklearn.impute)**: Implements k-Nearest Neighbours imputation to fill in missing values based on the values of neighbouring data points.
- **StandardScaler (from sklearn.preprocessing)**: Used for scaling features to a standard normal distribution, ensuring that each feature contributes equally to the model.

Visualization:

- **matplotlib.pyplot (plt)**: A versatile plotting library used to create static, interactive, and animated visualizations, including line charts, bar charts, histograms, and scatter plots.
- **seaborn (sns)**: Built on top of matplotlib, this library provides a high-level interface for creating informative and attractive statistical graphics, making it easier to interpret complex datasets.

Model Selection and Evaluation:

- **ParameterGrid (from sklearn.model_selection)**: Allows for systematic testing of hyperparameter combinations during grid search, enhancing model tuning efficiency.

- **sklearn.metrics**: Provides a suite of functions, including accuracy_score, precision_score, recall_score, f1_score, roc_auc_score, confusion_matrix, roc_curve, and auc, for evaluating the performance of ML models.
- **plot_tree (from sklearn.tree)**: Visualizes decision trees, providing insights into the structure and logic of tree-based models.
- **export_text (from sklearn.tree)**: Exports decision trees as text for easier inspection and interpretation of tree splits and decisions.
- **compute_sample_weight (from sklearn.utils)**: Computes weights for handling imbalanced datasets, helping to improve model performance on rare classes.

Handling Class Imbalance:

- **SMOTE (from imblearn.over_sampling)**: A synthetic minority over-sampling technique (SMOTE) that generates synthetic samples for underrepresented classes, helping to balance the dataset and improve model sensitivity to minority classes.

Machine Learning Models:

- **RandomForestClassifier (from sklearn.ensemble)**: A meta-estimator that fits multiple decision trees on sub-samples of the dataset and averages predictions to improve accuracy and control overfitting.
- **MLPClassifier (from sklearn.neural_network)**: A multi-layer perceptron classifier, an artificial neural network optimized using backpropagation.
- **xgboost (xgb)**: An optimized gradient boosting library designed to handle large datasets efficiently and achieve high predictive accuracy with non-linear relationships, implemented in this study through its Python API.

Model Serialization:

- **joblib**: Facilitates saving and loading trained models and other Python objects. Joblib is especially useful for large objects such as ML models, ensuring they can be quickly saved and reused without retraining.

All the work conducted in Jupyter Notebooks is documented in the repository (Lavado, 2024). The publicly accessible repository includes the CSV files used for both training and testing, alongside over 3 500 lines of code spanning tasks such as data preprocessing, feature engineering, and model evaluation.

3.5.METHODOLOGY APPLIED

Before diving into the methodological specifics, here’s a brief summary: We began by importing fire events data from 2015 to 2023 from ICNF into ArcGIS Pro. With the help of the CAOP dataset, we focused only on wildfires within the study area and used query filtering techniques to only keep wildfires that burned at least 100 hectares (ha). We also imported a Digital Elevation Model (DEM) from ESRI Portugal and process it to derive not only elevation raster data but also slope and aspect data rasters. We then integrated Sentinel-2 satellite imagery captured days before the fires. In some cases, imagery captured just before the wildfire event was unusable due to cloud cover; thus, we utilized images from 7-8 days prior. Using spatial analysis tools, we distinguished between burned and unburned areas. For each point, representing a 20 m x 20 m raster pixel, spectral data from nine Sentinel-2 bands, elevation, slope and aspect data were extracted. Points outside burned areas were labelled as ‘0’ (Not Burned), while those inside burned areas were labelled as ‘1’ (Burned). To improve the model’s ability to identify wildfire-prone areas, we maintained a 1:2 ratio of burned to unburned points by randomly removing unburned points. This balance prioritized reducing false negatives, a critical aspect of wildfire management.

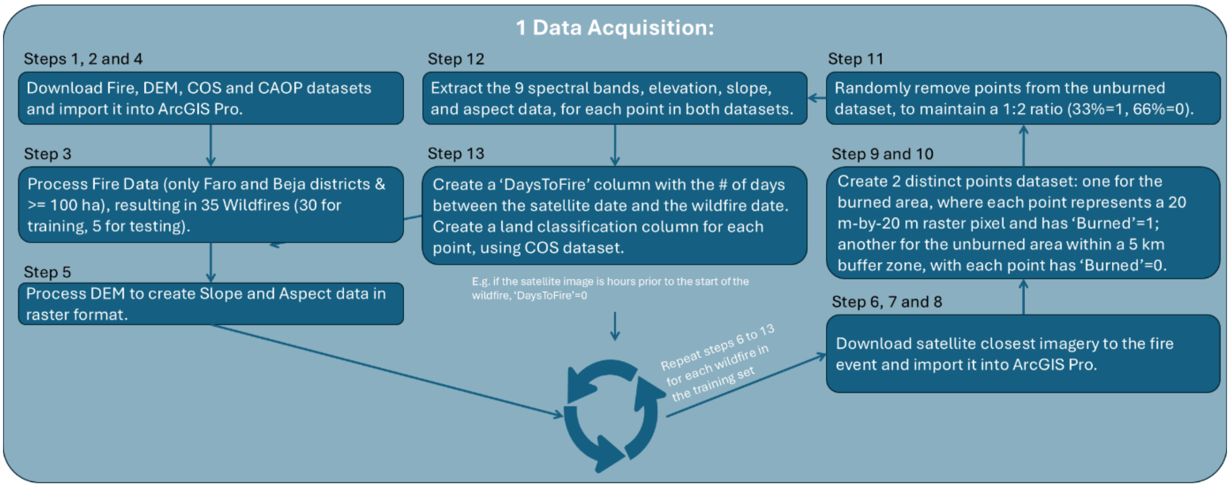
After importing the data in Jupyter Notebook, we began by adjusting data types (correcting the ‘Elevation’, ‘Slope’ and ‘Aspect’ columns), filling missing values, and analysing outliers using Z-score. We created two subsets based on the median days between the satellite image capture and the wildfire event: subset 1 focused on data closer to the wildfire event, while subset 2 focused on data further away from it. Vegetation indices such as NDVI, EVI, and others were added, and features that were either weakly correlated with the target variable or too highly correlated with other features were removed. We employed a time series strategy to prevent data leakage by training on data from earlier years and testing on subsequent years, ensuring that each model learns from past data without being exposed to future information.

Three models—Random Forest (RF), eXtreme Gradient Boosting (XGBoost), and Multi-Layer Perceptron (MLP)—were applied to 6 different DataFrames combinations (different outlier treatments with different ‘Aspect’ configurations) for each subset (1 and 2). The best-performing DataFrame combination for each subset for each model was fine-tuned, and the final selection was based on recall performance to ensure the highest sensitivity to wildfire-prone areas. Ultimately, we trained three models (RF, XGB, and MLP) for each subset (1 and 2), resulting in six unique models.

Using unseen test data from 2023, consisting of five wildfires, and applying the same preprocessing steps in ArcGIS Pro and Jupyter Notebook as for the training data, predictions were drawn for each model for each wildfire. The best model for each subset was chosen based on the highest recall. Then, using ArcGIS Pro, through spatial analysis and tools, we created dynamic fire risk maps using these best performance models.

In the beginning of each process (Data Acquisition, Data Preprocessing; Data Evaluation; and Model Evaluation), a diagram provides a clearer visual representation of each step taken in that process (Figure 3-6, Figure 3-8, Figure 3-9 and Figure 3-10).

3.5.1. Data acquisition



Download Fire, DEM, COS and CAOP datasets and import it into ArcGIS Pro.

Extract the 9 spectral bands, elevation, slope, and aspect data, for each point in both datasets.

Randomly remove points from the unburned dataset, to maintain a 1:2 ratio (33%=1, 66%=0).

Process Fire Data (only Faro and Beja districts & >= 100 ha), resulting in 35 Wildfires (30 for training, 5 for testing).

Create a 'DaysToFire' column with the # of days between the satellite date and the wildfire date. Create a land classification column for each point, using COS dataset.

Create 2 distinct points dataset: one for the burned area, where each point represents a 20 m-by-20 m raster pixel and has 'Burned'=1; another for the unburned area within a 5 km buffer zone, with each point has 'Burned'=0.

Process DEM to create Slope and Aspect data in raster format.

E.g. if the satellite image is hours prior to the start of the wildfire, 'DaysToFire'=0

Repeat steps 6 to 13 for each wildfire in the training set

Download satellite closest imagery to the fire event and import it into ArcGIS Pro.

Figure 3-6 Data Acquisition Diagram

This section details each step from downloading data and importing it into ArcGIS Pro to processing it for Jupyter Notebook. Below is a breakdown of the steps involved:

Step 1 - Download and import fire event data:

We downloaded the GIS-polygon fire data for all fire events from 2015 to 2023 from the ICNF site and uploaded it as a shapefile into ArcGIS Pro. Using the 'Merge' tool, we combined data from different years into a unified dataset.

Step 2 - Download and import land classification and delimitation data:

As referenced in section 3.2.4, we downloaded the 2018 Portugal Mainland Land Classification GIS-polygon dataset (COS 2018 dataset, latest available) and the District and Municipality GIS-polygons from the CAOP dataset. These files were then uploaded into ArcGIS Pro.

Step 3 - Process fire event data:

We filtered the wildfire dataset, using query techniques in ArcGIS Pro, to include only fire events equal to or greater than 100 hectares (ha). To narrow the analysis to the Faro district (Algarve) and the southern part of the Beja district, we used the previously imported CAOP dataset with the 'Intersect' tool to isolate the areas of interest. This resulted in Table 0-1 (Appendix A) containing only wildfires of at least 100 hectares in the specified region, sorted in ascending order by date. Each wildfire is represented as a GIS polygon shape.

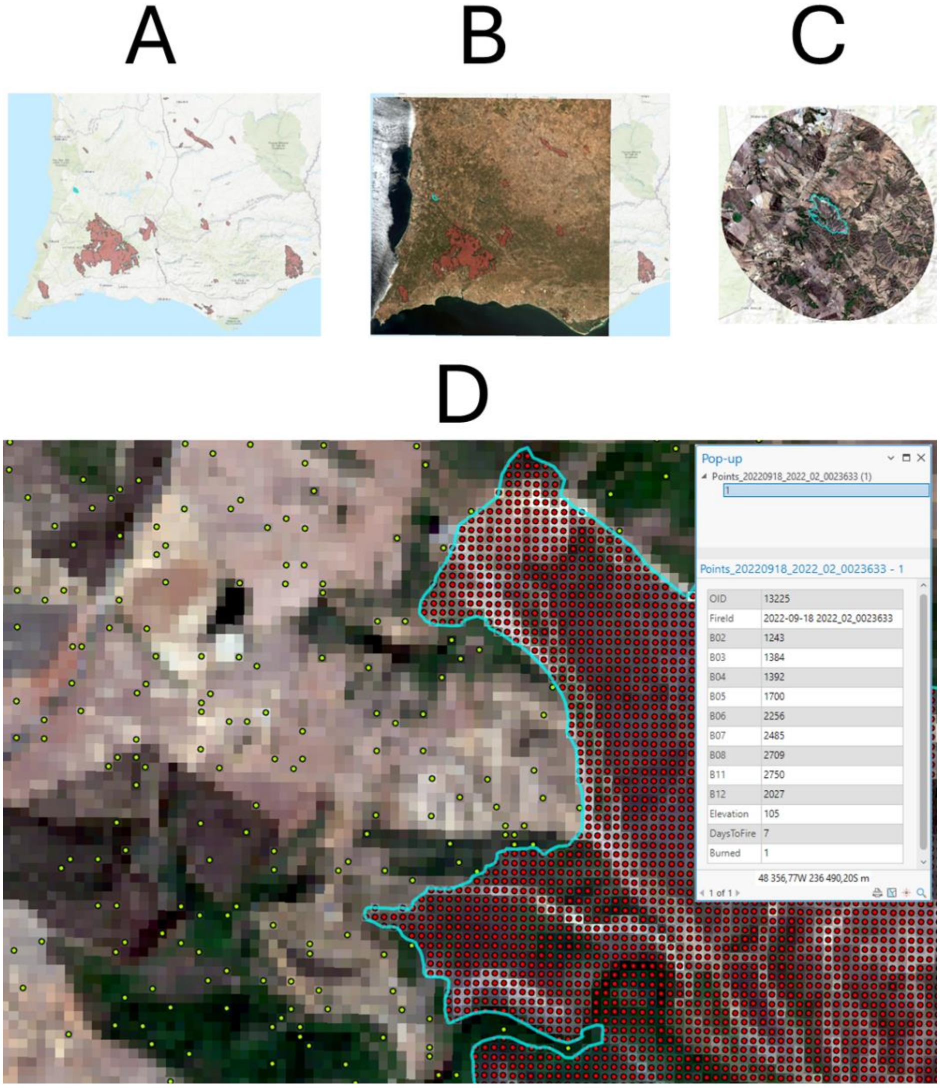
Step 4 - Download and import DEM data from ESRI Portugal:

We downloaded the Digital Elevation Model (DEM) data in raster format from the ESRI Portugal site and uploaded it into ArcGIS Pro.

Step 5 - Processing DEM data to obtain slope and aspect:

Using the elevation raster data, with the tools 'Slope' and 'Aspect', we were able to create those in the raster format as well.

Step 6 to Step 12 will be applied to every individual wildfire used for training (2015 to 2022) that is in our wildfire filtered train table. Here is a breakdown in images:



Points_20220918_2022_02_0023633 (1)
1

Points_20220918_2022_02_0023633 - 1

OID	13225
FireId	2022-09-18 2022_02_0023633
B02	1243
B03	1384
B04	1392
B05	1700
B06	2256
B07	2485
B08	2709
B11	2750
B12	2027
Elevation	105
DaysToFire	7
Burned	1

48 356,77W 236 490,20S m

1 of 1

Figure 3-7 A: Select a specific wildfire from the training set (highlighted in blue). B: Download and import the corresponding satellite imagery for the selected wildfire. C: Create a 5 km buffer around the burned area polygon. D: Generate points for each 20 m-by-20 m pixel and randomly remove points from the unburned area (green points). Extract feature values for each point (shown in pop-up).

Step 6 - Download and upload satellite imagery closest to wildfire date from Copernicus:

As shown in Figure 3-7 B, we obtained the closest satellite imagery relative to the date of a given wildfire event (present in our filtered wildfire train table) from the Copernicus Program. Depending on availability, the images ranged from a few hours before the fire to a maximum of 18 days prior. For early Sentinel-2 data (2015-2016), the satellite cycle was different from current times. Now, a full cycle is achieved every 4 days in the study area due to latitudinal variations (the revisit time of 5 days applies to the equator but is faster at higher latitudes). In cases where the closest image was unusable due to cloud cover, we selected the second closest image, typically 7-8 days prior to the fire. After downloading the TIF file, we uploaded it into ArcGIS Pro for further processing.

Step 7 - Convert TIF data to RGB format:

Although the model only requires spectral values, we used the ‘Composite Bands’ tool to combine the spectral bands and convert the imagery to RGB format for easier human visualization and interpretation. This step helped us better understand the data visually, even though the model itself does not rely on colour.

Step 8 - Generate a TIF area for the burned area:

To focus on the burned area within the larger satellite image, we used the ‘Extract by Mask’ tool to isolate the future burned area, creating a TIF image of just the relevant region.

Step 9 - Generate a buffer zone for the burned area:

For this thesis, we created a 5 km buffer zone around each burned area using the ‘Buffer’ tool. This produced a polygon with the same shape and centre as the burned area, but larger (Figure 3-7 C). The primary purpose of this buffer zone is to analyse nearby unburned areas and understand why certain vegetated zones remained unaffected by the fire. Non-flammable surfaces like sand, water and roads are expected not to burn, but we were particularly interested in nearby vegetated areas that remained unburned. Several factors, explained in detail in section 2.1.2 Spreading behaviour., might explain this, such as shifts in wind direction, higher moisture retention in the vegetation, or north-facing slopes with lower drying rates. By including this buffer zone, we hope our model can detect subtle differences in spectral signatures and terrain data, helping to explain why certain vegetated areas were spared from the fire.

Step 10 - Generate points for the burned area and buffer zone:

Using the ‘Raster to Point’ tool, we generated points for each 20 m-by-20 m pixel, centred within each pixel, for both the burned and buffer zones.

Step 11 - Randomly reduce buffer zone points:

For the burned area, all points remained untouched. However, the buffer zone, being significantly larger, required random removing techniques to maintain a balanced dataset. Using the ‘Create Random Points’ tool, we reduced the buffer zone points to achieve a 1:2 ratio of burned to unburned data (meaning for each burned point, there are two unburned points). Without this reduction, approximately 98% of the data would have been labelled as ‘Not Burned’ (target=0) and only 2% as ‘Burned’ (target=1), making it challenging for the model to recognize patterns associated with burned areas. With a 1:2 ratio, the model can better capture the characteristics of burned areas (target=1) (Figure 3-7 D). For instance, if a burned area contained 3 746 points, the buffer zone, which might initially include 270 643 points, would be randomly reduced to 7 492 points.

Step 12 - Extract spectral and terrain data for each point:

We used the ‘Extract Multi Values to Points’ tool to extract spectral band data (bands B02, B03, B04, B05, B06, B07, B08, B11, and B12) along with terrain data (elevation, slope, and aspect) for each point. The resulting dataset contained 13 columns: 9 spectral bands, 3 terrain data variables, and ‘Burned’ column labelled ‘1’ for burned points and ‘0’ for unburned points.

Using the example above, if a burned area included 3746 points, each row in this subset contained values for the 13 features, with ‘Burned’ marked as 1. Additionally, there were 7492 rows for unburned points with the same feature columns but ‘Burned’ marked as 0.

Step 13 - Add ‘DaysToFire’ column and land classification information for each point:

We added a ‘DaysToFire’ column to indicate the number of days prior to the wildfire event that the satellite image was captured. If the image was captured on the same day, hours prior to the wildfire event, we assigned ‘DaysToFire’=0.

Using the ‘Intersect’ tool, we also incorporated the first (broadest) level of land classification from the COS dataset to each point. This process added a new column specifying the class label (e.g., ‘Wetlands’, ‘Agriculture’). Although COS data is not used as input for model training, it helps analyse which classes are more prone to burning based on known burned areas. Additionally, this information serves as a reference for evaluating model predictions on new data, aiding in performance assessment and identifying areas for improvement.

Finally, we introduced a ‘FireId’ column (combining the wildfire date and ICNF code, in the format {date}_{icnf_code}) for easy identification of points by wildfire.

Step 14 - Repeat steps 6 to 13 for each wildfire:

This process was repeated for every wildfire event used in training (every wildfire in the dataset excluding year 2023) in the filtered dataset (wildfires larger than or equal to 100 ha within the study area).

Step 15 - Merge all wildfires for training:

Using the ‘Merge’ tool, we combined all wildfire points datasets from 2015 to 2022 into a single points dataset.

Step 16 - Export final dataset as a CSV file:

Using ‘Export Table’ tool, we exported the final training points dataset as a CSV (Comma-separated values) file, to be imported into a Jupyter Notebook.

3.5.2. Data preprocessing

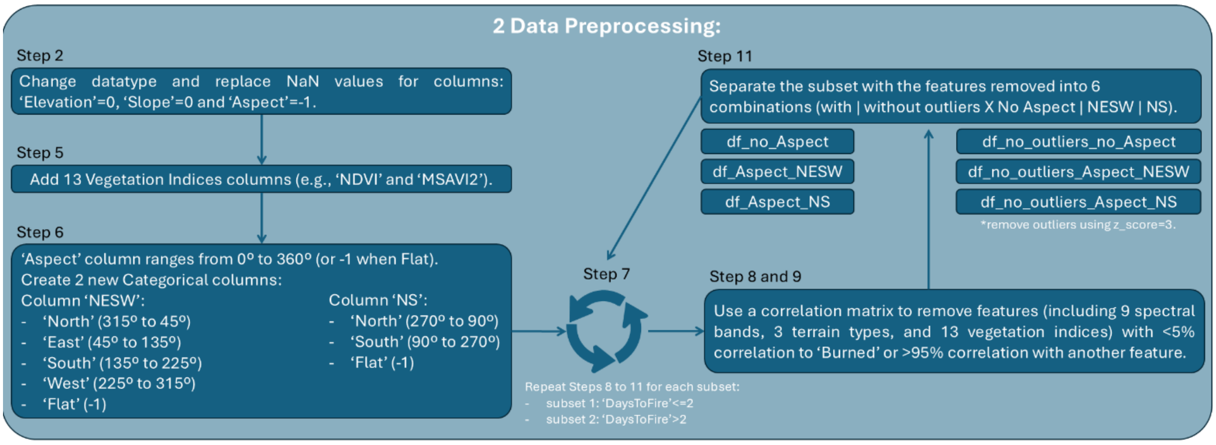


Figure 3-8 Data Preprocessing Diagram

This section details the data preprocessing steps applied to the training dataset in Jupyter Notebook. Below is a detailed breakdown of the process:

Step 1 - Importing the CSV file into Jupyter notebook:

With data acquisition in ArcGIS Pro complete, we imported the dataset as a CSV file into Jupyter Notebook, setting up for further preprocessing, feature engineering, and model preparation.

Step 2 - Cleaning the DataFrame by changing columns data type:

Next, we refined the dataset by converting the ‘Elevation’, ‘Slope’, and ‘Aspect’ columns from strings (CSV file format) to numerical values. This was necessary because these values were stored as strings, though they represent decimal numbers. All other columns were already in the desired format.

Step 3 - Cleaning the DataFrame by replacing missing values:

After adjusting the data types for the ‘Elevation’, ‘Slope’ and ‘Aspect’ columns, we checked for any missing values. ‘Elevation’ column had some, particularly for sea-adjacent points where elevation data was unavailable. Consequently, ‘Slope’ and ‘Aspect’ columns (both derived from elevation) also had missing values for these points. Instead of using imputation techniques such as filling with the median or nearest neighbours, we adopted a more logical approach. For sea areas, we replaced missing ‘Elevation’ values with 0 (representing sea level) and set missing ‘Slope’ values to 0 and missing ‘Aspect’ values to -1 (representing flat terrain), as the sea is considered flat. With this data cleaning process complete, we proceeded to add more features.

Step 4 - Initial exploratory analysis using FireId and COS columns:

We conducted an initial exploratory analysis by examining the ‘FireId’ and COS columns to gain insights. For instance, we plotted fire magnitudes based on ‘FireId’ and assessed the COS classes with the highest burn-prone areas based on ‘COS18n1_C’ and ‘COS18n1_L’.

Step 5 - Feature engineering by adding vegetation indices:

We enriched the dataset with new features by adding vegetation indices. As discussed in section 3.2.2.1, multispectral indices can be particularly effective for capturing soil, moisture, and vegetation correlations in areas prone to burn. We added columns for NDVI, NDRE, NDWI, EVI, LCI, CVI, GCI, ARI, MSAVI2, BSI, NDMI, NBR and NDSI. Rather than just adding every possible multispectral index, we planned to assess these features through correlation matrices to determine which would best contribute to model performance.

Step 6 - Creation of aspect categories:

Since slope direction significantly influences fire susceptibility—for example, increased exposure to sunlight reduces moisture retention, leading to drier vegetation—we experimented with different encoding schemes for the ‘Aspect’ column to assess how directional specificity impacts the prediction of areas prone to burn. As described in section 3.2.3 Terrain data, the ‘Aspect’ column ranges from 0 to 360 degrees, representing the slope direction, much like a compass. For flat areas, the slope is recorded as -1 degrees.

To incorporate slope direction into the dataset, we created a categorical column, ‘Aspect_NESW’, which classifies slopes into four main compass directions: North, East, South, and West. The encoding schema was defined as follows:

- -1:’Flat’;
- 0 to 45 and 315 to 360: ‘North’;
- 45 to 135: ‘East’;
- 135 to 225: South;
- 225 to 315: West;

However, recognizing that excessive granularity might hinder model performance, we also created a simplified version of the slope direction, ‘Aspect_NS’ column, which reduced the directional categories to just three: North, South, and Flat. The encoding schema for ‘Aspect_NS’ was as follows:

- -1: ‘Flat’;
- 0 to 90 and 270 to 360: ‘North’;
- 90 to 270: ‘South’;

After achieving our goal with the Aspect transformation, we dropped the original ‘Aspect’ column. Additionally, ‘FireId’, ‘COS18n1_C’, and ‘COS18n1_L’, which were helpful for initial exploratory analysis, were also removed to streamline the dataset for model training. To conclude, our final training DataFrame consisted of 28 columns, including 9 spectral bands, 2 terrain features (Elevation, Slope), 13 vegetation indices, 2 aspect configuration columns (Aspect_NESW, Aspect_NS), the ‘DaysToFire’ variable, and the target variable ‘Burned’.

Step 7 - Separating the DataFrame into two subsets based on ‘DaysToFire’ column:

Given the variation in the timing of satellite imagery relative to the wildfires event, from a few hours (i.e. ‘DaysToFire’ = 0) to up to 18 days prior, we divided the dataset into two subsets based on the median ‘DaysToFire’ value.

Step 8 - Analysing numerical features for each subset:

Analysing numerical features in each subset separately helped us identify outliers and understand how spectral distributions varied with temporal proximity to wildfire events. Visualizations such as scatter plots, violin plots, and histograms for each spectral band revealed some outliers, likely due to clouds or buildings. While these outliers can be removed to enhance model performance, they may still occur in the test data, reflecting real-world conditions.

Step 9 - Removal of outliers for both subsets:

To handle outliers, we created two versions of each subset: one including outliers and one excluding them, resulting in four distinct DataFrames: Subset 1, subset 1 without outliers, subset 2, and subset 2 without outliers. We used Z-scores to determine which outliers should be removed, setting a threshold to ensure that less than 1% of the data was removed.

Step 10 - Removing uncorrelated or highly correlated features using a correlation matrix:

To determine which features should be retained for their predictive value, we constructed a correlation matrix for each of the four DataFrames. This helped us identify features with either low correlation to the target variable or high correlation with other features, which could introduce redundancy. We applied the following rules:

a) Drop any feature with a correlation of less than 5% (rounded) with the target variable.
b) For features with more than 95% correlation to another feature, retain only the one with a stronger correlation to the target variable.
c) For features with more than 95% correlation to another feature and the same correlation with the target variable, we retained only the feature with the lowest average correlation to the remaining features.

As a result, each DataFrame might now contain fewer than the 28 features listed in Step 6 - Creation of aspect categories:, offering a streamlined feature set specific to each DataFrame. This approach enhances model performance by focusing on the most relevant data, reducing complexity and redundancy.

Step 11 - Creation of feature combinations for optimal model performance:

Having already created four different DataFrames (two subsets with and without outliers) and removing the unnecessary columns of each, we now introduced the three Aspect encoding options (no Aspect information, NESW encoding, and NS encoding) into each DataFrame. This yielded 12 combinations, as shown in Table 3-1 and Table 3-2. The combinations included:

- **Without Aspect:** DataFrames without any slope orientation information.
- **With N and S:** DataFrames including only the North/South/Flat encoding.
- **With N, E, S, and W:** DataFrames with full North/East/South/West/Flat encoding.

