

POLITECNICO DI TORINO

MASTER's Degree in Data Science and Engineering



**Politecnico
di Torino**

MASTER's Degree Thesis

Machine Learning Approaches for Estimating Wildfire Propagation

Supervisors

Prof. Paolo GARZA

Dr. Luca BARCO

Dr. Andrea BRAGAGNOLO

Candidate

Nicola BAVARO

OCTOBER 2025

"The world is yours"
Scarface, 1983

Abstract

Wildfires are a growing environmental and socio-economic concern, with climate change driving more frequent and severe events. Remote sensing has enabled accurate post-event burned area mapping, yet predictive approaches that anticipate fire impact before ignition remain limited and highly challenging. The complex interactions between environmental, climatic, and topographic factors hinder the development of robust pre-event models. This thesis investigates the feasibility of proactive wildfire prediction through a custom multimodal deep learning architecture designed to estimate burned area extent from pre-fire observations. The framework integrates diverse data sources, including Sentinel and Landsat optical imagery, Copernicus Digital Elevation Model, landcover information, and meteorological variables from ERA5, each processed through dedicated encoders tailored to the nature of the input. The feature representations are subsequently fused through specialized fusion blocks, enabling the model to capture complementary aspects of the wildfire dynamics within a unified predictive structure. Conducted as an experimental study, this work does not aim to deliver an operational system but to explore the opportunities and challenges of shifting from post-event segmentation to pre-event prediction. The best configuration achieved a maximum Intersection over Union (IoU) of **38.59%**, demonstrating both the potential and the limitations of the approach. The results enhance the study of employing multimodal deep learning for forecasting wildfire risk.

Keywords: Wildfire prediction; Burned area estimation; Remote sensing; Multimodal deep learning; Pre-fire modeling; Anticipatory risk assessment

ACKNOWLEDGMENTS

I would first like to express my sincere gratitude to my supervisor, Paolo Garza, for his invaluable guidance and support throughout the development of this thesis.

A special thanks goes to Luca Barco and Andrea Bragagnolo from LINKS Foundation for their constant support during the research activities and for their precious help, from the project phase to the writing process. I am also thankful to EDISU, whose financial aid greatly supported me and made it possible to live and study in this city.

I am deeply grateful to my family, who have always been close to me, even from a distance, with unwavering support and encouragement.

My heartfelt thanks also go to all my friends, scattered across the globe, who have enriched me both socially and technically. Thank you for the fun, the inspiration, and the unforgettable moments we shared. I am grateful to my longtime friends from my hometown for always being a safe harbor in times of challenge. I want to especially thank my girlfriend, Paola, for her unwavering support during this past year, which was particularly challenging mentally.

I would be lying if I said this journey was simple, but thanks to the support of everyone who cares for me, I always found a way to keep going and overcome the only real enemy: my own mind, which sometimes works against me. Finally, I would like to thank myself, for always being hungry for more and never giving up.

Table of Contents

1	Introduction	1
2	Related Works	4
2.1	Wildfire Risk Assessment Frameworks	4
2.1.1	From Fire Danger to Wildfire Risk.	4
2.1.2	Danger: Fire Ignition and Propagation.	4
2.1.3	Exposure: Assets and Ecosystems at Risk.	6
2.1.4	Vulnerability: Susceptibility and Resilience.	6
2.1.5	Integrating the Dimensions: From Indicators to Indices.	6
2.2	Machine Learning for Fire Mapping	7
2.2.1	Deep Learning Approaches for Burned Area Mapping	8
2.2.2	Multi-Task Learning for Burned Area and Land Cover Integration	9
2.2.3	Predictive Modeling of Fire Behavior and Severity	10
2.2.4	Multimodal Approaches and Fire Spread Prediction	11
2.2.5	Infrastructure Integration and Spatial Analysis	11
2.2.6	Final Burned Area Prediction and Early Fire Characteristics	12
2.2.7	Positioning and Novel Contributions of This Work	12
3	Dataset	14
3.1	Data Sources	14
3.1.1	Piedmont Civil Protection Data	14
3.1.2	Sentinel-2 Level 2A	15
3.1.3	Landsat Collection 2 Level-2	17
3.1.4	Copernicus DEM GLO-30	18
3.1.5	Copernicus ERA5	19
3.1.6	Piedmont Streets Data	20
3.1.7	Dynamic World Land Cover (IO-LULC Annual v2)	20
3.1.8	Ground Truth Mask	21
3.2	Data Processing	21
3.2.1	Sentinel-2 Image Selection and Preparation	21
3.2.2	Landsat Image Selection and Preparation	23
3.2.3	Digital Elevation Model (DEM) Data	24
3.2.4	ERA5-Land Processing	25
3.2.5	Land Cover Data	26

3.2.6	Streets Rasterization and Preparation	27
3.2.7	Cloud Masking with CloudSen12	28
3.3	Dataset Information	29
3.3.1	Fire Size Distribution	30
4	Methodology	31
4.1	Problem Statement	31
4.2	Multi-Modal Approach	31
4.3	Network Architectures	33
4.3.1	Feature Pyramid Network	34
4.3.2	ResNet	35
4.3.3	Multi-Layer Perceptron for Tabular Data Processing	37
4.3.4	Multi-Modal Fusion Blocks	38
4.3.4.1	Kernel Size Configuration	38
5	Experiment and Results	40
5.1	Implementation Details	40
5.1.1	Experimental Setup	40
5.1.2	Training Process	41
5.2	Loss Functions	44
5.3	Evaluation Metrics	45
5.4	Results	46
6	Conclusion	51
	Bibliography	53

List of Figures

2.1	Conceptual integration of wildfire risk assessment components within the FirEUriSk project. This scheme could also be eventually applied to risk reduction and adaptatio [12]	7
3.1	Sentinel-2 image patch (256×256 pixels at 10 m resolution) centered on the fire perimeter.	22
3.2	Landsat image patch (256×256 pixels at 30 m resolution) centered on the fire perimeter.	24
3.3	Landsat image patch after resampling (10 m resolution).	24
3.4	Dem image patch (256×256 pixels at 10 m resolution) centered on the fire perimeter.	25
3.5	Landcover image patch (256×256 pixels at 10 m resolution) centered on the fire perimeter.	27
3.6	Streets raster image patch (256×256 pixels at 10 m resolution) centered on the fire perimeter.	28
3.7	Example of Sentinel-2 image (b) and corresponding CloudSen12 output mask (a).	29
3.8	Distribution of burned area sizes (in pixels).	30
4.1	Burned Scars for the 1998 summer forest fires on the Elba Island, overlaid on a Landsat 5 TM false colour composite. [38]	32
4.2	Schematic representation of the proposed multimodal architecture	33
4.3	Feature Pyramid Network architecture (adapted from Lin et al. [39]). The bottom-up pathway extracts hierarchical features, while the top-down pathway with lateral connections (1×1 convolutions) enables multi-scale feature fusion.	35
5.1	Examples of qualitative comparison: Sentinel-2 pre-fire acquisition (left), ground truth burned area (center), and predicted burned area (right).	50

List of Tables

3.1	Sentinel-2 L2A spectral bands characteristics	16
3.2	Landsat Collection 2 Level-2 bands used in this research (*resampled from 100m)	18
3.3	CloudSen12 mask classes, corresponding description and color. . . .	28
3.4	Statistical summary of fire sizes (in pixels, 10 m resolution).	30
4.1	ResNet backbone configurations for different input modalities	36
5.1	Python packages and versions used in the experiments	41
5.2	Summary of training hyperparameters and configuration settings . .	43
5.3	Data augmentation parameters used in the proposed pipeline.	44

Chapter 1

Introduction

Catastrophic hazards are of different nature, like floods, earthquakes, and severe storms. These events represent potential threats from many perspectives: humanitarian, economic, and environmental [1].

Among them, forest fires have always stood out for their destructive power and frequency. Floods, earthquakes and violent storms certainly represent a significant threat, but wildfires in particular have the capability to affect multiple areas simultaneously. Beside environmental destruction, they also cause a devastating impact on the economy. An additional level of difficulty is their propagation and spreading in remote or difficult-to-access areas: their detection and timely intervention to extinguish them can be problematic.

From an environmental perspective, local flora and fauna suffer from the destruction of the land, radically altering native ecosystems and endangering biodiversity. Therefore, the combustion of the forests emits a huge quantity of CO₂ in the atmosphere, accelerating climate change. Recent data demonstrates the magnitude of this impact: during the 2023-2024 fire season, global fire carbon emissions were 16% above average, totaling 2.4 Pg C [2]. The Canadian boreal forests alone experienced unprecedented emissions, with fires emitting nine times more carbon than in recent decades, releasing about 640 million metric tons of carbon [3, 4]. Additionally, the reduction of vegetation has long-term environmental impacts: the area's ability to absorb CO₂ is compromised, and exposed soil without a healthy tree cover increases the risk of erosion, then causing events such as landslides and floods.

Finally, smoke and particulate emissions caused by fire contribute to air pollution, with potentially harmful effects on air quality and human health. Such phenomena when they occur can even last for days. Winds can push smoke into remote areas, even far away from the fires. Recent examples demonstrate this impact vividly: the devastating January 2025 Los Angeles wildfires not only caused unprecedented destruction to property but also created unhealthy air quality conditions for millions of people across Southern California [5]. Similarly, in early June 2023, New York City was shrouded in smoke from Canadian wildfires, demonstrating the transboundary nature of wildfire impacts [6].

From an economic aspect, fires also cause extensive damage to private property,

causing significant economic impacts to the affected area. For Southern Europe alone, wildfires cause an average annual production loss of 13-21 billion euros per season [7]. The recent Los Angeles wildfires exemplify the extreme scale of potential destruction, with estimated total property and capital losses ranging between \$76 billion and \$131 billion [8]. In addition, the efforts required to extinguish the forest fire and the subsequent recovery and restoration after the event must be taken into account. Local heritage and historic places may also suffer irreparable losses.

The restoration of an area after a fire could take a long time. Depending on the extent, damage severity, and the specific characteristics of the soil and ecosystem, a full recovery can take decades or even centuries [9]. Therefore, fire prevention becomes even more important: prompt intervention to extinguish fires in their early stages and monitor closely burned areas after the flames have been extinguished.

Global awareness of these kinds of global issues is steadily increasing, underlined by the continuing reality of escalating wildfire seasons worldwide. 2024 was the most extreme year for forest fires on record, with at least 13.5 million hectares of forest burned [10]. As nations grapple with increased fire risk and escalating drought conditions, the need for robust, coordinated international responses becomes ever more urgent. In Europe, 2023 was among the worst wildfire years in this century, producing some 20 megatonnes of CO₂ emissions [11].

Facing this challenge, this thesis involves the implementation of a predictive model that can anticipate burned areas before fires occur through the analysis of pre-fire satellite imagery and environmental data. Unlike traditional post-fire assessment approaches, this research focuses on developing a proactive system that can predict the potential extent of fire spread based on pre-fire conditions.

In this context, the application of deep learning models is proving to be a promising approach for wildfire prevention and management. The objective of this thesis is intended to add a contribution to the promotion of a more proactive and resilient approach to fire risk management.

Specifically, the developed model aims to predict the final burned area using pre-fire imagery from satellites (primarily Sentinel, with Landsat when available) combined with additional environmental data sources including Digital Elevation Models (DEM), street network data, ERA5 meteorological data, landcover and ignition point information. Through this multisource approach, a novel multimodal neural network architecture is developed, employing different encoders for each data modality to capture and integrate the complex relationships between environmental conditions, topography, infrastructure, and fire behavior for predicting the potential extent of fire spread.

The implementation of such a predictive model represents an experimental research approach to traditional reactive fire monitoring methods from literature. This research explores the potential for fire management strategies, particularly in scenarios where a fire has just ignited and rapid assessment of potential spread is critical for emergency response coordination. By attempting to predict fire spread patterns at ignition, emergency services could potentially improve resource allocation and evacuation

planning decisions.

In this thesis, a custom multimodal deep learning architecture is developed and evaluated for its performance in predicting burned area extent by processing pre-fire satellite acquisitions and environmental data through specialized encoders for each data type. This experimental research represents an exploratory investigation into the feasibility of wildfire management through predictive modeling, acknowledging the inherent complexity and challenges of predicting natural phenomena. This approach signifies a new research trajectory that shifts focus from evaluating events after they occur to anticipating them beforehand, thereby enriching the expanding knowledge base in computational wildfire management.

Chapter 2

Related Works

2.1 Wildfire Risk Assessment Frameworks

Wildfire risk assessment has evolved from a predominantly reactive and hazard-centered approach to a more comprehensive and systemic understanding of wildfire impacts. In particular, recent frameworks seek to integrate not only the likelihood of fire occurrence and propagation, but also the exposure and vulnerability of human and ecological assets. One of the most influential contributions in this domain is the work of Chuvieco et al. [12], who propose a structured and harmonized approach aligned with terminology adopted in the broader natural hazards literature. Their framework explicitly embraces the conceptual triad of **hazard (or danger)**, **exposure**, and **vulnerability**, thereby aligning wildfire risk science with frameworks used in flood risk, earthquake resilience, and climate adaptation.

2.1.1 From Fire Danger to Wildfire Risk.

Historically, the wildfire community has used terms like *fire danger* and *fire hazard* somewhat interchangeably, often focusing on immediate weather and fuel conditions to estimate fire behavior potential. However, this terminology lacked consistency and limited the scope of risk assessment to short-term forecasting. Chuvieco et al. argue for the adoption of a broader definition, where wildfire risk is conceived as “*the potential for adverse consequences or impacts due to the interaction between one or more natural or human-induced hazards, the exposure of humans, infrastructure, and ecosystems, and the systems’ vulnerabilities*” [12]. In this view, wildfire risk is not a unidimensional output of fuel and weather conditions, but a function of the complex interplay between fire occurrence, landscape characteristics, societal values, and resilience capacities.

2.1.2 Danger: Fire Ignition and Propagation.

The first core dimension, referred to as *Danger* (or *Hazard*, though the term “danger” is preferred in wildfire science), represents the likelihood that a fire will both ignite and spread. Danger encompasses two interconnected processes: ignition and propagation.

Ignition sources can be broadly categorized into natural and anthropogenic causes. Human ignitions are responsible for approximately 90% of wildfires globally, according to FAO estimates, and their spatial distribution is shaped by variables such as population density, proximity to roads and urban edges, livestock activity, and socio-political factors. Modeling human ignition probability often requires regionally-specific proxies and, increasingly, data-driven methods such as Random Forest classifiers using historical fire occurrence data.

In contrast, natural ignitions are primarily caused by lightning. Of particular interest are lightning discharges with long continuing current (LCC) — typically lasting over 40 milliseconds — and those of positive polarity, which are more likely to initiate combustion. The assessment of natural ignition potential remains challenging due to the so-called *holdover phase* — a time-lag between lightning strike and visible flame — which complicates both detection and prediction. Nevertheless, statistical models using lightning density from Numerical Weather Prediction (NWP) systems, combined with fuel and moisture indicators, are showing promise in estimating ignition likelihood.

Propagation potential, the second subcomponent of danger, reflects how fire spreads across the landscape, its speed, intensity, and direction. It is controlled by a combination of topography, meteorological conditions, and fuel characteristics.

- **Weather and climate** factors — including temperature, humidity, precipitation, and wind — directly affect fuel flammability and fire spread rates. Climate extremes, in particular, are increasingly recognized as key drivers of large and destructive fire events. Long-term shifts in climate patterns are also altering fuel loads, fire season duration, and lightning regimes.

- **Topography** plays a critical role by influencing wind patterns, solar radiation, and moisture distribution. Slope steepness accelerates uphill fire spread through pre-heating, while concave landforms such as canyons can amplify fire intensity due to wind funneling effects.

- **Fuel properties** are fundamental to propagation dynamics. Fuels are characterized by their biomass load, vertical and horizontal structure, moisture content, and flammability traits. Two specific measures of fuel moisture are commonly used: *Dead Fuel Moisture Content (DFMC)* is a short-term indicator influenced by weather conditions. It determines how much energy is required to vaporize moisture before ignition, and is a core component of many meteorological fire danger indices, *Live Fuel Moisture Content (LFMC)* varies on longer time scales and is influenced by plant physiology and climatic conditions. While more difficult to measure, LFMC is increasingly estimated using remote sensing data (e.g., multispectral sensors on MODIS, Sentinel-2, or microwave instruments) in combination with radiative transfer models.