Table 3-1 Subset 1 DataFrame combinations.

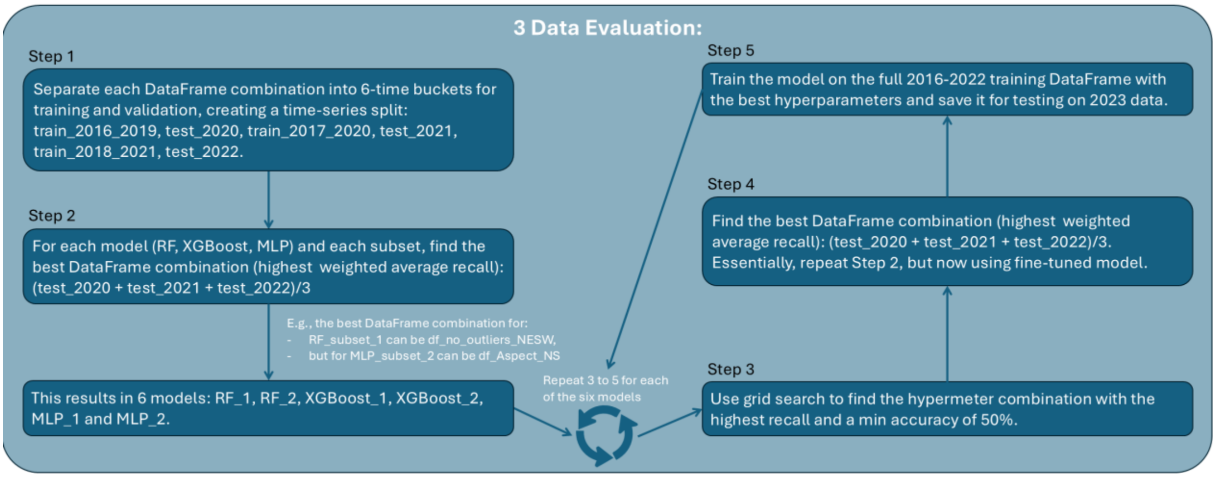
Subset 1	With Outliers	Without Outliers
Without Aspect	df_subset_1_no_Aspect	df_subset_1_no_outliers_no_Aspect
With N and S	df_subset_1_Aspect_NS	df_subset_1_no_outliers_NS
With N, E, S, and W	df_subset_1_Aspect_NESW	df_subset_1_no_outliers_Aspect_NESW

Table 3-2 Subset 2 DataFrame combinations.

Subset 2	With Outliers	Without Outliers
Without Aspect	df_subset_2_no_Aspect	df_subset_2_no_outliers_no_Aspect
With N and S	df_subset_2_Aspect_NS	df_subset_2_no_outliers_Aspect_NS
With N, E, S, and W	df_subset_2_Aspect_NESW	df_subset_2_no_outliers_Aspect_NESW

This approach allowed us to explore which combination of DataFrames would result in better performance across the three machine learning (ML) models—RF, XGBoost, and MLP. The combination of different DataFrames with varying levels of Aspect information provided valuable insights into the impact of slope orientation on fire risk prediction and helped determine the most effective feature set for each model.

3.5.3. Data evaluation



Step 1: Separate each DataFrame combination into 6-time buckets for training and validation, creating a time-series split: train_2016_2019, test_2020, train_2017_2020, test_2021, train_2018_2021, test_2022.

Step 2: For each model (RF, XGBoost, MLP) and each subset, find the best DataFrame combination (highest weighted average recall): (test_2020 + test_2021 + test_2022)/3

E.g., the best DataFrame combination for:
- RF_subset_1 can be df_no_outliers_NESW,
- but for MLP_subset_2 can be df_Aspect_NS

This results in 6 models: RF_1, RF_2, XGBoost_1, XGBoost_2, MLP_1 and MLP_2.

Step 3: Use grid search to find the hypermeter combination with the highest recall and a min accuracy of 50%.

Step 4: Find the best DataFrame combination (highest weighted average recall): (test_2020 + test_2021 + test_2022)/3. Essentially, repeat Step 2, but now using fine-tuned model.

Step 5: Train the model on the full 2016-2022 training DataFrame with the best hyperparameters and save it for testing on 2023 data.

Figure 3-9 Data Evaluation Diagram

With six DataFrame combinations prepared for each subset (1 and 2), the evaluation process aimed to identify the best-performing combination for each model.

Defining the best DataFrame combination - Importance of Recall in wildfire prediction:

In wildfire management, recall—also known as the true positive rate—is crucial as it measures the model's ability to correctly identify actual wildfire events. High recall is essential to capture the majority of real events, as missing even a single wildfire event can result in significant losses to life, property, and the environment. Therefore, this study prioritizes recall, instead of accuracy, to minimize the risk of undetected wildfires and improve response times. While emphasizing recall can lead to a trade-off with accuracy, resulting in more false alarms, the consequences of these false positives are less severe than those of missing a fire event. To implement this prioritization of recall and identify the best-performing DataFrame combination, we structured the evaluation process into 5 distinct steps.

Step 1 - Creation of a time series dataset:

To prevent data leakage - a common issue in ML where test data inadvertently influences the model during training, resulting in overly optimistic results that fail to generalize - we restructured each DataFrame combination to follow a time series logic.

Initially, we used a simple train-test split, where data from the same wildfire event appeared in both training and testing datasets, although individual points were not duplicated. This approach seemed promising at first, as the models initially achieved remarkably high accuracy in predicting wildfire-prone areas. However, when applied to new regions, the performance of models plummeted. This revealed that the models were overfitting to localized wildfire patterns, failing to generalize to new wildfire events.

To mitigate this issue, we revised our approach, training the models on data from 2015 to 2021 and validating it on 2022 data. While this improved validation performance, the models still struggled with unseen data from 2023. This highlighted a deeper issue: the models were overly tuned to the conditions of a single year, limiting their ability to generalize across different temporal and spatial contexts.

To solve this, we implemented a robust time series methodology. The rationale was to train the model on a rolling window of historical data and validate it on future years, thereby preventing the model from overfitting to a particular time period or location. We began by discarding the 2015 data, as it contained only one wildfire, and the satellite imagery for that fire was captured 18 days before the event, making it less reliable for our analysis.

Each DataFrame combination was divided into six time buckets, with three allocated for training and three for validation. The first bucket consisted of data from 2016 to 2019 for training, while the second bucket contained data from 2020 for validation. The third bucket included data from 2017 to 2020 for training, and the fourth bucket used data from 2021 for validation. The fifth bucket comprised data from 2018 to 2021 for training, while the sixth bucket contained data from 2022 for validation.

While this approach was computationally heavier and required more memory, it enabled fine-tuning the model’s hyperparameters based on performance across multiple time periods, rather than relying on a single large training dataset spanning multiple years. This strategy offered several advantages, such as exposing the model to diverse regions, environmental conditions, and wildfire dynamics by training it on specific periods and validating it on the following year. While certain models’ hyperparameter settings could perform better for specific validation time buckets, our focus was on configurations that performed consistently well across all validation time buckets.

Weighted Recall Average:

To ensure a fair evaluation, we used a weighted average for recall instead of a simple arithmetic average during the validation stage. This approach gave larger validation time buckets (2020, 2021, and 2022) proportional influence on the overall recall values, preventing smaller validation time buckets with higher recall values from skewing the results. The weighted average (Equation 3-14$Weighted\ Recall\ Average = \frac{\sum_i^n (Validation\ Dataset\ Size_i * Recall_i)}{\sum_i^n (Validation\ Dataset\ Size_i)}$ (3-14), which accounts for validation time buckets size, provided a more robust and reliable assessment of the model's performance:

$$Weighted\ Recall\ Average = \frac{\sum_i^n (Validation\ Dataset\ Size_i * Recall_i)}{\sum_i^n (Validation\ Dataset\ Size_i)} \qquad (3\text{-}14)$$

With the time series dataset established and using weighted recall average, we evaluated the six DataFrame combinations for each subset and model in the next step.

Step 2 - Evaluating the six DataFrame combinations for each subset for each model:

Using the time series dataset we were able to evaluate the recall for each of the six DataFrame combinations across both subsets (1 and 2) and for all three models (RF, XGBoost, and MLP). For the MLP Classifier, we scaled the training data prior to validation. The combination with the highest weighted recall was selected as the 'winner' combination for each model in each subset, resulting in six best-performing models (three models per subset).

Step 3 - Fine tuning each model for the selected combination:

After identifying the best DataFrame combination for each model and subset, we proceeded with a grid search to fine-tune each model's hyperparameters. This process involved testing how changes to hyperparameter values—either increasing or decreasing—affected recall. Two distinct trends emerged during this testing. The first trend involved correlated hyperparameters, which demonstrated a clear relationship with recall, as changes in their values directly affected performance. The second trend involved non-correlated hyperparameters, which showed no significant impact on recall regardless of the adjustments made. For non-correlated hyperparameters, we selected values better suited for handling more complex and larger datasets. For correlated, we continued testing until we observed a convergence in recall values.

Occasionally, some hyperparameter settings achieved a perfect score of 1, which seemed ideal at first, making them as the best setting. However, these settings also showed low accuracy scores (around 0.33), indicating an over-prediction of burned areas - suboptimal outcome. To address this, we set a minimum accuracy threshold of 0.50 (50%), selecting only hyperparameter combinations that met this criterion.

Step 4 - Re-evaluating DataFrame combinations with fine-tuned models:

Before training the six models (three models for each subset) on the full dataset, we validated whether the previously selected DataFrame combinations remained optimal with the updated hyperparameters. This step essentially repeats Step 2, but with the fine-tuned hyperparameters applied to each model. While fine-tuning was initially performed on specific combinations, other combinations might now perform better due to the refined hyperparameters. To ensure the best choice, all combinations were re-evaluated for each model, and the one achieving the highest recall, while maintaining an accuracy threshold of at least 0.5, was selected.

Step 5 - Training models with the full training dataset:

After selecting the best-performing DataFrame combination from Step 4 (with fine-tuned hyperparameters from Step 3) for each of the six models (RF, XGBoost, and MLP for subsets 1 and 2), we trained these models on the complete dataset spanning 2016 to 2022.

This step prepared the models for testing with unseen data from 2023 to identify the best-performing model for each subset.

3.5.4. Model evaluation

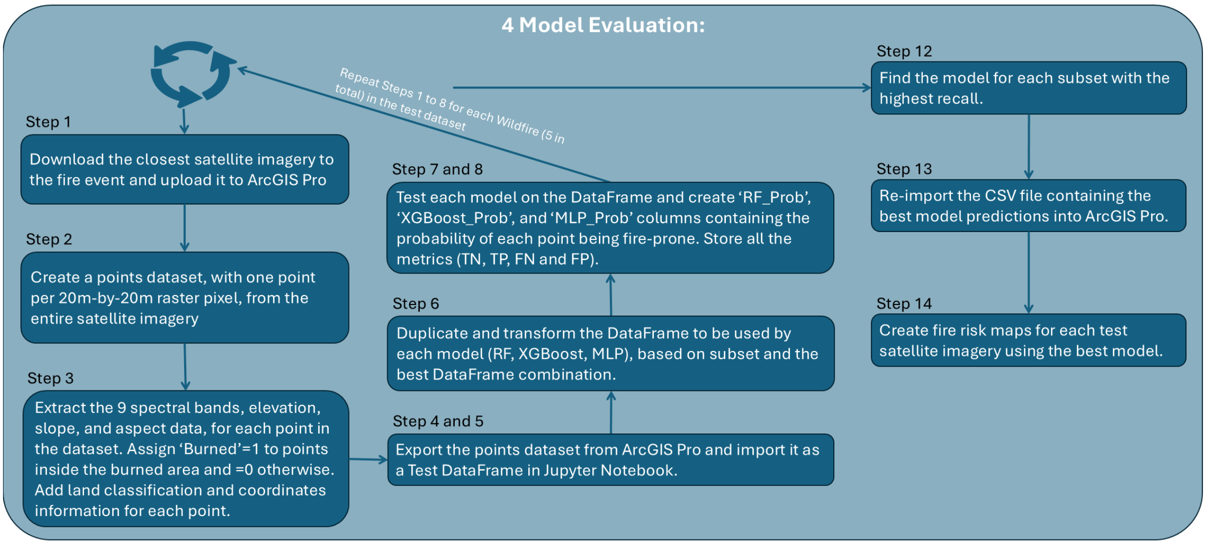


Figure 3-10 Model Evaluation Diagram

With the six models trained (RF_subset1, RF_subset2; XGBoost_subset1; XGBoost_subset2; MLP_subset1; MLP_subset2), the next phase is to evaluate them using unseen data from 2023. This testing phase will determine the best-performing model for each subset (1 and 2).

Steps 1 through 11 were applied to each wildfire in the test dataset, i.e. those from 2023 that burned at least 100 hectares in the study area. A preliminary note: besides the wildfire test dataset, ArcGIS Pro already contains the imported terrain data as well as land classification and delimitation data. For each wildfire test case, we only need to download the relevant satellite imagery, apply necessary transformations, and preprocess the data to make it ready for testing with each of the three models (RF, XGBoost, and MLP).

Step 1 - Download and upload satellite imagery closest to wildfire date:

As described in Step 6 - Download and upload satellite imagery closest to wildfire date from Copernicus: of section 3.5.1 Data acquisition, we obtained the closest satellite imagery relative to the date of each wildfire event (present in our filtered test wildfire table) from the Copernicus Program. After downloading the TIF file, it was uploaded into ArcGIS Pro for further processing.

Step 2 - Prepare the TIF satellite imagery data for use in Jupyter Notebook:

Following Step 7 - Convert TIF data to RGB format: to Step 13 - Add 'DaysToFire' column and land classification information for each point: of section 3.5.1 Data acquisition, with one key difference: instead of using a 5 km buffer to define unburned areas, we processed the entire satellite imagery. This approach tests the model's ability to predict areas prone to burn across the imagery, with the expectation that one of these areas will overlap with the actual burned region. The outcome was a points dataset where each point represented a 20 m-by-20 m raster pixel, containing data for 9 spectral bands, 3 terrain variables, 1 land classification column, and a 'Burned' column (1 for burned areas, 0 for unburned areas).

Step 3 - Add XY coordinates to the points dataset:

To facilitate re-importing the points dataset from Jupyter Notebook back to ArcGIS Pro, we needed a unique identifier for each point. We found that using the 'Add XY Coordinates' tool was the most reliable approach for ensuring proper geolocation when re-importing. This process added two columns, 'POINT_X' and 'POINT_Y', to the points dataset, alongside with the previously 14 added data columns.

Step 4 - Export the points dataset as a CSV file:

With the 'Export Table' tool, we saved the resulting geospatial points dataset as a CSV file.

Step 5 - Import the CSV file into Jupyter notebook as a test DataFrame:

Switching to Jupyter Notebook, we imported the previously exported CSV file as a Test DataFrame, for generating the model's predictions.

Step 6 - Transform the Test DataFrame based on the subset (1 or 2) and the model:

After importing the CSV file, we applied transformations to match the specific DataFrame combinations (e.g., Aspect settings, outlier handling) used during the training phase for each subset and each model (different models may have different winner combinations regarding Step 3 - Fine tuning each model for the selected combination: of section 3.5.3Data evaluation). If the difference between satellite imagery date and the fire event date was smaller than or equal to the median of 'DaysToFire' column calculated in Step 7 - Separating the DataFrame into two subsets based on 'DaysToFire' column: of section 3.5.2 Data preprocessing, the Test DataFrame was marked as subset 1; otherwise, it was marked as subset 2. The Test DataFrame was thus transformed accordingly to ensure compatibility with each model and subset.

Step 7 - Test each model (RF, XGBoost, and MLP) on the Test DataFrame:

After the Test DataFrame was transformed for each model, three columns were added, each reflecting one model's predicted burn likelihood for each point in the dataset: 'RF_Prob', 'XGBoost_Prob', and 'MLP_Prob'. These columns store the predicted probability (ranging from 0 to 1) that each point (i.e., each 20 m-by-20 m raster pixel) is prone to burning, with values closer to 0 indicating lower likelihood and values closer to 1 indicating higher likelihood.

Step 8 - Store models metrics:

Using the 'Burned' column for true test data and the prediction columns ('RF_Prob', 'XGBoost_Prob', and 'MLP_Prob'), we calculated metrics such as True Positives (TP), False Positives (FP), True Negatives (TN), False Negatives (FN), Recall, and Accuracy. All metrics were calculated using a probability threshold of 0.4, meaning a point was classified as prone to burn if its probability prediction was equal to or greater than 0.4. This approach ensured consistency across models and emphasized true positives to prioritize recall. These metrics were stored for later evaluation.

Step 9 - Evaluate each model using land classification (COS):

We conducted an initial review of each model's results by examining the distribution of points classified as '1' (prone to burn) across COS first level classes. We, specifically, checked for high percentages in classes where prone to burn predictions should be minimal or absent, such as water bodies. This evaluation helped ensure that the models accurately identified vegetation areas as prone to burn, while avoiding non-flammable classes.

Step 10 - Export the Test DataFrame as a CSV file:

The Test DataFrame, now containing model predictions, was exported as a CSV file.

Step 11 - Repeat Steps 1 to 10 for each wildfire in the test dataset:

This process (Steps 1 through 10) was repeated for each wildfire in the test dataset (wildfires from 2023 in the study area that burned at least 100 hectares).

Step 12 - Identify the best model for each subset:

Still in Jupyter Notebook, using the previously stored metrics, we calculated the total recall for each model (RF, XGBoost, and MLP) within each subset (subset 1 and 2). The model with the highest recall for each subset was selected as the "winner".

Step 13 - Re-import the CSV file into ArcGIS Pro:

After adding the CSV files previously exported in Step 10 to ArcGIS Pro, we used the 'XY Table To Point' tool to generate a geographic points dataset. This generated points dataset now included the model prediction columns (RF_Prob, XGBoost_Prob, and MLP_Prob), distinguishing it from the dataset in Step 5.

Step 14 - Create fire risk raster maps for each test satellite imagery using the best model:

Using the 'Points to Raster' tool, we generated fire risk raster maps for each test points dataset, which encompassed the entire test satellite imagery, based on the best model for each subset. Unlike previous filtering steps, this time, we included all data points, without applying the 40% threshold, to visualize the areas prone to burn more comprehensively.

This concludes the Methodology chapter, where we detailed the study area, the datasets, and the technological tools employed to address the research question posed in the Introduction chapter and to bridge the research gap identified in the 2.3 Related works section. In the following chapter, Results and discussion, we will present and analyse the outcomes of this methodology, including all relevant findings, visualizations, and metrics. Finally, we will evaluate whether this approach effectively addresses the identified challenges and offers a meaningful contribution to mitigating the wildfire problem.

4. RESULTS AND DISCUSSION

This chapter presents the results obtained through the application of the methodology, encompassing all stages from Data acquisition to Model evaluation. This chapter is pivotal in providing insights and conclusions drawn from the process. Each result is accompanied by a concise description, analysis, and interpretation to facilitate understanding and contextualization. The chapter is structured into six main sections: training data, training models, improving models, testing data, testing models, and comparing studies.

4.1. TRAINING DATA

As outlined in section 3.2.1 Fire events data, the training and validation process (2015-2022) relied on data from 30 distinct wildfires events within the study area. These wildfires cover major forested regions and some highly populated areas, as illustrated in Figure 4-1. It provides a visual representation of the training dataset, highlighting wildfire polygons and the associated 5 km buffer zones that represent unburned areas.

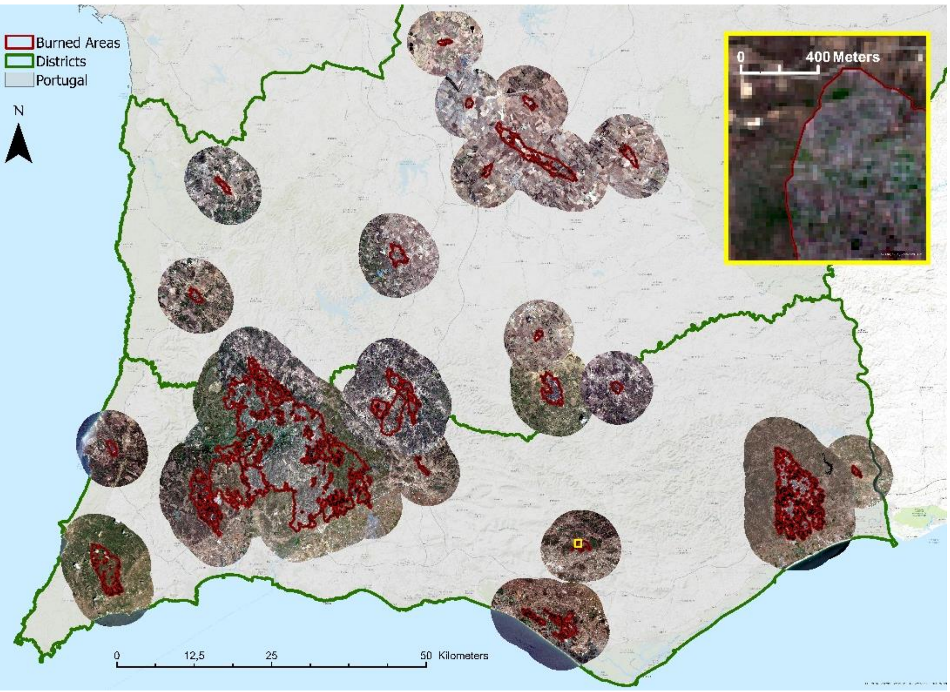


Figure 4-1 Training dataset of burned (red polygons) and unburned areas (buffer) (2015-2022)

In Figure 4-1, the yellow square highlights a close up view of a portion of a burned area (prior to burning), showcasing the 20 m resolution of the dataset. For reference, the 400-meter scale covers 20 pixels. This illustrates the complexity of the classification problem, as the models must differentiate between vegetation near burned areas (classified as unburned).

The training dataset contains 32 unique wildfire records, as shown in Table 4-1. This slight discrepancy arises because two of the 30 wildfire events occurred over multiple days, resulting in two separate satellite imagery datasets for each, labelled as “1st Fire” and “2nd Fire”. This method ensures that temporal variations are adequately represented in the training data.

Table 4-1 Wildfire records in the training set

Fire Id (Date + ICNF Code)	**Days to fire**	**Unburned points**	**Burned points**	**Total points**
2015-10-01 AG115581	18	10 156	5 078	15 234
2016-08-08	0	90 432	45 216	135 648
2016-09-03 1st Fire	6	19 758	9 879	29 637
2016-09-03 2nd Fire	0	171 578	85 789	257 367
2016-09-09	2	6 684	3 342	10 026
2017-08-10	2	7 492	3 746	1 1238
2017-09-16	4	5 034	2 517	7 551
2018-08-03 AG118728 1st Fire	0	1 125 050	562 525	1 687 575
2018-08-03 AG118728 2nd Fire	0	21 8130	109 065	327 195
2018-08-07 AG118766	2	5 488	2 744	8 232
2018-09-01 AT318350	4	51 568	25 784	77 352
2019-06-10 AT319256	1	27 452	13 726	41 178
2019-06-17 AT319291	8	17 012	8 506	25 518
2019-06-17 AT319292	8	5 616	2 808	8 424
2019-07-19 AG119811	0	10 368	5 184	15 552
2019-07-26 AG119842	2	5 626	2 813	8 439
2019-09-10 AT319505	3	9 794	4 897	14 691
2020-06-19 AG120550	1	111 782	55 891	167 673
2020-07-03 AT320155	0	5 868	2 934	8 802
2020-07-06 AG120614	3	26 758	13 379	40 137
2020-07-13 AT320181	0	116 670	58 335	175 005
2021-05-22 AG121869	1	4 996	2 498	7 494
2021-07-12 AT321215	4	9 978	4 989	14 967
2021-07-17 AG1211131	4	96 638	48 319	144 957

2021-08-04 AG1211218	7	10 944	5 472	16 416
2021-08-16 AG1211287	2	331 374	165 687	497 061
2021-08-18 AG1211309	1	9 842	4 921	14 763
2021-08-18 AT321313	1	45 658	22 829	68 487
2022-07-12 2022_08_0035308	4	36 910	18 455	55 365
2022-07-25 2022_08_0038271	2	65 712	32 856	98 568
2022-08-22 2022_02_0021224	0	5 832	2 916	8 748
2022-09-18 2022_02_0023633	7	12 792	6 396	19 188

‘Days to fire’ column indicates the number of days between the satellite imagery date (Table 0-3 in Appendix A) and the wildfire’s start date (Table 0-1 in Appendix A), with a value of 0 meaning the imagery was taken just hours before the wildfire began.

Each point in the dataset corresponds to a pixel of 20 m-by-20 m, with an area of 400 m² (0.04 ha). This granularity allows the dataset to accurately reflect the reported burned areas. For instance, the wildfire from 2015-10-01 AG115581 has 5 078 points inside the burned polygon. Multiplying 5 078 by 0.04 ha, we obtain 203.12 ha, closely aligning with the reported burned area of 203.06 ha in Table 0-1 in Appendix A. This consistency demonstrates that the 20 m resolution training data effectively captures actual wildfire events.

However, the same level of accuracy cannot be claimed for the unburned buffer zones. To maintain a balanced dataset with a 1:2 ratio of burned to unburned points, random sampling techniques were applied. As evident in Table 4-1, the number of points classified as Burned=0 is consistently double that of Burned=1 across all wildfire events. This ensures that the model prioritizes sensitivity to burned areas while preserving computational efficiency.

4.1.1. Training DataFrame

Following the methodology outlined in the Data acquisition section, we generated a comprehensive points dataset in ArcGIS Pro, which was then imported into Jupyter Notebook as a DataFrame (Table 0-1 in Appendix B). This DataFrame contains over 4 million records (4 018 488 in total), with each record corresponds to a 20 m-by-20 m raster pixel from the satellite imagery. The DataFrame includes spectral data from 9 Sentinel-2 bands, terrain data derived from 3 variables (elevation, slope, and aspect), and wildfire-related data. The ‘Burned’ column indicates whether a pixel burned during a specific wildfire event (target = 1) or remained unburned (target = 0). This DataFrame served as the foundation for later preprocessing, analysis, and model training.

4.1.2. Cleaning DataFrame

Following the methodology outlined in section 3.5.2 Data preprocessing (Step 3 - Cleaning the DataFrame by replacing missing values:), we identified a total of 12 415 points (0.31% of the entire training DataFrame) with missing values for the ‘Elevation,’ ‘Slope,’ and ‘Aspect’ features. To address this, we replaced missing values in ‘Elevation’ with 0, in ‘Slope’ with 0, and in ‘Aspect’ with -1. This adjustment aligns with the characteristics of sea and ocean areas, which are considered flat.