To translate vegetation heterogeneity into operational metrics, fuels are often abstracted into *fuel models* — such as those developed by Rothermel, Scott and Burgan, or the FCCS system — which capture generalized combinations of fuel types, loads, and moisture scenarios.

2.1.3 Exposure: Assets and Ecosystems at Risk.

The second major dimension is *Exposure*, which refers to the presence of people, infrastructure, or ecological assets in areas that could be affected by wildfire. Exposure may be direct (e.g., damage to buildings or forests by fire) or indirect (e.g., smoke exposure, disruption of hydrological cycles, or soil degradation). Exposure assessment typically involves overlaying modeled fire behavior (e.g., flame length, burn probability, fireline intensity) with geospatial datasets describing asset locations — such as census data, land use maps, or habitat distributions.

A particularly important concept is the *Wildland-Urban Interface (WUI)*, which designates the spatial boundary between developed areas and wildland vegetation. WUI regions are highly exposed due to the convergence of ignition sources (e.g., human activity) and combustible landscapes. Mapping WUI zones has become a critical component of fire risk modeling and emergency planning.

2.1.4 Vulnerability: Susceptibility and Resilience.

Vulnerability represents the degree to which exposed elements — whether human, economic, or ecological — are susceptible to damage, and their capacity to resist or recover. A region with high-value assets and low recovery capacity is considered highly vulnerable.

Vulnerability encompasses several subdimensions:

- **Socio-economic vulnerability** includes both tangible losses (e.g., destroyed homes, infrastructure, economic disruption) and intangible values (e.g., cultural heritage, public health). These can be monetized using asset valuations or indirect indicators such as population density and social deprivation indices.
- **Ecological vulnerability** refers to the potential degradation of ecosystem services and biodiversity. Wildfires can compromise timber production, carbon sequestration, grazing land, and recreational resources. Ecological values may also include biodiversity indices (species richness, rarity), habitat connectivity, and conservation status (e.g., presence of protected areas).
- **Resilience**, closely related to vulnerability, denotes the capacity of a system to resist damage and return to its pre-fire state. In ecological systems, resilience is influenced by species traits (e.g., resprouting ability, bark thickness), abiotic conditions (e.g., soil moisture), and disturbance history (e.g., fire return interval). In human systems, social resilience is linked to governance, infrastructure, and community adaptive capacity.

2.1.5 Integrating the Dimensions: From Indicators to Indices.

A critical challenge in wildfire risk assessment is the integration of heterogeneous indicators — ranging from physical to social variables — into synthetic, actionable indices. Chuvieco et al. describe several strategies to harmonize and combine these components:

- **Normalization and scaling:** Raw indicators are converted into a common scale

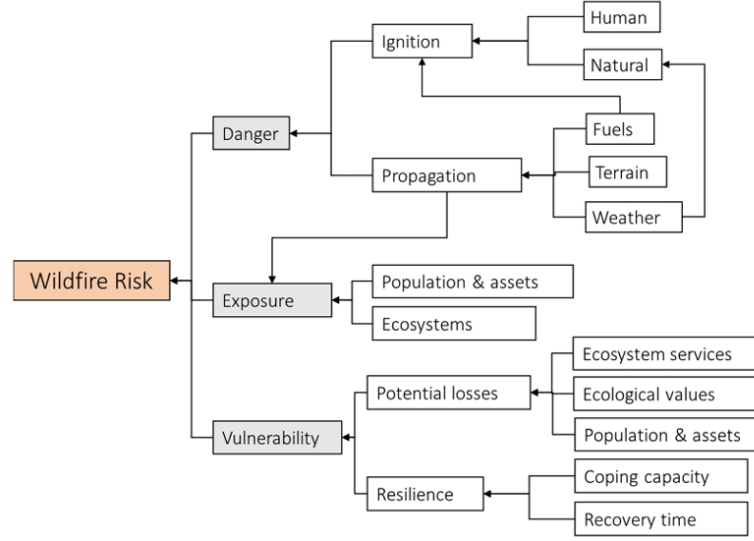


Figure 2.1: Conceptual integration of wildfire risk assessment components within the FirEUriSk project. This scheme could also be eventually applied to risk reduction and adaptatio [12]

(e.g., 0–1 or 0–100), using min-max rescaling, qualitative classes, or probabilistic transformations.

- **Index aggregation:** Once standardized, components can be aggregated using conceptual formulas (e.g., $\text{danger} \times \text{vulnerability}$), multicriteria decision analysis, or statistical models.
- **Probabilistic modeling:** Techniques such as logistic regression, geographically weighted regression, or machine learning algorithms (e.g., Random Forests, SVMs, MaxEnt) can be trained on historical fire occurrence data to derive risk surfaces.
- **Ranking approaches:** Pareto-based ranking systems can preserve ordinal relationships among risk values, as used in some continental-scale assessments.
- **Weighted averaging:** Inspired by systems like the Environmental Performance Index, indicators are combined using expert-derived weights reflecting their perceived importance. An example of an operational system implementing this integrated framework is the European Union’s **FirEUriSk** project. FirEUriSk provides risk assessments at multiple spatial resolutions — from continental-scale (1 km²) to high-resolution pilot sites (0.01 km²) — and temporal scales, supporting both near-real-time and historical scenario analysis.

2.2 Machine Learning for Fire Mapping

The intersection of wildfire research and machine learning has evolved dramatically over the past decade, driven by the increasing availability of high-resolution satellite imagery, advances in computational methods, and the urgent need for more effective wildfire management strategies. While traditional approaches to wildfire analysis have predominantly focused on post-event assessment and damage quantification, the field

is experiencing a paradigm shift toward predictive modeling that leverages pre-fire environmental conditions and real-time data to support proactive decision-making and emergency response operations.

2.2.1 Deep Learning Approaches for Burned Area Mapping

The application of deep learning techniques to burned area mapping represents one of the most significant methodological advances in wildfire remote sensing. Knopp et al. (2020) [13] pioneered the development of a fully automated processing chain based on deep learning for rapid burned area segmentation using mono-temporal Sentinel-2 imagery. Their approach addresses critical limitations of traditional methods that typically require extensive preprocessing, multi-temporal data acquisition, or complex rule-based classification systems—factors that significantly hinder operational deployment in time-critical scenarios. The core innovation lies in their implementation of a U-Net architecture for semantic segmentation, originally developed for biomedical image analysis but adapted for remote sensing applications due to its ability to learn effectively from limited training data while maintaining a relatively compact parameter space.

The methodological significance of Knopp et al.’s work extends beyond the neural architecture itself. Their systematic evaluation of spectral band combinations led to the identification of an optimal six-band configuration (Blue, Green, Red, Near Infrared, and two Shortwave Infrared bands - specifically B2, B3, B4, B8, B11, and B12 from Sentinel-2) that balances classification accuracy with computational efficiency. Particularly relevant to the present work is their demonstration that Level L1C data—imagery that has not undergone atmospheric correction—can achieve high accuracy levels while dramatically reducing preprocessing time and complexity. This finding challenges conventional wisdom in remote sensing applications and opens possibilities for near real-time processing scenarios. Furthermore, their processing chain incorporates sophisticated handling of invalid pixels through a separate CNN module that identifies and masks clouds, cloud shadows, snow, and ice, providing a comprehensive framework for managing atmospheric interference in operational settings. Building upon these foundational developments, Ban et al. (2020) [14] introduced Burnt-Net, an end-to-end deep learning framework that combines morphological operations with multi-scale residual blocks to enhance burned area detection accuracy. Their approach demonstrates the value of incorporating domain-specific knowledge about fire-affected landscape morphology into neural network architectures, achieving improved performance in identifying burned area boundaries and reducing false positive classifications in spectrally similar land cover types such as agricultural fields and exposed soil. The synergistic integration of multi-sensor data has emerged as another important development in deep learning-based burned area mapping. Farasin et al. (2020) [15] demonstrated significant performance improvements through the combined use of Sentinel-1 SAR and Sentinel-2 optical data, leveraging the complementary information provided by active and passive

remote sensing systems. Their work illustrates how SAR data can penetrate cloud cover and provide structural information about vegetation changes, while optical data contributes detailed spectral signatures of burned surfaces. This multi-sensor approach is particularly valuable for operational applications where cloud coverage or smoke plumes may compromise optical imagery quality.

The transferability of deep learning models across different sensor platforms has been investigated by Hu et al. (2021) [16], who demonstrated successful knowledge transfer between Landsat and Sentinel-2 sensors using U-Net architectures. Their findings suggest that models trained on one sensor platform can maintain acceptable accuracy levels when applied to data from different sensors, provided appropriate spectral band mapping and radiometric adjustments are implemented. This cross-sensor transferability is crucial for operational systems that must integrate historical archives from multiple satellite missions.

2.2.2 Multi-Task Learning for Burned Area and Land Cover Integration

The integration of auxiliary tasks in deep learning architectures for burned area mapping has emerged as a promising approach to enhance model performance through shared feature learning. Recent work by Sdraka et al. (2024) [17] proposed a robust multitask learning framework that incorporates land cover classification as an auxiliary task to enhance the robustness and performance of burned area segmentation models, demonstrating improved accuracy by leveraging the inherent relationships between fire-affected landscapes and underlying land cover characteristics.

The theoretical foundation for multi-task learning in wildfire applications rests on the observation that burned area detection and land cover classification share common spectral and spatial feature representations. Fire-affected areas exhibit characteristic signatures that are intrinsically linked to pre-fire vegetation types and land cover conditions. By training networks simultaneously on both tasks, models develop more robust feature extractors that capture these underlying relationships while reducing overfitting to task-specific noise patterns. This shared learning paradigm enables the optimization process to benefit from complementary information provided by each objective, resulting in enhanced performance on both primary and auxiliary tasks.

Recent implementations of multi-task architectures typically employ shared encoder networks with task-specific decoder heads. The common feature extraction backbone processes input imagery through convolutional layers, developing hierarchical representations that capture both spectral patterns and spatial context relevant to both burned area delineation and land cover classification. Separate decoding pathways then transform these shared representations into task-specific predictions, with the combined loss function enabling simultaneous optimization across multiple objectives. Multi-scale convolutional approaches (Wang et al., 2021) [18] and transferable deep models (Liu et al., 2019) [19] have shown particular promise for land cover classification tasks, providing architectural foundations that can be adapted

for multi-task learning scenarios.

The practical advantages of multi-task learning extend beyond improved accuracy metrics. Models trained with auxiliary land cover tasks demonstrate enhanced generalization across diverse geographic regions and ecosystem types, as the land cover classification objective forces the network to learn discriminative features for various vegetation classes that subsequently improve burned area boundary delineation. This approach particularly addresses common challenges in burned area detection, such as confusion between burned areas and spectrally similar land cover types including bare soil, agricultural fields, and naturally sparse vegetation. By explicitly modeling land cover relationships, these networks better distinguish between actual fire-affected areas and naturally occurring spectral patterns that might otherwise generate false positive classifications.

2.2.3 Predictive Modeling of Fire Behavior and Severity

The transition from reactive burned area mapping to predictive fire behavior modeling represents a fundamental shift in wildfire research methodology. Klimas et al. (2025) [20] developed a comprehensive machine learning framework for predicting potential burn severity in unburned forest areas, addressing the critical need for proactive fuel management and erosion prevention strategies. Their Random Forest-based model incorporates twenty-three environmental variables spanning topographic characteristics, canopy and surface fuel properties, environmental conditions, meteorological parameters, and geographic positioning. Through systematic feature selection, seventeen predictors were retained in the final model, with Net Primary Productivity, elevation, and canopy fuel characteristics emerging as the most influential variables.

The methodological sophistication of Klimas et al.’s approach is evident in their treatment of meteorological data, particularly their calculation of the Energy Release Component (ERC) as a 31-day mean centered on fire ignition dates. This temporal aggregation approach captures the cumulative effects of weather conditions on fuel moisture and fire potential, providing a more robust predictor than instantaneous meteorological measurements. Their model’s ability to predict continuous differenced Normalized Burn Ratio (dNBR) values at 30-meter resolution enables fine-scale assessment of potential fire impacts, supporting precision management interventions and risk assessment protocols.

The performance characteristics of the Klimas et al. model reveal important insights about spatial scale effects in fire behavior prediction. While pixel-level predictions achieved an out-of-bag R^2 of 0.67, aggregation to hydrologic sub-catchment scales substantially improved performance ($R^2 = 0.75$), suggesting that fire behavior prediction benefits from spatial averaging that captures landscape-level patterns while reducing pixel-level noise. Their analysis also demonstrated the overriding influence of extreme meteorological conditions, showing that under high ERC scenarios, local vegetation characteristics become less predictive as fire behavior becomes increasingly weather-driven across diverse landscape types.

2.2.4 Multimodal Approaches and Fire Spread Prediction

The integration of heterogeneous data sources through multimodal machine learning architectures represents an emerging frontier in wildfire prediction research. Hodges and Lattimer (2019) [21] pioneered the application of CNNs for fire propagation prediction by integrating topographic, meteorological, and fuel data within unified neural network architectures. Their work demonstrated that CNNs can effectively learn spatial patterns of fire spread by processing gridded environmental data as multi-channel images, analogous to RGB image analysis but with environmental variables as input channels.

More recently, Huot et al. (2022) [22] developed sophisticated deep learning models that combine satellite imagery, weather data, and topographic information for next-day wildfire spread prediction. Their multimodal architecture employs separate encoder networks for different data modalities—a design philosophy that allows specialized feature extraction from heterogeneous data types while enabling cross-modal information fusion through shared decoder networks. This architectural approach is conceptually aligned with the methodology adopted in the present work, where different environmental data sources require specialized processing before integration into unified predictive models.

The work of Prapas et al. (2024) [23] further advanced multimodal approaches by demonstrating wildfire spread prediction using deep neural networks that integrate geospatial data from multiple sources. Their framework successfully incorporates static environmental variables (topography, land cover, climate normals) with dynamic meteorological data (weather forecasts, fuel moisture estimates) to predict fire spread patterns up to several days in advance. The success of their approach underscores the value of combining data sources with different temporal characteristics—static geographic features that provide baseline environmental context with dynamic variables that capture changing fire weather conditions.

Radke et al. (2019) [24] proposed FireCast, a system that combines satellite imagery with meteorological data for fire risk prediction, demonstrating the operational potential of integrated data approaches for supporting fire management decisions. Their work illustrates how machine learning models can synthesize information from disparate sources to provide actionable intelligence for fire suppression resource deployment and evacuation planning.

2.2.5 Infrastructure Integration and Spatial Analysis

The incorporation of human infrastructure data into fire prediction models has received increasing attention as researchers recognize the complex interactions between fire behavior and anthropogenic landscape modifications. Alcasena et al. (2018) [25] conducted comprehensive analyses of how road networks influence both fire ignition patterns and subsequent propagation dynamics. Their findings demonstrate that roads can serve dual roles as both fire barriers—through the creation of fuel breaks and providing access for suppression activities—and fire corridors—by facilitating

ignition from human activities and creating wind channels that accelerate fire spread. This nuanced understanding of infrastructure effects on fire behavior is particularly relevant for the present work’s integration of road network data as a predictive variable.

The spatial analysis methodologies developed by Alcasena et al. extend beyond simple distance-to-road calculations to include network topology measures, road density metrics, and accessibility indices that capture the complex ways infrastructure influences fire risk and suppression effectiveness. Their work provides a theoretical foundation for understanding how transportation networks should be incorporated into predictive fire models, emphasizing the importance of considering both positive and negative infrastructure effects on fire outcomes.

2.2.6 Final Burned Area Prediction and Early Fire Characteristics

Specific research addressing final burned area prediction using early fire characteristics or ignition-point data represents a relatively specialized but critically important subdomain of wildfire prediction. Artés et al. (2019) [26] developed statistical models that utilize topographic, meteorological, and vegetation characteristics measured at fire ignition points to predict ultimate fire size. Their approach recognizes that certain environmental conditions at the time and location of ignition can be strongly predictive of subsequent fire growth, providing valuable intelligence for initial attack resource allocation and strategic planning.

Recent advances in this area include the work of Silva et al. (2022) [27], who demonstrated the effectiveness of ensemble machine learning methods for burned area prediction using comprehensive environmental variable sets. Their comparative analysis of multiple algorithms—including Random Forest, Support Vector Machines, and Gradient Boosting approaches—revealed that ensemble methods consistently outperform single-algorithm approaches, suggesting that the complexity of fire-environment interactions benefits from diverse modeling perspectives.

2.2.7 Positioning and Novel Contributions of This Work

The present research occupies a unique position within the broader landscape of wildfire prediction studies through several distinctive methodological and conceptual innovations. Unlike existing approaches that focus primarily on post-fire burned area segmentation (such as Knopp et al.) or hypothetical severity assessment under various scenarios (such as Klimas et al.), this work addresses the specific challenge of predicting final burned area extent using exclusively pre-fire environmental data. This predictive focus, combined with the integration of ignition point coordinates as a key input variable, positions the research at the intersection of fire initiation studies and spread prediction modeling.

The methodological innovation lies in the comprehensive integration of multi-sensor satellite imagery (Sentinel and Landsat platforms), high-resolution Copernicus DEM data, ERA5 meteorological reanalysis information, ignition point coordinates,

and road network data within a unified multimodal neural architecture. This data fusion approach draws methodological elements from multiple streams of existing literature while combining them in a novel configuration specifically designed for pre-fire prediction scenarios. The use of separate encoders for different data modalities allows for specialized feature extraction that respects the unique characteristics of each data source while enabling sophisticated cross-modal information integration through shared decoder networks.

A distinctive aspect of this work is the implementation of multi-task learning through the integration of land cover classification as an auxiliary task alongside primary burned area prediction. This approach leverages the complementary relationship between land cover characteristics and fire behavior patterns, enabling the model to develop more robust feature representations that benefit both tasks simultaneously. The shared encoder-decoder architecture with task-specific output heads allows the network to learn common underlying patterns while maintaining specialized prediction capabilities for each objective. This multi-task formulation not only improves burned area prediction accuracy through enhanced feature learning but also provides additional operational value by generating land cover information useful for post-fire assessment and management applications.

The temporal focus on pre-fire prediction distinguishes this work from the majority of existing research that either analyzes post-fire conditions or predicts hypothetical scenarios without reference to actual ignition events. By conditioning predictions on known ignition points and pre-fire environmental states, this approach provides a framework for near real-time decision support that could inform emergency response protocols, resource allocation strategies, and evacuation planning during the critical early hours of fire events when intervention effectiveness is typically highest.

Chapter 3

Dataset

This chapter focuses on the description and construction of the dataset. The first section presents the data sources and services used to download the required information, while the second section details the review and preprocessing steps applied to the data.

3.1 Data Sources

3.1.1 Piedmont Civil Protection Data

One of the core sources used to build the wildfire dataset was provided by the Civil Protection of the Piedmont Region [28]. The data package included the following files:

- **Shapefiles:**
 - `INC_SDO_PL_INCENDI.shp`: a point shapefile containing the ignition locations of wildfires.
 - `INC_SDO_PL_INCENDIO.shp`: a polygon shapefile contains the final burned area geometries for only a subset of the fires — specifically those for which the perimeter was large enough to be mapped or recorded.
- **CSV file:**
 - A tabular file containing metadata for each fire, such as ID, start date, area, and location. The CSV and shapefiles share a common identifier, `ID_INCENDI`, which allows for joining and relational querying across files.

After filtering the dataset to the 2012–2024 period, a total of 2,152 fire records were retained. Among these, 2,150 were associated with ignition points, while only 1,233 included a corresponding burned area polygon. This means that just over half of the recorded fires had spatial geometries representing the final burned perimeter, whereas the remainder were excluded from the analysis—either because they were too small, extinguished rapidly, or lacked sufficient information for perimeter mapping.

Using the common identifier `ID_INCENDI`, the area shapefile was joined with the corresponding metadata in the CSV. The resulting dataset was exported as a GeoJSON file named `piedmont_fa.geojson`, reprojected into EPSG:3857 (Web Mercator). Each of the 1,233 entries includes:

- `id`: Fire event ID
- `initialdate`: Fire start date
- `finaldate`: Set to `null` (not available)
- `area_ha`: Burned area in hectares
- `point_x`, `point_y`: Ignition coordinates projected to EPSG:3857
- `geometry`: Polygon or multipolygon of the burned area
- `country`, `iso2`, `iso3`: Static metadata (e.g., `Italia`, `IT`, `ITA`)
- `admlvl1`, `admlvl2`, `admlvl5`: Administrative region descriptors (e.g., `Nord-Ovest`, `Piemonte`, local toponym from `LOCALITA`)

Additional fields following the EFFIS schema—such as vegetation class percentages and Natura2000 coverage—were retained as placeholders for schema compatibility but not populated, as this data was not available from the Civil Protection source.

This curated dataset, served as the spatial foundation for the entire project. It enabled the automated querying and retrieval of pre- and post-fire satellite imagery from multiple platforms, including:

- **Sentinel-2** and **Landsat**: multispectral imagery for vegetation and burn index computation
- **Impact Observatory Land Cover**: thematic land cover data
- **Copernicus ERA5 Data**: wind, temperature and pressure data
- **Copernicus Digital Elevation Models (DEM)**: terrain and elevation analysis

The spatial and temporal precision of the `piedmont_fa.geojson` file was crucial for aligning satellite data to fire events, allowing for targeted image extraction, masking, and further analysis described in the Methodology chapter.

3.1.2 Sentinel-2 Level 2A

The Sentinel-2 mission consists of two polar-orbiting satellites that capture multispectral imagery at 10-60 meter spatial resolution across 13 spectral bands [29]. These satellites are designed primarily for land monitoring applications, including agricultural assessment, forest management, and environmental change detection. Sentinel-2

data is distributed in two main processing levels. Level-1C products contain top-of-atmosphere reflectance values with geometric correction but without atmospheric correction. Level-2A products undergo additional processing through the Sen2Cor algorithm, which removes atmospheric effects to provide bottom-of-atmosphere reflectance values. This atmospheric correction process eliminates interference from water vapor, aerosols, and other atmospheric components, resulting in surface reflectance data that is more suitable for quantitative analysis and multi-temporal studies. The Sentinel-2 instrument captures data across 13 spectral bands with varying spatial resolutions. Four bands (Blue, Green, Red, Near-Infrared) are provided at 10m resolution, six bands (Red Edge and SWIR bands) at 20m resolution, and three atmospheric bands at 60m resolution. In the L2-A version of the dataset there are 12 bands which are presented in the following table:

Band	Description	Wavelength (nm)	Resolution (m)
B01	Coastal Aerosol	443	60
B02	Blue	490	10
B03	Green	560	10
B04	Red	665	10
B05	Vegetation Red Edge	705	20
B06	Vegetation Red Edge	740	20
B07	Vegetation Red Edge	783	20
B08	Near Infrared	842	10
B8A	Vegetation Red Edge	865	20
B09	Water Vapour	945	60
B11	SWIR	1610	20
B12	SWIR	2190	20

Table 3.1: Sentinel-2 L2A spectral bands characteristics

For this research, Sentinel-2 L2A data was acquired through Microsoft Planetary Computer [30], a cloud-based geospatial data platform that provides free access to analysis-ready Earth observation datasets. The platform offers Sentinel-2 data in Cloud Optimized GeoTIFF format with STAC-compliant metadata, enabling efficient data discovery and access.

In each query to Planetary Computer the following parameters are needed:

- *collections*: Specifies the STAC collection to search, in this case Sentinel-2 L2A products.

- *bbox*: Defines the bounding box coordinates derived from the fire geometry with search buffer, specifying the geographic boundaries for image retrieval.
- *datetime*: Sets the temporal constraints for the search using a time range string relevant to the study objectives.
- *limit*: Maximum number of search results to return, set to 500 to ensure comprehensive coverage of available imagery within the specified criteria.

Data retrieval was implemented through the Planetary Computer STAC API, which provides programmatic access to the complete Sentinel-2 archive with standardized query capabilities and metadata structure.

3.1.3 Landsat Collection 2 Level-2

The Landsat program represents the longest continuous record of Earth observation from space, with the Collection 2 Level-2 products providing analysis-ready surface reflectance and surface temperature data from multiple Landsat sensors [31]. These products are derived from Landsat 4-5 Thematic Mapper (TM), Landsat 7 Enhanced Thematic Mapper Plus (ETM+), and Landsat 8-9 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) instruments, offering consistent 30-meter spatial resolution multispectral imagery suitable for long-term environmental monitoring and change detection studies.

Landsat Collection 2 represents a comprehensive reprocessing of the entire Landsat archive, incorporating improved geometric and radiometric calibration, updated Digital Elevation Models, and enhanced cloud detection algorithms. The Level-2 products undergo atmospheric correction to convert top-of-atmosphere reflectance to surface reflectance values. For Landsat 4-5 TM and Landsat 7 ETM+ data, the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) algorithm is applied, while Landsat 8-9 OLI/TIRS products utilize the Land Surface Reflectance Code (LaSRC) algorithm. These corrections remove atmospheric interference from water vapor, aerosols, and scattering effects, providing bottom-of-atmosphere reflectance values that enable accurate quantitative analysis and multi-temporal comparisons.

The Landsat Collection 2 Level-2 products include both surface reflectance and surface temperature data products. Surface reflectance bands include coastal aerosol, blue, green, red, near-infrared, and shortwave infrared bands at 30m resolution, while surface temperature products include thermal infrared bands and associated intermediate parameters for temperature derivation. The specific bands utilized in this research are presented in the following table:

Band	Description	Wavelength (m)	Resolution (m)
coastal	Coastal Aerosol	0.43-0.45	30
blue	Blue	0.45-0.51	30

Band	Description	Wavelength (m)	Resolution (m)
green	Green	0.53-0.59	30
red	Red	0.64-0.67	30
nir08	Near Infrared	0.85-0.88	30
swir16	SWIR 1	1.57-1.65	30
swir22	SWIR 2	2.11-2.29	30
lwir11	Thermal Infrared	10.6-11.19	30*
atran	Atmospheric Transmittance	-	30
cdist	Distance to Cloud	-	30
drad	Downwelled Radiance	-	30
emis	Emissivity	-	30
emsd	Emissivity Standard Deviation	-	30
trad	Thermal Radiance	-	30
urad	Upwelled Radiance	-	30

Table 3.2: Landsat Collection 2 Level-2 bands used in this research (*resampled from 100m)

For this research, Landsat Collection 2 Level-2 data was acquired through Microsoft Planetary Computer, a cloud-based platform that provides free access to analysis-ready Earth observation datasets. The platform offers Landsat data in Cloud Optimized GeoTIFF format with STAC-compliant metadata, enabling efficient data discovery and programmatic access to the complete Landsat archive.

Data retrieval was conducted using the Planetary Computer STAC API, following a similar approach as described for Sentinel-2, including bounding box and temporal constraints tailored to each fire event. The "landsat-c2-l2" collection was queried using STAC-compliant requests to obtain both surface reflectance and thermal data relevant to this study.

3.1.4 Copernicus DEM GLO-30

The Copernicus Digital Elevation Model (DEM) GLO-30 is a global Digital Surface Model (DSM) providing worldwide elevation data at 30-meter spatial resolution [32]. The dataset represents the Earth's surface including natural topography,

buildings, infrastructure, and vegetation, making it particularly suitable for terrain analysis, hydrological modeling, and topographic characterization in remote sensing applications.

The Copernicus DEM is derived from the WorldDEM product, which is based on radar interferometry data acquired during the TanDEM-X Mission conducted by the German Aerospace Center (DLR). The original WorldDEM underwent comprehensive editing processes including flattening of water bodies, consistent flow modeling for rivers, shoreline and coastline refinement, and correction of implausible terrain structures. These processing steps ensure hydrological consistency and improve the dataset's reliability for quantitative geospatial analysis.

The dataset provides global coverage with notable vertical accuracy characteristics. The Copernicus DEM GLO-30 achieves a global Root Mean Square Error (RMSE) of approximately 4 meters, with the underlying TanDEM-X data demonstrating absolute vertical accuracy of 3.49 meters and relative vertical accuracy of 0.99 meters on flat terrain (slopes $< 20^\circ$) and 1.37 meters on steep terrain (slopes $> 20^\circ$). This accuracy specification makes the dataset suitable for applications requiring precise elevation information at regional and global scales.

For this research, Copernicus DEM GLO-30 data was accessed through Microsoft Planetary Computer using the "cop-dem-glo-30" collection identifier. The dataset is distributed in Cloud Optimized GeoTIFF format with STAC-compliant metadata, enabling efficient spatial queries and seamless integration with other Earth observation datasets in the analysis workflow.

3.1.5 Copernicus ERA5

ERA5-Land is a climate reanalysis dataset produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) as part of the Copernicus Climate Change Service (C3S)[33]. It provides a consistent and high-resolution reconstruction of land-surface variables over time by combining model data with observations through a data assimilation process governed by physical laws. Unlike traditional observations, reanalysis generates spatially and temporally complete climate fields over multiple decades, enabling robust climate and environmental studies.

ERA5-Land is generated by replaying the land component of the ERA5 reanalysis, using corrected atmospheric variables such as temperature, humidity, and pressure to drive land-surface simulations. While observations are not directly assimilated into ERA5-Land, they influence the dataset indirectly via the atmospheric forcing fields. The higher spatial resolution (0.1°) and improved representation of land processes make ERA5-Land especially suitable for localized analysis.

For this research, we used the post-processed *Derived ERA5-Land Daily Statistics* product, which aggregates the original hourly data to daily time steps using statistical summaries such as daily mean. This version was downloaded via the Copernicus Climate Data Store (CDS).

The following daily variables were selected to capture key atmospheric and soil

conditions relevant to fire dynamics:

- **2m Temperature (t2m)**: Daily average air temperature at 2 meters above ground.
- **10m Wind Components (u10, v10)**: Zonal and meridional components of wind at 10 meters height.

These variables are particularly relevant to wildfire risk assessment, as they influence fuel dryness, ignition potential, and fire behavior. Spatial and temporal extraction procedures, including reprojection and resampling to match the spatial extent of fire patches, are detailed in the Data Processing section.

3.1.6 Piedmont Streets Data

The road network information used in this work was obtained from the Global Roads Inventory Project (GRIP4) dataset [34]. GRIP was developed to provide a harmonized and up-to-date global representation of road infrastructure, with the specific aim of supporting environmental and biodiversity modelling frameworks such as GLOBIO. Unlike datasets intended for navigation, GRIP emphasizes consistency and completeness at global and regional scales, integrating nearly sixty individual data sources (including OpenStreetMap) under the UNSDI-Transportation data-model. This harmonization process ensures cross-border continuity and attribute compatibility across different regions of the world.

The GRIP4 release provides data in both vector and raster formats: regional and global road networks in ESRI File Geodatabase and Shapefile formats, and global rasters of road density at a resolution of 5 arcminutes (approximately 8 km). The dataset is openly available under a Creative Commons Zero (CC-0) license, which allows unrestricted scientific use. Access to the data is provided through the official GLOBIO portal [34, 35].

For the purposes of this study, the **European shapefile** was extracted from GRIP4 and subsequently restricted to the **Piedmont region (Italy)**. This produced a regional vector dataset containing the full road network of Piedmont, including both major and minor roads, as represented in the global inventory. This cropped dataset serves as the road network input for the wildfire analysis, and in later steps it was converted into raster layers aligned with the spatial patches of each fire event (see Section 3.2.6).

3.1.7 Dynamic World Land Cover (IO-LULC Annual v2)

The Dynamic World Land Cover dataset (IO-LULC Annual v2) provides global annual maps of land cover at 10-meter spatial resolution [36]. The dataset characterizes the Earth’s surface across multiple land cover classes, including water, trees, grass, crops, shrubland, built-up areas, bare ground, and snow/ice. Its high spatial detail and annual temporal granularity make it well suited for monitoring landscape changes,

analyzing human–environment interactions, and supporting applications in ecology, agriculture, and wildfire risk assessment.

The product is generated through machine learning models trained on multispectral observations from Sentinel-2 imagery, incorporating extensive reference data to ensure robust classification performance across diverse geographic regions. The methodology emphasizes consistency over time, enabling the detection of gradual as well as abrupt land cover transitions. This design makes the dataset particularly relevant for long-term environmental monitoring and for studies requiring harmonized global land cover information.