4.1.3. COS information

Leveraging the COS (Land Classification) data alongside the Burned column, we calculated the distribution of burned points (Burned = 1) across different COS classes, as shown in Table 0-2 in Appendix B.

As expected, vegetation-dense classes such as Forests and Scrublands show the highest percentage of burned points, while water and wetland-related classes, which are less flammable, exhibit lower burned percentages.

It is essential to acknowledge that both COS data and ICNF burned area polygons are not flawless. The COS dataset, last updated in 2018, is prone to misclassification due to land-use changes; for instance, an area classified as ‘Wetlands’ in 2018 might now be an agriculture field. Similarly, ICNF burned area polygons, created through manual delineation, may occasionally include non-flammable regions such as lakes, roads, buildings, or bare soil. While COS data is not utilized during model training, it will have its use later to evaluate the model’s predictions. For example, if the model predicts that water-dominated areas are more prone to burning than vegetation-dense areas, it would indicate poor model performance. Despite the limitations outlined above, such result would remain unacceptable.

4.1.4. Aspect information

Using the schema described in section 3.5.2 Data preprocessing Step 6 - Creation of aspect categories:, we replaced the original ‘Aspect’ column (which ranged from 0 to 360 degrees) with cardinal directions: North, East, South, and West. Table 4-2 shows that slopes facing East and South exhibited a higher likelihood of burning compared to those facing North or West. This aligns with geographical expectations: East- and South-facing slopes experience higher aridity due to prolonged sun exposure, while North-facing slopes are cooler and retain more moisture, reducing fire risk.

Table 4-2 Percentage of burned points (target=1) across main cardinal directions

Direction	Burned (%)	Dataset (%)
East	35.07	25.53
South	34.48	28.15
West	32.66	24.58
North	31.55	21.06

Flat	0.19	0.68

Table 4-2 highlights that, if the aspect of a slope were uncorrelated with the target variable (‘Burned’), all the values in the ‘Burned (%)’ column would be closer to 33.33%, reflecting the 1:2 ratio of burned to unburned areas in the dataset. However, the higher percentages for East and South-facing slopes suggest a meaningful correlation between slope direction and fire occurrence. Flat areas account for only 0.68% of the dataset, with 0.31% (12 415 missing values from section 4.1.2) attributed to sea. Since these areas are not prone to burning, they minimally impact the analysis of burned points.

A further simplification of the categorization into just North and South directions yielded similar insights. Table 4-3 demonstrates that South-facing slopes again showed a higher likelihood of burning (34.41%) compared to North-facing slopes (32.53%).

Table 4-3 Percentage of burned points (target=1) only across North and South

Direction	Burned (%)	Dataset (%)
South	34.41	54.40
North	32.53	44.92
Flat	0.19	0.68

Later in this chapter, we will evaluate which Aspect encoding configuration performed best in the models: no Aspect information, full cardinal directions (North, East, South, West), or the simplified North-South categorization.

4.1.5. Splitting the training DataFrame into two subsets

Our model is designed to identify areas prone to burn but not the timing of such events. Ideally, all satellite imagery would be captured at the same temporal distance from wildfire events, however, this is not the case, as showed in Figure 4-2.

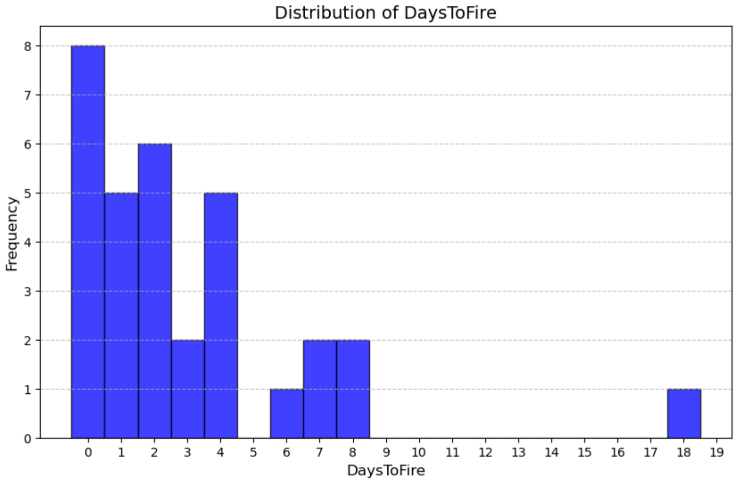


Figure 4-2 Histogram showing the distribution of 'DaysToFire' feature.

To address this uneven distribution, we divided the training DataFrame based on the median value of the 'Days to fire' column, calculated from the 32 wildfire records in Table 4-1. The median value was determined to be 2, leading to the creation of two subsets:

- **Subset 1 DataFrame**: Includes records where 'DaysToFire' is less than or equal to 2 days, capturing data that was taken closer to the fire event. This subset is expected to contain spectral patterns that are more immediately linked to the fire's occurrence.
- **Subset 2 DataFrame**: Includes records where 'DaysToFire' is greater than 2 days, capturing data that reflects broader environmental and vegetation conditions less immediately linked to the wildfire event.

By creating these subsets based on temporal proximity ('DaysToFire'), we enabled the models to analyse variations in spectral data from different timeframes. This approach may help uncover time-sensitive conditions that influence fire susceptibility. However, this division introduced a significant imbalance between the subsets. The six largest wildfires (as shown in Table 4-1) are all part of subset 1, resulting in Subset 1 containing 3 549 051 records, while subset 2 contains only 469 437 records.

From this point onward, the project will consistently work with two distinct datasets—subset 1 and subset 2.

4.1.6. Outliers' analysis and removal

We examined outliers for every numeric feature in both subsets and observed non-smooth distributions primarily in the spectral band data. Figure 0-1 (left: subset 1, right: subset2) in Appendix C illustrates the presence of outliers across the spectral bands.

As mentioned in Step 9 - Removal of outliers for both subsets: of 3.5.2 Data preprocessing section, these outliers are not corrupt or incorrect data; rather, they reflect actual measurements captured by the satellite bands. Consequently, similar patterns are expected in the test dataset. However, removing outliers might enhance the overall performance of the ML models. To investigate this, we decided to create separate DataFrames for each subset: one with outliers removed and one with outliers retained.

Figures (Figure 0-3 for subset 1 and Figure 0-4 for subset 2) in the Appendix C showcase the outlier removal process for one band per subset (‘B11’ for subset 1 and ‘B04’ for subset 2). Although, the same procedure, using a Z-score threshold of 3, was applied to all nine spectral bands in both subsets.

In total, 37 821 points were removed from subset 1 (1.07% of the dataset), and 5 593 points were removed from subset 2 (1.19% of the dataset). As a result, we created four distinct DataFrames:

- **subset_1**: Original subset 1 with outliers retained
- **subset_1_no_outliers**: subset 1 with 37 821 points removed
- **subset_2**: Original subset 2 with outliers retained
- **subset_2_no_outliers**: subset 2 with 5 593 points removed

This approach allows us to evaluate the impact of outlier removal on the performance of the ML models in subsequent steps.

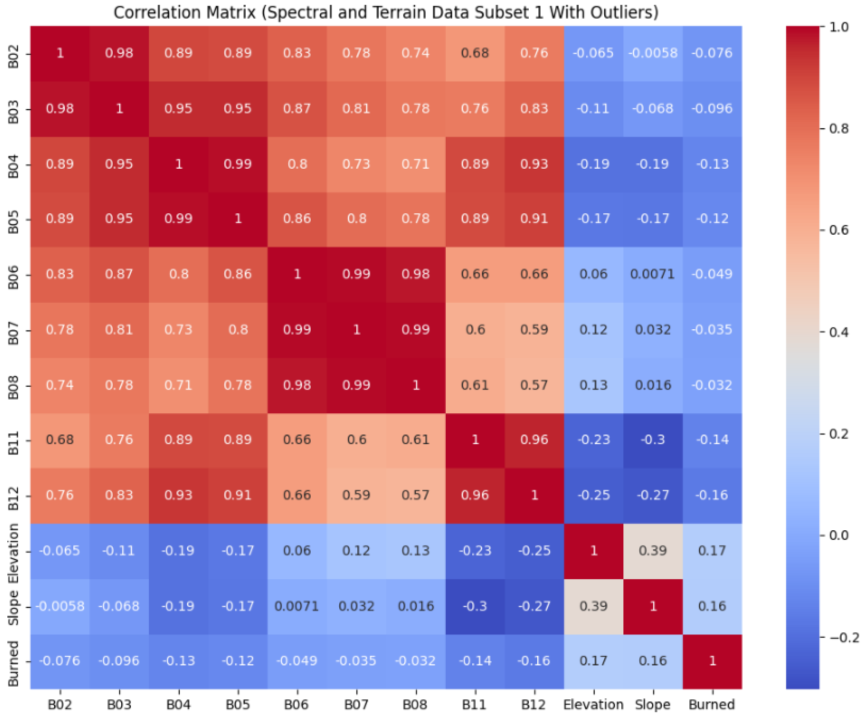
For a detailed breakdown of the process, including visualizations, please refer to Appendix C.

4.1.7. Feature removal using correlation matrix

While dimensionality reduction techniques such as PCA, t-SNE, and UMAP were initially considered, correlation matrices were selected as the simplest and most effective approach for this study. Unlike dimensionality reduction, which transforms the dataset into a lower-dimensional space, correlation matrices facilitate feature selection by examining relationships between features and the target variable. This method enabled the removal of redundant features or those with limited correlation to the target variable, ensuring a more interpretable and streamlined dataset for model training.

We will show the rationale for one DataFrame (subset_1), which had the most nuanced details. Additional visualizations and explanations for other DataFrames are provided in Appendix D.

For subset_1 DataFrame, we observed the following results (Figure 4-3, Figure 4-4):



	B02	B03	B04	B05	B06	B07	B08	B11	B12	Elevation	Slope	Burned
B02	1	0.98	0.89	0.89	0.83	0.78	0.74	0.68	0.76	-0.065	-0.0058	-0.076
B03	0.98	1	0.95	0.95	0.87	0.81	0.78	0.76	0.83	-0.11	-0.068	-0.096
B04	0.89	0.95	1	0.99	0.8	0.73	0.71	0.89	0.93	-0.19	-0.19	-0.13
B05	0.89	0.95	0.99	1	0.86	0.8	0.78	0.89	0.91	-0.17	-0.17	-0.12
B06	0.83	0.87	0.8	0.86	1	0.99	0.98	0.66	0.66	0.06	0.0071	-0.049
B07	0.78	0.81	0.73	0.8	0.99	1	0.99	0.6	0.59	0.12	0.032	-0.035
B08	0.74	0.78	0.71	0.78	0.98	0.99	1	0.61	0.57	0.13	0.016	-0.032
B11	0.68	0.76	0.89	0.89	0.66	0.6	0.61	1	0.96	-0.23	-0.3	-0.14
B12	0.76	0.83	0.93	0.91	0.66	0.59	0.57	0.96	1	-0.25	-0.27	-0.16
Elevation	-0.065	-0.11	-0.19	-0.17	0.06	0.12	0.13	-0.23	-0.25	1	0.39	0.17
Slope	-0.0058	-0.068	-0.19	-0.17	0.0071	0.032	0.016	-0.3	-0.27	0.39	1	0.16
Burned	-0.076	-0.096	-0.13	-0.12	-0.049	-0.035	-0.032	-0.14	-0.16	0.17	0.16	1

Figure 4-3 Correlation matrix of spectral and terrain features for subset_1 DataFrame.

Analysing Figure 4-3, we began by applying the rules outlined in Step 10 - Removing uncorrelated or highly correlated features using a correlation matrix: of 3.5.2 Data preprocessing section to discard features. Applying rule a) we discarded the features ‘B07’ and ‘B08’. Feature ‘B06’ was retained as it had a correlation of -0.049 with ‘Burned,’ which rounds to -0.05, satisfying the 5% threshold. Applying rule b) to ‘B02’ and ‘B03’ (which had a correlation of 0.98 with each other), we retained only one. Since ‘B02’ had a correlation of -0.076 with the ‘Burned’ variable compared to ‘B03’s -0.096, we removed B02. Similarly, we applied the same approach to ‘B04’ and ‘B05’, removing ‘B05’, and to ‘B11’ and ‘B12’, removing ‘B11’.

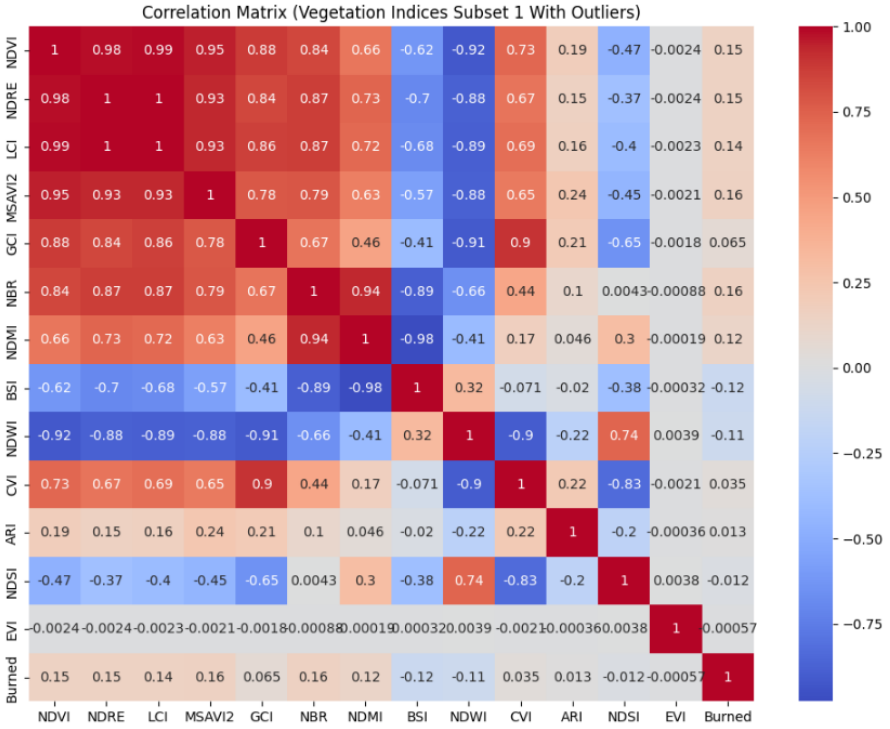


Figure 4-4 Correlation matrix of multispectral indices for subset_1 DataFrame.

Analysing Figure 4-4 and applying rule a) we discarded the features ‘CVI’, ‘ARI’, ‘NDSI’ and ‘EVI’. Since ‘NDVI’, ‘NDRE’ and ‘LCI’ all had correlations greater than 95% with each other’s, only one of these three features could remain (as state in rule b)). We discarded ‘LCI’ because it exhibited the weakest correlation with the ‘Burned’ variable, however, as ‘NDVI’ and ‘NDRE’ were tied with a 15% correlation with the target variable, we needed to apply rule c). Rule c) required us to evaluate each feature’s average correlation with the remaining features (excluding those already discarded). In this case, regarding feature ‘MSAVI2’, ‘NDVI’ had a 95% correlation, while ‘NDRE’ had a 93%, giving ‘NDRE’ a slight advantage by 2%. After applying this process to all remaining features, ‘NDRE’ was found to have a higher average correlation, and was therefore discarded. Finally, we needed to choose between ‘NDMI’ and ‘BSI’. Since both had the same correlation with the target variable, we reapplied rule c) and removed ‘NDMI’.

From this point onward, the process proceeded more efficiently as the rules outlined in Step 10 - Removing uncorrelated or highly correlated features using a correlation matrix: of section 3.5.2 Data preprocessing were already established. The remaining correlation matrices and their analyses are provided in Appendix D.

To simplify the visualization of feature removal, we created a matrix (Table 4-4,

Table 4-5) where each row corresponds to a specific DataFrame, and each column represents a feature. Each cell in the matrix indicates the outcome for a given feature-DataFrame combination: dark grey cells represent features that were retained, while white cells indicate features that were removed.

Table 4-4 Matrix of spectral bands and terrain features for each DataFrame combination. Dark grey cells represent retained features, while white cells represent removed features.

	B02	B03	B04	B05	B06	B07	B08	B11	B12	Elevation	Slope
Subset 1		■	■		■				■	■	■
Subset 1 no outliers		■	■		■				■	■	■
Subset 2			■			■			■		
Subset no outliers			■		■				■		

Table 4-5 Matrix of Multispectral indices for each DataFrame combination. Dark grey cells represent retained features, while white cells represent removed features.

	NDVI	NDRE	LCI	MSAVI2	GCI	NBR	NDMI	BSI	NDWI	CVI	ARI	NDSI	EVI
Subset 1	■			■	■	■		■	■				
Subset 1 no outliers				■	■	■			■		■		
Subset 2				■					■		■	■	
Subset no outliers				■					■			■	

Analysing the results of feature selection across subset 1 and 2, we observed significant differences, which were expected given the variations in data magnitudes between subsets. The treatment of outliers further influenced feature configurations, underscoring how retaining or removing outliers can have a notable impact on both feature selection and model evaluation.

Certain features, such as ‘B07’, ‘B08’, ‘CVI’, ‘NDSI’, and ‘EVI’, were consistently removed for subset 1 (both with and without outlier treatment). Similarly, for subset 2, the features ‘Elevation’, ‘Slope’, ‘GCI’, ‘NBR’, ‘NDMI’, ‘BSI’, ‘CVI’, and ‘EVI’ were excluded across both outlier treatments. Additionally, features ‘B02’, ‘B05’, ‘B11’, ‘NDRE’, and ‘LCI’ were removed from all four DataFrame configurations due to high correlations with other features that had stronger correlations with the target variable.

Despite these exclusions, spectral bands ‘B04’, ‘B06’, and ‘B12’ - corresponding to spectral resolutions of 665 nm, 740 nm and 2190 nm, respectively, as detailed in Table 0-2 in Appendix A- emerged as the most significant spectral features for analysing wildfire events. While multispectral indices such as ‘NDVI’ consistently demonstrated strong correlations with the target variable, ‘MSAVI2’ stood out as the top performer, exhibiting the highest correlation across all configurations. Furthermore, ‘NDWI’ also was retained consistently across all DataFrame configurations. These findings not only provide a clearer understanding of feature relevance but also offer valuable insights for future research, particularly in refining spectral and multispectral indices for wildfire prediction models.

4.2. TRAINING MODELS

As detailed in the Step 11 - Creation of feature combinations for optimal model performance: of 3.5.2 Data preprocessing section, each of the four previous DataFrames (subset_1, subset_1_no_outliers, subset_2, subset_2_no_outliers) was further divided into three configurations based on Aspect information: No Aspect, North, East, South, and West (NESW), and North and South (NS). This process resulted in twelve distinct DataFrame combinations—six for each subset (1 and 2).

Three ML models, Random Forest (RF), eXtreme Gradient Boosting (XGBoost), and Multi-Layer Perceptron (MLP), were applied to each subset. For each model, we evaluated which of the six DataFrame combinations (two outlier treatments with three Aspect configurations) achieved the highest recall, leading to six unique sets of results (three models with two subsets).

As previously explained in Step 1 - Creation of a time series dataset: of 3.5.3 Data evaluation section, a time-series approach was used for training and validation, ensuring robust testing across temporal windows. Additionally, a weighted recall average was applied to account for differences in validation dataset sizes. For instance, the validation time bucket for 2020 contained 351 489 records, while 2021 had 587 815 records, and 2022 had only 107 316 records. This weighting prevented smaller validation time buckets, such as 2022, from disproportionately skewing the results, while larger validation time buckets, like 2021, had a greater influence on the final performance metrics.

4.2.1. RF subset 1

The initial hyperparameter values for the RF Classifier for subset 1 were as follows:

- **n_estimators**=150
- **max_depth**=10
- **min_samples_split**=0.002
- **min_samples_leaf**=0.001
- **max_features**=None

Table 0-1 (Appendix E) presents the performance metrics for the RF model on subset 1. These results reflect the initial hyperparameter settings and highlight the trade-off between recall and accuracy across the different combinations. As recall increases, accuracy decreases. This trade-off underscores the challenge of achieving a model that optimizes both metrics simultaneously. Additionally, removing outliers appears to improve recall, suggesting the importance of this preprocessing step.

The df_subset_1_no_outliers_NESW combination currently exhibits the highest Recall, with 0.361122 for subset 1, and is identified as the best-performing combination at this stage.

4.2.2. RF subset 2

The same hyperparameter values used for RF subset 1 were applied to subset 2. Table 0-2 (Appendix E) outlines the performance metrics for subset 2, where higher recall values are observed compared to subset 1. This higher recall values for subset 2 are likely influenced by the fact that subset 2 contains significantly fewer data points, covering a much smaller area, and is therefore less complex.

The df_subset_2_no_outliers_NESW combination is identified as the best-performing combination for subset 2 at this stage, achieving the highest recall of 0.581467.

4.2.3. XGBoost subset 1

The initial hyperparameter values for the XGBoost for subset 1 were as follows:

- n_estimators=125
- eta=0.1
- max_depth=10
- min_child_weight=0.001
- subsample=0.6
- colsample_bytree=1

The performance results for XGBoost on subset 1 are summarized in Table 0-3 (Appendix E). Compared to RF, XGBoost showed a noticeable decline in Recall. This could be attributed to the chosen hyperparameters, suggesting potential for improvement during fine-tuning.

The current best-performing combination for subset 1 is df_subset_1_no_outliers_NESW, achieving the highest recall value of 0.295829.

4.2.4. XGBoost subset 2

The same hyperparameter values were applied to subset 2. The performance results for XGBoost on subset 2 are summarized in Table 0-4 (Appendix E). Subset 2 exhibited relatively strong accuracy values across all combinations compared to RF subset 2. This will provide an interesting comparison when evaluating performance on unseen test data from 2023.

The current best-performing combination for subset 2 is df_subset_2_no_outliers_no_Aspect, which achieved the highest recall value of 0.457076.

4.2.5. MLP subset 1

The initial hyperparameter values for the MLP Classifier for subset 1 were as follows:

- **hidden_layer_sizes**=(32, 16, 8)
- **max_iter**=200
- **activation**=‘relu’
- **solver**=‘adam’
- **alpha**=0.00001

The performance results for MLP on subset 1 are summarized in Table 0-5 (Appendix E). MLP performance in terms of recall for subset 1 is lower compared to RF. However, the model demonstrated relatively stable Accuracy values across combinations. These results indicate the potential for improvement during the fine-tuning phase.

The current best-performing combination for subset 1 is df_subset_1_no_outliers_NESW, which achieved the highest recall value of 0.308916.

4.2.6. MLP subset 2

The same hyperparameter values were applied to subset 2. The performance results for MLP on subset 2 are summarized in Table 0-6 (Appendix E). For subset 2, MLP achieved better accuracy values compared to other models tested on this subset. These results will be further analysed using unseen test data from 2023 to assess their robustness.

The current best-performing combination for subset 2 is df_subset_2_no_Aspect, which achieved the highest recall value of 0.504278.

4.2.7. Evaluation summary

In summary, six models were trained and evaluated: three machine learning models (RF, XGBoost, and MLP) for each subset (1 and 2). Each model was tested on six distinct DataFrame combinations, resulting from two outlier treatments (with and without) and three Aspect configurations (No Aspect, NESW, and NS). Table 4-6 summarizes the best-performing DataFrame combination for each of the six models in terms of weighted recall:

Table 4-6 Summary of the best-performing DataFrame combinations for each model.

Model	Best DataFrame Combination	Weighted Recall	Weighted Accuracy
RF_subset_1	df_subset_1_no_outliers_NESW	0.361122	0.587648
RF_subset_2	df_subset_2_no_outliers_NESW	0.581467	0.584443
XGBoost_subset_1	df_subset_1_no_outliers_NESW	0.295829	0.595489
XGBoost_subset_2	df_subset_2_no_outliers_no_Aspect	0.457076	0.624022
MLP_subset_1	df_subset_1_no_outliers_NESW	0.308916	0.592715
MLP_subset_2	df_subset_2_no_Aspect	0.504278	0.631306

These results emphasize the importance of testing various DataFrame combinations. Notably, removing outliers and applying the NESW Aspect configuration consistently improved the performance of most models, particularly for subset 1. Further analysis and fine-tuning will confirm whether these combinations remain optimal across broader testing conditions.

4.3. IMPROVING MODELS

In this section, we fine-tune each of the six models (rf_1, rf_2, xgboost_1, xgboost_2, mlp_1, mlp_2) using grid search techniques. As detailed in Step 3 - Fine tuning each model for the selected combination: of 3.5.3 Data evaluation section, we conducted multiple grid searches to identify hyperparameter values that converge to optimal recall performance. To simplify the presentation of results, we focus on the total range of values tested for each hyperparameter and the final conclusions, rather than detailing every individual grid search iteration.

4.3.1. RF subset 1

The primary objective was to determine whether reducing the number of estimators (**n_estimators**) or adjusting the depth of the trees (**max_depth**) could improve recall. Given the significant impact of **max_depth**, we excluded extremely shallow depths (1-3) to avoid oversimplifying the model. Additionally, we tested modifications to **min_samples_split** and **min_samples_leaf** by evaluating both higher and lower thresholds. The impact of using 'sqrt' for **max_features** was also investigated. Beyond these hyperparameters, we evaluated different values for **bootstrap** (default = True) and **criterion** (default = 'gini').

The grid search evaluated the following hyperparameter values:

- **n_estimators**=[125, 150]
- **max_depth**=[4, 5, 6, 7, 8, 9, 10]
- **min_samples_split**=[0.0005, 0.001, 0.0015, 0.002, 0.004]
- **min_samples_leaf**=[0.0005, 0.001, 0.002]
- **max_features**=[None, 'sqrt']
- **bootstrap**=[True, False]
- **criterion**=['gini', 'entropy', 'log_loss']

Among the 2 520 grid search combinations, hyperparameter **max_depth** had the most significant impact on recall, underscoring its importance in controlling the model's granularity. A value of **max_depth**=4 emerged as the optimal choice. The **max_features** hyperparameter also proved influential, with the sqrt value consistently favouring recall. In contrast, **n_estimators** and **min_samples_leaf** showed minimal correlation with recall scores. As a result, we opted for lower values for these hyperparameters to enhance performance on larger datasets and conserve computational resources. The **min_samples_split** hyperparameter had a minor impact overall, but a value of 0.0015 provided a slight advantage. For the **criterion** hyperparameter, both 'entropy' and 'log_loss' values performed equally well, but 'entropy' was selected due to its faster computational efficiency. Lastly, enabling **bootstrap**=True improved recall compared to disabling it (False), further supporting its inclusion in the final model configuration.