The dataset is distributed as annual global mosaics in Cloud Optimized GeoTIFF format, accompanied by STAC-compliant metadata to support efficient search and retrieval in cloud-based environments. For this research, the IO-LULC Annual v2 collection was accessed through Microsoft Planetary Computer using the identifier `io-lulc-annual-v02`, allowing seamless integration with other Earth observation datasets in the workflow.

3.1.8 Ground Truth Mask

A binary raster image was generated to represent the fire extent. The raster contains a single band, where pixels with a value of 1 correspond to the fire-affected area and pixels with a value of 0 correspond to the background. The fire perimeter used to produce this binary image was derived from the vector geometry provided in the file `piedmont_fa.geojson`.

3.2 Data Processing

3.2.1 Sentinel-2 Image Selection and Preparation

The Sentinel-2 component of the dataset was constructed starting from the fire perimeter information contained in the `piedmont_fa.geojson` file, which provided the spatial reference for each fire event. For every fire, the corresponding geometry was projected into a target coordinate system (EPSG:32632) and buffered in order to ensure that the area surrounding the fire was also included in the analysis. The buffered geometry was then re-projected into WGS84 coordinates, which are required to define the bounding box for the image search through the STAC catalogue.

Since the objective was to capture the conditions prior to each fire, a temporal filter was applied to retrieve images acquired in the days immediately before the ignition date. In this work, a configurable temporal window of seven days was used. Fire events that occurred before the availability of Sentinel-2 data (June 2015) were discarded. The Microsoft Planetary Computer STAC API was then queried for Sentinel-2 Level-2A products matching both the spatial and temporal constraints, and only those acquisitions that effectively intersected the fire geometry were retained for further processing.

Once the candidate images had been identified, they were converted into fixed-size patches of 256×256 pixels, corresponding to an area of approximately $2.56 \text{ km} \times 2.56 \text{ km}$ at 10 m spatial resolution. Each patch was centered on the centroid of the fire perimeter in order to guarantee that the fire area was always included in the extracted image. Sentinel-2 spectral bands, which are provided at different native resolutions (10 m, 20 m and 60 m), were resampled to a uniform 10 m grid to ensure consistency across the dataset.

To improve data quality, a similarity check was carried out on images acquired on the same date. Patches were compared at the band level, and whenever two acquisitions differed by less than 0.5% in relative pixel values, the redundant image was discarded. This procedure avoided storing duplicate information and ensured that only unique acquisitions were preserved.

In parallel with the image extraction, a binary ground truth mask was generated for each fire event. The fire geometry was rasterized onto the same spatial grid as the Sentinel-2 patches, assigning a value of 1 to all pixels inside the fire perimeter and 0 to the background. This produced a single mask per fire, spatially aligned with the Sentinel-2 imagery, which serves as the reference for supervised learning tasks.

Through this pipeline, every fire event was thus associated with one or more Sentinel-2 image patches and a corresponding ground truth mask, forming a consistent and spatially aligned dataset for further analysis and model training.



Figure 3.1: Sentinel-2 image patch (256×256 pixels at 10 m resolution) centered on the fire perimeter.

3.2.2 Landsat Image Selection and Preparation

The Landsat component of the dataset was derived using the same fire perimeter information from the `piedmont_fa.geojson` file. As for Sentinel-2, each fire geometry was buffered and re-projected to define the area of interest and then used to query the Microsoft Planetary Computer STAC API [30]. In this case, the search targeted the Landsat Collection 2 Level-2 Science Products [31], which provide atmospherically corrected surface reflectance data for the Landsat 8 and Landsat 9 missions. Fires occurring before 2013, i.e. before the availability of Landsat 8 data, were discarded.

To maintain temporal consistency with Sentinel-2, images were retrieved within a seven-day window preceding the ignition date of each fire. Candidate images were then subjected to two quality filters. First, acquisitions with more than 30% of pixels flagged as nodata were removed. Second, products with an unexpected number of bands (i.e. not equal to 15, corresponding to the surface reflectance bands and ancillary quality layers) were discarded. These steps ensured that only complete and usable images were retained.

Because of the different native resolutions of the two missions, an additional resampling step was necessary. Landsat bands, originally acquired at 30 m, were resampled to 10 m resolution so that they could be made consistent with the Sentinel-2 grid. Patches of 256×256 pixels were then extracted from the resampled imagery, centered on the centroid of the fire perimeter. Before resampling, each patch represented an area of approximately $7.68 \text{ km} \times 7.68 \text{ km}$, which was subsequently aligned with the Sentinel-2 patch size of $2.56 \text{ km} \times 2.56 \text{ km}$.

Unlike Sentinel-2, for Landsat no independent ground truth masks were produced. Instead, the binary masks derived from Sentinel-2 were used directly after resampling, ensuring that both datasets shared the same reference geometry. This choice avoided inconsistencies that would arise from generating separate rasterized masks at different resolutions.

Through this procedure, the Landsat dataset was built to mirror the Sentinel-2 dataset in structure, while accounting for the differences in spatial resolution and sensor design. This alignment allows the two datasets to be used jointly in subsequent analyses and supports comparative and multi-sensor approaches to burned area mapping.

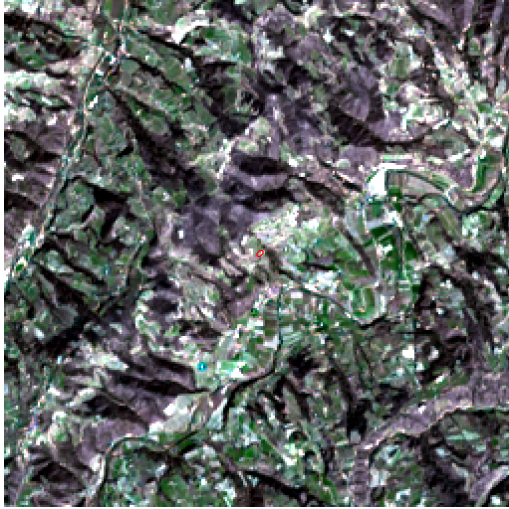


Figure 3.2: Landsat image patch (256×256 pixels at 30 m resolution) centered on the fire perimeter.



Figure 3.3: Landsat image patch after resampling (10 m resolution).

3.2.3 Digital Elevation Model (DEM) Data

In addition to the optical imagery, a digital elevation model was extracted for each fire in order to provide topographic context to the dataset (see 3.1.4). The DEM patches were generated following the same spatial logic adopted for Sentinel-2 and Landsat. For each fire geometry, the centroid of the burned area was used as the reference point and a square patch of 256×256 pixels was defined. At the native DEM resolution, this corresponds to an area of approximately $7.68 \text{ km} \times 7.68 \text{ km}$. To ensure spatial consistency, only fires with an area larger than 200 m^2 were considered suitable for DEM extraction, as smaller events would not provide meaningful information at 30 m.

DEM tiles intersecting each fire geometry were identified via STAC queries, downloaded, and cropped to the target extent. The extracted patches were then reprojected to the target UTM CRS (EPSG:32632) to match the reference system used for the optical imagery. Finally, as in the case of Landsat, a resampling step was applied in order to bring the DEM from its native 30 m resolution to 10 m, thereby ensuring compatibility with the Sentinel-2 grid. This resampling allowed a direct overlay between DEM, Sentinel-2 and Landsat patches, facilitating multi-source analysis.

Each resulting DEM patch was stored as a single-band GeoTIFF with no-data values set to -9999 . The files were named according to the convention `fire_{id}_dem.tif` and placed in the corresponding fire folder within the dataset structure. This procedure guarantees that every fire event is associated with a DEM patch aligned in extent, resolution and reference system with the optical data, making the topographic information immediately usable for downstream applications such as burned area mapping or fire behavior analysis.

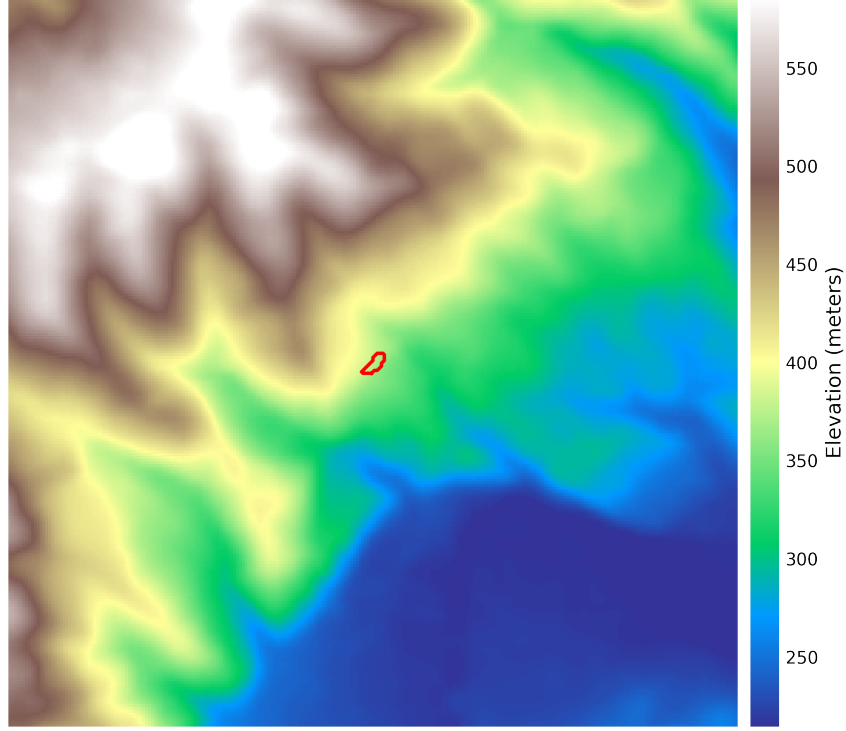


Figure 3.4: Dem image patch (256×256 pixels at 10 m resolution) centered on the fire perimeter.

3.2.4 ERA5-Land Processing

To integrate ERA5-Land variables with fire events, a dedicated processing pipeline was implemented. The procedure starts by selecting the optimal pre-fire Sentinel-2 image for each fire event, ensuring minimal cloud contamination. CloudSen12 was used to generate cloud masks, and a cloud contamination score was computed over the ground-truth fire footprint. The image with the lowest score was retained as the reference pre-fire scene. Its acquisition date and spatial reference provided the temporal and spatial anchors for ERA5-Land data retrieval.

For each selected event, the centroid of the Sentinel-2 fire patch was computed in WGS84 coordinates, and a bounding box of $\pm 0.5^\circ$ around the centroid was defined to query ERA5-Land data via the Copernicus Climate Data Store (CDS) API. The *Derived ERA5-Land Daily Statistics* dataset was downloaded for the exact acquisition day of the selected pre-fire image. Three variables were retrieved: 2m air temperature (t_{2m}), and the zonal (u_{10}) and meridional (v_{10}) components of 10m wind.

The retrieved data, originally provided in NetCDF format with a native spatial resolution of 0.1° , were reprojected from geographic coordinates (EPSG:4326) to the UTM zone of the corresponding Sentinel-2 patch (EPSG:32632 for the study area). Each variable was resampled to a fixed 256×256 grid with a target spatial resolution of 10 m, matching the spatial footprint of Sentinel-derived fire patches ($2.56 \text{ km} \times 2.56 \text{ km}$). Bilinear interpolation was used for the resampling step to preserve continuous fields. The three variables were then stacked into a single multi-band GeoTIFF file, with metadata annotations for each band. Invalid values were masked

and replaced with a consistent nodata flag.

Finally, a logging mechanism was employed to track processed fire IDs and prevent redundant downloads. The resulting multi-band ERA5 rasters are spatially and temporally aligned with the fire patches, enabling direct integration of meteorological predictors into subsequent fire modeling and analysis workflows.

3.2.5 Land Cover Data

In addition to topographic information, annual land cover data was extracted for each fire event to characterize the surrounding environment (see 3.1.7). The land cover patches were generated using the same spatial logic adopted for DEM and optical imagery. For each burned area, the centroid of the fire geometry was used as the reference point, and a square patch of 256×256 pixels was defined. At the native resolution of the IO-LULC Annual v2 dataset (10 m), this corresponds to an area of approximately $2.56 \text{ km} \times 2.56 \text{ km}$. To ensure the reliability of the extracted data, only fires with an area larger than 200 m^2 were considered, as smaller events would not provide meaningful information at this resolution.

Land cover tiles intersecting each fire geometry were identified through STAC queries to the Microsoft Planetary Computer. For each fire, all available items within a broad temporal window (from five years before to five years after the event, limited to 2012–2024) were retrieved. From these candidates, the land cover map corresponding to the year closest to the fire date was selected, ensuring temporal consistency between the fire occurrence and the environmental description.

Once the most suitable land cover item was identified, the corresponding raster was downloaded and cropped to the target patch extent. The data was reprojected to the reference UTM CRS (EPSG:32632), preserving the categorical nature of the land cover classes. This procedure ensured spatial alignment and resolution compatibility with Sentinel-2, Landsat, and DEM patches.

Each resulting land cover patch was stored as a single-band GeoTIFF with nodata values preserved from the source dataset. Files were named according to the convention `fire_{id}_landcover.tif` and placed within the corresponding fire folder in the dataset structure. This guarantees that every fire event is associated with a land cover raster aligned in extent, resolution, and reference system with the optical and DEM data, enabling integrated multi-source analysis.

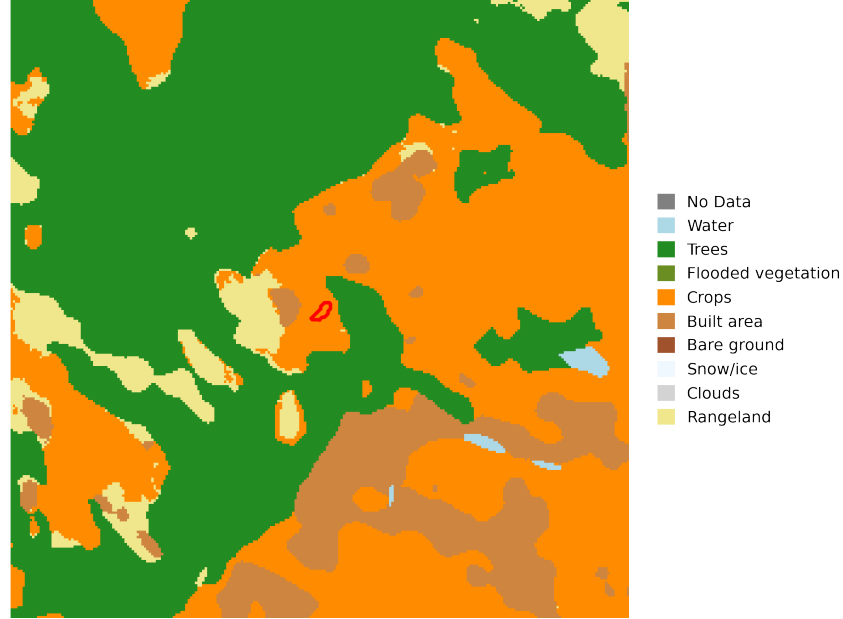


Figure 3.5: Landcover image patch (256×256 pixels at 10 m resolution) centered on the fire perimeter.

3.2.6 Streets Rasterization and Preparation

The road network information extracted from the GRIP4 dataset (see Section 3.1.6) was processed into raster layers aligned with the wildfire event patches. For each fire directory, the corresponding Sentinel-2 image was used as a spatial reference in order to guarantee identical extent, resolution, and coordinate system between satellite and ancillary layers.

The procedure began by identifying a Sentinel-2 patch within each fire folder and reading its spatial metadata (bounding box, affine transform, image dimensions, and CRS). The Piedmont road network vector layer was then spatially filtered using this bounding box, retaining only the geometries intersecting the area of interest. This filtering was accelerated by means of a spatial index, which allowed efficient queries even on the full regional dataset.

The selected road geometries were subsequently rasterized on the same grid as the Sentinel-2 patch. Each pixel intersected by a road segment was assigned the categorical value corresponding to the `GP_RTP` attribute of the GRIP dataset, which encodes the type of road (e.g., highway, primary, secondary). All other pixels were assigned a value of zero, indicating the absence of roads. The rasterization was performed with the `all_touched` option enabled, ensuring that all pixels touched by a road geometry were included.

The resulting product is a single-band raster, saved in GeoTIFF format as `fire_ID_streets.tif`, perfectly co-registered with the Sentinel-2 data and the other ancillary layers. This guarantees spatial consistency across modalities and allows the road network information to be seamlessly integrated into subsequent analyses and model training.

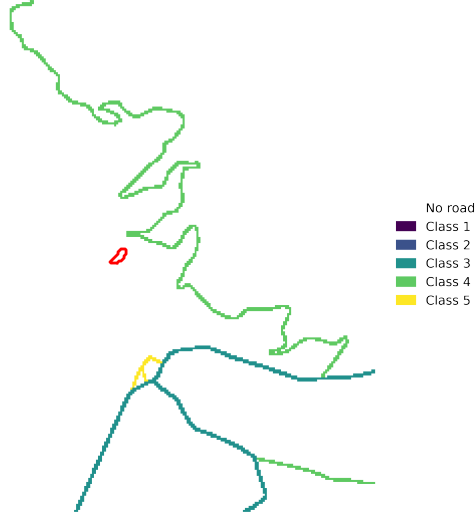


Figure 3.6: Streets raster image patch (256×256 pixels at 10 m resolution) centered on the fire perimeter.

3.2.7 Cloud Masking with CloudSen12

A recurring issue in the analysis of Sentinel-2 imagery is the presence of clouds and their shadows, which can obscure important land features and compromise the reliability of downstream analyses. In order to mitigate this problem, cloud masking was performed using the *CloudSen12* model [37], a semantic segmentation network trained on Sentinel-2 Level-2A data. The model produces a four-class mask distinguishing between clear sky, thick clouds, thin clouds (or haze/smoke), and cloud shadows.

The output mask assigns a categorical label to each pixel, as summarized in Table 3.3. An example of a CloudSen12 output mask is shown in Figure 3.7.

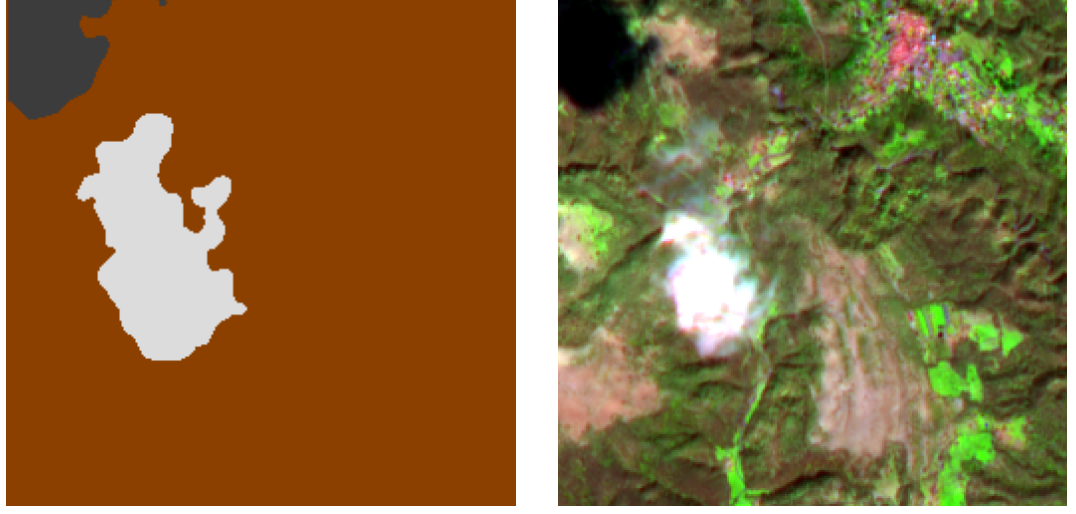
Table 3.3: CloudSen12 mask classes, corresponding description and color.

Class	Description	Pixel Value
0	Clear sky	
1	Thick cloud	
2	Thin cloud / haze / smoke	
3	Cloud shadow	

For each fire event, the pre-fire Sentinel-2 images were processed with CloudSen12 to generate the corresponding cloud masks. To ensure that the images used in training were minimally affected by cloud contamination, a quantitative *cloud score* was computed over the fire area. Specifically, the ground-truth (GT) fire mask was combined with the cloud mask so that only the pixels within the burned area (GT = 1) were considered. Their cloud values were then summed to obtain the score, with lower values indicating less cloud coverage.

Based on this metric, the best pre-fire image for each fire event was selected, i.e., the one with the lowest cloud contamination over the fire footprint. This step was

crucial to guarantee the reliability of the training dataset. Once the optimal image was identified, its acquisition date and spatial reference were extracted and used to retrieve the corresponding ERA5 meteorological data for the same location and time.



(a) CloudSen12 output mask

(b) Corresponding Sentinel-2 RGB image

Figure 3.7: Example of Sentinel-2 image (b) and corresponding CloudSen12 output mask (a).

3.3 Dataset Information

The dataset is composed of **1,090 directories**, each corresponding to a single wildfire event. These were derived from the original **1,231 fire polygons** contained in the official geojson database, covering the period **2012–2024**. Since **Sentinel-2 imagery** is only available from late 2015 onwards, events without valid Sentinel coverage were excluded, resulting in the final set of 1,095 fire samples.

Among these:

- **529 fires** include both Sentinel-2 and Landsat (resampled to 10 m resolution) data.
- **561 fires** include only Sentinel-2 data (Landsat is missing or represented by placeholders).

Each directory contains the following information:

- Satellite imagery (Sentinel-2, and Landsat when available),
- Digital Elevation Model (DEM),
- ERA5 climate variables,
- Landcover
- Streets data

- Ignition point data.

All raster patches are standardized to a fixed size of **256 × 256 pixels**, regardless of the actual fire extent. This ensures consistency across the dataset but also introduces strong variation in the proportion of burned area contained in each patch. The fires show a highly unbalanced distribution: the majority of events correspond to very small burned areas, while only a few represent very large fires. A statistical summary is reported in Table 3.4.

Table 3.4: Statistical summary of fire sizes (in pixels, 10 m resolution).

Statistic	Value
Total fires analyzed	1,090
Smallest fire (ID 7073)	4 pixels
Largest fire (ID 5440)	65,536 pixels
Mean fire size	1,473.42 pixels
Standard deviation	5,937.24 pixels

3.3.1 Fire Size Distribution

Figure 3.8 illustrates the distribution of burned area sizes across the dataset. The plot confirms the skewed nature of the data, with many small events and only a handful of extreme cases.

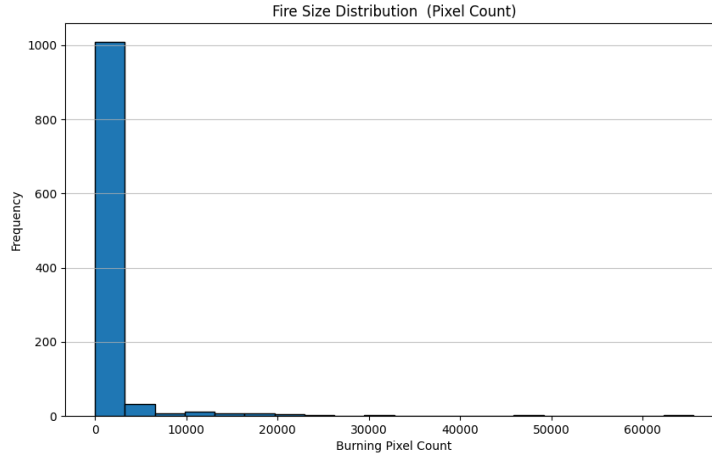


Figure 3.8: Distribution of burned area sizes (in pixels).

Chapter 4

Methodology

This chapter presents the methodology developed to address the research objectives, with a particular focus on the design of a multimodal deep learning framework. The approach integrates heterogeneous geospatial and environmental data to predict wildfire burned areas, and is described through the problem formulation, the model architecture, and the fusion methods employed.

4.1 Problem Statement

This thesis focuses on the delineation of wildfire burned areas in the Piedmont region using pre-fire imagery and ancillary data. The objective of the proposed model is to predict the final extent of a wildfire, given a set of heterogeneous inputs: digital elevation model (DEM) data, ERA5 meteorological variables corresponding to the date of the Sentinel-2 acquisition, pre-fire Sentinel-2 imagery, Landsat imagery when available, road network data, and the ignition point.

The model is therefore required to learn the expected propagation dynamics of a wildfire from these pre-event conditions. This task is particularly challenging, as it differs fundamentally from conventional burned area segmentation approaches, where post-fire satellite images are employed to directly identify affected regions. Instead, the aim here is predictive: to estimate the final burned area footprint before the fire occurs, based solely on the pre-fire state of the landscape and meteorological conditions.

4.2 Multi-Modal Approach

In traditional remote sensing applications, the task of mapping wildfires is usually framed as a post-fire semantic segmentation problem. The model receives satellite images collected after the event and learns to distinguish burned from unburned surfaces. This setup is particularly effective because spectral differences are pronounced: for example, combinations such as SWIR, NIR, and red bands make burned areas stand out in false-color composites.

In this thesis however, the problem is different. The aim is not to delineate



Figure 4.1: Burned Scars for the 1998 summer forest fires on the Elba Island, overlaid on a Landsat 5 TM false colour composite. [38]

fire scars after the event, but to predict the final burned extent using only pre-fire information. This changes the nature of the challenge significantly, since the obvious spectral signatures of burned surfaces are absent. Early trials using only pre-fire imagery confirmed this difficulty: the network often converged to degenerate solutions, for example by predicting no burned pixels at all, simply because no clear spectral evidence was available.

To overcome this, a multi-modal strategy was adopted, combining different sources of information into a single predictive framework. Instead of relying on a single encoder, the model is built around several parallel encoders, each processing a different modality but producing feature maps at the same spatial resolution. The main branch is dedicated to Sentinel-2 optical imagery, which captures vegetation and surface conditions across twelve spectral bands. A second encoder processes Landsat imagery, which contributes fifteen bands plus a binary presence flag; this flag is used to suppress spurious activations when Landsat data are unavailable or incomplete. Topographic information is added through a digital elevation model (DEM), which has a strong influence on fire behavior, especially in mountainous terrain. In addition, an ignition map is provided to the model, giving it a spatial prior on where the fire started. Meteorological conditions, which are key drivers of wildfire spread, are incorporated through ERA5 reanalysis data: wind components are treated as raster inputs and processed by a dedicated encoder, while tabular variables such as near-surface temperature are handled by a small multilayer perceptron and then broadcast across the spatial grid so that they can interact with the convolutional features.

All these streams of information are fused progressively during encoding. At each depth of the network, the features from Sentinel-2, Landsat, DEM + Streets, ignition point, ERA5 rasters, and the expanded tabular embeddings are concatenated and

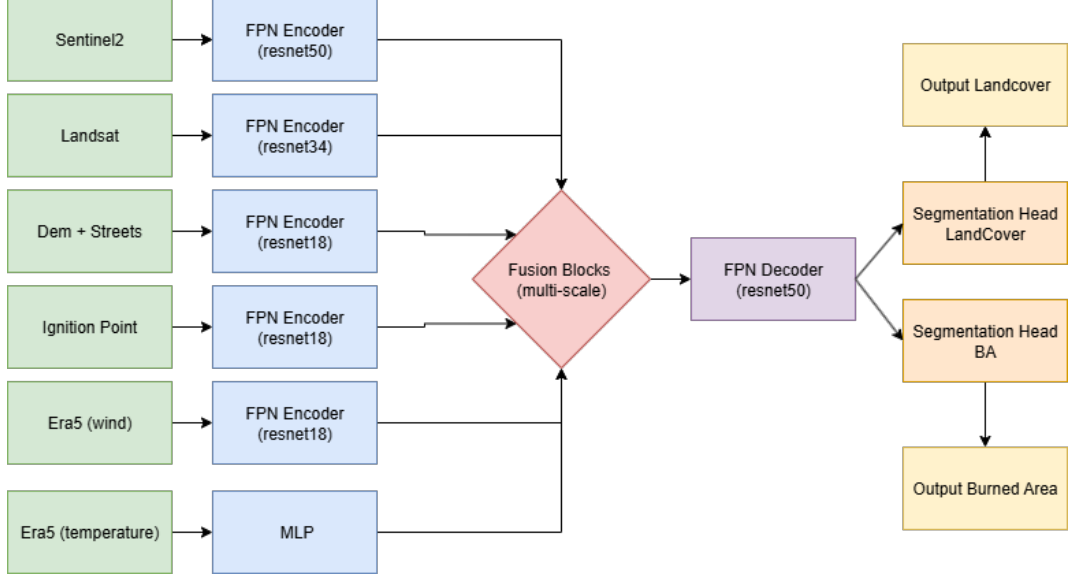


Figure 4.2: Schematic representation of the proposed multimodal architecture

passed through a fusion block. This ensures that the complementary information from different sources is integrated before being fed to the decoder. The decoder itself follows the design of a Feature Pyramid Network (FPN) [39], which upsamples the fused features and produces outputs through two separate segmentation heads: the primary head generates a burned-area probability mask, while a secondary head predicts land cover classes. This multi-task learning approach leverages the shared representations learned by the decoder to improve generalization, as the auxiliary land cover prediction task encourages the model to develop more robust feature representations of the underlying landscape characteristics.

4.3 Network Architectures

The predictive framework is implemented as a multi-modal convolutional neural network that extends the Feature Pyramid Network (FPN) design with multi-task learning capabilities. Each modality is processed by a dedicated FPN encoder with a ResNet backbone, which extracts hierarchical features at multiple spatial scales. These include Sentinel-2, Landsat, digital elevation and street layers, ignition maps, and ERA5 raster variables. In addition, tabular ERA5 variables are processed by a multilayer perceptron (MLP), transformed into low-dimensional embeddings, and spatially broadcast to align with the convolutional feature maps. At each resolution stage, the outputs of all encoders and the tabular embeddings are concatenated and passed through a fusion block, ensuring that complementary information is integrated before decoding.

The shared FPN decoder then upsamples the fused features and feeds them to two separate segmentation heads: the primary head produces the burned-area probability mask, which is the main objective of the model, while the auxiliary head performs land cover classification. This multi-task learning setup serves as a regularization

mechanism that encourages the shared decoder to learn more generalizable feature representations. The land cover prediction task, while not the primary focus of this work, helps the model develop a better understanding of the underlying landscape characteristics that influence fire behavior, ultimately improving the quality of burned area predictions.

4.3.1 Feature Pyramid Network

The Feature Pyramid Network (FPN) is a convolutional neural network architecture introduced by Lin et al. [39] designed to improve multi-scale feature representation in dense prediction tasks such as object detection and segmentation. Unlike standard convolutional backbones that naturally build a hierarchy of feature maps but typically utilize only the deepest layer for downstream tasks, FPN explicitly combines information across different levels of the feature hierarchy. This design enables the network to leverage both fine-grained spatial details from shallow layers and high-level semantic information from deeper layers.

The FPN architecture consists of two main components: a bottom-up pathway and a top-down pathway with lateral connections.

Bottom-up pathway: This corresponds to the feedforward computation of a conventional convolutional backbone, such as ResNet, which produces a sequence of feature maps at progressively decreasing spatial resolutions. At each stage of the backbone, spatial resolution is reduced through pooling operations or strided convolutions, while the semantic abstraction level of the features increases. This process generates a set of feature maps $\{C_2, C_3, C_4, C_5\}$ at multiple resolution levels, where the subscript indicates the pyramid level.

Top-down pathway and lateral connections: The top-down pathway reconstructs semantically strong feature maps at higher resolutions by progressively upsampling the deeper feature maps. At each resolution level, the upsampled features are combined with the corresponding bottom-up features of the same spatial scale through lateral connections. These lateral connections employ 1×1 convolutions to match channel dimensions before element-wise addition for feature fusion. This process yields a new set of enhanced feature maps $\{P_2, P_3, P_4, P_5\}$, where each pyramid level encodes both rich semantic information and fine spatial details.

This architectural design enables FPN to construct a feature pyramid that maintains semantic richness across all resolution levels, contrasting with traditional single-scale approaches that rely solely on the deepest feature layer. The multi-scale representation is particularly advantageous for tasks involving objects or regions of interest that exhibit significant scale variation, such as fire scars in our wildfire prediction application, where burned areas can range from small isolated patches to extensive regions.

In our implementation, FPN serves as the encoder backbone for each input modality (Sentinel-2, Landsat, DEM with street layers, ignition points, and ERA5 raster data). By adopting this multi-scale architecture across all modalities, the

model preserves hierarchical feature representations from each data source prior to cross-modal fusion. Additionally, our decoder architecture is also based on the FPN design, ensuring that the reconstruction of the final segmentation mask benefits from the same principle of combining semantic richness with spatial precision across multiple scales.

The original FPN architecture from [39] is illustrated in Figure 4.3, showing the bottom-up pathway, top-down pathway, and lateral connections that enable effective multi-scale feature fusion.

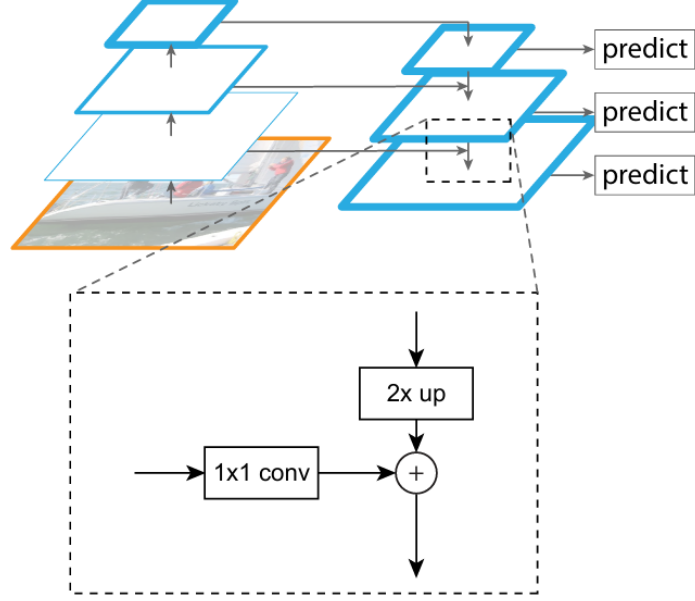


Figure 4.3: Feature Pyramid Network architecture (adapted from Lin et al. [39]). The bottom-up pathway extracts hierarchical features, while the top-down pathway with lateral connections (1×1 convolutions) enables multi-scale feature fusion.

4.3.2 ResNet

The Residual Network (ResNet) architecture, introduced by He et al. [he2016deep], serves as the backbone encoder for each modality-specific FPN in our implementation. ResNet addresses the vanishing gradient problem in deep networks through the introduction of residual connections, enabling the training of significantly deeper networks while maintaining gradient flow and representational capacity.

The core innovation of ResNet lies in its residual blocks, which implement skip connections that allow information to bypass one or more layers. Instead of learning a direct mapping $\mathcal{H}(\mathbf{x})$, each residual block learns the residual function $\mathcal{F}(\mathbf{x}) = \mathcal{H}(\mathbf{x}) - \mathbf{x}$, making the final output $\mathcal{H}(\mathbf{x}) = \mathcal{F}(\mathbf{x}) + \mathbf{x}$. This formulation transforms the optimization problem from learning an unreferenced mapping to learning perturbations around the identity function, which empirically proves easier to optimize and enables the construction of much deeper networks.

In our multi-modal architecture, we employ different ResNet variants tailored

to the complexity and information content of each input modality, as detailed in Table 4.1. This graduated approach to backbone selection reflects a careful balance between computational efficiency and representational capacity requirements for each data source.

Table 4.1: ResNet backbone configurations for different input modalities

Input Modality	Architecture	Layers	Pre-trained Weights
Sentinel-2	ResNet-50	50	ImageNet
Landsat	ResNet-34	34	ImageNet
DEM + Streets	ResNet-18	18	ImageNet
Ignition Maps	ResNet-18	18	None (random initialization)
ERA5 Raster	ResNet-18	18	None (random initialization)

The selection of ResNet-50 for Sentinel-2 data processing reflects the rich spectral information content and high dimensionality of this primary optical satellite imagery modality. Sentinel-2’s multispectral bands capture diverse vegetation indices and land surface characteristics that are crucial for fire risk assessment, necessitating a deeper network architecture to extract complex hierarchical features from this information-dense input stream.

For Landsat data, ResNet-34 provides sufficient representational capacity while maintaining computational efficiency. This intermediate complexity aligns with Landsat’s role as a complementary optical imagery source that, while valuable, isn’t always present. The 34-layer depth enables adequate feature extraction without the computational overhead required for the primary Sentinel-2 modality.

The auxiliary data modalities—digital elevation model with street layers, ignition point maps, and ERA5 raster variables—utilize ResNet-18 architectures. This choice reflects the more focused information content of these data sources compared to the comprehensive spectral imagery modalities. Digital elevation and infrastructure data provide topographical and accessibility constraints, ignition points offer discrete spatial indicators of fire initiation, and meteorological raster data contribute environmental conditioning factors. The lighter ResNet-18 architecture efficiently processes these specialized inputs while preserving computational resources for the more complex optical imagery modalities.

Our implementation adopts a selective approach to pre-trained weight initialization based on domain compatibility considerations. Landsat and DEM encoders leverage ImageNet pre-trained weights to benefit from learned low-level visual features that translate effectively to natural imagery and elevation data visualization. Conversely, Sentinel-2, ignition map, and ERA5 raster encoders are initialized from scratch, reflecting the domain-specific nature of these modalities where standard natural image features may not provide optimal initialization for specialized geospatial and meteorological data representations.

Each ResNet encoder within the FPN framework generates hierarchical feature representations at multiple spatial scales, producing feature maps at four different

resolution levels that correspond to progressively increasing levels of semantic abstraction. These multi-scale representations are essential for capturing fire-related patterns that manifest across various spatial scales, from localized ignition points and fine-grained vegetation characteristics to large-scale topographical influences and regional meteorological conditions affecting fire propagation dynamics.

4.3.3 Multi-Layer Perceptron for Tabular Data Processing

The Multi-Layer Perceptron (MLP) component of our architecture addresses the integration of tabular meteorological variables from the ERA5 dataset, specifically temperature measurements that provide crucial atmospheric conditioning information for fire propagation modeling. Unlike the raster-based ERA5 variables that maintain spatial structure and can be processed through convolutional layers, this tabular meteorological parameter represent point measurements or spatially-averaged values that require a different processing approach to extract meaningful feature representations.

The MLP architecture employed for tabular ERA5 data processing consists of a compact two-layer fully connected network designed to transform raw meteorological measurements into dense feature embeddings suitable for spatial integration with convolutional feature maps. The network begins with an input layer that accepts the tabular meteorological variables, followed by a first hidden layer with 64 neurons and ReLU activation. This intermediate representation undergoes further transformation through a second hidden layer that compresses the features into a 16-dimensional embedding space, again followed by ReLU activation to maintain non-linearity in the learned representations.

This architectural design reflects several key considerations for effective tabular data integration. The initial expansion from input variables to 64 dimensions allows the network to learn complex non-linear relationships between meteorological parameters, capturing interactions between temperature and pressure that may not be immediately apparent from raw measurements. The subsequent compression to 16 dimensions creates a compact yet informative representation that balances expressiveness with computational efficiency, particularly important given that these embeddings must be spatially broadcast across multiple resolution levels in the feature pyramid.

The choice of ReLU activation functions throughout the MLP ensures computational efficiency while providing sufficient non-linearity for learning complex meteorological relationships. ReLU activation also maintains gradient flow during backpropagation, which is crucial for end-to-end training of the entire multi-modal architecture. The relatively shallow depth of the MLP reflects the structured nature of tabular meteorological data, where excessive depth could lead to overfitting given the limited dimensionality of the input space compared to the rich spatial information content of the imagery modalities.

Following feature extraction, the 16-dimensional embeddings undergo spatial

broadcasting to align with the multi-resolution convolutional feature maps generated by the various FPN encoders. This broadcasting process replicates the tabular feature embeddings across all spatial locations at each resolution level, effectively providing every spatial location with access to the global meteorological context. The broadcast embeddings are then concatenated with the corresponding convolutional features at each pyramid level before fusion processing.

This integration strategy acknowledges that while temperature provides important regional atmospheric conditioning for fire behavior, the relatively homogeneous spatial distribution at our study scale makes spatial averaging an appropriate preprocessing step. By broadcasting the temperature-derived features spatially, the model can leverage this atmospheric conditioning information throughout the spatial domain while allowing the fusion blocks to learn how local topographical characteristics, vegetation patterns, and spatially-variable wind conditions interact with the regional temperature context to influence fire behavior and propagation dynamics.

4.3.4 Multi-Modal Fusion Blocks

The fusion blocks constitute a critical component of our multi-modal architecture, responsible for integrating feature representations from all input modalities at each resolution level of the feature pyramid. These blocks operate at every encoder stage, combining features from Sentinel-2, Landsat, DEM with street layers, ignition points, ERA5 raster data, and spatially-broadcast temperature embeddings before passing the unified representations to the shared FPN decoder.

At each pyramid level, the fusion process begins with feature concatenation along the channel dimension. The input channels to each fusion block consist of the sum of output channels from all five FPN encoders plus the 16-dimensional temperature embeddings, resulting in varying input dimensionalities across different pyramid stages. For example, at the deepest encoder stage (ResNet stage 5), the fusion block receives features with combined channel dimensions reflecting the full representational capacity of each modality-specific encoder.

The fusion blocks employ a straightforward yet effective architecture consisting of a single convolutional layer followed by batch normalization and ReLU activation. The choice of kernel size represents a key architectural decision that influences the spatial integration characteristics of the fusion process.

4.3.4.1 Kernel Size Configuration

Two alternative configurations are considered for the fusion convolution:

3×3 Convolution Approach: This configuration employs a 3×3 convolutional kernel with padding to maintain spatial dimensions. The larger receptive field enables each fused feature to incorporate information from a local neighborhood, facilitating the learning of spatial relationships between different modalities. This approach is particularly beneficial when modalities exhibit spatially-correlated patterns, such as the relationship between topographical features and vegetation characteristics, or

the interaction between wind patterns and local terrain effects on fire propagation.

1×1 Convolution Approach: The pointwise convolution alternative processes each spatial location independently, focusing purely on channel-wise feature integration without spatial mixing. This configuration emphasizes learning optimal linear combinations of multi-modal features at each pixel, making it computationally efficient while maintaining the spatial alignment of different modality inputs. The 1×1 approach is suitable when the primary integration challenge lies in balancing the contributions of different modalities rather than learning spatial cross-modal relationships.

Both configurations maintain identical output channel dimensions matching the Sentinel-2 encoder specifications, ensuring compatibility with the subsequent FPN decoder. The fusion blocks are replicated across all encoder stages, with each block tailored to handle the specific channel dimensions corresponding to its pyramid level. This staged fusion strategy ensures that multi-modal integration occurs at multiple scales, allowing the model to leverage both fine-grained spatial details and high-level semantic relationships across all input modalities.

The batch normalization component stabilizes training dynamics by normalizing the fused feature distributions, while ReLU activation introduces non-linearity essential for learning complex cross-modal interactions. This simple yet robust fusion architecture enables effective integration of diverse geospatial and meteorological information streams while maintaining computational efficiency throughout the feature pyramid.

Chapter 5

Experiment and Results

In this chapter, we outline the setup employed for training and evaluating the model. Subsequently, we provide a concise overview of the losses and optimization techniques applied in addressing the task. Lastly, a summary and discussion of the experimental outcomes are presented.

5.1 Implementation Details

The process of training neural networks is intricate and requires significant resources, demanding careful preparation of data and careful selection of hyperparameters. This section delivers a comprehensive overview of the experiments performed, detailing the methods used for both training and evaluation.

5.1.1 Experimental Setup

The purpose of this thesis is to evaluate the multi-modal multitask approach by training and comparing the performance of the proposed MultiModalFPN architecture against baseline models. The main model is based on a Feature Pyramid Network (FPN) design extended with multiple encoders to handle different data modalities. Each encoder utilizes ResNet backbones of varying complexity: ResNet50 for the primary Sentinel-2 encoder to handle the complexity of the 12-band optical data, ResNet34 for the Landsat encoder managing 15 bands plus a presence flag, and ResNet18 for the lighter modalities including DEM, streets, ignition points, and ERA5 raster data. This hierarchical approach balances computational efficiency with the representational capacity needed for each data source.

The multi-modal architecture receives as input multiple tensors of different dimensions: Sentinel-2 imagery ($256 \times 256 \times 12$), Landsat imagery ($256 \times 256 \times 16$), topographic data including DEM and street information ($256 \times 256 \times 2$), ignition point maps ($256 \times 256 \times 1$), ERA5 meteorological raster data ($256 \times 256 \times 2$), and tabular ERA5 variables processed through a multilayer perceptron. The model produces two outputs: the primary burned area probability mask and an auxiliary land cover classification mask, implementing a multitask learning approach that serves as regularization and improves feature representation.

The dataset has been divided into training and validation sets following an 80–20 split strategy. Specifically, the **training** set comprises **872** fire events (about 80% of the total 1090), covering diverse geographical locations across the Piedmont region. The remaining **218** events (20%) form the **validation** set, selected to maintain geographical diversity while ensuring an independent evaluation. This division ensures that the model is trained on a substantial portion of the available data while preserving sufficient samples for robust validation.

The experiments were conducted on one NVIDIA RTX 2080i with 11GB GDDR6 memory. The substantial computational resources were necessary due to the multi-modal nature of the data and the complexity of processing multiple high-resolution satellite imagery sources simultaneously. Data preprocessing and analysis were performed using Python with libraries including NumPy, pandas, and rasterio for geospatial data handling. The neural network models were developed and trained using the PyTorch framework with additional dependencies including segmentation-models-pytorch for baseline architectures and specialized remote sensing libraries for efficient data loading and augmentation. All Python packages and their versions are summarized in Table 5.1

Library	Version
python	3.10.12
torch	2.8.0
torchvision	0.23.0
pytorch-lightning	2.5.3
segmentation-models-pytorch	0.5.0
torchgeo	0.6.2
torchmetrics	1.8.1
rasterio	1.4.3
geopandas	1.1.1
xarray	2025.6.1
numpy	2.2.6
pandas	2.3.1
scipy	1.15.3
opencv-python-headless	4.12.0.88
albumentations	2.0.8
kornia	0.8.1
matplotlib	3.10.5
hydra-core	1.3.2
tqdm	4.67.1

Table 5.1: Python packages and versions used in the experiments

5.1.2 Training Process

The training methodology employed in this work incorporates several specialized components designed to handle the unique challenges of multimodal wildfire prediction and multi-task optimization. All input tiles are sampled at a resolution of 256×256 pixels, providing sufficient spatial context for local fire behavior patterns while

maintaining computational efficiency. The dataset exhibits significant class imbalance, with burned pixels representing a small fraction of the total pixel population for each image. There are also cases of large burned areas that occupy the majority of the 256×256 tile, but most fires in the dataset are relatively small in spatial extent.

The model optimization employs a Stochastic Gradient Descent (SGD) optimizer with momentum, configured with an initial learning rate of 0.01, momentum parameter of 0.9, and weight decay of 1×10^{-4} for regularization. This configuration provides robust convergence characteristics for the multi-task learning scenario while preventing overfitting through appropriate regularization. The learning rate scheduling follows a Cosine Annealing strategy with a minimum learning rate threshold of 1×10^{-4} . This scheduler provides a learning rate pattern that begins with higher learning rates for rapid initial convergence and gradually reduces the learning rate following a cosine function. The cosine annealing approach enables effective exploration of the parameter space during early training phases while allowing fine-grained optimization in later epochs, facilitating escape from local optima and improved final convergence.

Training is conducted with a batch size of 16 samples over a maximum of 100 epochs, with each epoch processing the entire training dataset. The relatively modest batch size accommodates the memory requirements of the multimodal architecture while ensuring stable gradient estimates for the multi-task learning objectives.

The training process incorporates a multi-task learning framework that simultaneously optimizes burned area prediction and land cover classification objectives through a combined loss function that balances both tasks. The specific loss function formulation and weighting strategies are detailed in Section 5.2. To prevent overfitting and ensure robust model generalization, the training process incorporates several regularization mechanisms. Beyond the weight decay regularization applied through the optimizer, an early stopping mechanism monitors validation performance to halt training when convergence is achieved.

The early stopping implementation tracks validation IoU (Intersection over Union) as the primary performance metric, with a patience parameter of 30 epochs and a minimum improvement threshold of 0.001. This configuration allows sufficient time for the model to recover from temporary performance plateaus while preventing unnecessary prolonged training that could lead to overfitting. The early stopping mechanism operates in 'maximum' mode, seeking to maximize validation IoU performance. Throughout training, the model with the highest validation IoU is automatically saved as the best checkpoint, ensuring that the final model represents the optimal performance state rather than the last training epoch.

Comprehensive training monitoring is implemented using TensorBoard logging to track loss progression, performance metrics, and model predictions throughout the training process. The monitoring system records separate loss values for both burned area and land cover tasks, enabling detailed analysis of multi-task learning dynamics. Visual monitoring includes periodic logging of prediction samples compared against ground truth data, providing qualitative assessment of model performance evolution. These visualizations display input satellite imagery using a false-color composite

(SWIR/NIR/Red bands) alongside ground truth burned area masks and model predictions, facilitating immediate assessment of prediction quality and potential failure modes. The monitoring framework tracks key metrics including training and validation loss for both tasks, IoU scores, and learning rate evolution, enabling detailed post-training analysis of convergence patterns, task interaction effects, and optimization dynamics specific to the multi-task learning framework.

The training implementation utilizes PyTorch as the primary deep learning framework, with GPU acceleration for efficient computation. The multimodal architecture requires substantial computational resources due to the processing of multiple high-resolution satellite imagery channels combined with auxiliary geospatial data streams. Data loading employs parallel processing with 4 worker threads and memory pinning for optimal GPU utilization, while the training process incorporates computational optimizations to manage memory requirements while maintaining numerical precision for the multi-task optimization objectives.

A comprehensive summary of all training hyperparameters and configuration settings employed in this work is provided in Table 5.2

Parameter	Value
Batch size	16
Epochs	100
Input resolution	256×256
Optimizer	SGD
Learning Rate	0.01
Momentum	0.9
Weight Decay	1×10^{-4}
Scheduler	Cosine Annealing
Min Learning Rate	1×10^{-4}
Early Stopping Patience	30 epochs
Early Stopping Threshold	0.001

Table 5.2: Summary of training hyperparameters and configuration settings

The Python library for data augmentation considered in this thesis is albumentations, which provides a wide range of techniques for image augmentation. In this work, the augmentation pipeline was specifically designed to improve the robustness of the model when dealing with multi-source geospatial data (Sentinel, Landsat, DEM, ERA5, ignition points, and landcover). The transformations are mainly geometric, aiming to simulate realistic variations that may occur in remote sensing data.

The pipeline applies horizontal and vertical flips, random 90° rotations, small affine transformations (shift, scale, and rotation), and random resized crops. These operations introduce variability in orientation, scale, and position of the input data, making the model more invariant to geometric transformations. All blank areas resulting from transformations are filled with reflection at the image borders, thus preserving local structures and continuity of visual features.

Table 5.3 summarizes the augmentations with their corresponding probabilities

and parameter values.

Transformation	Probability	Values
Horizontal Flip	0.6	–
Vertical Flip	0.6	–
Random Rotate 90°	0.6	–
Shift	0.5	± 0.15
Scale	0.5	± 0.1
Rotation	0.5	$\pm 15^\circ$
Random Resized Crop	0.7	Scale: [0.7, 0.9]

Table 5.3: Data augmentation parameters used in the proposed pipeline.

By using these different augmentation techniques, the training set is enriched with a greater diversity of possible image variations, allowing the model to increase its robustness and generalization capabilities when applied to real-world remote sensing scenarios.

5.2 Loss Functions

The choice of the loss function plays a fundamental role in guiding the optimization process of the network and, consequently, in determining its performance on the target tasks. In the proposed framework two tasks are learned jointly: the segmentation of the burned area and the classification of landcover categories. For each of these tasks, a different loss function has been adopted, selected according to the specific characteristics of the problem.

For the segmentation of burned areas, the network is required to discriminate between burned and unburned pixels, which corresponds to a binary segmentation problem. In this context, a *Dice loss* has been employed. The Dice coefficient, from which this loss is derived, measures the degree of overlap between the predicted mask and the ground truth. Its value ranges from 0 to 1, where 1 corresponds to a perfect overlap. The corresponding loss is defined as one minus the Dice coefficient, thus forcing the model to maximize the similarity between predictions and actual burned regions. This formulation is particularly effective when the classes are highly imbalanced, as is often the case in wildfire mapping: the burned area typically occupies only a small fraction of the input patch, and standard pixel-wise losses such as Cross-Entropy may lead the model to favor the majority class (non-burned). By focusing directly on the overlap, the *Dice loss* mitigates this problem and better reflects the quality of the segmentation in real-world applications.

In parallel, the model also learns to predict landcover classes, which provide additional contextual information about the type of surface affected by the fire. This is formulated as a multi-class classification problem, for which a *Cross-Entropy loss* has been used. This function measures the dissimilarity between the predicted probability distribution and the true class distribution. Specifically, for each pixel the prediction consists of a probability vector over the possible classes, and the loss

penalizes strongly those cases in which the probability assigned to the correct class is low. The logarithmic nature of the formulation ensures that misclassifications with high confidence are penalized more severely, while correct predictions with high probability contribute little to the loss. In order to prevent irrelevant background pixels from influencing the optimization, all pixels labeled as background are ignored in the computation of the *Cross-Entropy loss*. This design choice allows the model to concentrate on the areas of real interest, improving the quality of the landcover predictions.

The overall optimization objective of the network is obtained by combining the two contributions into a single scalar value, which the model minimizes during training. The total loss is expressed as a weighted sum of the individual losses:

$$L = \lambda_{\text{mask}} \cdot L_{\text{Dice}} + \lambda_{\text{lc}} \cdot L_{\text{CE}}. \quad (5.1)$$

In the experiments both weights were set to $\lambda_{\text{mask}} = \lambda_{\text{lc}} = 0.5$, giving equal importance to the two tasks. This balance reflects the dual objective of the framework: on one hand, to accurately delineate the extent of the burned areas, which represents the main output of the system, and on the other hand, to exploit auxiliary landcover classification in order to encourage the network to learn richer and more discriminative feature representations. By jointly optimizing these two complementary objectives, the model achieves improved robustness and generalization, which are crucial in the context of wildfire mapping across heterogeneous landscapes and varying environmental conditions.

5.3 Evaluation Metrics

The evaluation of model performance was carried out using the *Intersection over Union (IoU)*, also known as the *Jaccard Index*. This metric was chosen as the sole performance indicator since the primary objective of the framework is the accurate segmentation of burned areas. The auxiliary landcover prediction task, although present during training, was not evaluated with a dedicated metric. Its role was limited to serving as a form of regularization: it was sufficient to verify that the corresponding loss decreased during training, ensuring that the network was effectively learning also this auxiliary task.

Formally, the IoU between a predicted region $\hat{Y}$ and the ground truth region Y is defined as

$$\text{IoU}(Y, \hat{Y}) = \frac{|Y \cap \hat{Y}|}{|Y \cup \hat{Y}|}, \quad (5.2)$$

where $|Y \cap \hat{Y}|$ denotes the number of pixels correctly predicted as belonging to the target class, and $|Y \cup \hat{Y}|$ the total number of pixels belonging to either the ground truth or the prediction. The value of the IoU ranges from 0 to 1, where 1 indicates perfect agreement between predicted and reference regions. This metric

is particularly suitable in segmentation problems, as it provides a direct measure of spatial overlap and is robust to class imbalance, which is especially relevant in wildfire mapping where burned areas may represent only a small fraction of the scene.

5.4 Results

The experimental evaluation aimed at assessing the feasibility of predicting final burned area extent from pre-fire observations using the proposed multimodal deep learning architecture. Given the exploratory nature of this work, the focus was not on reaching operational performance, but rather on understanding the model’s behavior under different configurations and identifying promising directions for future research. The evaluation strategy centered on systematic ablation studies, where individual components were varied while maintaining all other factors constant, enabling a clearer understanding of each design choice’s contribution to overall performance.

The best-performing configuration, obtained with the selected hyperparameters, fixed data augmentations, balanced loss weights (0.5 for both land cover and mask losses), and 1×1 convolutions in the fusion blocks, achieved a maximum Intersection over Union (IoU) of **38.59%**. While modest in absolute terms, this represents a meaningful result considering the inherent complexity of the task and the experimental constraints of this study. This baseline configuration was established after preliminary hyperparameter tuning and served as the reference point against which all subsequent variations were compared. The equal weighting between the two loss components was motivated by the hypothesis that both environmental context (captured through land cover classification) and direct spatial information (encoded in the burned area mask) should contribute equally to the learning process.

To further explore the robustness of the model and investigate the sensitivity of the architecture to specific design decisions, several controlled variations were tested while keeping all hyperparameters constant and modifying only a single component at a time. This experimental protocol was designed to isolate the effect of each modification and provide insights into the relative importance of different architectural and training choices:

- **Fusion block convolution size.** The first variation examined the impact of kernel size within the fusion blocks, which are responsible for integrating information from different modalities. Replacing 3×3 convolutions with 1×1 kernels led to a decrease in performance, with a maximum IoU of 37.73%. While the difference appears relatively small, it suggests a consistent trend across the validation set. A plausible explanation is that larger kernels introduced additional parameters and a broader receptive field, which may have diluted the capacity of the fusion block to integrate modality-specific features effectively in the limited dataset scenario. The 1×1 convolutions, being more parameter-efficient, may have provided a better inductive bias for the cross-modal fusion task when training data is scarce. This finding highlights the importance of

balancing model capacity with dataset size, particularly in fusion architectures where different information streams must be combined effectively.

- Loss weighting.** The second experiment investigated whether altering the balance between the two supervised signals could improve performance by prioritizing one learning objective over the other. Adjusting the weights to 0.6 for land cover classification and 0.4 for burned area mask prediction resulted in a maximum IoU of 34.69%, representing a more substantial drop compared to the baseline. This suggests that deviating from equal weighting reduced the model’s ability to jointly optimize environmental context and burned area segmentation, potentially biasing the network toward one modality at the expense of overall predictive coherence. The land cover classification task, while providing valuable contextual information, may not directly encode the specific spatial patterns required for accurate burned area prediction. Over-emphasizing this auxiliary task appears to have diverted the model’s representational capacity away from the primary segmentation objective, demonstrating the delicate balance required in multi-task learning scenarios.
- Data augmentation strategy.** Given the limited size of the available dataset, data augmentation plays a crucial role in preventing overfitting and improving generalization. However, the third experiment revealed that augmentation strategies must be carefully calibrated to the specific characteristics of the data. Increasing the probability of geometric augmentations (e.g., flips up to 0.8) and introducing additional pixel-level perturbations such as brightness and contrast adjustments degraded performance to 29.74% IoU. This result indicates that overly aggressive augmentations may have distorted the already limited data distribution, reducing the model’s ability to learn meaningful spatial patterns of burned area propagation. Wildfire spread patterns are inherently tied to specific environmental conditions and topographical features, and excessive augmentation may have broken these natural associations, creating synthetic training examples that diverge too far from the underlying physical processes. This finding underscores the need for domain-aware augmentation strategies that respect the physical constraints of the prediction task.
- Stratified batch sampling.** An additional experiment attempted to address the class imbalance inherent in the dataset by enforcing stratified sampling within each batch. The dataset exhibits a strong imbalance in fire sizes, with the vast majority of events being small fires, while large fires are relatively rare but often of greater societal and ecological impact. To ensure that the model received balanced exposure to fires of different scales during training, fires were categorized by burned area extent into small (< 2000 pixels), medium (2000–10000 pixels), and large (> 10000 pixels), and batches were constructed to maintain equal proportions across these categories. The intent was to ensure balanced exposure to different fire scales during training, potentially improving

the model’s capacity to generalize across varying fire magnitudes and reducing the bias toward predicting smaller burned areas. However, this approach yielded a maximum IoU of 30.89%, representing a notable decline from the baseline performance. This suggests that artificially enforcing uniform representation may have disrupted the natural data distribution and prevented the model from learning scale-dependent patterns effectively. The forced balancing may have caused the model to encounter large fires with disproportionate frequency relative to their true occurrence, potentially leading to a misalignment between the learned distribution and the actual distribution of fire sizes encountered during validation.

- **Dataset oversampling by fire size.** To further investigate the model’s sensitivity to fire scale and explore an alternative approach to addressing class imbalance, an oversampling strategy was applied by selectively replicating training samples based on their burned area extent. Specifically, medium-sized fires were included twice and large fires three times in each training epoch, while keeping small fires at their original frequency. The rationale was to mitigate the dominance of small fires in the dataset and improve predictive performance on larger burned areas, which are often of greater operational interest for fire management agencies and pose more significant threats to communities and ecosystems. This approach was expected to provide the model with more learning opportunities from underrepresented fire scales without completely disrupting the natural distribution, as would occur with strict balanced sampling. Contrary to expectations, this strategy led to severe performance degradation, with a maximum IoU of only 20.32%, the lowest result across all tested configurations. The substantial drop suggests that the introduced redundancy may have caused overfitting to the limited set of large fire instances, while simultaneously reducing the model’s ability to capture the diversity of smaller events that constitute the majority of real-world cases. The repeated exposure to the same large fire examples likely led the model to memorize specific patterns rather than learning generalizable features, demonstrating the risk of naive oversampling approaches in scenarios where the number of available samples in minority classes is already very limited.

These experiments highlight how small design changes can substantially affect model performance in a task where signal-to-noise ratio is inherently low and where the relationship between input features and target outputs is highly complex and nonlinear. Notably, modifications to the data sampling and distribution strategy proved particularly detrimental, underscoring the challenge of balancing dataset representativeness with the need to address class imbalance in wildfire prediction. The failures of both stratified sampling and oversampling approaches suggest that the natural distribution of fire sizes, despite its imbalance, may encode important information about the underlying processes that the model needs to learn. Artificially altering this distribution, whether through balanced sampling or selective replication, appears

to interfere with the model’s ability to capture these patterns. The consistency of reasonable performance across different configurations (with most variations yielding IoU values above 30%) provides evidence that the core architecture is sound and capable of extracting meaningful patterns from the available data.

Qualitative inspection further illustrates the model’s behavior and reveals specific characteristics of its predictions. In many cases, predictions tend to favor compact and homogeneous shapes, in contrast with the often irregular and fragmented geometries of the true burned areas. This systematic bias toward smoother, more regular predictions likely reflects both the inductive biases of convolutional architectures, which naturally favor spatially coherent patterns, and the challenge of capturing the stochastic nature of fire spread from deterministic environmental inputs alone. The model appears to learn average fire spread tendencies but struggles to reproduce the fine-grained spatial variations that emerge from local heterogeneities in fuel distribution, micro-topography, and moment-to-moment changes in wind conditions. Figure 5.1 shows examples where pre-fire Sentinel-2 imagery is compared with the ground truth burned area and the predicted burned area, illustrating both cases where the model successfully captures the general fire extent and cases where it fails to predict irregular propagation patterns or fragmented burned areas.

The experimental results demonstrate that while the current approach shows promise, significant technical challenges remain. The moderate IoU scores reflect inherent limitations in predicting fire behavior from static environmental snapshots without dynamic meteorological information during the fire event itself. Weather conditions such as wind speed and direction, temperature, and humidity change throughout a fire’s progression and play crucial roles in determining final burned extent. The static nature of the pre-fire observations used in this study cannot capture these dynamic influences, which likely accounts for much of the prediction uncertainty. Additionally, the limited dataset size constrains the model’s ability to learn robust representations across the diverse conditions under which fires occur. The experiments with data augmentation and resampling revealed how sensitive performance is to the characteristics of the training distribution, suggesting that substantially larger and more diverse datasets would be necessary to achieve more reliable predictions across varied fire scenarios.

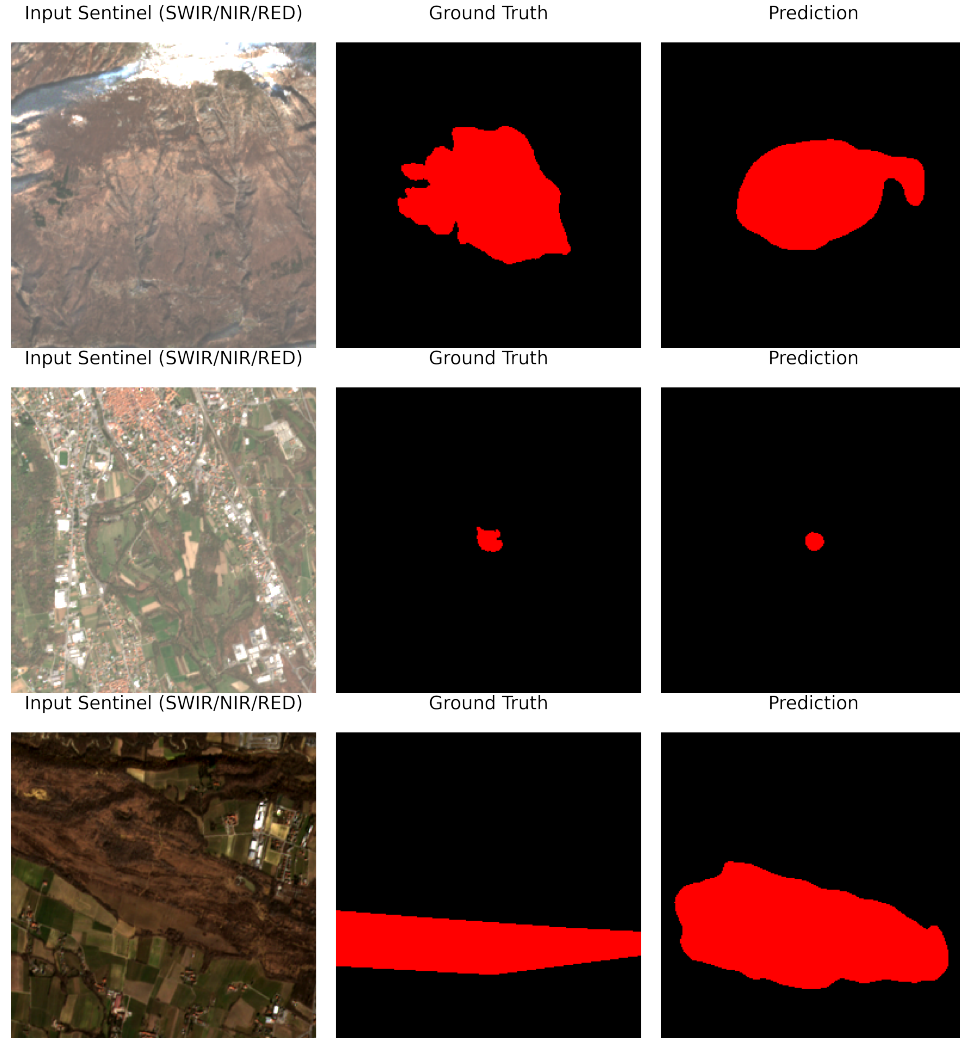


Figure 5.1: Examples of qualitative comparison: Sentinel-2 pre-fire acquisition (left), ground truth burned area (center), and predicted burned area (right).

Chapter 6

Conclusion

This thesis set out to explore whether it is possible to predict wildfire extent before the fire actually occurs, using only environmental conditions and the knowledge of where ignition takes place. This represents a significant departure from traditional approaches in wildfire remote sensing, which have predominantly focused on mapping burned areas after fires have concluded. The challenge was substantial: rather than identifying spectral changes caused by fire damage, the task required learning complex relationships between landscape characteristics and fire behavior patterns from environmental preconditions alone. The multimodal deep learning framework developed in this work achieved a maximum validation IoU of 38.59% for burned area prediction. While this performance level might initially seem modest, it must be understood within the context of predicting future fire behavior rather than classifying existing fire damage. The results demonstrate that environmental data contains meaningful predictive signals about fire spread patterns, establishing that pre-fire prediction is not only conceptually feasible but practically achievable with current remote sensing and machine learning technologies. One of the most important insights from this research concerns the critical role of ignition point information. The prediction task becomes fundamentally different when the model knows where a fire starts compared to scenarios where ignition location must also be inferred. With ignition coordinates provided, the network can focus entirely on learning how fires spread given specific environmental conditions, rather than simultaneously solving the problems of where fires are likely to start and how they subsequently propagate. This finding has important implications for operational applications, suggesting that early detection systems combined with predictive models could provide valuable decision support for emergency response teams. The systematic experimentation conducted throughout this work revealed how sensitive multimodal architectures can be to seemingly minor design choices. The optimal configuration emerged from careful balancing of architectural components, loss function weighting, and data augmentation strategies. These findings underscore the complexity of the prediction task and the importance of methodical experimental approaches when working in such unexplored research territory. Perhaps most importantly, this work establishes a methodological foundation for future research in pre-fire

wildfire prediction. The experimental framework, data integration strategies, and evaluation methodologies developed here provide a starting point for researchers who wish to advance this field further. The relatively modest performance achieved suggests significant room for improvement, making this an exciting area for continued investigation. Looking forward, numerous avenues exist for enhancing predictive capabilities. Larger datasets encompassing diverse geographic regions and fire regimes could improve model generalization. More sophisticated neural architectures might better capture the complex spatial-temporal dynamics of fire propagation. Additional environmental variables or higher temporal resolution data could provide richer context for prediction models. Each of these directions builds upon the foundation established in this thesis. The broader significance of this work extends beyond the specific technical contributions. By demonstrating the feasibility of pre-fire burned area prediction, this research opens possibilities for shifting wildfire management from reactive to proactive strategies. Rather than responding to fires after they have spread, future systems could potentially anticipate fire behavior patterns and inform preventive measures or optimized resource deployment. This represents a meaningful step toward more effective wildfire risk management in an era of increasing fire frequency and intensity.

Bibliography

- [1] IPCC. “Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change”. In: (2023), p. 184 (cit. on p. 1).
- [2] Matthew W Jones et al. “State of Wildfires 2023–2024”. In: *Earth System Science Data* 16.8 (2024), pp. 3601–3685. DOI: 10.5194/essd-16-3601-2024 (cit. on p. 1).
- [3] Xanthe J Walker et al. “Global rise in forest fire emissions linked to climate change in the extratropics”. In: *Science* 383.6684 (2024), ead15889. DOI: 10.1126/science.ad15889 (cit. on p. 1).
- [4] NASA Jet Propulsion Laboratory. *New NASA Study Tallies Carbon Emissions From Massive Canadian Fires*. <https://www.jpl.nasa.gov/news/new-nasa-study-tallies-carbon-emissions-from-massive-canadian-fires/>. Accessed: August 30, 2025. Aug. 2024 (cit. on p. 1).
- [5] CAL FIRE. *January 2025 Los Angeles County Wildfires*. California Department of Forestry and Fire Protection. Incident reports and damage assessments. Jan. 2025 (cit. on p. 1).
- [6] New York City Department of Environmental Protection. *Canadian Wildfire Smoke Impacts New York City Air Quality*. Air Quality Alert and Health Advisory. June 6-8, 2023 air quality emergency. June 2023 (cit. on p. 1).
- [7] Marco Turco, Sonia Jerez, Francisco J Doblas-Reyes, Amir AghaKouchak, Maria Carmen Llasat, and Antonello Provenza. “The regional economic impact of wildfires: Evidence from Southern Europe”. In: *Journal of Environmental Economics and Management* 118 (2023), p. 102787. DOI: 10.1016/j.jeem.2023.102787 (cit. on p. 2).
- [8] AccuWeather. *Los Angeles Wildfires Economic Impact Assessment*. Economic impact analysis. Total damage and economic loss estimate. Jan. 2025 (cit. on p. 2).
- [9] Jon E Keeley. “Fire intensity, fire severity and burn severity: a brief review and suggested usage”. In: *International Journal of Wildland Fire* 18.1 (2009), pp. 116–126. DOI: 10.1071/WF07049 (cit. on p. 2).

- [10] World Resources Institute. *New Data Confirms: Forest Fires Are Getting Worse*. <https://www.wri.org/insights/global-trends-forest-fires>. Accessed: August 30, 2025. 2024 (cit. on p. 2).
- [11] European Commission Joint Research Centre. *Wildfires: 2023 among the worst in the EU in this century*. https://joint-research-centre.ec.europa.eu/jrc-news-and-updates/wildfires-2023-among-worst-eu-century-2024-04-10_en. Accessed: August 30, 2025. Apr. 2024 (cit. on p. 2).
- [12] Emilio Chuvieco, Alexander Röder, Marj Tonini, Maria Lucrecia Pettinari, Maria J. López García, Hugo Storm, and Angelika Heil. “Towards an Integrated Approach to Wildfire Risk Assessment: When, Where, What and How May the Landscapes Burn”. In: *Fire* 6.5 (2023), p. 215. DOI: 10.3390/fire6050215 (cit. on pp. 4, 7).
- [13] Lukas Knopp, Marc Wieland, Maximilian Rättich, and Sandro Martinis. “A Deep Learning Approach for Burned Area Segmentation with Sentinel-2 Data”. In: *Remote Sensing* 12.15 (2020), p. 2422. DOI: 10.3390/rs12152422 (cit. on p. 8).
- [14] Yifang Ban, Puzhao Zhang, Andrea Nascetti, Alistair R Bevington, and Michael A Wulder. “Burnt-Net: Wildfire burned area mapping with single post-fire Sentinel-2 data and deep learning morphological neural network”. In: *Ecological Indicators* 140 (2022), p. 108999 (cit. on p. 8).
- [15] Alessandro Farasin, Jacopo Pegoraro, and Claudio Rossi. “Deep-learning-based burned area mapping using the synergy of Sentinel-1&2 data”. In: *Remote Sensing of Environment* 264 (2021), p. 112575 (cit. on p. 8).
- [16] Xiaowei Hu, Yifang Ban, and Andrea Nascetti. “Two scalable approaches for burned-area mapping using U-Net and Landsat imagery”. In: *arXiv preprint arXiv:2311.17368* (2023) (cit. on p. 9).
- [17] Maria Sdraka, Emmanouil B. Alexakis, and Konstantinos Karantzalos. “Robust Burned Area Delineation Through Multitask Learning”. In: *Proceedings of the European Conference on Computer Vision Workshops*. Springer, 2024, pp. 532–547. DOI: 10.1007/978-3-031-74633-8_32 (cit. on p. 9).
- [18] Li Wang, Jie Zhang, Pengfei Liu, Kim-Kwang Raymond Choo, and Fenghua Huang. “Land cover classification from remote sensing images based on multi-scale fully convolutional network”. In: *Geo-spatial Information Science* 25.2 (2021), pp. 278–294. DOI: 10.1080/10095020.2021.2017237 (cit. on p. 9).
- [19] Yan Liu, Yanfei Zhong, Fan Fei, Qing Zhu, and Qianqing Qin. “Land-cover classification with high-resolution remote sensing images using transferable deep models”. In: *Remote Sensing of Environment* 237 (2019), p. 111322. DOI: 10.1016/j.rse.2019.111322 (cit. on p. 9).
- [20] K. B. Klimas et al. “A machine learning model to predict wildfire burn severity for pre-fire risk assessments, Utah, USA”. In: *Fire Ecology* 21.1 (Feb. 2025), p. 8. DOI: 10.1186/s42408-024-00346-z (cit. on p. 10).

- [21] John L Hodges and Brian Y Lattimer. “Wildland fire spread modeling using convolutional neural networks”. In: *Fire Technology* 55.6 (2019), pp. 2115–2142 (cit. on p. 11).
- [22] Fabien Huot, Rui L Hu, Nadhir Goyal, Teja Sankar, Matthias Ihme, and Yi-Fan Chen. “Next day wildfire spread: A machine learning dataset to predict wildfire spreading from remote-sensing data”. In: *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), pp. 1–13 (cit. on p. 11).
- [23] Ioannis Prapas et al. “Wildfire spreading prediction using multimodal data and deep neural network approach”. In: *Scientific Reports* 14.1 (2024), p. 1990 (cit. on p. 11).
- [24] Derek Radke, Anja Hessler, and Derek Ellsworth. “FireCast: Leveraging deep learning to predict wildfire spread”. In: *Proceedings of the 28th International Joint Conference on Artificial Intelligence* (2019), pp. 4575–4581 (cit. on p. 11).
- [25] Ferran J Alcasena, Michele Salis, Alan A Ager, Rosa Castell, and Cristina Vega-García. “Wildfire risk transmission in the Colorado Front Range, USA”. In: *Risk Analysis* 38.2 (2018), pp. 441–459 (cit. on p. 11).
- [26] Teixeira Artés, Duarte Oom, Daniele de Rigo, Tracy Houston Durrant, Pieralberto Maianti, Guido Liberty, and Jesús San-Miguel-Ayanz. “A global wildfire dataset for the analysis of fire regimes and fire behaviour”. In: *Scientific Data* 6.1 (2019), pp. 1–11 (cit. on p. 12).
- [27] Pedro Silva, Ana Santos, and Miguel Costa. “Machine learning approaches for burned area prediction using environmental covariates”. In: *Environmental Modelling & Software* 147 (2022), p. 105245 (cit. on p. 12).
- [28] Regione Piemonte – Protezione Civile (A18120); ARPA Piemonte. *Regione Piemonte – Incendi boschivi: aree percorse e punti di innesco (1997–2024)*. https://www.geoportale.piemonte.it/geonetwork/auth/api/records/r_piemon:5d8d0dd5-c0ac-4403-afb9-c116ccefedc8. Dataset vettoriale, aggiornato al 23 maggio 2024; WGS84/UTM32N; accessibile via WMS o shapefile. 2024 (cit. on p. 14).
- [29] European Space Agency. *Copernicus Sentinel-2 Level-2A*. <https://planetarycomputer.microsoft.com/dataset/sentinel-2-l2a>. Accessed via Microsoft Planetary Computer. 2021 (cit. on p. 15).
- [30] Microsoft. *Microsoft Planetary Computer*. <https://planetarycomputer.microsoft.com/>. Accessed: 2025-09-05. 2025 (cit. on pp. 16, 23).
- [31] U.S. Geological Survey. *Landsat Collection 2 Level-2 Science Products*. <https://planetarycomputer.microsoft.com/dataset/landsat-c2-12>. Accessed via Microsoft Planetary Computer. 2021 (cit. on pp. 17, 23).
- [32] Copernicus European Union. *Copernicus Digital Elevation Model (DEM) - GLO-30*. <https://planetarycomputer.microsoft.com/dataset/cop-dem-glo-30>. Accessed via Microsoft Planetary Computer. 2020 (cit. on p. 18).

- [33] Copernicus Climate Change Service (C3S). *Derived ERA5-Land Daily Statistics*. <https://cds.climate.copernicus.eu/cdsapp!/dataset/derived-era5-land-daily-statistics>. Accessed via Copernicus Climate Data Store. 2024 (cit. on p. 19).
- [34] Johan R. Meijer, Mark A.J. Huijbregts, Kees C.G.J. Schotten, and Aafke M. Schipper. “Global patterns of current and future road infrastructure”. In: *Environmental Research Letters* 13.6 (2018). Data available via the GLOBIO portal, p. 064006. DOI: 10.1088/1748-9326/aabd42 (cit. on p. 20).
- [35] GLOBIO Project. *Global Roads Inventory Project (GRIP) dataset*. <https://www.globio.info/download-grip-dataset>. Accessed: 2025-08-26. 2018 (cit. on p. 20).
- [36] Esri Impact Observatory Microsoft. *IO-LULC Annual V2: Dynamic World Land Cover*. <https://planetarycomputer.microsoft.com/dataset/io-lulc-annual-v02>. Accessed: 2025-09-09. 2023 (cit. on p. 20).
- [37] C. Aybar, L. Ysuhuaylas, J. Loja, et al. “CloudSEN12, a global dataset for semantic understanding of cloud and cloud shadow in Sentinel-2”. In: *Scientific Data* 9 (2022), p. 782. DOI: 10.1038/s41597-022-01878-2. URL: <https://doi.org/10.1038/s41597-022-01878-2> (cit. on p. 28).
- [38] M. Paganini et al. “ITALSCAR, a regional burned forest mapping demonstration project in Italy”. In: *IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No.03CH37477)*. Toulouse, France: IEEE, 2003, pp. 1290–1292. DOI: 10.1109/IGARSS.2003.1294086 (cit. on p. 32).
- [39] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. “Feature Pyramid Networks for Object Detection”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, pp. 2117–2125 (cit. on pp. 33–35).