The final RF Classifier for subset 1 was configured with the following hyperparameters:

- **n_estimators**=125
- **max_depth**=4

- **min_samples_split**=0.0015
- **min_samples_leaf**=0.0005
- **criterion**=‘entropy’
- **max_features**=‘sqrt’
- **bootstrap**=True

After finalizing these hyperparameters, we re-evaluated every DataFrame to ensure the optimal combination was selected under the updated hyperparameter settings. As shown in Table 0-7 (Appendix E), the DataFrame combination df_subset_1_no_Aspect emerged as the best-performing DataFrame for subset 1, achieving a recall of 0.487750.

The RF model, referred to as rf_1, was subsequently trained using the df_subset_1_no_Aspect DataFrame. This training did not apply the time series rationale and utilized the finalized hyperparameters.

We will now apply the same grid search approach to optimize the model for subset 2.

4.3.2. RF subset 2

Building on the approach for subset 1, we began by evaluating whether reducing the number of estimators (**n_estimators**) and decreasing the maximum depth **(max_depth**) could improve recall. We also tested varying values for **min_samples_split**, **min_samples_leaf**, and **max_features**, including exploring the impact of setting **max_features** to ‘sqrt’. Following what was done with subset 1, we also evaluated different values for **bootstrap** and **criterion**.

The hyperparameter grid for subset 2 included the following ranges:

- **n_estimators**=[50, 100, 150]
- **max_depth**=[4, 5, 6, 7, 8, 9, 10]
- **min_samples_split**=[0.001, 0.002, 0.004, 0.006, 0.008]
- **min_samples_leaf**=[0.0005, 0.001, 0.002]
- **max_features**=[None, ‘sqrt’]
- **criterion**=[‘gini’, ‘entropy’, ‘log_loss’]
- **bootstrap**=[True, False]

A total of 3 780 hyperparameter combinations were tested. Models with **max_depth**=4 achieved high recall but exhibited accuracy scores below 50%, indicating overly simplistic models that misclassified most points as burned areas. Consequently, these configurations were discarded. Among the remaining configurations, **max_depth**=5 emerged as the most effective setting. The **n_estimators** parameter showed minimal impact on recall, with 100 consistently yielding optimal results. Similarly, **min_samples_split** displayed little variation, though 0.008 provided a slight edge. **min_samples_leaf** showed no significant trends and was fixed at 0.0005 for better scalability with larger datasets. For **max_features**, the ‘sqrt’ setting consistently achieved the highest recall and was retained. Finally, among top-performing configurations, both ‘entropy’ and ‘log_loss’ values for **criterion** provided comparable recall scores. We selected ‘entropy’ for its computational efficiency and interpretability, which are critical for larger datasets.

The finalized hyperparameters for subset 2 were:

- **n_estimators**=100
- **max_depth**=5
- **min_samples_split**=0.008
- **min_samples_leaf**=0.0005
- **criterion**=‘entropy’
- **max_features**=‘sqrt’
- **bootstrap**=True

Using these hyperparameters, we reran each DataFrame combination to evaluate any changes in performance. As shown in Table 0-8 (Appendix E), the df_subset_2_NESW DataFrame emerged as the best-performing combination, achieving a recall of 0.768374. Using this DataFrame, the RF model (rf_2) was trained with the finalized hyperparameters, now leveraging the entire training dataset (2016-2022) instead of the time series rationale.

Next, we present the XGBoost results from the grid search.

4.3.3. XGBoost subset 1

To enhance performance, we explored how adjustments to key hyperparameters such as the number of estimators (**n_estimators**), tree depth (**max_depth**), and the number of features sampled at each split (**colsample_bytree**) influenced recall. We also examined the effects of varying the learning rate (**eta**), **min_child_weight**, and **subsample**. Lastly, we tested the **gamma** hyperparameter, which controls the model’s conservativeness by requiring a higher reduction in loss for a split to occur.

The hyperparameter grid was as follows:

- **n_estimators**=[50, 75, 100, 125]
- **eta**=[0.005, 0.01, 0.015, 0.1, 0.3]
- **gamma**=[0, 5]
- **max_depth**=[4, 5, 6, 7, 8, 9, 10]
- **min_child_weight**=[0.0005, 0.001, 0.002]
- **subsample**=[0.4, 0.6, 0.8]
- **colsample_bytree**=[0.4, 0.6, 0.8, 1]

XGBoost’s faster computational speed allowed us to test a larger number of hyperparameter combinations within the same timeframe compared to RF. The top-performing combinations did not exceed the accuracy threshold of 0.50, therefore they were disregarded. Among the tested values, a **max_depth** of 8 consistently emerged as the best choice, as shallower depths (e.g., 4, 5, 6, or 7) produced lower accuracies, and deeper trees offered no additional improvements in recall. For **n_estimators**, a value of 75 consistently yielded optimal results.

Reducing the learning rate (**eta**) to 0.005 proved most effective, balancing precision and generalization. Additionally, limiting the features considered at each split (**colsample_bytree**) to 60% demonstrated a clear advantage over using all features. Hyperparameters such as **gamma**, **subsample**, and **min_child_weight** showed minimal influence on recall across the tested ranges. Therefore, we fixed **gamma** at 5 and **subsample** at 0.8 for a slightly more conservative model, while **min_child_weight** was set at 0.0005 to maintain scalability for larger datasets.

The final model configuration that balanced recall and accuracy while meeting thresholds was as follows:

- **n_estimators**=75
- **eta**=0.005
- **gamma**=5
- **max_depth**=8
- **min_child_weight**=0.0005
- **subsample**=0.8
- **colsample_bytree**=0.6

To validate the robustness of the selected hyperparameters, we reran all DataFrame combinations to confirm whether df_subset_1_no_outliers_NESW remained the top-performing option. As shown in Table 0-9 (Appendix E), the final top-performing combination changed to df_subset_1_no_outliers_no_Aspect, achieving a recall of 0.851477. This result highlights that including Aspect information does not necessarily enhance recall, contrary to initial expectations. The XGBoost model was trained using the complete training dataset (2016-2022) and is referred to as xgboost_1.

We will now proceed with the grid search hyperparameter tuning for subset 2.

4.3.4. XGBoost subset 2

Using the same rationale as the previous grid searches and leveraging XGBoost's faster computational speed with the smaller size of subset 2, we tested a total of 62 208 hyperparameter combinations. For this subset 2, we also introduced the **tree_method** hyperparameter The hyperparameter grid included the following ranges:

- **n_estimators**=[50, 75, 100, 125]
- **eta**=[0.005, 0.01, 0.1, 0.3]
- **gamma**=[0, 2.5, 5]
- **max_depth**=[4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15]
- **min_child_weight**=[0.0005, 0.001, 0.002]
- **subsample**=[0.6, 0.8, 1]
- **colsample_bytree**=[0.4, 0.6, 0.8, 1]
- **tree_method**=['exact, 'approx, 'hist']

The top-ranked combinations - those with **n_estimators** set to 50 or 75 and **max_depth** between 4 and 12 - achieved perfect recall values (recall = 1). However, their accuracy fell below the 0.50 threshold, leading to their exclusion. Among the remaining models, consistent trends emerged: hyperparameter combinations with **n_estimators** set to 100, **eta** set to 0.005, **max_depth** set to 13, and **colsample_bytree** set to 0.6 proved dominant. Parameters such as **gamma**, **min_child_weight** and **subsample** demonstrated weak correlations with recall, so their values were fixed based on both general recommendations for XGBoost training and the highest recall averages observed during testing. Among the options for **tree_method**, the 'hist' method performed best, aligning with its status as the default setting in XGBoost.

The final XGBoost model for subset 2 was configured with the following hyperparameters:

- **n_estimators**=100
- **gamma**=5
- **eta**=0.005
- **max_depth**=13
- **min_child_weight**=0.0005
- **tree_method**='hist'
- **subsample**=1
- **colsample_bytree**=0.6

After rerunning all the DataFrame combinations for subset 2 with the newly fine-tuned hyperparameters, we obtained the results in Table 0-10 (Appendix E). As shown in Table 0-10, The DataFrame combination with the highest recall, df_subset_2_no_outliers_NESW, fell below the accuracy threshold. As a result, df_subset_2_no_outliers_no_Aspect was selected as the best-performing combination for subset 2. The XGBoost model, referred to as xgboost_2, was subsequently trained using the entire training dataset (2016-2022).

Next, we present the results of the grid search applied to the MLP model.

4.3.5. MLP subset 1

As MLP was the slowest model to train, the grid search involved the fewest hyperparameter combinations compared to the other models. The optimization focused on key hyperparameters such as **activation**, **solver**, and **learning_rate**, as these parameters had fixed interactions (e.g., certain solvers were compatible only with specific activation functions). The hidden layer configuration (**hidden_layer_sizes**) was initially set to (32, 16, 8) during training. Given the strong correlation between hidden layer size and model performance, we tested various shapes and sizes to identify potential improvements. Additionally, we examined the impact of the **learning_rate_init** hyperparameter, which had been set to the default value of 0.001, and evaluated the effect of the shuffle parameter (whether to shuffle samples at each iteration). To keep computational costs manageable, we fixed several other hyperparameters: **max_iter** at 200, **alpha** at 0.00001, and **tol** at 0.0001.

The grid search tested the following combinations:

- **activation**=[‘identity’, ‘logistic’, ‘tanh’, ‘relu’]
- **solver**=[‘lbfgs’, ‘sgd’, ‘adam’]
- **learning_rate**=[‘constant’, ‘invscaling’, ‘adaptive’]
- **hidden_layer_sizes**=[(64, 32, 16, 8), (64, 32, 16), (64, 32), (32, 16, 8), (32, 16), (16, 8), (8)]
- **learning_rate_init**=[0.0001, 0.0005, 0.001, 0.002, 0.005, 0.01]
- **shuffle**=[True, False]

Some combinations reached the iteration limit of 200 while the training loss was still improving. While increasing the maximum iterations might have allowed further refinement, the computational cost would have been excessive, so **max_iter** parameter remained fixed. Combinations with **activation**=‘logistic’, **solver**=‘sgd’, and **learning_rate**=‘invscaling’, performed poorly with accuracy below the 0.5 threshold. Similarly, all models with **shuffle**=False classified all areas as prone to burn, leading to their immediate exclusion. Among the remaining combinations, **activation**=‘identity’ consistently delivered the best performance, while **solver**=‘adam’ and **learning_rate**=‘constant’ achieved the highest recall values. Interestingly, the original hidden layer configuration of (32, 16, 8) remained the best-performing structure for hyperparameter **hidden_layer_sizes**, with no clear trends emerging from alternative configurations. For **learning_rate_init**, a value of 0.0005 produced the highest recall, though no consistent patterns were observed across the range of values tested.

The final tuned MLP model for subset 1 featured the following hyperparameters:

- **hidden_layer_sizes**=(32, 16, 8)
- **max_iter**=200
- **activation**=‘identity’
- **learning_rate**=‘constant’
- **solver**=‘adam’
- **learning_rate_init**=‘0.0005
- **alpha**=0.00001
- **tol**=0.0001
- **shuffle**=True

After rerunning the DataFrame combinations to evaluate potential changes in performance, the results are summarized in Table 0-11 (Appendix E). The top-performing DataFrame combination shifted from df_subset_1_no_outliers_NESW to df_subset_1_no_outliers_no_Aspect, achieving a recall of 0.496683. This indicates that excluding Aspect information improved recall while maintaining competitive accuracy. Using this top-performing DataFrame combination and the fine-tuned hyperparameters, the MLP model (referred to as mlp_1) was trained with the entire training dataset (2016-2022).

Next, we will conduct the same analysis for subset 2.

4.3.6. MLP subset 2

Building on the approach used for Subset 1, we aimed to test various combinations of **activation**, **solver**, **learning_rate**, and **hidden_layer_sizes** to identify the best-performing configuration. Additionally, we examined the effect of **learning_rate_init** while keeping other hyperparameters fixed at their default values: **max_iter**=200, **alpha**=0.00001, and **tol**=0.0001. The grid search tested the following parameter ranges:

- **activation**=[‘identity’, ‘logistic’, ‘tanh’, ‘relu’]
- **solver**=[‘lbfgs’, ‘sgd’, ‘adam’]
- **learning_rate**=[‘constant’, ‘invscaling’, ‘adaptive’]
- **hidden_layer_sizes**=[(64, 32, 16, 8), (64, 32, 16), (64, 32), (32, 16, 8), (32, 16), (16, 8), (8)]
- **learning_rate_init**=[0.0001, 0.0005, 0.001, 0.002, 0.005]

Results showed that combinations using **activation**=‘logistic’ with **solver**=‘sgd’ generally failed to meet the accuracy threshold of 0.5. However, one combination with a simple hidden layer structure (**hidden_layer_sizes**) of (8) achieved a balance between recall and accuracy. The choice of **learning_rate** continued to exhibit no clear trends with recall, similar to Subset 1 results, though **learning_rate**=‘invscaling’ provided the highest recall. Regarding **learning_rate_init**, the default value of 0.001 emerged as the most effective, as lower values such as 0.0005 failed to meet the accuracy threshold when paired with the identified hyperparameters.

The final tuned MLP model for Subset 2 was configured with the following hyperparameters:

- **hidden_layer_sizes**=(8)
- **max_iter**=200
- **activation**=‘logistic’
- **learning_rate**=‘invscaling’
- **solver**=‘sgd’
- **learning_rate_init**=0.001
- **alpha**=0.00001
- **shuffle**=True
- **random_state**=42

When re-running the DataFrame combinations with the optimized model, the results showed, as visible in Table 0-12 (Appendix E), that most combinations achieved near-perfect recall but fell below the accuracy threshold of 0.5, indicating overfitting to the burned class. Among all DataFrames, only the original top performer, df_subset_2_no_Aspect, met the accuracy threshold with a recall of 0.824925 and an accuracy of 0.519166. This DataFrame was, therefore, selected for training the MLP model (mlp_2) on the entire training dataset.

4.3.7. Evaluation summary

After optimizing the six models (rf_1, rf_2, xgboost_1, xgboost_2, mlp_1, mlp_2), key insights were drawn. Table 4-7 summarizes the selected models for testing on unseen data, along with their respective DataFrame combinations and recall values from validation:

Table 4-7 Summary of the best-performing DataFrames for each fine-tuned model.

Model	Best DataFrame Combination	Weighted Recall	Weighted Accuracy
RF subset 1	df_subset_1_no_Aspect	0.487750	0.577985
RF subset 2	df_subset_2_NESW	0.768374	0.522991
XGBoost subset 1	df_subset_1_no_outliers_no_Aspect	0.851477	0.500532
XGBoost subset 2	df_subset_2_no_outliers_no_Aspect	0.842785	0.500523
MLP subset 1	df_subset_1_no_outliers_no_Aspect	0.496683	0.587756
MLP subset 2	df_subset_2_no_Aspect	0.824925	0.519166

Some models are near the upper limit for accuracy, reflecting the impact of fine-tuning. These models excel in recall; however, only testing on unseen 2023 data will determine if they are overfitted to the training data. Achieving a model where recall is greater than zero and accuracy exceeds 0.66(6) has proven elusive. This challenge arises because models classifying all points as unburned (recall = 0) inherently achieve 66% accuracy due to the 1:2 ratio of burned to unburned points. However, models that surpass 0.66 accuracy would yield strong F1-scores and precision, making them valuable targets for future works.

For subset 1, a notable trend emerged: excluding Aspect information consistently improved recall across all models. Additionally, removing outliers proved beneficial for subset 1 in both the XGBoost and MLP models. Among the models, XGBoost stood out with recall values exceeding 0.85 in validation, significantly outperforming RF and MLP, which remained below 0.49. This performance may indicate that XGBoost is highly adept at predicting areas prone to burning—or it could suggest potential overfitting to the training data, despite our efforts to mitigate this risk using a time series rationale. The test data will ultimately confirm whether XGBoost's high recall is a strength or a limitation.

For subset 2, the results were more varied, with three different DataFrame combinations selected. Models performed better with outliers included, and, similar to subset 1, Aspect configurations did not enhance recall. RF had the lowest recall among the subset 2 models, while both XGBoost and MLP achieved higher values, emphasizing their ability to generalize better under these conditions.

These findings underscore the critical balance between selecting features, tuning hyperparameters, and ensuring that high recall does not come at the expense of generalizability, which will be further evaluated on unseen test data.

4.4.TEST DATA

This section focuses on evaluating the models using the test data, which comprises five distinct wildfire events, each representing unique conditions. These wildfires provide a diverse set of scenarios for model validation.

Among the five events (Figure 4-5), three fall within subset 1, where satellite imagery was captured within the two-day threshold defined in section 4.1.5 (‘DaysToFire’ ≤ 2 for subset 1 and ‘DaysToFire’ > 2 for subset 2) before the wildfire event (Figure 4-5 A, D and E). The subset 1 models will be tested on these events to determine which performs best. Notably, one of these events (Figure 4-5 A) occurred in the eastern region of the study area, while the remaining two (Figure 4-5 D and E) are in the western region.

For subset 2 (Figure 4-5 B and C), two wildfire events are included, both located in the western region. The three models associated with subset 2 will be tested to assess their predictive performance.

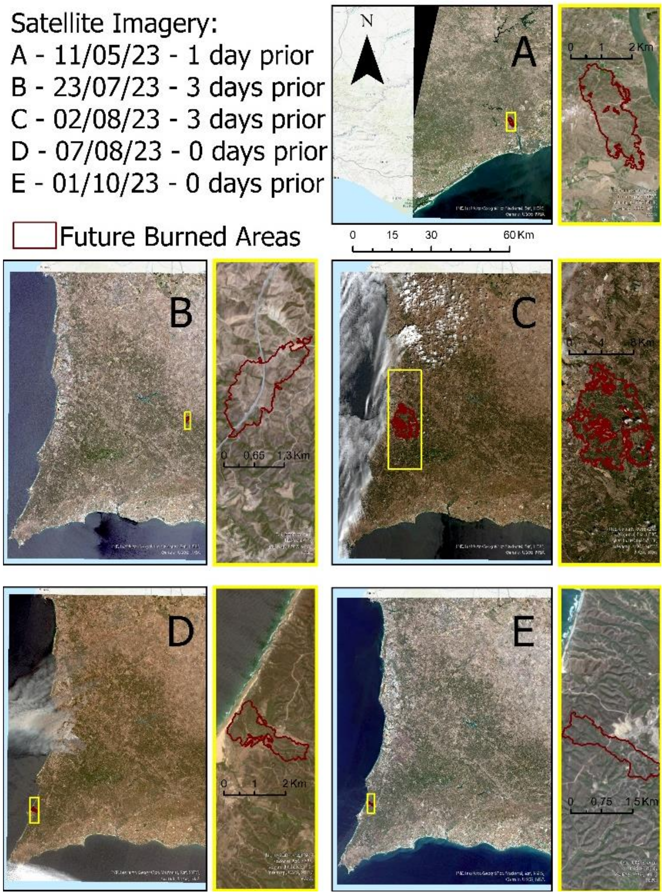


Figure 4-5 Satellite imagery of the five wildfire events used as test data. Each panel (A-E) represents a specific wildfire event, with red polygons marking the future burned areas. Panels A, D, and E correspond to subset 1 (‘DaysToFire’ ≤ 2), while panels B and C

correspond to subset 2 (‘DaysToFire’ > 2). The yellow boxes highlight zoomed-in areas, showing detailed spatial resolution (20 m) and the proximity of vegetation near burned areas.

Among these, the wildfire event on 08/05 (Figure 4-5 C) presents a unique challenge, as the closest available satellite image (from 08/02, three days prior) is partially obscured by clouds. This offers an opportunity to evaluate the models’ robustness in handling cloud coverage.

For the wildfire on 08/07 (Figure 4-5 D), the satellite image was taken on the same day, just hours before the fire started (0 days prior). This image includes large columns of smoke from the 08/05 wildfire. This scenario is particularly valuable as it mirrors some training conditions where smoke and burned areas from nearby wildfires were visible. Ideally, the models should classify recently burned areas as not prone to burn, showcasing their ability to distinguish between unburned and recently burned regions.

The remaining three wildfires—12/05/23 (Figure 4-5 A), 26/07/23 (Figure 4-5 B), and 01/10/23 (Figure 4-5 E)—have satellite imagery captured under clear skies, providing ideal conditions for testing model performance.

These images highlight the severe class imbalance inherent in the test dataset. Less than 0.1% of the entire study area actually burned, which means a naïve model classifying everything as “not burned” would achieve an accuracy of 99.90%. This demonstrates why focusing on improving accuracy is not appropriate for this problem. However, models that optimize solely for recall risk generating excessive false positives. While some false positives are expected, the goal is to achieve a reasonable balance between predicting areas prone to burn and accurately identifying burned areas, ensuring practical utility in wildfire management.

4.5.EVALUATING MODELS

This section presents the results for the three fine-tuned models - Random Forest (RF), eXtreme Gradient Boosting (XGBoost), and Multi-Layer Perceptron (MLP) - for each subset, as described in section 4.3. The objective is to identify the best model for each subset based on recall, using a prediction threshold of 0.4 for evaluation (as stated in Step 9 - Evaluate each model using land classification (COS): of section 3.5.4).

4.5.1. Subset 1

The best model for subset 1 was determined based on the aggregated results from three test wildfires, where satellite imagery was captured at most two days prior to the event (23/05/12, 23/08/07, and 23/10/01). Detailed results of each test wildfire can be found in Appendix F.

After compiling the metrics and results from the three subset 1 test wildfires, the aggregated outcomes are presented below in Table 4-8.

Table 4-8 Complete Test DataFrame for subset 1 with the total number of True Negatives (pixels unburned) and True Positives (pixels burned).

Wildfire Date	Total pixels	Pixels unburned (target = 0)	Pixels burned (target = 1)	Burned/Unburned Ratio (%)
2023/05/12	7 943 946	7 935 764	8 182	0.10310
2023/08/07	24 027 162	24 022 501	4 661	0.01940
2023/10/01	24 027 573	24 024 620	2 953	0.01229
Total	55 998 681	55 982 885	15 796	0.02822

In our training we kept a 33% of Burned/Unburned ratio. If we used the entire satellite imagery, and not just the buffer of 5km (and randomly removed the excess), we would have ended with ratios around 0.02% - just to show how great is the imbalance of this problem when using high spatial resolution.

In Table 4-9, we have the aggregated metrics:

Table 4-9 Aggregated performance metrics of each model (RF, XGBoost, MLP) for three wildfire events (12/05/2023, 07/08/2023 and 01/10/2023) within subset 1.

Model	TP	FP	TN	FN	Recall	Accuracy
RF	850	7 011 066	48 971 819	14 946	0.053811	0.874533
XGBoost	0	223 626	55 759 259	15 796	0	0.995725
MLP	1 297	7 127 495	48 855 390	14 499	0.082109	0.872461

If accuracy were the primary evaluation metric, XGBoost would appear to be the best-performing model, achieving near-perfect accuracy of 99.5725% due to its extremely conservative predictions. However, as recall is the chosen metric for identifying areas prone to burn, MLP emerges as the best model for subset 1, with a recall of 8.2109%.

Despite MLP's higher number of false positives and occasional misclassification of humid areas as prone to burn, it achieved the highest recall, detecting correctly more burned areas compared to RF and XGBoost. Additionally, MLP was the only model to assign predicted probabilities above 0.55, indicating greater confidence in its classifications. In contrast, RF's top probabilities hovered around 0.5, suggesting less certainty in its predictions.

The results suggest that while MLP offers the best recall, it also requires careful interpretation due to its tendency to overestimate burn-prone areas.

4.5.2. Subset 2

Similarly to subset 1, the best model for subset 2 was determined based on the aggregated results from two wildfires, where satellite imagery was captured at least three days prior to the wildfire event (23/07/26 and 23/08/05). Detailed results of each test wildfire can be found in Appendix F.

After compiling the metrics and results from the two test wildfires in subset 2, the aggregated outcomes are presented below in Table 4-10.

Table 4-10 Complete Test DataFrame for subset 2 with the total number of True Negatives (pixels unburned) and True Positives (pixels burned).

Wildfire Date	Total pixels	Pixels unburned (target = 0)	Pixels burned (target = 1)	Burned/Unburned Ratio (%)
2023/07/28	24 027 162	24 023 107	4 055	0.01688
2023/08/07	24 027 162	23 839 153	188 009	0.78866
Total	48 054 324	47 862 260	192 064	0.40129

In the following table (Table 4-11) we have the aggregated metrics:

Table 4-11 Aggregated performance metrics of each model (RF, XGBoost, MLP) for three wildfire events (12/05/2023, 07/08/2023 and 01/10/2023) within subset 1.

Model	TP	FP	TN	FN	Recall	Accuracy
RF	106 434	9 949 017	37 913 243	85 630	0.554159	0.791181
XGBoost	42 646	4 805 517	43 056 743	149 418	0.222041	0.896889
MLP	0	0	47 862 260	192 064	0	0.996003

RF correctly identified more than half of the burned areas, achieving a recall of 55.4159%. Each correctly predicted point corresponds to a pixel of 20 m-by-20 m, which equates to an area of 0.04 hectares (ha). Thus, RF’s true positives correspond to a total burned area of 4 257.36 hectares (~4 km²). However, RF also predicted a total of 397 960.68 hectares (~400 km²) as prone to burn. This larger area could reflect regions at actual risk due to vegetation and terrain conditions that, fortunately, did not burn. While we cannot verify whether these false positives were genuinely at risk, the results suggest RF's potential utility in creating dynamic wildfire risk maps.

XGBoost, by comparison, was more conservative, predicting fewer false positives but also identifying far fewer true positives, with a recall of only 22.2041%. This conservative approach might have its utility in scenarios where minimizing false positives is more critical than achieving higher recall.

When comparing subset 1's MLP model to subset 2's RF model, we observe that despite subset 2 having less training data, RF achieved a much higher recall (55.42%) compared to MLP's recall in subset 1 (8.2109%). However, determining which model is "better" is challenging because the subsets are fundamentally different. If we had to choose, we would express greater confidence in the results of subset 1, given the dataset's larger scope and MLP's better performance in the context of that subset.

Having drawn the results and conclusions for both subsets, it is now time to compare these findings with related studies discussed in section 2.3 Related works and evaluate whether our work has successfully addressed the identified research gaps.

4.6. COMPARISON WITH RELATED STUDIES

In this section, we compare our work and results with the studies discussed in section 2.3 Related works. While our primary objective was to predict areas prone to burn, our models also showed potential for creating dynamic fire risk maps. Both applications will be analysed in the context of the most relevant studies in the literature.

4.6.1. Fire Risk Maps

Although creating fire risk maps was not the primary objective of this study, the models demonstrated their potential for this application, as illustrated Figure 4-6. By leveraging Sentinel-2 satellite imagery with a revisit time of four days, dynamic fire risk maps were generated using the MLP (subset 1) and RF (subset 2) models. These two maps (Figure 4-6 B and A, respectively) could be further refined through cross-checking and filtering to reduce false positives and identify genuine high-risk areas. Such refinements could enable government agencies to allocate resources more effectively by focusing on smaller, precise hotspots.

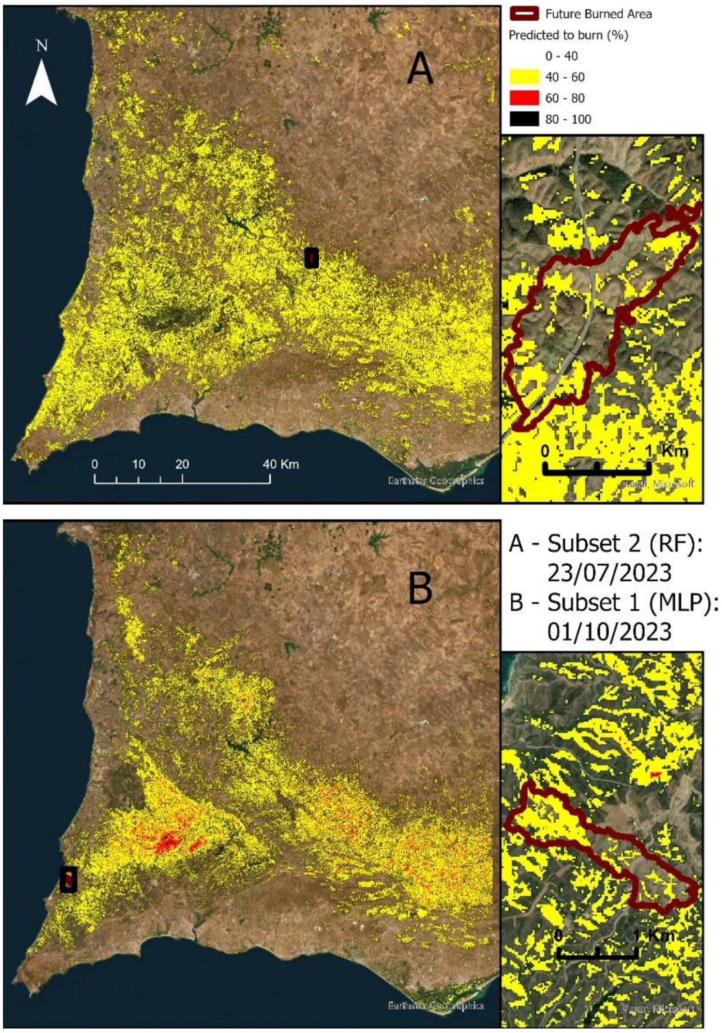


Figure 4-6 Fire Risk Maps generated by machine learning models. Top (A): RF model prediction for 23/07/2023 (three days before the wildfire event, hence subset 2). Bottom (B): MLP model prediction for 01/10/2023 (on the same day as the wildfire event, hence subset 1).

However, it is crucial to acknowledge that wildfires are inherently unpredictable events, with most being caused by human activities. This unpredictability limits the influence of environmental factors, such as vegetation and terrain, which were the primary inputs for our models. Despite these limitations, the models performed well compared to those in similar studies (Dimuccio et al., 2011; Ermitão et al., 2023; Sivrikayaq et al., 2024), even without incorporating weather data or land classification features during training.

4.6.2. Predicting Fire no-Fire

Our objective, shared by related studies, was to predict wildfires and mitigate their risks by identifying areas prone to burning. This involves analysing vegetation and environmental conditions, such as moisture retention, elevation, and temperature, to understand correlations that indicate fire risk. While methodologies and data sources varied across studies, the overarching goal remains consistent: enhancing wildfire prediction to reduce impacts.

Table 4-12 presents study area, data and other relevant topics used across the different studies:

Table 4-12 Comparison of data, methods and results between this work and related studies.

Topics	**Our Work**	**Study A** **(Kaur et al., 2023)**	**Study B** **(Ntinopoulos et al., 2023)**	**Study C** **(Cheruku et al., 2022)**	**Study D** **(Malik et al., 2021)**
Study Area	Algarve, Portugal	Alberta, Canada	Attica, Greece	Western US	San Diego, California
Data Sources	Fire Events Sentinel-2 Topographic Multispectral Indices	Fire Events Landsat Weather Topographical Temperature	Fire Events MODIS, Sentinel-2 Weather Thermal Multispectral Indices	Fire Events LFMC Land Vegetation Classification	Fire Events Landsat (GEE) Weather Multispectral Indices Land Vegetation Classification
Spatial Resolution	20 metres	10 kilometres	1 kilometres	8 kilometres	1 kilometres
ML Models	RF XGBoost MLP	RF XGBoost MLP	ANN RBF	KNN SVM RF CNN	Weather-based fire prediction model: SVM XGBoost RF Remote sensing-based fire prediction model: MLP CNN LSTM
Score	RF: 55.42% recall (subset 2); MLP: 8.21% recall (subset 1)	XGBoost achieved a ROC-AUC of 87% and 75% Recall	RBF achieved a TPR-based accuracy of 83.33% and ANN of 66.67%	RF achieved the highest accuracy of 75.07%	Ensemble model (weather-based + RS-based) with 100% accuracy

Data Sources:

As highlighted throughout this thesis, we used fire events data from ICNF, spectral data from Sentinel-2, which also propelled to the usage of multispectral indices, and terrain data from ESRI Portugal. While all studies reviewed used fire data—sourced from government authorities or public databases—the choice of additional data sources varied significantly, highlighting key differences in scope and methodology.

A key distinction between our work and others is the use (or lack thereof) of weather data. As explained in section 3.2.5 Weather data, we could not find weather data in our study area with spatial resolution comparable to the rest of our data (20 m). As a result, we did not include it in our study. Conversely, other works that used much lower spatial resolutions (e.g., 1 km, 10 km) were able to incorporate weather data into their models. However, since our buffer zone had a radius of 5 km, those weather values would have represented a single cell, making it impossible for weather and climate features to distinguish between unburned and burned land. In contrast, Study A used meteorological data from both the Alberta Climate Information Service and Copernicus ERA5-Land, both with a resolution of 10 km. Study B included the use of MODIS satellite data instead of Sentinel-2, resulting in a temporal resolution of 1 day (as opposed to 4-5 days). While Study B used temperature data with a 1 km spatial resolution, which might have been acceptable for our study (although not ideal), their climatic variables (e.g., 38 km resolution) would not have been suitable. We could not find the spatial or temporal resolution of the National Center for Environmental Information (NCEI) weather data used in Study D; however, since it is from governmental entities, it was not available in our study area. Study C did not directly use any satellite or spectral data; however, Live Fuel Moisture Content (LFMC) - the mass of water per unit dry biomass in vegetation (Rao et al., 2020) - maps are built using Sentinel-1 and Landsat 8 data, with a temporal resolution of 15 days and a spatial resolution of 250 m.

Another notable difference lies in the use of land classification data. While we previously explained why we did not include it, it is worth reiterating that if COS datasets were updated annually and could maintain a spatial resolution comparable to 20 m, we would have incorporated them into our analysis, as Studies C and D did.

Methodology:

Our methodology aimed to predict wildfire-prone areas using single satellite imagery acquired days—or, when feasible, hours—before wildfire events. The term ‘when feasible’ is emphasized because some satellite imagery was affected by cloud coverage. Study D employed rigorous preprocessing techniques on remote sensing data, excluding images with cloud cover exceeding 10% and applying atmospheric corrections to the Red and NIR bands. In contrast, our methodology opted to utilize older satellite imagery with clear skies over the burned area or buffer zone, should the most temporally proximate imagery contain clouds. This approach was taken even when such imagery was captured 7-8 days before the wildfire event rather than 3-4 days prior. Atmospheric corrections were not applied in this study.

Data were collected for 35 wildfire events, with a 5 km buffer zone delineating unburned areas. Study A addressed class imbalance by restricting data within provincial boundaries (excluding areas with negligible fire risk) and employing techniques such as Near Miss. Similarly, our methodology mitigated class imbalance by randomly removing unburned points, maintaining a 1:2 ratio between burned and unburned classes. Study B approached imbalance by generating 100 random points within its study area, categorizing each as ‘fire’ or ‘no fire’ based on its spatial relationship to wildfire events.

Our dataset was stratified into two subsets based on the temporal proximity of satellite imagery to the wildfire event: subset 1 comprised imagery captured closer to the event, while subset 2 utilized more temporally distant imagery. Study C implemented a comparable stratification, conducting two distinct tests: one with data closest to the wildfire event and another incorporating data from three temporal intervals (closest, one period prior, and two periods prior). For each subset, we evaluated six feature configurations (DataFrame combinations) using Random Forest (RF), XGBoost, and Multi-Layer Perceptron (MLP) models to identify the optimal approach. All reviewed studies (A-D) employed at least one of these models, with Studies B, C, and D incorporating more advanced deep learning architectures compared to the MLP utilized in our analysis.

A time-series split was implemented for training, ensuring model generalization across both temporal and spatial domains. Given the high imbalance of the dataset, the evaluation prioritized recall over accuracy to minimize false negatives. This emphasis contrasts with related works, which frequently prioritize accuracy, but aligns more closely with our objective of identifying wildfire-prone areas. Related studies employed divergent validation and testing strategies. Study A utilized stratified 10-fold cross-validation, while Study B partitioned its dataset into 10 batches of 10 points each (with the first nine batches used for training and optimizing weights and biases for the ANN, and the last batch for testing). For the RBF network, Study B used 50 points split into five 10-point batches (four for training and one for testing). Study D applied an 80:20 train-test split alongside a 4:1 training-validation ratio. The absence of a time-series rationale in predicting wildfire-prone areas, as elaborated in Section 3.4.3 (Data Evaluation, Step 1), represents a significant limitation in these methodologies.

Our analysis utilized a high spatial resolution of 20 m-by-20 m, offering significantly greater detail than Study A’s 10 km-by-10 km resolution and Study C’s 8 km-by-8 km resolution. Even the 1 km-by-1 km resolution used in Studies B and D does not compare, as a single 20 m-by-20 m pixel represents an area 2 500 times smaller than 1 km^2 and 250 000 times smaller than 10 km^2. This granularity enabled even small wildfires (~200 hectares) to be represented by approximately 5 000 pixels (0.04 hectares per pixel), facilitating precise differentiation between burned and unburned areas.

Results:

Our analysis identified Sentinel-2 spectral bands ‘B04,’ ‘B06,’ and ‘B12,’ and the multispectral indices ‘NDWI’ and ‘MSAVI2’ as having the highest correlation with the ‘Burned’ target, making them strong predictors of fire susceptibility. Creating a new index was not our focus, unlike Study B, which achieved accuracy rates of 83.33% and 66.67% with RBF and ANN models, respectively, leveraging a newly introduced index, FWIveg, which combines the Fire Weather Index and the Vegetation Health Index.

Among our models tested, MLP achieved the highest recall for subset 1 (8.21%) but occasionally misclassified wetlands as prone to burn. Compared to Study A, their data and methodology seem to be more effective for wildfire predictions than ours, achieving a ROC-AUC of 87% and a recall of 75% using XGBoost. However, it is important to note that our study utilized a 20 m resolution, offering 250 000 times greater detail than their 10 km resolution. Additionally, Study A employed a stratified train-test split, which, while effective in their context, may not generalize well to unseen data from a year excluded from training. Nonetheless, techniques like the Near Miss 3 undersampling method, which proved effective in Study A for addressing data imbalance, could potentially enhance our approach in future iterations.

For subset 2, RF achieved the highest recall (55.42%), but its fire probability predictions did not exceed 60%, marking it as a conservative model. Similarly, Study C implemented a two-subset approach, where one subset incorporated data from a single temporal point close to the wildfire, while the other included data from three time intervals: just before the fire and two prior periods. This approach led to improved predictions, with recall increasing from 82% to 84% when temporal information from multiple intervals was included. We could have implemented something similar by combining subsets 1 and 2 and adding more satellite imagery data spaced temporally away from the wildfire events. However, such an approach would require tripling the data, posing substantial computational challenges given the current limitations of our hardware.

Our models were evaluated on wildfire events from 2023—unseen data from both a temporal and spatial perspective. In this regard, Study D employed an innovative ensemble model. They combined separate models for fire and weather data (SVM, XGBoost, and RF) with models for fire and remote sensing data (MLP, CNN, and LSTM) using a majority voting approach, also known as cellular automata. This strategy resulted in an impressive 100% accuracy and recall for their predictions. The success of their methodology was likely due to the inclusion of atmospheric corrections, weather data, land classification data, and the use of more advanced and complex models. However, this achievement was based on a 1 km-by-1 km grid resolution. While higher-resolution data might reduce accuracy and recall, we believe that part of their perfect accuracy stems from their data-splitting approach instead of a more realistic time-series split approach, as mentioned above. Although it is far from perfect, it prioritizes adaptability to predict fire-prone areas effectively under changing conditions.

Our models achieving reasonable recall under unseen temporal and spatial conditions underscores the models' ability to generalize effectively. Moreover, the models avoided overgeneralization by not classifying all vegetation as prone to burn and accurately identifying susceptibility for specific classes. This highlights their robustness in handling complex and highly imbalanced data.

In the next section, we will discuss the limitations of our approach and explore potential improvements for future work.

4.7. LIMITATIONS

As discussed earlier when comparing our work with related studies, several limitations have been identified. Additionally, we acknowledge that the wildfire crisis is a highly complex issue that cannot be resolved with a single approach or study.

One limitation of this study was the use of a 1:2 ratio for unburned areas. While this approach effectively helped the model focus on learning the characteristics of the “Burned” class (target=1), the random sampling method employed to achieve this ratio may not have been optimal. Although random sampling is theoretically unbiased, it could inadvertently introduce bias if certain areas were sampled more densely than others.

Another constraint was the absence of atmospheric corrections in our preprocessing pipeline. Although we relied on a broad set of satellite images (27 in total spanning from 2016 to 2022) to help the models learn across diverse conditions, atmospheric corrections could potentially enhance results, particularly for images taken under less favourable lighting conditions. We assumed that Sentinel-2, being a sensor rather than a camera, minimized the need for such adjustments, but incorporating corrections might further improve the model’s performance.

The resolution of fire polygons provided by ICNF was another challenge. These polygons, serving as ground truth data, often lacked accuracy in older datasets, with broader margins around wildfires. While 2022 and 2023 datasets showed significant improvements, earlier data introduced potential bias and inaccuracies in training. Future advancements in burned area detection could result in more precise polygon data, enhancing model reliability.

The time series training approach effectively promoted model generalization; however it had limitations. Each time-bucket retraining session began from scratch, preventing the models from leveraging knowledge gained in prior time-buckets training. Implementing recursive learning with a warm-start mechanism could address this limitation by allowing models to retain and build upon previously learned knowledge, reducing computational costs and improving efficiency.

Finally, while we used a weighted average performance across three temporal buckets to fine-tune models before training on the full dataset, there is no definitive metric to confirm whether these final models generalize effectively. Hyperparameters optimized for the weighted average might not perform as effectively when applied to the entire dataset. Fine-tuning hyperparameters using test data introduces the risk of overfitting to specific years, potentially undermining generalization. Recursive learning and warm-start techniques, as mentioned earlier, could mitigate this challenge, enabling models to generalize more effectively across diverse conditions without sacrificing predictive accuracy.

5. CONCLUSIONS AND FUTURE WORKS

This study faced significant challenges due to the inherent complexity of predicting wildfire-prone areas, a task requiring the integration of diverse and detailed data sources. While we could not address every challenge or avoid the temptation to over-engineer aspects of the methodology, we believe this work offers valuable insights and lays the groundwork for future research. We are optimistic that this study, along with related works, will contribute to better strategies for combating the wildfire crisis. Future works could explore the following:

Expanding the dataset is crucial for improving model performance. Incorporating data from before 2015 would require using alternative satellite imagery, such as Landsat, as Sentinel-2 was launched only in 2015. Additionally, extending the study to other regions, particularly northern Portugal, where the wildfire crisis is even more severe, could provide new insights.

Advanced Techniques: Initially, we aimed to apply advanced computer vision techniques to imagery data but opted for tabular data to simplify integration with additional variables, such as wind and humidity. With modern advancements in Transformers, future studies could directly process multiple data sources in binary or channel formats. This approach could reduce preprocessing requirements and improve predictive accuracy.

Weather and Environmental Data: Incorporating weather data could enhance predictions by integrating atmospheric and climatic variables. Ideally, through IoT equipment, such as localized weather sensors, deployed in forested areas and their surroundings, it could be possible to leverage weather data granular sufficient for 20 m resolution. Additionally, Live Fuel Moisture Content (LFMC) maps, though unavailable for our study area, hold great potential with their 250 m resolution. With further research and preprocessing, LFMC maps could be adapted to Portugal, enhancing the model’s ability to predict wildfire risks.

Our methodology demonstrated the potential of using 20-metre spatial resolution for fire risk prediction but fell short of effectively identifying wildfire occurrences due to low recall values and high false positives. While dynamic fire risk maps could be generated using this approach, the results highlight the need for further refinement to achieve reliable predictions. The results confirmed that the models could differentiate between land classes based on spectral bands, multispectral indices, and terrain variables. However, limitations in accuracy, particularly with wetlands misclassified as prone to burning, suggest areas for improvement. Sentinel-2 spectral bands B04, B06, and B12, along with the multispectral indices NDWI and MSAVI2, emerged as the strongest predictors of fire susceptibility. We hope these findings will be leveraged by other researchers and government entities to advance wildfire management strategies.

While this study did not fully achieve its objective of reliably predicting wildfire occurrences, it offers critical insights into feature importance, methodology, and areas for improvement. The research highlights the challenges of balancing high-resolution data with computational limitations and emphasizes the need for adaptable models capable of handling complex and dynamic conditions. We hope this work inspires further research and contributes to the collective effort to address the wildfire crisis in Portugal and beyond.

BIBLIOGRAPHICAL REFERENCES

Adhikari, D., Chen, W., Guo, Y., Huang, L., & Gao, J. (2023). Wildfire Progression Prediction and Validation Using Satellite Data and Remote Sensing in Sonoma, California. *2023 IEEE International Conference on Service-Oriented System Engineering (SOSE)*, 262-271. https://doi.org/10.1109/SOSE58276.2023.00037

Afonso, R., Neves, A., Viegas Damásio, C., Moura Pires, J., Birra, F., & Santos, M. Y. (2020). Assessment of Interventions in Fuel Management Zones Using Remote Sensing. *ISPRS International Journal of Geo-Information*, *9*(9), 533. https://doi.org/10.3390/ijgi9090533

ArcGIS. (2023, November 11). Digital Terrain Model Portugal (ETRS89). https://www.arcgis.com/home/item.html?id=f80bcdbe65a048a78115f8aeee9091eb

Avetisyan, D., Stankova, N., & Dimitrov, Z. (2023). Assessment of Spectral Vegetation Indices Performance for Post-Fire Monitoring of Different Forest Environments. *Fire*, *6*(8), 290. https://doi.org/10.3390/fire6080290

Bee2Fire. (2024, October 29). Bee2Fire Detection. http://www.bee2firedetection.com/

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

Castrejon, D. J., Wang, C., Osmak, D., Kukadiya, B., Liu, L., Giraldo, M., & Jiang, X. (2023). Machine Learning-based California Wildfire Risk Prediction and Visualization. *2023 International Conference on Machine Learning and Applications (ICMLA)*, 1212-1217. https://doi.org/10.1109/ICMLA58977.2023.00182

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794. https://doi.org/10.1145/2939672.2939785

Cheruku, R., Kohli, A., Kodali, P., Kavati, I., & E, S. (2022). Predicting Wildfire using Live Fuel Moisture Content with Machine Learning. *2022 IEEE 19th India Council*

International Conference (INDICON), 1-6. https://doi.org/10.1109/INDICON56171.2022.10039960

Copernicus. (2023, December 20). S2 Mission. https://sentiwiki.copernicus.eu/web/s2-mission

Copernicus Browser. (2024, May 5). Copernicus Browser. https://browser.dataspace.copernicus.eu/

Costa, L., Badeck, F., & Kirsten, T. (2007). *Sensitivity of Portuguese Forest Fires to Land Use and Climatic Factors.pdf*. https://www.researchgate.net/publication/228519614_Sensitivity_of_Portuguese_Forest_Fires_to_Land_Use_and_Climatic_Factors/references

Cunha, L. (2021). (Physical) Geography and Environmental Issues in Portugal. In R. C. Lois-González (Ed.), *Geographies of Mediterranean Europe* (pp. 105-120). Springer International Publishing. https://doi.org/10.1007/978-3-030-49464-3_5

Dados.gov. (2023a, November 14). Carta de Uso e Ocupação Do Solo - 2018. https://dados.gov.pt/pt/datasets/carta-de-uso-e-ocupacao-do-solo-2018/

Dados.gov. (2023b, November 14). Carta Administrativa Oficial de Portugal - CAOP2022 (Continente). https://dados.gov.pt/pt/datasets/carta-administrativa-oficial-de-portugal-caop2022-continente-1/

Das, A. C., Shahriar, S. A., Chowdhury, M. A., Hossain, M. L., Mahmud, S., Tusar, M. K., Ahmed, R., & Salam, M. A. (2023). Assessment of remote sensing-based indices for drought monitoring in the north-western region of Bangladesh. *Heliyon*, *9*(2), e13016. https://doi.org/10.1016/j.heliyon.2023.e13016

de Castro, C. F., Serra, G., Parola, J., & Reis, J. (2006). *Combate a incêndios florestais (vol. XIII)*. https://www.enb.pt/admin/docs/repositorio/Combate%20a%20Inc%C3%AAndios%20Florestais.pdf

De Luca, G., Silva, J. M. N., & Modica, G. (2021). A workflow based on Sentinel-1 SAR data and open-source algorithms for unsupervised burned area detection in Mediterranean ecosystems. *GIScience & Remote Sensing*, *58*(4), 516-541. https://doi.org/10.1080/15481603.2021.1907896

De Luca, G., Silva, J., Oom, D., & Modica, G. (2021). Combined Use of Sentinel-1 and Sentinel-2 for Burn Severity Mapping in a Mediterranean Region. In O. Gervasi & B. Murgante (Eds.), *Computational Science and Its Applications - ICCSA 2021: 21st International Conference, Cagliari, Italy, September 13-16, 2021, Proceedings, Part VII* (Vol. 12955, pp. 139-154). Springer International Publishing. https://link.springer.com/10.1007/978-3-030-87007-2

dgTerritório. (2024, February 2). PT-TM06/ETRS89. https://www.dgterritorio.gov.pt/geodesia/sistemas-referencia/portugal-continental/PT-TM06-ETRS89

Dimuccio, L. A., Ferreira, R., Cunha, L., & Campar De Almeida, A. (2011). Regional forest-fire susceptibility analysis in central Portugal using a probabilistic ratings procedure and artificial neural network weights assignment. *International Journal of Wildland Fire*, *20*(6), 776. https://doi.org/10.1071/WF09083

dmlc, xgboost. (2021). *Dmlc*. GitHub. https://github.com/dmlc/xgboost/tree/master

Dos Santos, S. M. B., Duverger, S. G., Bento-Gonçalves, A., Franca-Rocha, W., Vieira, A., & Teixeira, G. (2023). Remote Sensing Applications for Mapping Large Wildfires Based on Machine Learning and Time Series in Northwestern Portugal. *Fire*, *6*(2), 43. https://doi.org/10.3390/fire6020043

Ermitão, T., Páscoa, P., Trigo, I., Alonso, C., & Gouveia, C. (2023). Mapping the Most Susceptible Regions to Fire in Portugal. *Fire*, *6*(7), 254. https://doi.org/10.3390/fire6070254

ESA. (2023, November 11). Sentinel 2 Applications. https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-2

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5). https://doi.org/10.1214/aos/1013203451

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, *36*(4), 193-202. https://doi.org/10.1007/BF00344251

Gabinete de Estratégia e Estudos da República Portuguesa, E. do D. de B. (2023). *Gabinete de Estratégia e Estudos da República Portuguesa—Economia—Distrito—Beja.pdf*. Ministério da Economia e do Mar. https://www.gee.gov.pt/pt/docs/doc-o-gee-2/estatisticas-regionais/distritos-concelhos/beja/2978-distrito-de-beja/file

Gabinete de Estratégia e Estudos da República Portuguesa, E. do D. de F. (2023). *Gabinete de Estratégia e Estudos da República Portuguesa—Economia—Distrito—Faro.pdf*. Ministério da Economia e do Mar. https://www.gee.gov.pt/pt/docs/doc-o-gee-2/estatisticas-regionais/distritos-concelhos/faro/3093-distrito-de-faro/file

Gómez, I., & Martín, M. P. (2011). Prototyping an artificial neural network for burned area mapping on a regional scale in Mediterranean areas using MODIS images. *International Journal of Applied Earth Observation and Geoinformation*, *13*(5), 741-752. https://doi.org/10.1016/j.jag.2011.05.002

Guede-Fernández, F., Martins, L., De Almeida, R. V., Gamboa, H., & Vieira, P. (2021). A Deep Learning Based Object Identification System for Forest Fire Detection. *Fire*, *4*(4), 75. https://doi.org/10.3390/fire4040075

Guria, R., Mishra, M., Silva, R. M. D., Mishra, M., & Santos, C. A. G. (2024). Predicting forest fire probability in Similipal Biosphere Reserve (India) using Sentinel-2 MSI data and machine learning. *Remote Sensing Applications: Society and Environment*, *36*, 101311. https://doi.org/10.1016/j.rsase.2024.101311

ICNF. (2024, January 31). [Territórios Ardidos]. https://sig.icnf.pt/portal/home/item.html?id=983c4e6c4d5b4666b258a3ad5f3ea5af

INE. (2024, September 29). Instituto Nacional de Estatística. https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_cont_inst&INST=6251038&xlang=pt

Kaur, P., Naik, K., Purcell, R., Sampalli, S., Lung, C.-H., Zaman, M., & Mutakabbir, A. (2023). A Data Integration Framework with Multi-Source Big Data for Enhanced Forest Fire Prediction. *2023 IEEE International Conference on Big Data (BigData)*, 344-351. https://doi.org/10.1109/BigData59044.2023.10386816

Khan, A. A., Chaudhari, O., & Chandra, R. (2024). A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation. *Expert Systems with Applications*, *244*, 122778. https://doi.org/10.1016/j.eswa.2023.122778

Lavado, A. (2024, December 2). *Wildfire ML Prediction PT*. GitHub. https://github.com/afonsoblavado/wildfire-ml-pred-pt

Lazaj, L. (2016). *Burnt areas detection and mapping using vegetation indices*.

Leitek. (2024, October 29). Leitek. https://www.leitek.co/

Liu, S., Zheng, Y., Dalponte, M., & Tong, X. (2020). A novel fire index-based burned area change detection approach using Landsat-8 OLI data. *European Journal of Remote Sensing*, *53*(1), 104-112. https://doi.org/10.1080/22797254.2020.1738900

Liu, Z., Yang, J., Chang, Y., Weisberg, P. J., & He, H. S. (2012). Spatial patterns and drivers of fire occurrence and its future trend under climate change in a boreal forest of

Northeast China. *Global Change Biology*, *18*(6), 2041-2056. https://doi.org/10.1111/j.1365-2486.2012.02649.x

Lourenço, M., Estima, D., Oliveira, H., Oliveira, L., & Mora, A. (2023). Automatic Rural Road Centerline Detection and Extraction from Aerial Images for a Forest Fire Decision Support System. *Remote Sensing*, *15*(1), 271. https://doi.org/10.3390/rs15010271

Lyon, J. G. (1998). *A Change Detection Experiment Using Vegetation Indices*.

Malik, A., Jalin, N., Rani, S., Singhal, P., Jain, S., & Gao, J. (2021). Wildfire Risk Prediction and Detection using Machine Learning in San Diego, California. *2021 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/IOP/SCI)*, 622-629. https://doi.org/10.1109/SWC50871.2021.00092

Martins, L., Guede-Fernández, F., Valente De Almeida, R., Gamboa, H., & Vieira, P. (2022). Real-Time Integration of Segmentation Techniques for Reduction of False Positive Rates in Fire Plume Detection Systems during Forest Fires. *Remote Sensing*, *14*(11), 2701. https://doi.org/10.3390/rs14112701

Mcculloch, W. S., & Pitts, W. (1943). *A LOGICAL CALCULUS OF THE IDEAS IMMANENT IN NERVOUS ACTIVITY*. https://doi.org/10.1007/BF02478259

Meira Castro, A. C., Nunes, A., Sousa, A., & Lourenço, L. (2020). Mapping the Causes of Forest Fires in Portugal by Clustering Analysis. *Geosciences*, *10*(2), 53. https://doi.org/10.3390/geosciences10020053

Mhawej, M., Faour, G., & Adjizian-Gerard, J. (2015). Wildfire Likelihood's Elements: A Literature Review. *Challenges*, *6*(2), 282-293. https://doi.org/10.3390/challe6020282

Ministério Público Portugal. (2024, January 12). Ministério Público Portugal. https://www.pgdlisboa.pt/leis/lei_mostra_articulado.php?nid=2542&tabela=leis

Minsky, M., & Papert, S. A. (1969). *Perceptrons: An Introduction to Computational Geometry*.

Mira, M., & Lourenço, L. (2013). Os incêndios florestais em Portugal têm solução. In *Grandes Incêndios Florestais, Erosão, Degradação e Medidas de Recuperação dos Solos*. https://www.researchgate.net/publication/298302652_OS_INCENDIOS_FLORESTAIS_EM_PORTUGAL_TEM_SOLUCAO/citations

Moreira, C. O. (2018). Portugal as a tourism destination: Paths and trends. *Méditerranée*, *130*. https://doi.org/10.4000/mediterranee.10402

NASA. (2023, November 12). Landsat 8. https://landsat.gsfc.nasa.gov/satellites/landsat-8/

Novo, A., Fernández, C., Míguez, C., & Suárez-Vidal, E. (2024). Analysing the capacity of multispectral indices to map the spatial distribution of potential post-fire soil losses based on soil burn severity. *Ecological Informatics*, *83*, 102793. https://doi.org/10.1016/j.ecoinf.2024.102793

Ntinopoulos, N., Sakellariou, S., Christopoulou, O., & Sfougaris, A. (2023). Fusion of Remotely-Sensed Fire-Related Indices for Wildfire Prediction through the Contribution of Artificial Intelligence. *Sustainability*, *15*(15), 11527. https://doi.org/10.3390/su151511527

Nunes, A. N., Lourenço, L., & Meira, A. C. C. (2016). Exploring spatial patterns and drivers of forest fires in Portugal (1980-2014). *Science of The Total Environment*, *573*, 1190-1202. https://doi.org/10.1016/j.scitotenv.2016.03.121

Oliveira, A., Matos-Carvalho, J. P., Moutinho, F., & Fachada, N. (2023). *Multispectral Indices for Wildfire Management* (arXiv:2309.01751). arXiv. http://arxiv.org/abs/2309.01751

Pacheco, A. D. P., Da Silva Junior, J. A., Ruiz-Armenteros, A. M., Henriques, R. F. F., & De Oliveira Santos, I. (2023). Analysis of Spectral Separability for Detecting Burned Areas Using Landsat-8 OLI/TIRS Images under Different Biomes in Brazil and Portugal. *Forests*, *14*(4), 663. https://doi.org/10.3390/f14040663

Pereira-Pires, J. E., Aubard, V., Baldassarre, G., Fonseca, J. M., Silva, J. M. N., & Mora, A. (2022). Fuel Break Monitoring with Sentinel-2 Imagery and GEDI Validation. In L. M. Camarinha-Matos, G. Heijenk, S. Katkoori, & L. Strous (Eds.), *Internet of Things. Technology and Applications: 4th IFIP International Cross-Domain Conference, IFIPIoT 2021, Virtual Event, November 4-5, 2021, Revised Selected Papers* (Vol. 641, pp. 67-85). Springer International Publishing. https://doi.org/10.1007/978-3-030-96466-5

Pereira-Pires, J. E., Aubard, V., Ribeiro, R. A., Fonseca, J. M., Silva, J. M. N., & Mora, A. (2020). Semi-Automatic Methodology for Fire Break Maintenance Operations Detection with Sentinel-2 Imagery and Artificial Neural Network. *Remote Sensing, 12*(6), 909. https://doi.org/10.3390/rs12060909

Pereira-Pires, J. E., Aubard, V., Silva, J. M. N., Ribeiro, R. A., Pereira, J. M. C., Fonseca, J. M., Campagnolo, M. L., & Mora, A. (2020). Pixel-based and object-based change detection methods for assessing fuel break maintenance. *2020 International Young Engineers Forum (YEF-ECE)*, 49-54. https://doi.org/10.1109/YEF-ECE49388.2020.9171818

Popescu, M.-C., Balas, V. E., Perescu-Popescu, L., & Mastorakis, N. (2009). *Multilayer Perceptron and Neural Networks*. *8*(7).

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning, 1*(1), 81-106. https://doi.org/10.1007/BF00116251

Rao, K., Williams, A. P., Flefil, J. F., & Konings, A. G. (2020). SAR-enhanced mapping of live fuel moisture content. *Remote Sensing of Environment, 245*, 111797. https://doi.org/10.1016/j.rse.2020.111797

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*(6), 386-408. https://doi.org/10.1037/h0042519

Rostami, A., Shah-Hosseini, R., Asgari, S., Zarei, A., Aghdami-Nia, M., & Homayouni, S. (2022). Active Fire Detection from Landsat-8 Imagery Using Deep Multiple Kernel Learning. *Remote Sensing*, *14*(4), Article 4. https://doi.org/10.3390/rs14040992

Rouse, W., & Haas, R. H. (1973). *MONITORING VEGETATION SYSTEMS IN THE GREAT PLAINS WITH ERTS*.

Santos, F. L. M., Couto, F. T., Dias, S. S., Ribeiro, N. D. A., & Salgado, R. (2023). Vegetation fuel characterization using machine learning approach over southern Portugal. *Remote Sensing Applications: Society and Environment*, *32*, 101017. https://doi.org/10.1016/j.rsase.2023.101017

Scikit learn. (2023a, December 17). MLPClassifier. https://scikit-learn.org/dev/modules/generated/sklearn.neural_network.MLPClassifier.html

Scikit learn. (2023b, December 20). RandomForestClassifier. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

Sivrikaya, F., Günlü, A., Küçük, Ö., & Ürker, O. (2024). Forest fire risk mapping with Landsat 8 OLI images: Evaluation of the potential use of vegetation indices. *Ecological Informatics*, *79*, 102461. https://doi.org/10.1016/j.ecoinf.2024.102461

Statista. (2023, October 11). Statista. https://www.statista.com/statistics/1260777/area-burned-by-wildfire-in-european-countries/

Sun, J., Qi, W., Huang, Y., Xu, C., & Yang, W. (2023). Facing the Wildfire Spread Risk Challenge: Where Are We Now and Where Are We Going? *Fire*, *6*(6), 228. https://doi.org/10.3390/fire6060228

Sun, Y., Jiang, L., Pan, J., Sheng, S., & Hao, L. (2023). A satellite imagery smoke detection framework based on the Mahalanobis distance for early fire identification and positioning. *International Journal of Applied Earth Observation and Geoinformation*, *118*, 103257. https://doi.org/10.1016/j.jag.2023.103257

Tedim, F., & Leone, V. (2020). The Dilemma of Wildfire Definition: What It Reveals and What It Implies. *Frontiers in Forests and Global Change*, *3*, 553116. https://doi.org/10.3389/ffgc.2020.553116

Tedim, F., Remelgado, R., Martins, J., & Carvalho, S. (2013). Os grandes incêndios florestais em Portugal. Desafios para a gestão do risco [Largeforest fires in Portugal. Challenges for risk management]. In A. B. Gonçalves & A. Vieira (Eds.), *Grandes Incêndios Florestais, Erosão, Degradaçãoe Medidas de Recuperação dos Solos [Large ForestFires, Erosion, Degradation and Soil RestorationMeasures]* (pp. 75-86). https://www.riscos.pt/wp-content/uploads/2018/Outras_Pub/outros_livros/VII_EGFA/vii_egfa_artigo_06.pdf

Tonbul, H., Yilmaz, E. O., & Kavzoglu, T. (2023). Comparative Analysis of Deep Learning and Machine Learning Models for Burned Area Estimation Using Sentinel-2 Image: A Case Study in Muğla-Bodrum, Turkey. *2023 10th International Conference on Recent Advances in Air and Space Technologies (RAST)*, 1-5. https://doi.org/10.1109/RAST57548.2023.10197926

Zekić-Sušac, M., Pfeifer, S., & Šarlija, N. (2014). A Comparison of Machine Learning Methods in a High-Dimensional Classification Problem. *Business Systems Research Journal*, *5*(3), 82-96. https://doi.org/10.2478/bsrj-2014-0021

APPENDIX A

Table 0-1 Wildfires with burned areas ≥ 100 ha in the study area (2015-2023), including GIS code, start data, municipality, fire type, cause, and burned area.

GIS Code	Start Date (YYYY-MM-DD)	Municipality	Type	Cause	Burned Area (ha)
AG115581	2015-10-01	Monchique	Forestry	Negligent	203.06
(no info)*	2016-08-08	Silves	(no info)	(no info)*	1 809.55
(no info)*	2016-09-03	Silves	(no info)	(no info)*	133.72
(no info)*	2016-09-09	Portimão	(no info)	(no info)*	3 831.01
AT317376	2017-08-10	Castro Verde	Forestry	Negligent	151.00
AG117658	2017-09-16	Loulé	Forestry	Negligent	100.92
AG118728	2018-08-03	Monchique	Forestry	Unknown	26 885.45
AG118766	2018-08-07	Castro Marim	Forestry	Unknown	109.67
AT318350	2018-09-01	Almodôvar	Forestry	Negligent	1 032.22
AT319256	2019-06-10	Ourique	(no info)	Negligent	549.21
AT319291	2019-06-17	Castro Verde	(no info)	Negligent	341.18
AT319292	2019-06-17	Almodôvar	(no info)	Negligent	112.36
AG119811	2019-07-19	Aljezur	(no info)	Unknown	207.39
AG119842	2019-07-26	Silves	(no info)	Negligent	112.93
AT319505	2019-09-10	Odemira	(no info)	Negligent	195.68
AG120550	2020-06-19	Vila do Bispo	Forestry	Negligent	2 236.42
AT320155	2020-07-03	Aljustrel	Agricultural	Negligent	117.63
AG120614	2020-07-06	Silves	Forestry	Negligent	535.63
AT320181	2020-07-13	Castro Verde	Agricultural	Negligent	2 333.96
AG121869	2021-05-22	Vila Real de Santo António	(no info)	Negligent	100.15
AT321215	2021-07-12	Castro Verde	(no info)	Negligent	199.58
AG1211131	2021-07-17	Portimão	(no info)	Negligent	1 933.79
AG1211218	2021-08-04	Loulé	(no info)	Intentional	218.56
AG1211287	2021-08-16	Tavira	(no info)	Intentional	6 629.60
AT321313	2021-08-18	Odemira	(no info)	Intentional	913.86
AG1211309	2021-08-18	Odemira	(no info)	Intentional	197.03
2022_08_0035308	2022-07-12	Loulé	Forestry	Negligent	742.06
2022_08_0038271	2022-07-25	Silves	Forestry	Unknown	1 326.94
2022_02_0021224	2022-08-22	Aljustrel	Agricultural	Negligent	117.20
2022_02_0023633	2022-09-18	Odemira	Forestry	Negligent	256.84
20230563519	2023-05-12	Castro Marim	(no info)	Intentional	327.78

20230901589	2023-07-26	Almodôvar	(no info)	Negligent	162.49
20230950499	2023-08-05	Odemira	(no info)	Unknown	7 526.19
20230960910	2023-08-07	Aljezur	(no info)	Negligent	187.06
20231214806	2023-10-01	Aljezur	(no info)	Intentional	118.32

*ICNF did not provide any information regarding 2016 wildfires. Through news channels we were able to complete start date and municipality data.

Table 0-2 Sentinel-2 Bands - source: (*Copernicus*, 2023).

Band	**Spatial Resolution (m)**	**Central Wavelength (nm)**	**Description**
B1	60	443	Ultra Blue (Coastal and Aerosol)
B2	10	490	Blue
B3	10	560	Green
B4	10	665	Red
B5	20	705	Visible and Near Infrared (VNIR)
B6	20	740	Visible and Near Infrared (VNIR)
B7	20	783	Visible and Near Infrared (VNIR)
B8	10	842	Visible and Near Infrared (VNIR)
B8a	20	865	Visible and Near Infrared (VNIR)
B9	60	945	Short Wave Infrared (SWIR)
B10	60	1 375	Short Wave Infrared (SWIR)
B11	20	1 610	Short Wave Infrared (SWIR)
B12	20	2 190	Short Wave Infrared (SWIR)

Table 0-3 List of downloaded satellite imagery TIF files for the study area, including file names, capture dates, and corresponding zones.

File Name	**Satellite Image Date (YYYY/MM/DD HHMMSS)**	**Zone**
S2A_MSIL2A_20150913T112116_N0213_R037_T29SNB_20210124T041110.SAFE	2015/09/13 11h21m16s	West side
S2A_MSIL2A_20160808T112122_N0213_R037_T29SNB_20210120T204918.SAFE	2016/08/08 11h21m22s	West side
S2A_MSIL2A_20160828T112122_N0213_R037_T29SNB_20210120T220436.SAFE	2016/08/28 11h21m22s	West side
S2A_MSIL2A_20160907T112112_N0213_R037_T29SNB_20210120T220436.SAFE	2016/09/07 11h21m12s	West side
S2B_MSIL2A_20170808T112109_N0208_R037_T29SNB_20221207T225803.SAFE	2017/08/08 11h21m09s	West side

S2A_MSIL2A_20170912T112111_N0213_R037_T29SNB_20210106T234957.SAFE	2017/09/12 11h21m11s	West side
S2B_MSIL2A_20180803T112109_N0500_R037_T29SNB_20230731T111322.SAFE	2018/08/03 11h21m09s	West side
S2A_MSIL2A_20180805T110621_N0500_R137_T29SPB_20230802T045945.SAFE	2018/08/05 11h06m21s	East side
S2A_MSIL2A_20180808T112111_N9999_R037_T29SNB_20221105T211910.SAFE	2018/08/08 11h21m11s	West side
S2A_MSIL2A_20180828T112111_N0208_R037_T29SNB_20180828T155805.SAFE	2018/08/28 11h21m11s	West side
S2B_MSIL2A_20190609T112119_N0207_R037_T29SNB_20190609T140510.SAFE	2019/06/09 11h21m19s	West side
S2B_MSIL2A_20190719T112119_N0213_R037_T29SNB_20190719T131332.SAFE	2019/07/19 11h21m19s	West side
S2A_MSIL2A_20190724T112121_N0213_R037_T29SNB_20190724T123805.SAFE	2019/07/24 11h21m21s	West side
S2B_MSIL2A_20190907T112119_N0500_R037_T29SNB_20230705T125842.SAFE	2019/09/07 11h21m19s	West side
S2A_MSIL2A_20200618T112121_N0500_R037_T29SNB_20230428T155508.SAFE	2020/06/18 11h21m21s	West side
S2B_MSIL2A_20200703T112119_N0214_R037_T29SNB_20200703T140926.SAFE	2020/07/03 11h21m19s	West side
S2B_MSIL2A_20200713T112119_N0214_R037_T29SNB_20200713T133039.SAFE	2020/07/13 11h21m19s	West side
S2A_MSIL2A_20210521T110621_N0500_R137_T29SPB_20230224T213045.SAFE	2021/05/21 11h06m21s	East side
S2B_MSIL2A_20210708T112119_N0500_R037_T29SNB_20230203T064127.SAFE	2021/07/08 11h21m19s	West side
S2A_MSIL2A_20210713T112121_N0500_R037_T29SNB_20230222T075804.SAFE	2021/07/13 11h21m21s	West side
S2B_MSIL2A_20210728T112119_N0500_R037_T29SNB_20230220T224708.SAFE	2021/07/28 11h21m19s	West side
S2B_MSIL2A_20210814T110619_N0500_R137_T29SPB_20230213T031209.SAFE	2021/08/14 11h06m19s	East side
S2B_MSIL2A_20210817T112119_N0500_R037_T29SNB_20230216T070441.SAFE	2021/08/17 11h21m19s	West side
S2A_MSIL2A_20220708T112131_N0400_R037_T29SNB_20220708T173530.SAFE	2022/07/08 11h21m31s	West side
S2B_MSIL2A_20220723T112119_N0400_R037_T29SNB_20220723T125549.SAFE	2022/07/23 11h21m19s	West side
S2B_MSIL2A_20220822T112119_N0400_R037_T29SNB_20220822T125445.SAFE	2022/08/22 11h21m19s	West side

S2B_MSIL2A_20220911T112119_N0400_R037_T29SNB_20220911T140223.SAFE	2022/09/11 11h21m19s	West side
S2A_MSIL2A_20230511T110621_N0509_R137_T29SPB_20230511T170516.SAFE	2023/05/11 11h06m21s	East side
S2A_MSIL2A_20230723T112121_N0509_R037_T29SNB_20230723T173156.SAFE	2023/07/23 11h21m21s	West side
S2A_MSIL2A_20230802T112121_N0509_R037_T29SNB_20230802T172558.SAFE	2023/08/02 11h21m21s	West side
S2B_MSIL2A_20230807T112119_N0509_R037_T29SNB_20230807T125106.SAFE	2023/08/07 11h21m19s	West side
S2A_MSIL2A_20231001T112121_N0509_R037_T29SNB_20231002T124603.SAFE	2023/10/01 11h21m19s	West side

APPENDIX B

Table 0-1 Sample of the first and last two rows in the Training DataFrame, displaying all initial features used in the analysis.

Point Id	1	2	…	4 018 487	4 018 488
Fire Id	2015-10-01 AG115581	2015-10-01 AG115581	…	2022-09-18 2022_02_0023633	2022-09-18 2022_02_0023633
B02	496	431	…	1 305	1 359
B03	762	626	…	1 472	1 508
…	…	…	…	…	…
B11	2 249	1 827	…	2 715	2 735
B12	1 310	1 058	…	2 055	2 112
Elevation	338.18	104.92	…	176.50	173.01
Slope	6.22	11.81	…	2.84	8.59
Aspect	283.03	310.84	…	253.35	159.64
Land class	Forest	Forest	…	Pastures	Forest
Days to fire	18	18	…	7	7
Burned	0	0	…	0	0

Table 0-2 Percentage of Burned Points (target=1) across Land Classification classes.

Land Classification Class	Burned (%)	Dataset (%)
Wetlands	3.08	0.34
Surface water bodies	4.54	1.59
Artificialized territories	7.37	2.36
Uncovered spaces or spaces with little vegetation	11.93	0.16
Pastures	13.35	4.11
Agriculture	20.09	17.22
Agroforestry areas (pt: SAF)	22.77	2.69
Scrublands	39.03	28.01
Forests	40.23	43.52

Dataset (%): This column indicates the percentage of the dataset belonging to each COS class. For example, wetlands account for 0.34% of the total training dataset.

Burned (%): This column represents the percentage of burned points (Burned = 1) within each COS class. For instance, 40.23% of all points in the Forests class are labelled as burned.

APPENDIX C

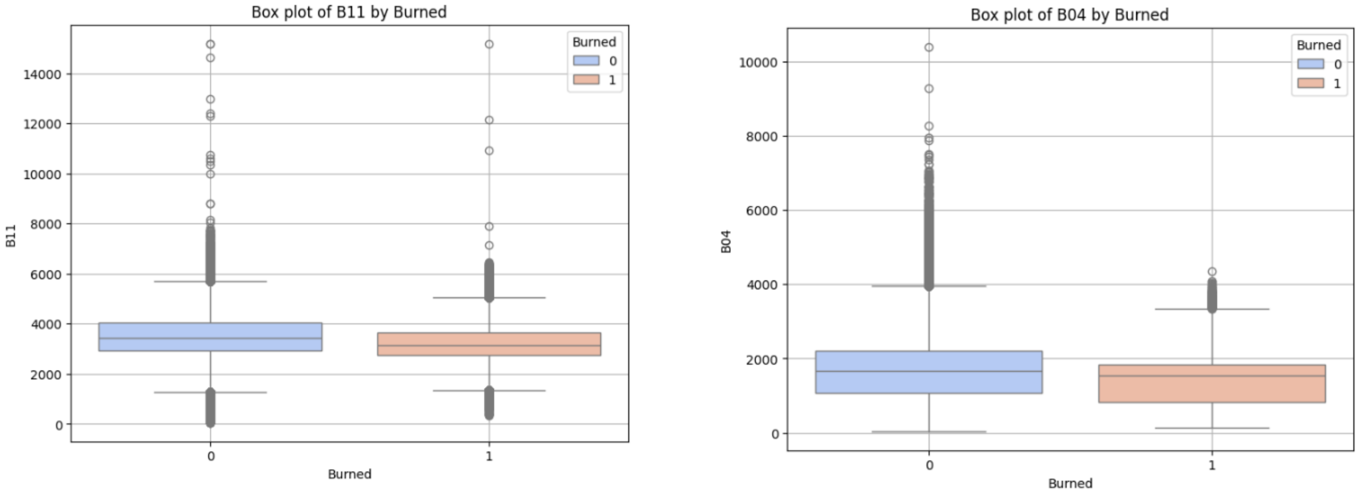


Figure 0-1 Box plots showing outliers for the ‘B11’ feature in subset 1 (left) and the ‘B04’ feature in subset 2 (right), categorized by Burned classes (0 = unburned, 1 = burned).

While we selected these figures (Figure 0-1) as representative examples, almost every spectral band in both subsets displayed visible outliers.

To determine a suitable threshold for removing outliers, we analysed the Z-score distributions for each spectral band. Figure 0-2(left: subset 1, right: subset 2) display the Z-score distributions for ‘B11’ and ‘B04’ features, respectively. The red dashed line represents the 1% threshold for total records—a visual guide to ensure minimal data loss during the outlier removal process.

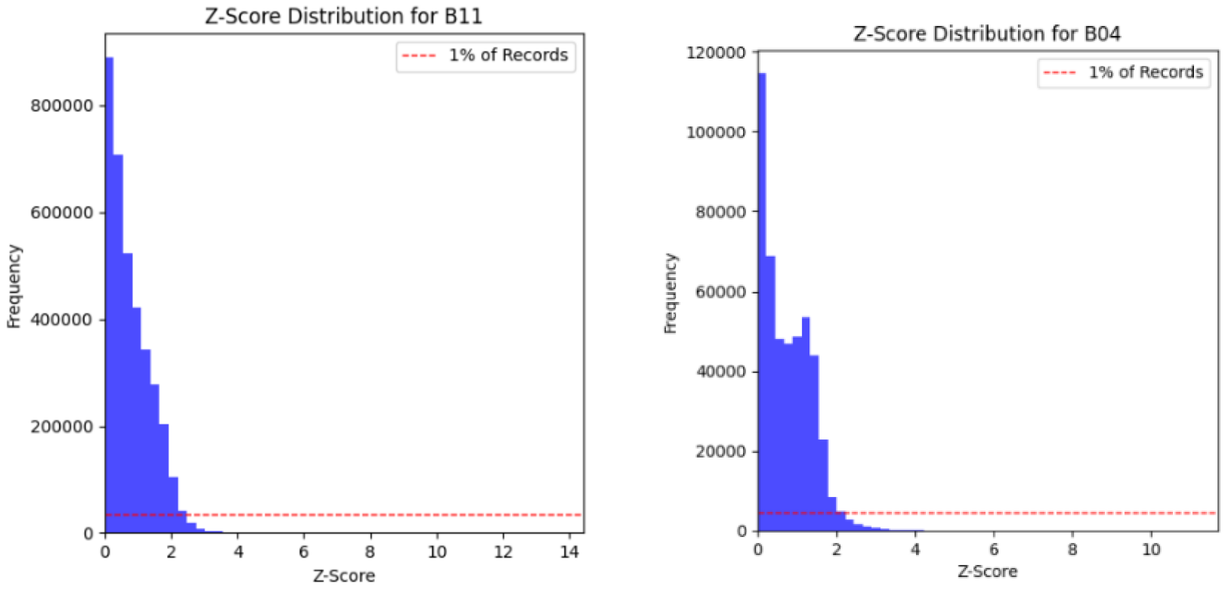


Figure 0-2 Z-score distributions for the ‘B11’ feature in subset 1 (left) and the ‘B04’ feature in subset 2 (right). The red dashed line represents the 1% threshold of total records.

As shown, the majority of values for ‘B11’ feature in subset 1 have Z-scores below 3, with only a small proportion exceeding this threshold. These outliers constitute less than 1% of the total data in subset 1, aligning with our objective of minimal data loss. A similar analysis for the ‘B04’ feature in subset 2 reveals a comparable pattern. Therefore, we decided to remove outliers using a Z-score threshold of 3 for both subsets, which had the following impact, as shown in Figure 0-3 (subset 1) Figure 0-4 (subset 2):

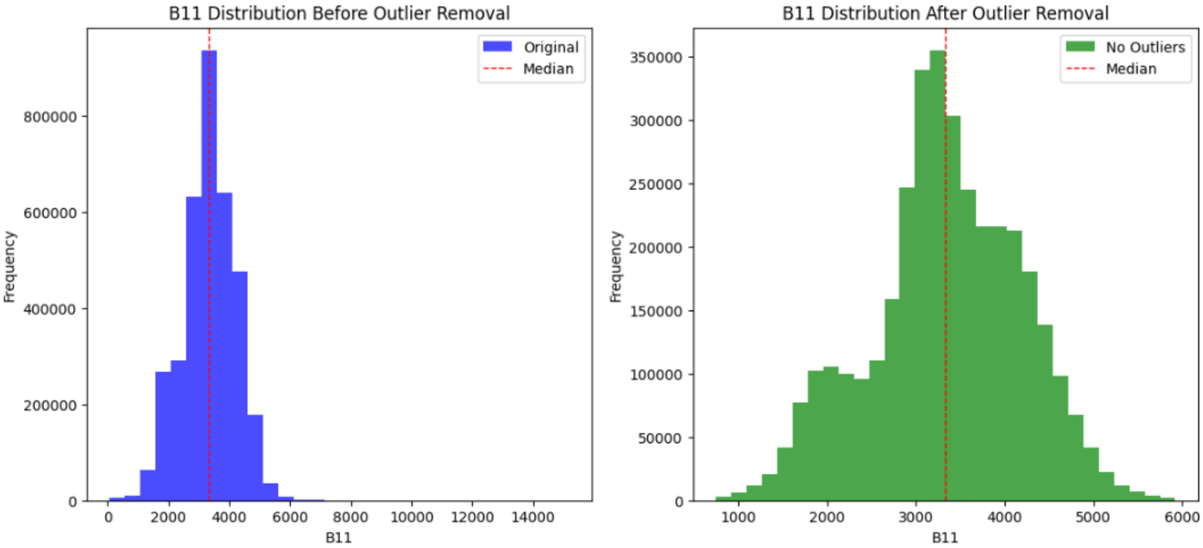


Figure 0-3 Distribution of the ‘B11’ feature in subset 1. Left: Original distribution, showing the presence of outliers. Right: Distribution after outlier removal, demonstrating a smoother and more balanced dataset.

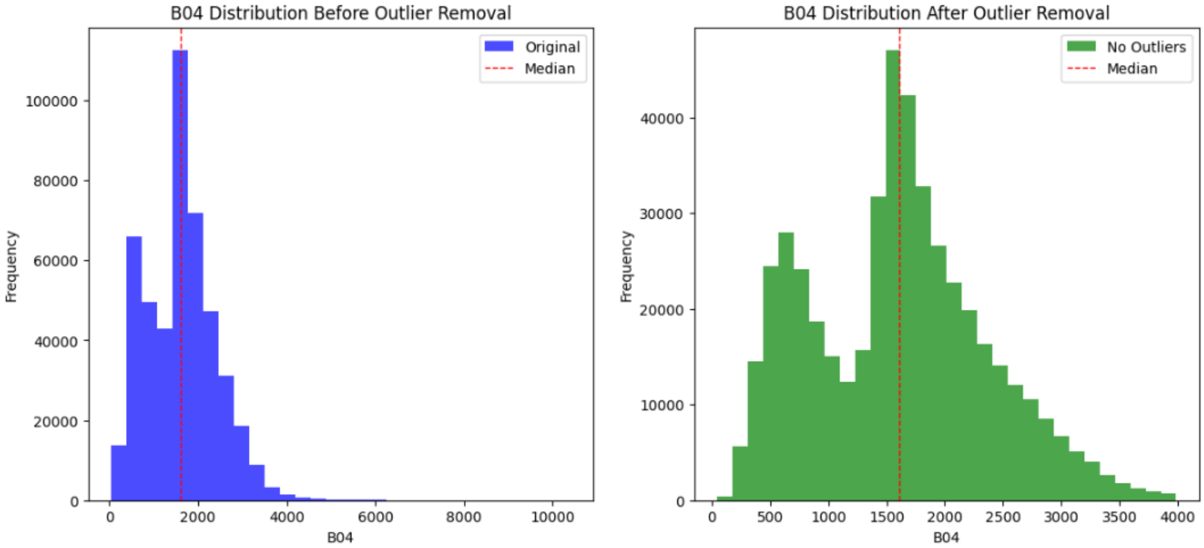


Figure 0-4 Distribution of the ‘B04’ feature in subset 1. Left: Original distribution, showing the presence of outliers. Right: Distribution after outlier removal, demonstrating a smoother and more balanced dataset.

For band ‘B11’ in subset 1 (Figure 0-3), the median Z-score was 0.6353, and the mean was 0.7846. Based on a Z-score threshold of 3, a total of 11 731 rows, representing 0.33% of the dataset, were identified as outliers and subsequently removed.

For band ‘B04’ in subset 2 (Figure 0-4), the median Z-score was 0.6828, and the mean was 0.7780. Using a Z-score threshold of 3, a total of 2 705 rows, representing 0.58% of the dataset, were identified as outliers and subsequently removed.

Additionally, Figure 0-3 and Figure 0-4 compare the distribution of band values before and after outlier removal. The left graph shows the original distribution, where outliers with high Z-scores are clearly visible. The right graph depicts the smoother distribution after outlier removal. The red dashed line in both graphs represents the median.

Although these figures illustrate the process for one band per subset (‘B11’ for subset 1 and ‘B04’ for subset 2), the same outlier removal procedure was applied to all nine spectral bands in both subsets.

APPENDIX D

For **subset_1_no_outliers** DataFrame, the correlation matrix for the spectral and terrain features is shown in Figure 0-1 and for the multispectral indices in Figure 0-2:

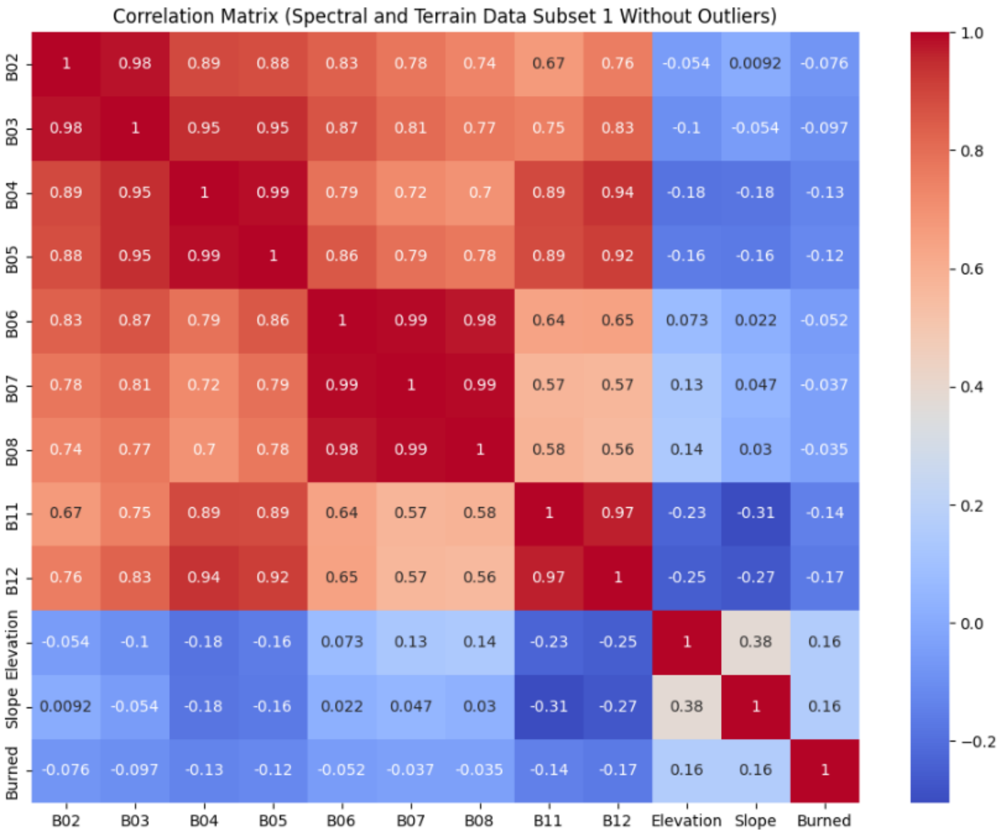
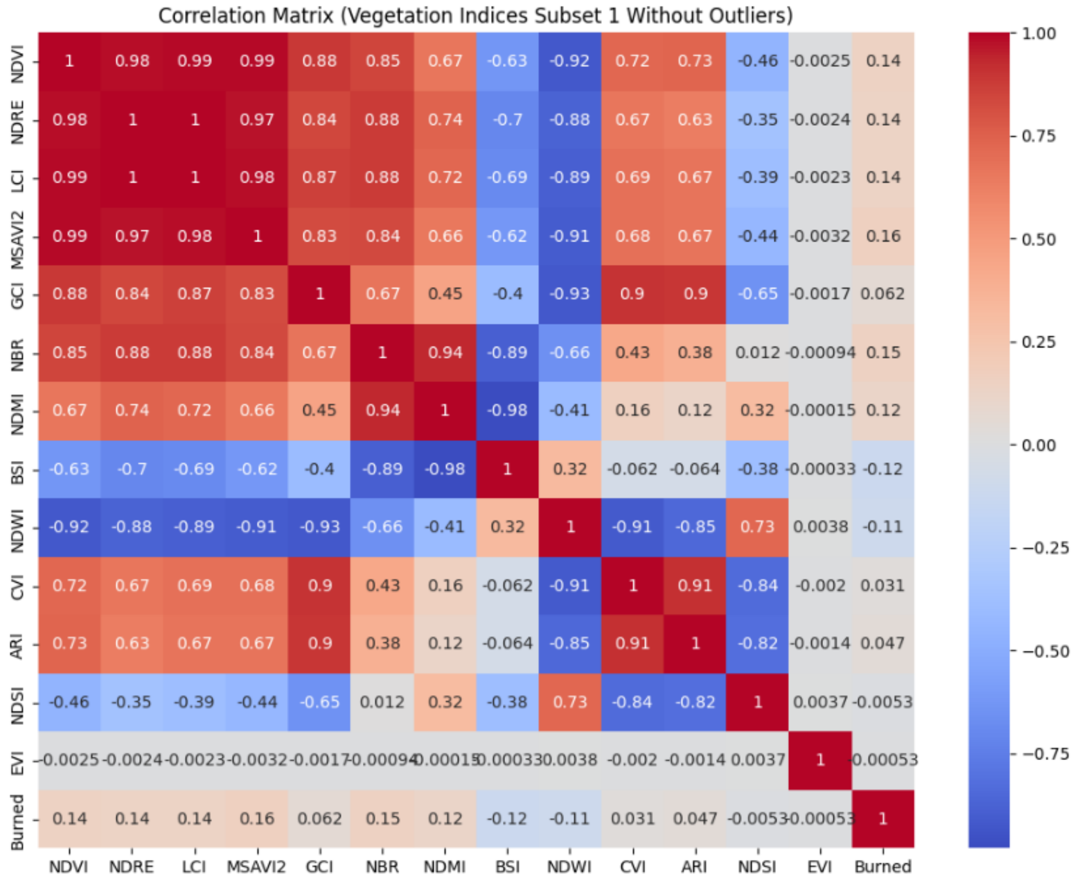


Figure 0-1 Correlation matrix of spectral and terrain features for subset_1_no_outliers DataFrame

By applying rule a) we discarded ‘B07’ and ‘B08’. We applied rule b) to ‘B02’ and ‘B03’, removing ‘B02’; to ‘B04’ and ‘B05’, removing ‘B05’; and to ‘B11’ and ‘B12’, removing ‘B12’.



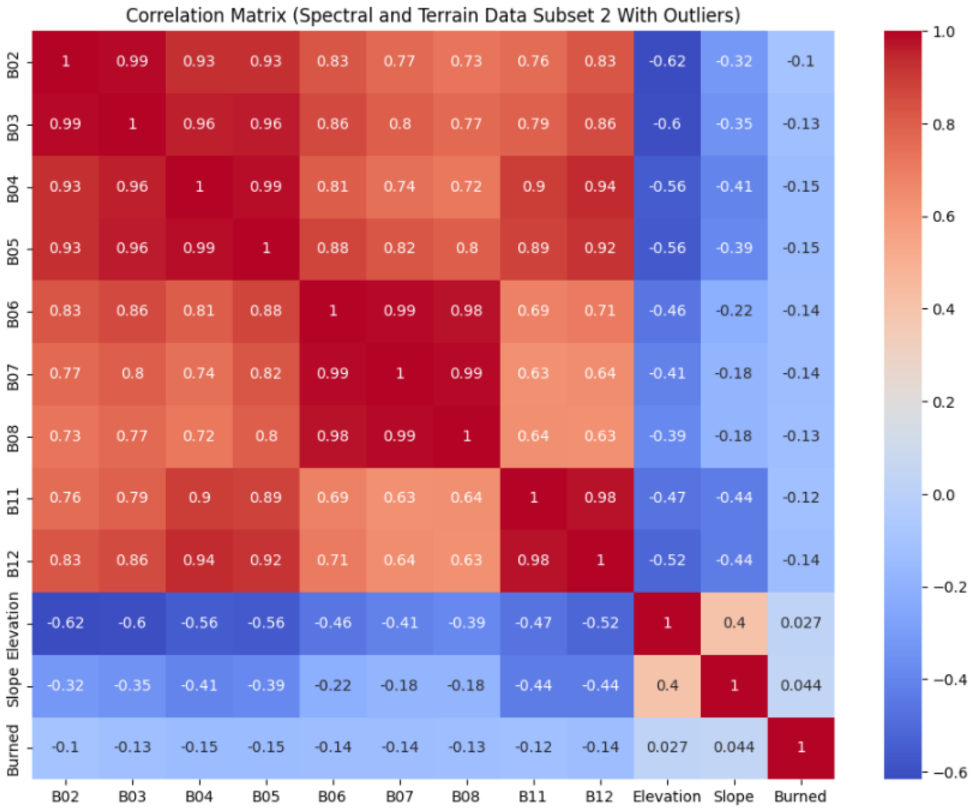
Correlation Matrix (Vegetation Indices Subset 1 Without Outliers)

	NDVI	NDRE	LCI	MSAVI2	GCI	NBR	NDMI	BSI	NDWI	CVI	ARI	NDSI	EVI	Burned
NDVI	1	0.98	0.99	0.99	0.88	0.85	0.67	-0.63	-0.92	0.72	0.73	-0.46	-0.0025	0.14
NDRE	0.98	1	1	0.97	0.84	0.88	0.74	-0.7	-0.88	0.67	0.63	-0.35	-0.0024	0.14
LCI	0.99	1	1	0.98	0.87	0.88	0.72	-0.69	-0.89	0.69	0.67	-0.39	-0.0023	0.14
MSAVI2	0.99	0.97	0.98	1	0.83	0.84	0.66	-0.62	-0.91	0.68	0.67	-0.44	-0.0032	0.16
GCI	0.88	0.84	0.87	0.83	1	0.67	0.45	-0.4	-0.93	0.9	0.9	-0.65	-0.0017	0.062
NBR	0.85	0.88	0.88	0.84	0.67	1	0.94	-0.89	-0.66	0.43	0.38	0.012	-0.00094	0.15
NDMI	0.67	0.74	0.72	0.66	0.45	0.94	1	-0.98	-0.41	0.16	0.12	0.32	-0.00015	0.12
BSI	-0.63	-0.7	-0.69	-0.62	-0.4	-0.89	-0.98	1	0.32	-0.062	-0.064	-0.38	-0.00033	-0.12
NDWI	-0.92	-0.88	-0.89	-0.91	-0.93	-0.66	-0.41	0.32	1	-0.91	-0.85	0.73	0.0038	-0.11
CVI	0.72	0.67	0.69	0.68	0.9	0.43	0.16	-0.062	-0.91	1	0.91	-0.84	-0.002	0.031
ARI	0.73	0.63	0.67	0.67	0.9	0.38	0.12	-0.064	-0.85	0.91	1	-0.82	-0.0014	0.047
NDSI	-0.46	-0.35	-0.39	-0.44	-0.65	0.012	0.32	-0.38	0.73	-0.84	-0.82	1	0.0037	-0.0053
EVI	-0.0025	-0.0024	-0.0023	-0.0032	-0.0017	-0.00094	-0.00015	-0.00033	0.0038	-0.002	-0.0014	0.0037	1	-0.00053
Burned	0.14	0.14	0.14	0.16	0.062	0.15	0.12	-0.12	-0.11	0.031	0.047	-0.0053	-0.00053	1

Figure 0-2 Correlation matrix of multispectral indices for subset_1_no_outliers DataFrame

Analysing Figure 0-2 and applying rule a), we discarded ‘CVI’, ‘NDSI’ and ‘EVI’. We retained ‘ARI’ because it had a 5% correlation when rounded, meeting the threshold defined in rule a). Since ‘NDVI’, ‘NDRE’, ‘LCI’ and ‘MSAVI2’ all had correlations greater than 95% with each other’s, only one feature among them could remain, as per rule b). As a result, we retained ‘MSAVI2’. We applied rule c) to ‘NDMI’ and ‘BSI’, removing ‘NDMI’.

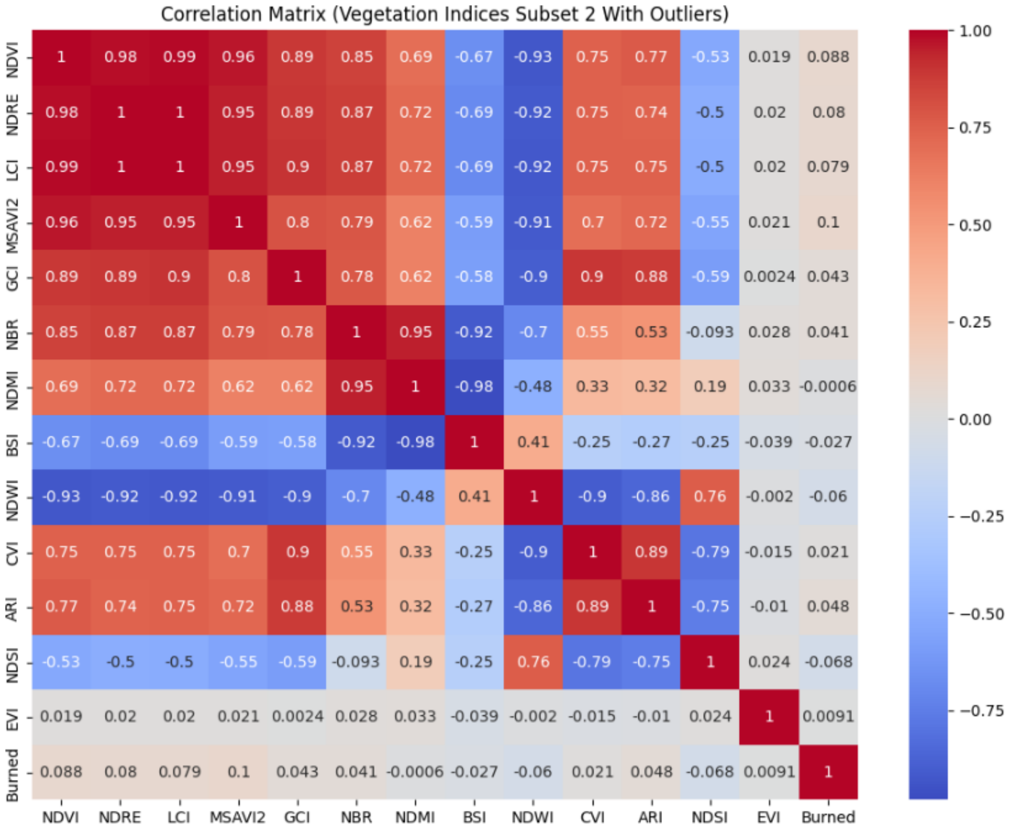
For **subset_2** DataFrame, the correlation matrix for the spectral and terrain features is shown in Figure 0-3 and for the multispectral indices in Figure 0-4:



	B02	B03	B04	B05	B06	B07	B08	B11	B12	Elevation	Slope	Burned
B02	1	0.99	0.93	0.93	0.83	0.77	0.73	0.76	0.83	-0.62	-0.32	-0.1
B03	0.99	1	0.96	0.96	0.86	0.8	0.77	0.79	0.86	-0.6	-0.35	-0.13
B04	0.93	0.96	1	0.99	0.81	0.74	0.72	0.9	0.94	-0.56	-0.41	-0.15
B05	0.93	0.96	0.99	1	0.88	0.82	0.8	0.89	0.92	-0.56	-0.39	-0.15
B06	0.83	0.86	0.81	0.88	1	0.99	0.98	0.69	0.71	-0.46	-0.22	-0.14
B07	0.77	0.8	0.74	0.82	0.99	1	0.99	0.63	0.64	-0.41	-0.18	-0.14
B08	0.73	0.77	0.72	0.8	0.98	0.99	1	0.64	0.63	-0.39	-0.18	-0.13
B11	0.76	0.79	0.9	0.89	0.69	0.63	0.64	1	0.98	-0.47	-0.44	-0.12
B12	0.83	0.86	0.94	0.92	0.71	0.64	0.63	0.98	1	-0.52	-0.44	-0.14
Elevation	-0.62	-0.6	-0.56	-0.56	-0.46	-0.41	-0.39	-0.47	-0.52	1	0.4	0.027
Slope	-0.32	-0.35	-0.41	-0.39	-0.22	-0.18	-0.18	-0.44	-0.44	0.4	1	0.044
Burned	-0.1	-0.13	-0.15	-0.15	-0.14	-0.14	-0.13	-0.12	-0.14	0.027	0.044	1

Figure 0-3 Correlation matrix of spectral and terrain features for subset_2 DataFrame

By applying rule a), which led to the removal of ‘Elevation’ and ‘Slope’. Next, we applied rule b) to ‘B02’ and ‘B03’, removing ‘B02’. We then applied rule b) to ‘B03’, ‘B04’, and ‘B05’, removing ‘B03’. Finally, rule c) was applied to ‘B04’ and ‘B05’, leading to the removal of ‘B05’ due to its higher average correlation with the remaining features. Applying the same methods to ‘B06’, ‘B07’ and ‘B08’, we only retained ‘B07’. Lastly, rule b) was applied to ‘B11’ and ‘B12’, thus removing ‘B11’.

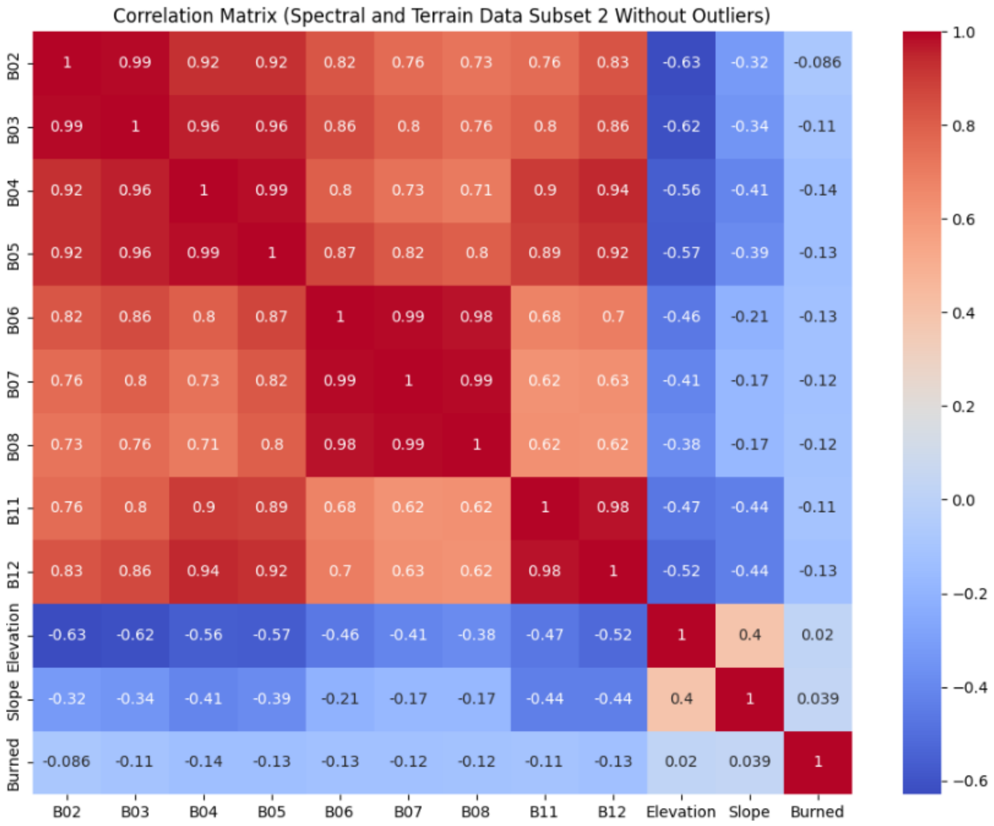


	NDVI	NDRE	LCI	MSAVI2	GCI	NBR	NDMI	BSI	NDWI	CVI	ARI	NDSI	EVI	Burned
NDVI	1	0.98	0.99	0.96	0.89	0.85	0.69	-0.67	-0.93	0.75	0.77	-0.53	0.019	0.088
NDRE	0.98	1	1	0.95	0.89	0.87	0.72	-0.69	-0.92	0.75	0.74	-0.5	0.02	0.08
LCI	0.99	1	1	0.95	0.9	0.87	0.72	-0.69	-0.92	0.75	0.75	-0.5	0.02	0.079
MSAVI2	0.96	0.95	0.95	1	0.8	0.79	0.62	-0.59	-0.91	0.7	0.72	-0.55	0.021	0.1
GCI	0.89	0.89	0.9	0.8	1	0.78	0.62	-0.58	-0.9	0.9	0.88	-0.59	0.0024	0.043
NBR	0.85	0.87	0.87	0.79	0.78	1	0.95	-0.92	-0.7	0.55	0.53	-0.093	0.028	0.041
NDMI	0.69	0.72	0.72	0.62	0.62	0.95	1	-0.98	-0.48	0.33	0.32	0.19	0.033	-0.0006
BSI	-0.67	-0.69	-0.69	-0.59	-0.58	-0.92	-0.98	1	0.41	-0.25	-0.27	-0.25	-0.039	-0.027
NDWI	-0.93	-0.92	-0.92	-0.91	-0.9	-0.7	-0.48	0.41	1	-0.9	-0.86	0.76	-0.002	-0.06
CVI	0.75	0.75	0.75	0.7	0.9	0.55	0.33	-0.25	-0.9	1	0.89	-0.79	-0.015	0.021
ARI	0.77	0.74	0.75	0.72	0.88	0.53	0.32	-0.27	-0.86	0.89	1	-0.75	-0.01	0.048
NDSI	-0.53	-0.5	-0.5	-0.55	-0.59	-0.093	0.19	-0.25	0.76	-0.79	-0.75	1	0.024	-0.068
EVI	0.019	0.02	0.02	0.021	0.0024	0.028	0.033	-0.039	-0.002	-0.015	-0.01	0.024	1	0.0091
Burned	0.088	0.08	0.079	0.1	0.043	0.041	-0.0006	-0.027	-0.06	0.021	0.048	-0.068	0.0091	1

Figure 0-4 Correlation matrix of multispectral indices for subset_2 DataFrame

Analysing Figure 0-4 and applying rule a), we discarded ‘GCI’, ‘NBR’, ‘NDMI’, ‘BSI’, ‘CVI’ and ‘EVI’. However, we retained ‘ARI’ as it had a 5% correlation when rounded (as stated in rule a)). Applying rule b) to ‘NDVI,’ ‘NDRE,’ ‘LCI,’ and ‘MSAVI2,’ since all had correlations with each other’s exceeding 95%, ‘MSAVI2’ was retained, and the remaining three features were removed.

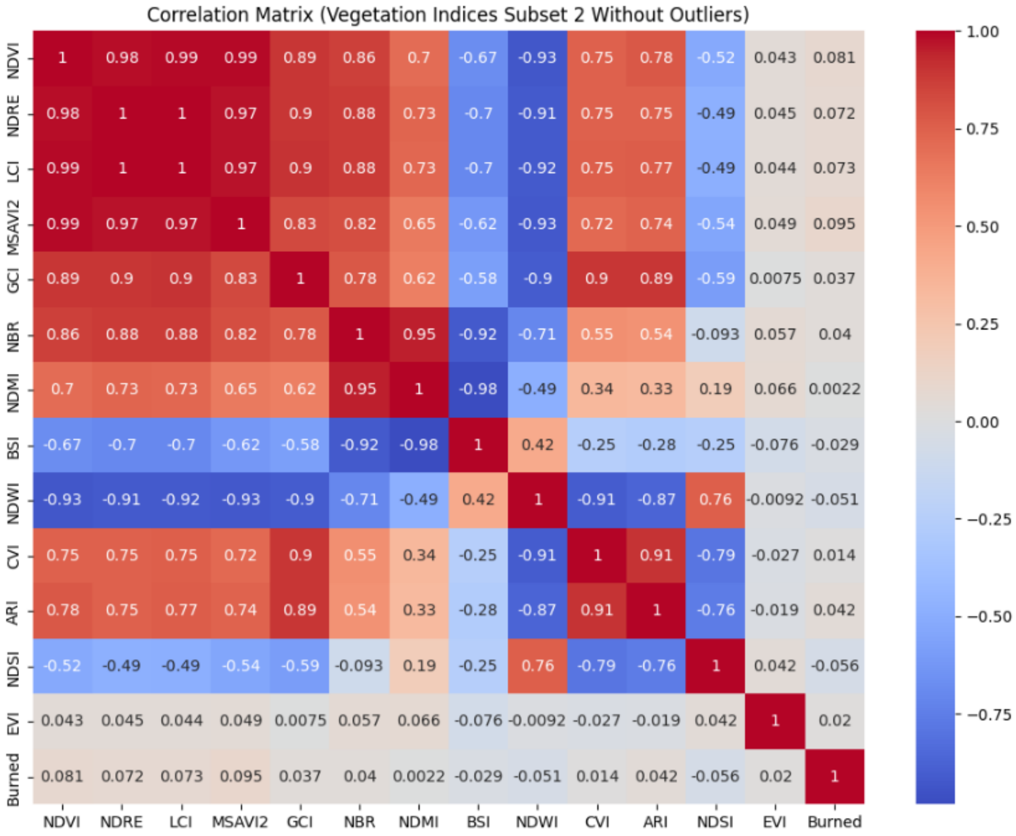
For **subset_2_no_outliers** DataFrame, the correlation matrix for the spectral and terrain features is shown in Figure 0-5 and for the multispectral indices in Figure 0-6:



	B02	B03	B04	B05	B06	B07	B08	B11	B12	Elevation	Slope	Burned
B02	1	0.99	0.92	0.92	0.82	0.76	0.73	0.76	0.83	-0.63	-0.32	-0.086
B03	0.99	1	0.96	0.96	0.86	0.8	0.76	0.8	0.86	-0.62	-0.34	-0.11
B04	0.92	0.96	1	0.99	0.8	0.73	0.71	0.9	0.94	-0.56	-0.41	-0.14
B05	0.92	0.96	0.99	1	0.87	0.82	0.8	0.89	0.92	-0.57	-0.39	-0.13
B06	0.82	0.86	0.8	0.87	1	0.99	0.98	0.68	0.7	-0.46	-0.21	-0.13
B07	0.76	0.8	0.73	0.82	0.99	1	0.99	0.62	0.63	-0.41	-0.17	-0.12
B08	0.73	0.76	0.71	0.8	0.98	0.99	1	0.62	0.62	-0.38	-0.17	-0.12
B11	0.76	0.8	0.9	0.89	0.68	0.62	0.62	1	0.98	-0.47	-0.44	-0.11
B12	0.83	0.86	0.94	0.92	0.7	0.63	0.62	0.98	1	-0.52	-0.44	-0.13
Elevation	-0.63	-0.62	-0.56	-0.57	-0.46	-0.41	-0.38	-0.47	-0.52	1	0.4	0.02
Slope	-0.32	-0.34	-0.41	-0.39	-0.21	-0.17	-0.17	-0.44	-0.44	0.4	1	0.039
Burned	-0.086	-0.11	-0.14	-0.13	-0.13	-0.12	-0.12	-0.11	-0.13	0.02	0.039	1

Figure 0-5 Correlation matrix of spectral and terrain features for subset_2_no_outliers DataFrame

We applied rule a) and discarded ‘Elevation’ and ‘Slope’. We applied rule b) to ‘B02’ and ‘B03’, removing ‘B02’.We then reapplied rule b) to ‘B03’, ‘B04’, and ‘B05’, retaining only ‘B04’. Applying the same principles to ‘B06’, ‘B07’ and ‘B08’, we retained only ‘B06’. Similarly, for ‘B11’ and ‘B12’, we retained ‘B12’.



	NDVI	NDRE	LCI	MSAVI2	GCI	NBR	NDMI	BSI	NDWI	CVI	ARI	NDSI	EVI	Burned
NDVI	1	0.98	0.99	0.99	0.89	0.86	0.7	-0.67	-0.93	0.75	0.78	-0.52	0.043	0.081
NDRE	0.98	1	1	0.97	0.9	0.88	0.73	-0.7	-0.91	0.75	0.75	-0.49	0.045	0.072
LCI	0.99	1	1	0.97	0.9	0.88	0.73	-0.7	-0.92	0.75	0.77	-0.49	0.044	0.073
MSAVI2	0.99	0.97	0.97	1	0.83	0.82	0.65	-0.62	-0.93	0.72	0.74	-0.54	0.049	0.095
GCI	0.89	0.9	0.9	0.83	1	0.78	0.62	-0.58	-0.9	0.9	0.89	-0.59	0.0075	0.037
NBR	0.86	0.88	0.88	0.82	0.78	1	0.95	-0.92	-0.71	0.55	0.54	-0.093	0.057	0.04
NDMI	0.7	0.73	0.73	0.65	0.62	0.95	1	-0.98	-0.49	0.34	0.33	0.19	0.066	0.0022
BSI	-0.67	-0.7	-0.7	-0.62	-0.58	-0.92	-0.98	1	0.42	-0.25	-0.28	-0.25	-0.076	-0.029
NDWI	-0.93	-0.91	-0.92	-0.93	-0.9	-0.71	-0.49	0.42	1	-0.91	-0.87	0.76	-0.0092	-0.051
CVI	0.75	0.75	0.75	0.72	0.9	0.55	0.34	-0.25	-0.91	1	0.91	-0.79	-0.027	0.014
ARI	0.78	0.75	0.77	0.74	0.89	0.54	0.33	-0.28	-0.87	0.91	1	-0.76	-0.019	0.042
NDSI	-0.52	-0.49	-0.49	-0.54	-0.59	-0.093	0.19	-0.25	0.76	-0.79	-0.76	1	0.042	-0.056
EVI	0.043	0.045	0.044	0.049	0.0075	0.057	0.066	-0.076	-0.0092	-0.027	-0.019	0.042	1	0.02
Burned	0.081	0.072	0.073	0.095	0.037	0.04	0.0022	-0.029	-0.051	0.014	0.042	-0.056	0.02	1

Figure 0-6 Correlation matrix of multispectral indices for subset_2_no_outliers DataFrame

Analysing Figure 0-6 and applying rule a), we discarded ‘GCI’, ‘NBR’, ‘NDMI’, ‘BSI’, ‘CVI’, ‘ARI’ and ‘EVI’. Since ‘NDVI’, ‘NDRE’, ‘LCI’ and ‘MSAVI2’ all had correlations exceeding 95% with each other’s, only one of these features could remain (as stated in rule b)). Ultimately, ‘MSAVI2’ was retained.

APPENDIX E

Table 0-1 Performance metrics of the RF model for each DataFrame within subset 1.

DataFrame Combinations	Weighted Recall	Weighted Accuracy
df_subset_1_no_outliers_NESW	0.361122	0.587648
df_subset_1_no_outliers_no_Aspect	0.360463	0.588056
df_subset_1_no_outliers_NS	0.359445	0.588465
df_subset_1_NESW	0.353762	0.592929
df_subset_1_no_Aspect	0.349503	0.593906
df_subset_1_NS	0.349133	0.593934

Table 0-2 Performance metrics of the RF model for each DataFrame within subset 2.

DataFrame Combinations	Weighted Recall	Weighted Accuracy
df_subset_2_no_outliers_NESW	0.581467	0.584443
df_subset_2_no_outliers_no_Aspect	0.562163	0.587699
df_subset_2_no_Aspect	0.547624	0.622864
df_subset_2_no_outliers_NS	0.546852	0.583380
df_subset_2_NESW	0.531471	0.613366
df_subset_2_NS	0.507814	0.611865

Table 0-3 Performance metrics of the XGBoost model for each DataFrame within subset 1.

DataFrame Combinations	Weighted Recall	Weighted Accuracy
df_subset_1_no_outliers_NESW	0.295829	0.595489
df_subset_1_no_outliers_NS	0.293609	0.591735
df_subset_1_no_outliers_no_Aspect	0.291925	0.590557
df_subset_1_NS	0.291429	0.599560
df_subset_1_NESW	0.290704	0.601685
df_subset_1_no_Aspect	0.287032	0.597422

Table 0-4 Performance metrics of the XGBoost model for each DataFrame within subset 2.

DataFrame Combinations	Weighted Recall	Weighted Accuracy
df_subset_2_no_outliers_no_Aspect	0.457076	0.624022
df_subset_2_no_outliers_NS	0.445241	0.623763
df_subset_2_no_outliers_NESW	0.441989	0.625694
df_subset_2_NS	0.432089	0.629382
df_subset_2_NESW	0.431821	0.630073
df_subset_2_no_Aspect	0.429512	0.628114

Table 0-5 Performance metrics of the MLP model for each DataFrame within subset 1.

DataFrame Combinations	Weighted Recall	Weighted Accuracy
df_subset_1_no_outliers_NESW	0.308916	0.592715
df_subset_1_NESW	0.245042	0.609713
df_subset_1_no_outliers_NS	0.232408	0.599655
df_subset_1_no_outliers_no_Aspect	0.224069	0.597095
df_subset_1_NS	0.198222	0.617358
df_subset_1_no_Aspect	0.193005	0.622644

Table 0-6 Performance metrics of the MLP model for each DataFrame within subset 2.

DataFrame Combinations	Weighted Recall	Weighted Accuracy
df_subset_2_no_Aspect	0.504278	0.631306
df_subset_2_no_outliers_NESW	0.499149	0.618148
df_subset_2_NS	0.447418	0.624111
df_subset_2_NESW	0.441820	0.625008
df_subset_2_no_outliers_no_Aspect	0.424751	0.617309
df_subset_2_no_outliers_NS	0.416608	0.626795

Table 0-7 Fine-tuned RF model performance metrics for each DataFrame within subset 1.

DataFrame Combinations	Weighted Recall	Weighted Accuracy
df_subset_1_no_Aspect	0.487750	0.577985
df_subset_1_no_outliers_no_Aspect	0.479761	0.574233
df_subset_1_NS	0.474063	0.578897
df_subset_1_no_outliers_NESW	0.470123	0.574735
df_subset_1_NESW	0.464815	0.575824
df_subset_1_no_outliers_NS	0.463340	0.580054

Table 0-8 Fine-tuned RF model performance metrics for each DataFrame within subset 2.

DataFrame Combinations	Weighted Recall	Weighted Accuracy
df_subset_2_NESW	0.768374	0.522991
df_subset_2_no_Aspect	0.765426	0.534831
df_subset_2_NS	0.744954	0.540339
df_subset_2_no_outliers_NESW	0.685045	0.513181
df_subset_2_no_outliers_no_Aspect	0.597763	0.516975
df_subset_2_no_outliers_NS	0.524547	0.518692

Table 0-9 Performance metrics of the fine-tuned XGBoost model for each DataFrame within subset 1.

DataFrame Combinations	Weighted Recall	Weighted Accuracy
df_subset_1_no_outliers_no_Aspect	0.851477	0.500532
df_subset_1_no_outliers_NESW	0.835629	0.500478
df_subset_1_no_Aspect	0.815761	0.505961
df_subset_1_NESW	0.815136	0.503872
df_subset_1_no_outliers_NS	0.801954	0.510955
df_subset_1_NS	0.785858	0.512237

Table 0-10 Performance metrics of the fine-tuned XGBoost model for each DataFrame within subset 2.

DataFrame Combinations	Weighted Recall	Weighted Accuracy
df_subset_2_no_outliers_NESW	0.844179	0.493690
df_subset_2_no_outliers_no_Aspect	0.842785	0.500523
df_subset_2_no_Aspect	0.825698	0.519881
df_subset_2_no_outliers_NS	0.821476	0.506331
df_subset_2_NESW	0.799598	0.523238
df_subset_2_NS	0.795176	0.523283

Table 0-11 Fine-tuned MLP model performance metrics for each DataFrame within subset 1.

DataFrame Combinations	Weighted Recall	Weighted Accuracy
df_subset_1_no_outliers_no_Aspect	0.496683	0.587756
df_subset_1_no_Aspect	0.473338	0.593152
df_subset_1_NS	0.470517	0.596051
df_subset_1_NESW	0.468425	0.597279
df_subset_1_no_outliers_NESW	0.467686	0.595167
df_subset_1_no_outliers_NS	0.462595	0.594788

Table 0-12 Fine-tuned MLP model performance metrics for each DataFrame within subset 2.

DataFrame Combinations	Weighted Recall	Weighted Accuracy
df_subset_2_no_outliers_NESW	0.999930	0.340553
df_subset_2_NS	0.999887	0.339920
df_subset_2_no_outliers_no_Aspect	0.999750	0.341682
df_subset_2_NESW	0.999227	0.345006
df_subset_2_no_outliers_NS	0.991410	0.377730
df_subset_2_no_Aspect	0.824925	0.519166

APPENDIX F

SUBSET 1

Test Wildfire - 2023/05/12

This test wildfire is the only one located in the eastern region of the study area.

Dataset Overview:

- Pixels labelled as not burned (target = 0): 7 935 764
- Pixels labelled as burned (target = 1): 8 182
- Burned/unburned ratio: 0.10310%

Table 0-1 compares the actual percentage of burned pixels (‘Burned (%)’ column) with the predictions of each model (‘RF (%)’, ‘XGBoost (%)’ and ‘MLP (%)’ columns) across different land classifications. For instance, RF identified some areas prone to burn within forests and scrublands, while XGBoost was overly conservative, detecting nearly no areas prone to burn. MLP exhibited a peculiar tendency to classify wetlands as prone to burn, despite their typically low fire risk.

Table 0-1 Percentage of Burned Predicted Points across Land Classification classes for each model, regarding on the wildfire event of 12/05/2023.

Land Classification Class	Burned (%)	RF (%)	XGBoost (%)	MLP (%)	Dataset (%)
Wetlands	0	0	0	15.62	1.15
Surface water bodies	0	1.54	0	1.37	3.82
Artificialized territories	0	0.13	0	0.20	1.96
Uncovered spaces or spaces with little vegetation	0	0	0	0.02	0.60
Pastures	0.01	0.12	0	0.12	6.40
Agriculture	0.01	1.24	0	0.68	17.98
Agroforestry areas	0.06	1.33	0.01	1.25	16.72
Scrublands	0.19	10.80	0.05	13.77	19.80
Forests	0.17	11.37	0.3	10.53	31.57

Burned (%): This column represents the percentage of points that actual burned within each COS class. RF (%), XGBoost (%) and MLP (%) columns represent the percentage of points predicted to burn within each COS class.

The performance metrics for each model are summarized in Table 0-2. RF achieved minimal recall but detected few true positives, whereas XGBoost and MLP failed entirely to identify any burned pixels. While RF displayed limited success, its recall of 0.1467% is still extremely low.

Table 0-2 Performance metrics of each model for the wildfire event on 12/05/2023

Model	TP	FP	TN	FN	Recall	Accuracy
RF	12	495 836	7 439 928	8 170	0.001467	0.936555
XGBoost	0	8 480	7 927 284	8 182	0	0.997903
MLP	0	526 318	7 409 446	8 182	0	0.932716

TP = True Positives; FP = False Positives; TN = True Negatives; FN = False Negatives

Test Wildfire - 2023/08/07

As mentioned earlier, the satellite imagery for this wildfire captures an ongoing wildfire event, characterized by large columns of grey clouds. This is the second test wildfire for subset 1 and the first located in the western region of the study area, which is the most extensively trained and largest side of the dataset.

Dataset Overview:

- Pixels labelled as not burned (target = 0): 24 022 501
- Pixels labelled as burned (target = 1): 4 661
- Burned/unburned ratio: 0.01940%

Table 0-3 compares the actual percentage of burned pixels (‘Burned (%)’ column) with the predictions of each model (‘RF (%)’, ‘XGBoost (%)’ and ‘MLP (%)’ columns) across different land classifications. Interestingly, MLP, which previously misclassified water classes as prone to burn, performed closer to RF in this scenario, reducing its misclassification rate for water-related land classes.

Table 0-3 Percentage of Burned Predicted Points across Land Classification classes for each model, regarding on the wildfire event of 07/08/2023.

COS18n1_L	**Burned (%)**	**RF (%)**	**XGBoost (%)**	**MLP (%)**	**Dataset (%)**
Wetlands	0	0	0	0	0.65
Surface water bodies	0	0.47	0.01	0.23	4.14
Artificialized territories	0	0.23	0.01	0.22	3.16
Uncovered spaces or spaces with little vegetation	0	0.84	0.32	1.19	0.33
Pastures	0.01	0.32	0	0.18	9.62
Agriculture	0	1.20	0.11	0.55	24.10
Agroforestry areas	0	0.81	0	0.83	9.45
Scrublands	0.09	18.92	0.76	17.61	13.34
Forests	0.02	24.69	1.87	22.45	35.21

Burned (%): This column represents the percentage of points that actual burned within each COS class. RF (%), XGBoost (%) and MLP (%) columns represent the percentage of points predicted to burn within each COS class.

Table 0-4 Performance metrics of each model for the wildfire event on 07/08/2023

Model	**TP**	**FP**	**TN**	**FN**	**Recall**	**Accuracy**
RF	124	2 797 290	21 225 211	4 537	0.026604	0.883389
XGBoost	0	189 187	23 833 314	4 661	0	0.991932
MLP	137	2 522 996	21 499 505	4 524	0.029393	0.894806

The performance metrics for each model are summarized in Table 0-4. XGBoost, as noted during fine-tuning, section 4.3.3, struggles with generalization. While its low false positive rate could indicate caution, the model fails to correctly predict areas prone to burning.

At this stage, the competition appears to be between RF and MLP. RF displayed a slightly lower recall but performed well overall, while MLP achieved the highest recall of 0.029393, showing potential for refinement.

Test Wildfire - 2023/10/01

This test wildfire represents the last subset 1 evaluation, captured under clear skies in the western region of the study area.

Dataset Overview:

- Pixels labelled as not burned (target = 0): 24 024 620
- Pixels labelled as burned (target = 1): 2 953
- Burned/unburned ratio: 0.01229%

Table 0-5 compares the actual percentage of burned pixels (‘Burned (%)’ column) with the predictions of each model (‘RF (%)’, ‘XGBoost (%)’ and ‘MLP (%)’ columns) across different land classifications. Unfortunately, MLP misclassified wetlands again as prone to burn, a recurring issue that might indicate the need for a more complex architecture or additional iterations during training. RF performed comparably to its earlier evaluations, while XGBoost remained overly conservative, detecting almost no burned areas.

Table 0-5 Percentage of Burned Predicted Points across Land Classification classes for each model, regarding on the wildfire event of 01/10/2023.

COS18n1_L	**Burned (%)**	**RF (%)**	**XGBoost (%)**	**MLP (%)**	**Dataset (%)**
Wetlands	0	0	0	12.79	0.65
Surface water bodies	0	0.70	0	1.04	4.14
Artificialized territories	0	0.85	0.01	0.57	3.16
Uncovered spaces or spaces with little vegetation	0	1.03	0.11	1.39	0.33
Pastures	0	0.63	0	0.64	9.62
Agriculture	0	2.72	0.03	0.98	24.10
Agroforestry areas	0	1.75	0	1.91	9.45
Scrublands	0.08	23.43	0.07	27.89	13.34
Forests	0	32.41	0.26	35.88	35.21

Burned (%): This column represents the percentage of points that actual burned within each COS class. RF (%), XGBoost (%) and MLP (%) columns represent the percentage of points predicted to burn within each COS class.

The performance metrics for each model are summarized in Table 0-6. Both MLP and RF delivered reasonable results given the challenges presented by the dataset’s extreme imbalance and the small burned area. The high percentage of false positives was anticipated due to the dataset’s nature and the large study area relative to the burned region.

Table 0-6 Performance metrics of each model for the wildfire event on 01/10/2023

Model	TP	FP	TN	FN	Recall	Accuracy
RF	714	3 717 940	20 306 680	2 239	0.241788	0.845170
XGBoost	0	25 959	23 998 661	2 953	0	0.998797
MLP	1 160	4 078 181	19 946 439	1 793	0.392821	0.830196

SUBSET 2

Test Wildfire - 2023/07/26

The first test wildfire for subset 2 is located in the Beja district, with clear skies in the satellite imagery.

Dataset Overview:

- Pixels labelled as not burned (target = 0): 24 023 107
- Pixels labelled as burned (target = 1): 4 055
- Burned/unburned ratio: 0.01688%

Table 0-7 compares the actual percentage of burned pixels (‘Burned (%)’ column) with the predictions of each model (‘RF (%)’, ‘XGBoost (%)’ and ‘MLP (%)’ columns) across different land classifications. The MLP model failed to make any predictions, which suggests overfitting during training, resulting in its inability to generalize to this test scenario. While XGBoost remained conservative, it predicted significantly more areas prone to burn compared with XGBoost subset 1. RF performed well in vegetation-dense areas, but unfortunately also classified some wetland areas as prone to burn.

Table 0-7 Percentage of Burned Predicted Points across Land Classification classes for each model, regarding on the wildfire event of 26/07/2023.

COS18n1_L	Burned (%)	RF (%)	XGBoost (%)	MLP (%)	Dataset (%)
Wetlands	0	5.14	0.91	0	0.65
Surface water bodies	0	2.79	1.14	0	4.14
Artificialized territories	0.03	1.50	0.51	0	3.16
Uncovered spaces or spaces with little vegetation	0	1.63	0.64	0	0.33
Pastures	0.04	2.50	3.83	0	9.62
Agriculture	0	1.85	2.39	0	24.10
Agroforestry areas	0	3.05	2.56	0	9.45
Scrublands	0.04	45.40	25.92	0	13.34
Forests	0.02	40.41	16.43	0	35.21

Burned (%): This column represents the percentage of points that actual burned within each COS class. RF (%), XGBoost (%) and MLP (%) columns represent the percentage of points predicted to burn within each COS class.

Table 0-8 summarizes the performance metrics for each model. RF detected twice as many true positives as XGBoost but at the cost of more than double of the false positives. Based on this test alone, both models exhibit reasonable performance, with RF taking a riskier approach and XGBoost adopting a more cautious strategy. Let’s proceed to the next wildfire test to draw broader conclusions.

Table 0-8 Performance metrics of each model for the wildfire event on 26/07/2023

Model	TP	FP	TN	FN	Recall	Accuracy
RF	1 088	5 155 470	18 867 637	2 967	0.268311	0.785308
XGBoost	544	2 522 105	21 501 002	3 511	0.134155	0.894885
MLP	0	0	24 023 107	4 055	0	0.999831

Test Wildfire - 2023/08/05

This second and final test wildfire for subset 2 is located along the west coast of Portugal. Although the satellite imagery contains scattered clouds, the actual burned area has minimal cloud cover, making it suitable for testing the models' performance under slightly adverse conditions.

Dataset Overview:

- Pixels labelled as not burned (target = 0): 23 839 153
- Pixels labelled as burned (target = 1): 188 009
- Burned/unburned ratio: 0.78866%

Table 0-9 compares the actual percentage of burned pixels ('Burned (%)' column) with the predictions of each model ('RF (%)', 'XGBoost (%)' and 'MLP (%)' columns) across different land classifications. The patterns observed in this COS land classification table align with those from the first test wildfire for subset 2.

Table 0-9 Percentage of Burned Predicted Points across Land Classification classes for each model, regarding on the wildfire event of 05/08/2023.

COS18n1_L	**Burned (%)**	**RF (%)**	**XGBoost (%)**	**MLP (%)**	**Dataset (%)**
Wetlands	0	4.33	0.65	0	0.65
Surface water bodies	0.01	2.57	0.98	0	4.14
Artificialized territories	0.04	1.49	0.49	0	3.16
Uncovered spaces or spaces with little vegetation	0	1.07	0.33	0	0.33
Pastures	0.35	3.03	3.75	0	9.62
Agriculture	0.19	2.14	2.15	0	24.10
Agroforestry areas	0	4.25	2.93	0	9.45
Scrublands	1.03	41.57	23.02	0	13.34
Forests	1.61	38.20	15.31	0	35.21

Burned (%): This column represents the percentage of points that actual burned within each COS class. RF (%), XGBoost (%) and MLP (%) columns represent the percentage of points predicted to burn within each COS class.

Table 0-10 summarizes the performance metrics for each model. Both RF and XGBoost performed well in ignoring the cloud and shadow-covered areas, correctly not classifying them as prone to burn. However, RF achieved significantly better recall, demonstrating its ability to detect a larger number of true positives, though at the cost of a higher false positive rate. XGBoost, while more conservative, struggled to step up in this test scenario, maintaining lower recall than RF. Unfortunately, MLP completely failed again, predicting no burned areas.

Table 0-10 Performance metrics of each model for the wildfire event on 05/08/2023

Model	TP	FP	TN	FN	Recall	Accuracy
RF	105 346	4 793 547	19 045 606	82 663	0.560324	0.797054
XGBoost	42 102	2 283 412	21 555 741	145 907	0.223936	0.898893
MLP	0	0	23 839 153	188 009	0	0.992175

While the aggregated results will finalize the analysis, it is already evident that RF outperformed the other models in this scenario.

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa