

Development of a wildfire risk prediction system based on Deep Learning methods and Remote Sensing

Jhony Alexander Sánchez Vargas

Matriculation Number: 554283

Thesis Supervisors:

Prof. Dr. Mana Gharun

Johannes Heisig

Prof. Dr. Marco Painho

In Partial Fulfillment of the Requirements
for the Degree of
Master of Science in Geospatial Technologies

University of Münster
Institute for Geoinformatics
IFGI Graduate School

Winter Semester 2025

Declaration of Originality

I, Jhony Alexander Sanchez Vargas, hereby declare that this thesis, titled "Development of a wildfire risk prediction system based on Deep Learning methods and Remote Sensing", is my original work and has been written independently. Any ideas, data, or works from other authors have been properly acknowledged and cited in accordance with academic standards.

I affirm that this research has not been previously submitted, in whole or in part, to any other institution for any degree or qualification. Additionally, I confirm that this work complies with the ethical guidelines and academic integrity policies of The University of Münster.

Signed in Münster, 20-02-2024

Jhony Alexander Sanchez Vargas

University of Münster

Abstract

Wildfires pose a significant threat to ecosystems, human life, and infrastructure, particularly in South America, where diverse climatic and environmental factors contribute to their occurrence. Climate change has intensified extreme weather conditions, leading to an increase in wildfire frequency and severity. Countries such as Brazil have experienced escalating wildfire damage, highlighting the need for predictive models that can accurately assess fire risk and support mitigation efforts. This thesis develops a wildfire risk prediction system leveraging deep learning methods and remote sensing data to address this challenge.

The objectives of this research are threefold: (1) explore how Earth Observation (EO) data cubes can be leveraged to efficiently process and analyze spatiotemporal data for wildfire danger prediction; (2) assess the suitability of various deep learning algorithms for detecting, assessing, and mitigating wildfire impacts using remote sensing data; and (3) examine how the integration of static and dynamic variables can enhance the accuracy of near real-time wildfire risk predictions. Key research questions focus on the role of EO data cubes in improving spatiotemporal analysis, identifying the most effective deep learning models for wildfire risk assessment, and evaluating the benefits of combining static and dynamic variables for enhanced predictive performance.

To tackle the stochastic nature of wildfire occurrences, this study evaluates the performance of Random Forest (RF), Long Short-Term Memory networks (LSTM), Convolutional Neural Networks (CNN), and Convolutional LSTM (ConvLSTM) models. The results indicate that LSTM is the most balanced model, achieving an AUROC of 89% and an F1-score of 72%, effectively capturing temporal dependencies while maintaining a trade-off between false positives and false negatives. CNN exhibited the highest sensitivity, with an 84% recall, though at the cost of lower precision (67%). ConvLSTM did not outperform LSTM significantly, suggesting that its advantage depends on the scale and structure of the input data. The RF model achieved the highest AUROC (91%) but struggled with sensitivity. Feature importance analysis identified human modification (gHM), land surface temperature (LST), and temperature as the most influential variables, reflecting the combined impact of anthropogenic and environmental drivers on fire occurrence. Seasonal analysis further demonstrated the influence of temperature, vegetation, and precipitation conditions on fire activity.

The findings highlight the effectiveness of integrating EO data cubes and deep learning for wildfire risk prediction. The LSTM model's ability to balance spatiotemporal dependencies makes it a strong candidate for real-time fire forecasting. This study contributes to advancing scalable, data-driven wildfire prediction models, providing valuable insights for fire management strategies in South America.

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my parents, whose unwavering support, encouragement, and belief in me have been the foundation of my academic journey. Their love and sacrifices have shaped the person I am today, and for that, I am eternally grateful.

I am also profoundly thankful to my supervisors for their invaluable guidance, patience, and insights throughout this research. Their expertise and continuous support have been instrumental in shaping this work, and I deeply appreciate the time and effort they have dedicated to mentoring me.

A special thank you to my friends and colleagues who have accompanied me on this journey. Their encouragement, thought-provoking discussions, and moments of shared frustration and triumph have made this experience both enriching and rewarding.

Finally, I extend my appreciation to everyone who contributed to this research in any capacity, whether through technical assistance, inspiring conversations, or simply offering a listening ear. This thesis would not have been possible without the collective support of all these incredible individuals.

Thank you all.

Contents

List of Figures	vi
List of Tables	vii
Abbreviations	viii
1 Introduction	1
2 Literature Review	4
2.1 Wildfire influencing factors	4
2.2 Risk of wildfires	5
2.3 Remote sensing for Wildfire monitoring	6
2.4 Existing wildfire monitoring platforms	7
2.5 Deep Learning in Wildfire related studies	9
2.5.1 Wildfire Risk Modelling in Germany	10
3 Dataset and Study Area	11
3.1 Study Area	11
3.2 Dataset	13
3.2.1 Target variable: MODIS Burned Area	13
3.2.2 Predictive variables	13
4 Methodology	20
4.1 Experimental Set-up	20
4.2 Google Earth Engine API	21
4.3 Active Fire Sampling	21
4.4 Non-Fire Sampling	23
4.5 Variable Extraction	23
4.5.1 Earth Observation Data Cubes:	23
4.5.2 Data Preprocessing:	25
4.6 Models Architecture	26
4.6.1 Random Forest (RF)	26
4.6.2 Long Short-Term Memory (LSTM)	26
4.6.3 Convolutional Neural Network (CNN)	27
4.6.4 Convolutional Long Short-Term Memory (ConvLSTM)	28
4.6.5 Model evaluation	30
4.6.6 Proposed models architecture	31
5 Results and Discussion	32
5.1 Sampling distribution	32
5.2 Environmental and Climatic Differentiation	33
5.2.1 Surface and Vegetation Conditions	33

5.2.2	Climatic Factors and Fire Susceptibility	34
5.2.3	Topographical and Anthropogenic Influences	35
5.2.4	Analysis of Variable Correlations	36
5.3	Wildfire Risk Prediction	38
5.3.1	Influence of Dynamic and Static variables	39
5.4	Real-World Application of Wildfire Risk Models	41
5.5	Wildfire risk prediction	41
5.5.1	Zone 1 - Dry season (february)	42
5.5.2	Zone 1 - Rain season (September)	44
5.5.3	Zone 2 - Dry season (September)	45
5.5.4	Zone 2 - Rain season (February)	46
5.6	Web Application	48
5.6.1	Web Application	48
5.6.2	Model Implementation	48
5.6.3	Performance and Scalability	50
5.7	Discussion	50
6	Conclusion	52
	References	60
A	Appendix	61
A.1	AI Use Declaration	61
A.2	Web Application documentation	61

List of Figures

1	Major climate regions of South America	12
2	Methodology Flowchart	22
3	Patch Images	25
4	Long short-term memory (LSTM)	27
5	Convolutional Long short-term memory (ConvLSTM)	29
6	Samples spatial distribution	32
7	LULC distribution for fire and non fire points	33
8	Distribution of Surface and Vegetation variables	34
9	Distribution of climatic variables	35
10	Distribution of Topographical and Anthropogenic variables	36
11	Variables correlation	37
12	Impact of varaibles in Wildfire Risk prediction	40
13	Fire activity time-series	41
14	Fire Risk prediction zone 1 - Dry season	43
15	Fire Risk prediction zone 1 - Rain season	44
16	Fire Risk prediction zone 2 - Dry season	46
17	Fire Risk prediction zone 2 - Rain season	47
18	Webapp - date selector	49
19	Webapp - draw Area of Interest	49
20	Webapp - calculate and visualize Wildfire Risk	50

List of Tables

1	Datacube input sources	14
2	ERA5-Land variables description	18
3	Metrics for wildfire risk prediction models	39
4	AI Use Declaration	61

Abbreviations

AHP	Analytical Hierarchy Process
ANN	Artificial Neural Network
API	Application Programming Interface
AUROC	Area Under the Receiver Operating Characteristic Curve
CNN	Convolutional Neural Network
DL	Deep Learning
dNBR	Differenced Normalized Burn Ratio
EFFIS	European Forest Fire Information System
FDRS	Fire Danger Rating Systems
FWI	Fire Weather Index
GEE	Google Earth Engine
GWIS	Global Wildfire Information System
INEGI	National Institute of Statistics and Geography (Mexico)
INPE	National Institute for Space Research (Brazil)
LWIR	Long-Wave Infrared
LSTM	Long Short-Term Memory
ML	Machine Learning
MLP	Multilayer Perceptron
MODIS	Moderate Resolution Imaging Spectroradiometer
MWIR	Mid-Wave Infrared
RNN	Recurrent Neural Networks
NBR	Normalized Burn Ratio
NDVI	Normalized Difference Vegetation Index
NIR	Near-Infrared
RdNBR	Relative Differenced Normalized Burn Ratio
SKNet	Selective Kernel Network
SLA-ConvLSTM	Static Location-Aware Convolutional Long Short-Term Memory
SVM	Support Vector Machine
SWIR	Short-Wave Infrared
VIIRS	Visible Infrared Imaging Radiometer Suite
VIS	Visible Spectrum

1 Introduction

Forests are considered an essential natural resource, which is key to maintaining environmental stability. The state of forests serves as a primary indicator of environmental health and overall ecosystem condition. They are also economically significant for many communities living near forested areas, holding social value and contributing to favorable weather patterns. Historically, forests covered around 5.9 billion hectares of the Earth's surface before industrialization, but they now span approximately 4 billion hectares, representing 31% of the global land area (Piralilou et al. 2022). However, Forest fires are a natural occurrence in many ecosystems. When these fires happen at regular intervals and with typical intensity, they play an important ecological role by clearing dead organic matter and parasites. Additionally, they can enhance soil fertility through nutrient release from burned biomass and create open spaces that allow more sunlight to reach the ground. Factors associated with wildfires, including meteorological conditions, have been shown to affect the population density of various species, serving as an additional indicator of ecosystem health (Ntinopoulos et al. 2023).

Unfortunately, Fire behavior is changing because the conditions under which it occurs are being altered. While not all fires are harmful or need to be extinguished, as they play essential ecological roles, large wildfires that burn for weeks and impact millions of people across vast areas pose a significant challenge we are currently poorly equipped to handle. Although lightning strikes and human negligence have always been triggers for uncontrolled fires and will continue to be, factors such as climate change, land-use shifts, and inadequate land and forest management have created conditions where wildfires are more likely to find the fuel and weather patterns needed to become catastrophic. Fires are now burning longer and more intensely in regions where they have historically occurred and are also emerging in unexpected places, including drying peatlands and thawing permafrost. (United Nations Environment Programme 2022)

Wildfires are a global challenge, causing extensive damage across regions worldwide. Sub-Saharan Africa experiences the highest number of annual fires per square kilometer, with some areas reporting between 10,000 and 20,000 fires each year. In the Americas, wildfires remain a persistent issue, with the United States witnessing between 1.5 million and 4.2 million hectares of land burned annually since 2000. Europe has also faced increasing wildfire damage, with over 550,000 hectares burned in 2021, rising to 750,000 hectares in 2022. Mediterranean countries such as Greece, Italy, Spain, and Portugal are among the most affected. South-East Asia and Oceania also suffer from highly destructive wildfires, with the 2019–2020 Australian bushfires being particularly severe, burning over 19 million hectares of land. (Mandala Partners 2024)

South America is experiencing unprecedented wildfires, with record-breaking blazes even affecting typically moist ecosystems like the Amazon rainforest. Elevated temperatures and prolonged drought, intensified by the El Niño event that concluded earlier in 2024, along with ongoing deforestation and climate change, have fueled fires across millions of hectares

and multiple countries. These fires pose significant threats to lives, property, and air quality, creating hazardous pollution levels in cities such as São Paulo and Rio de Janeiro (Root [2024](#)). While the alarming scale of this year’s fires is concerning, it reflects a troubling long-term pattern. Data from 2001 to 2023 shows that fires in the Amazon now consume at least twice as much forest annually as they did two decades ago. These fires are part of a dangerous feedback loop, largely driven by human activity (World Resources Institute [2024](#)).

Unlike boreal forests, where fire is a natural ecological process, fires in humid tropical forests like the Amazon are rare and primarily human-induced. Many are linked to deforestation practices, where land is intentionally burned to clear space for agriculture and ranching. Research indicates that deforestation contributes to regional shifts in weather patterns, resulting in more intense and prolonged droughts, which in turn increases forest vulnerability to fire. During hot and dry conditions, fires can rapidly spread, consuming vast stretches of forest (X. Xu et al. [2020](#)).

In addition to these direct causes, climate change exacerbates wildfire risk by raising global temperatures and altering rainfall distribution. The past decade includes the ten warmest years since 1985, and 2024 is on track to surpass 2023 as the hottest year on record. As forests burn, they release significant amounts of carbon dioxide and other pollutants into the atmosphere, worsening climate change and degrading air quality over vast distances. While regrowing forests can eventually recapture some of this carbon, it often takes decades to recover what is lost in a single year of severe fires (Hausfather [2024](#)). As South American countries face record-breaking fires this year and those to come, the focus must be on addressing both the direct and indirect drivers of forest fires.

Although wildfires play a natural role in ecosystem processes, they also pose significant environmental and socioeconomic risks worldwide, particularly in the context of climate change, and can have devastating impacts on local communities. When large wildfires occur, suppression efforts often dominate attention, leading to a historical focus on firefighting as the primary wildfire management strategy.

However, there is growing recognition that prevention is the most effective approach to managing wildfires. Wildfire prevention involves a variety of strategies implemented at various levels, including policy development, forest management practices, education, and training programs. These measures all aim to reduce the likelihood of fire ignition and control the intensity and spread of fire events (EFI’s Mediterranean Facility (EFIMED) [2024](#)). Traditionally, Fire Danger is predicted by state-wide Fire Danger Rating systems (FDRS), such as the Canadian FDRS or the Australian FDRS. In Europe, the European Forest Fire Information System (EFFIS) uses a similar approach to the Canadian FDRS, which is based on a meteorologically derived Fire Weather Index (FWI) which relies solely on meteorological conditions and disregards the status of other fire drivers related to vegetation and human factors (Prapas et al. [2021](#)).

Wildfire is intrinsically stochastic and the result of complex interactions between all

the fire drivers operating at different spatial and temporal scales, namely climate, vegetation, and human activity, thus it is challenging to model. To mitigate this difficulty, many studies have focused on identifying the links between fire weather and wildfire activity, following the assumption that weather has the most important control of fire (Kondylatos et al. 2022). Early warning systems rely on identifying high-risk fire spots, a high probability of fire outbreaks characterizes these zones and poses a significant threat to surrounding areas as well. Accurate mapping of these zones is essential for reducing the potential consequences of forest fires. To effectively predict, identify, and prioritize fire-prone zones, there is a need for a robust framework that integrates geospatial technologies with collaborative strategies involving local communities. This approach can significantly enhance wildfire mitigation efforts and safeguard forest ecosystems (Talukdar et al. 2024).

Wildfire risk is currently mapped through Fire Weather Index (FWI), statistical models, traditional machine learning, and deep learning techniques. Fire Danger Rating Systems (FDRS), widely used in countries like Canada, the USA, and Australia, integrate empirical and mathematical models based on principles of fluid dynamics, combustion, energy transfer, and factors such as fuels (canopy, surface, fine, coarse, etc.), weather, and terrain conditions (Rezaie et al. 2023). Traditional machine learning methods, including Random Forests, SVM, ensemble models, and Bayesian techniques, enhance statistical models by leveraging data-driven approaches and manual feature engineering to uncover patterns in historical fire data. These non-parametric methods are well-suited for handling non-linear relationships (Oliveira et al. 2021), but they pose challenges in interpreting variables compared to statistical approaches. In contrast, deep learning models, while requiring larger datasets, can automatically capture complex features through hierarchical architectures, minimizing the need for manual feature engineering and improving generalizability.

Initially, Convolutional Neural Networks (CNNs) were explored due to their sensitivity to spatial patterns and relationships (Zhang et al. 2019; Huot et al. 2020). More recent studies have also investigated Recurrent Neural Networks (RNNs), which are better suited for analyzing temporal sequences. Furthermore, hybrid models that combine CNN and RNN have been proposed to create systems capable of capturing spatial and temporal wildfire dynamics (Prapas et al. 2021).

This study builds on existing methods and advances the estimation of wildfire likelihood through several key contributions. First, it integrates Earth Observation Data Cubes API to track wildfires in near real time, providing access to global fire driver variables and burned area data. South America faces a significant gap in wildfire research and datasets, as most available openly accessible wildfire datasets are concentrated in the United States, while other wildfire-prone regions like the Mediterranean, Canada, Australia, and South America remain underrepresented. (Z. Xu et al. 2024)

Second, the study rigorously defines wildfire danger forecasting as a machine learning (ML) problem. Fire events linked to large burned areas are identified, along with their ignition dates, enabling the modeling of the joint probability of fire occurrence and becoming large.

To address this, wildfire prediction is framed as a forecasting task, using a temporal split to create training, validation, and test datasets. In addition, careful negative sampling is applied to effectively address data set imbalance. Third, the study proposes and evaluates a range of deep learning (DL) models capable of capturing spatial, temporal, and spatiotemporal patterns. An operational prototype is presented, offering daily national-scale wildfire danger predictions tailored to the unique environmental conditions of South America.

2 Literature Review

2.1 Wildfire influencing factors

Wildfire behavior is influenced by a complex interaction of weather, fuel, topography, and anthropogenic factors. Fire weather conditions, such as wind speed, directly affect fire spread, while variables such as relative humidity indirectly influence wildfire activity by altering fuel moisture levels (Dios et al. 2015). Fuel characteristics, including moisture content, fuel load, and fuel type, are critical in determining fire ignition, spread, and combustion. Structural forest attributes, such as canopy height, crown base height, tree cover, and crown bulk density, are particularly significant in assessing the likelihood of fires transitioning from surface fires to crown fires (Quan et al. 2021).

Effective fuel treatments, such as burning, pruning, and thinning of fuel, are crucial for the management of wildfires. These methods aim to increase crown base height and reduce canopy continuity, lowering the potential for crown fires and mitigating the risk of high-intensity fires (Collins et al. 2023).

Topography plays a pivotal role by influencing local climate and vegetation distribution, which in turn affects fire behavior. Fires spread more rapidly uphill than downhill due to the preheating of vegetation on slopes (Gharakhanlou et al. 2021). Topographic factors such as slope, elevation, and aspect are particularly important in mountainous or complex terrains (Pourtaghi et al. 2016). Slope steepness accelerates the fire’s rate of spread, while elevation, aspect, and relief significantly modulate fire weather and fuel conditions (Resco de Dios 2020). Topography also correlates with key variables like accessibility to infrastructure, lightning distribution, fuel types, and recreational activities (Jaafari et al. 2017).

Anthropogenic factors further shape fire regimes by altering ignition frequency and landscape composition, thereby affecting global productivity (Prasad et al. 2008). Infrastructure such as roads and residential areas play a dual role: they serve as artificial fuel breaks to prevent fire spread and also increase ignition frequency, making proximity to these features a critical factor in wildfire probability modeling (Jaafari et al. 2022). Natural landscape features, including rivers, lakes, and wildland-urban interfaces, act as natural fuel breaks, isolating fire-prone regions and aiding fire management efforts (Thomas et al. 2021).

Together, these factors form a highly interconnected system where weather, fuels, topography, and human activity collectively influence wildfire occurrence, spread, and intensity.

2.2 Risk of wildfires

Wildfire risk refers to the likelihood of a wildfire igniting or spreading, influenced by factors such as fuel availability, terrain characteristics, and weather conditions. In this context, wildfire risk is essentially synonymous with wildfire probability (Yang et al. 2004). This section explores explicit definition methods for wildfire risk, which involve directly utilizing historical observation data, determining risk based on the weighted influence of various factors, and integrating these factors to model wildfire distribution and derive risk from simulation outcomes.

Wildfire ignition and spread are inherently random processes, characterized by a strong nonlinear relationship with numerous influencing factors. Data-driven machine learning methods are widely applied in wildfire risk prediction due to their capacity to model complex nonlinear functions using large-scale datasets. However, these approaches often lack the ability to explicitly represent the contribution of individual factors to overall wildfire risk. Typically, machine learning-based wildfire risk models rely on binary historical wildfire event datasets as labels, with multiple features serving as input parameters. The resulting binary or probabilistic outputs are then interpreted as wildfire risk (Z. Xu et al. 2024). For instance, Yahia et al. 2023 utilized the normalized binary burn ratio as an indicator of wildfire risk probability. This probability, combined with proxy variables such as temperature vegetation dryness, vertical drought index, differential water index, and indices derived from the optical trapezoid model, was employed to train a Gaussian naive Bayes classifier.

Ager et al. 2014 investigated the spatiotemporal patterns of wildland fire occurrence and size in Sardinia, Italy, and Corsica, France, using historical fire data and the Fire Weather Index (FWI) to assess fire risk conditions. The study analyzed key factors influencing fire distribution, including climatic variables, vegetation types, and human activity, to better understand regional fire dynamics. By incorporating FWI data, the research provided insights into how weather conditions drive fire behavior and identified high-risk fire zones. The findings emphasized the importance of integrating fire weather indices into wildfire risk assessment and management strategies, offering valuable information for improving regional fire prevention and mitigation efforts.

However, the question of which techniques to adopt is also a matter of concern. Researchers applied models such as AHP (Nikhil et al. 2021), fuzzy logic (Gheshlaghi et al. 2020), logistic regression, artificial neural network (ANN) (Goldarag et al. 2016), support vector machine (SVM) (Shao et al. 2022), naïve Bayes (W. Chen et al. 2021), and random forest (Janiec et al. 2020) for mapping wildfire risk zones. Moreover, it is important to mention that the wildfire risk predicted by machine learning models represents a pseudo-probability rather than a true probability density (Pelletier et al. 2023). For example, in XGBoost, the pseudo-probability is derived from the inverse log transformation of the sum of terminal leaf weights. In Support Vector Machines (SVM), it is estimated by measuring the distance to the classification decision boundary and transforming these distances into probability estimates. In Random Forest, the pseudo-probability is calculated as the average of the voting outcomes from multiple decision trees. Meanwhile, in models like Multilayer Perceptron (MLP), Long Short-Term Memory

(LSTM), and Convolutional Neural Networks (CNN), the probability estimate typically originates from the final layer’s activation function (Z. Xu et al. 2024).

2.3 Remote sensing for Wildfire monitoring

Traditionally, fire monitoring has depended on ground-based observations, such as direct visual assessments from watchtowers or more structured monitoring conducted by fire management agencies, including aerial reconnaissance. Although these methods offered useful in-situ data at particular sites, they were inherently restricted in their coverage and could not deliver the consistent and extensive monitoring needed on a continental or global scale (Y. Chen et al. 2024).

Remote sensing technology has revolutionized the way we observe and analyze wildfires on a global scale. The origins of space-based fire detection and mapping date back to the 1970s with the launch of the first Landsat satellite. Since then, advancements in remote sensing have led to successive generations of more sophisticated instruments, each enhancing fire detection capabilities. These sensors have provided critical data on fire locations, burned areas, and fire emissions, generating valuable datasets for assessing fire impacts. Ongoing improvements in remote sensing technology, combined with progress in data analysis and computational power, present significant opportunities to deepen our understanding of evolving fire patterns and address fire management challenges in an increasingly warm and fire-prone world.

Wildfires have profound effects on Earth’s ecosystems, atmosphere, and climate. Remote sensing plays a crucial role in monitoring every stage of a wildfire’s life cycle to support effective management and mitigation efforts. In the pre-fire phase, remote sensing is employed to evaluate fuel loads, fuel conditions (e.g., live versus dead vegetation, moisture levels), and fire weather, aiding in predictions of fire likelihood and intensity. During active wildfires, remote sensing provides real-time data on fire locations and intensity. In the post-fire phase, it helps assess changes in burned landscapes, such as surface reflectance, roughness, and temperature, as well as vegetation characteristics like biomass and structure, to quantify burn extent and severity. Additionally, remote sensing captures data on atmospheric composition during and after wildfires, contributing to air quality and pollution assessments (Y. Chen et al. 2024).

Thermal images, which detect infrared radiation emitted by fires, are essential tools for real-time monitoring of active fires (AFs). Vegetation burning at temperatures between 500°C and 1,200°C releases significantly more infrared radiation than the surrounding environment, making fires identifiable through their infrared signatures. The mid-wave infrared (MWIR; 3–5 μm) range is particularly effective for detecting fires as it captures the peak energy emitted by vegetation fires. To improve detection accuracy, MWIR data are often combined with long-wave infrared (LWIR; 8–14 μm) data, which enhances contrast and fire identification. Additionally, short-wave infrared (SWIR; 1.4–2.3 μm) data are used, especially during daylight hours, to minimize interference from reflected sunlight or when MWIR sensors are unavailable (Natural Resources Canada n.d.). For instance, Schroeder et al. 2016 evaluated an active fire detection algorithm designed for Landsat 8 OLI, which expands on the Enhanced Thematic Mapper+

(ETM+) active fire algorithm. While the ETM+ algorithm primarily relies on the short-wave infrared (SWIR) band, the updated Landsat 8 algorithm incorporates visible and near-infrared (NIR) bands to enhance the accuracy of active fire detection.

Active fires can be detected using a variety of spectral sensors, but those best suited for monitoring active fires typically have high temporal resolution. While sensors onboard satellites like Landsat and Sentinel-2 are effective for detecting active fires, their relatively long revisit intervals (5–16 days) limit their suitability for real-time monitoring. In contrast, sensors such as MODIS and VIIRS are preferred for generating active fire products due to their short revisit times (less than 2 days), enabling frequent updates and timely fire monitoring (Szpakowski et al. 2019). The above-mentioned sensors also provide a means for estimating and mapping burned areas at local, regional, and global scales. The methods used for burned area estimates differ depending on the scale and purpose of the assessment. These sensors are typically used for change detection via spectral index generation and image differencing. Roy et al. 2006 developed a method for mapping burned areas on regional and global scales using data from the MODIS sensor. This approach, known as MCD45, employs a bi-directional reflectance model-based change detection algorithm applied to 500 m resolution data. The resulting product delivers burned area estimates that align with active fire detection data.

Measurements across the solar spectrum, including visible (VIS; 0.4–0.7 μm), near-infrared (NIR; 0.8–1.4 μm), and SWIR wavelengths, are vital for analyzing surface and atmospheric changes after a fire. These changes—such as vegetation charring, exposed soil, and ash deposits—affect reflectance patterns in these bands. Reflectance in the NIR band generally decreases post-fire, while SWIR and red (0.6–0.7 μm) bands often exhibit increased reflectance. Spectral indices comparing NIR with SWIR or red bands such as the Normalized Burn Ratio (NBR) and the Normalized Difference Vegetation Index (NDVI) are also valuable tools for assessing burn severity and identifying burned areas (BAs) (Y. Chen et al. 2024).

The NBR is usually computed before and after the fire to find regions of notable change by fire activity, and the difference is then calculated and denominated as dNBR (Escuin et al. 2008). However, Miller et al. 2007 highlighted a key limitation of the differenced Normalized Burn Ratio (dNBR) in assessing stand-replacing fires. They observed that dNBR, which measures absolute change across an image, struggles to accurately capture burn severity in areas with sparse vegetation or diverse vegetation types. This is because dNBR does not account for relative changes specific to each location. As a solution, the Relative Differenced NBR (RdNBR) was introduced to better classify high burn severity in landscapes with heterogeneous vegetation. Their findings showed that RdNBR outperformed dNBR in identifying high burn severity under these conditions.

2.4 Existing wildfire monitoring platforms

Wildfire monitoring and risk reduction platforms leverage a variety of advanced technologies, including remote sensing, artificial intelligence, and real-time data collection systems, to address the growing threat of wildfires. Among these, OroraTech’s Wildfire Solution offers global

coverage by deploying nanosatellites equipped with thermal-infrared sensors. This system integrates data from over 20 satellites, including Sentinel and Landsat missions, to detect active fires and map burnt areas with high precision. The platform's advanced algorithms analyze near-infrared (NIR) and short-wave infrared (SWIR) spectra to assess fire severity and vegetation damage. By providing real-time detection and post-event analysis, OroraTech supports early interventions and long-term fire management strategies (OroraTech 2023).

The FireAId initiative, implemented in Türkiye, uses artificial intelligence and machine learning to predict wildfire risks based on historical, meteorological, and geographical data. Achieving an 80% accuracy rate, this system can forecast wildfire occurrences 24 hours in advance, enabling authorities to prepare and respond proactively. Its interactive risk maps are vital for preemptive planning and resource allocation, demonstrating the potential of AI-driven tools to complement remote sensing technologies (Forum n.d.). Similarly, California's ALERTCalifornia program employs a network of over 1,000 wildfire monitoring cameras placed on the trees. These cameras, equipped with advanced sensors, provide real-time data on active wildfires and other natural hazards across the state, offering crucial information for emergency responders and policymakers (Scull 2023).

Other platforms, such as the EPRI Wildfire Tool Inventory, focus on providing comprehensive evaluations of existing tools and methodologies for wildfire risk management. While not exclusively reliant on remote sensing, these platforms employ diverse techniques, including vegetation management, infrastructure hardening, and risk assessment modeling. By tailoring solutions to specific regional needs or broader applications, these tools serve as valuable resources for stakeholders seeking to mitigate wildfire risks. Collectively, these platforms exemplify the integration of cutting-edge technologies in wildfire management, offering spatial insights, predictive capabilities, and actionable data to reduce wildfire impacts effectively (Alto 2024).

Globally, wildfire monitoring platforms and technologies are unevenly distributed, with a significant concentration in North America, Europe, and Asia. In North America, platforms like ALERTCalifornia and OroraTech's global operations demonstrate the integration of advanced remote sensing and real-time monitoring systems. Europe is home to the European Forest Fire Information System (EFFIS), a comprehensive platform that provides pan-European wildfire monitoring using satellite data and predictive tools. EFFIS delivers timely information on active fires, burned areas, and fire risk assessments, supporting coordinated firefighting efforts across European countries (Copernicus 2024). Asia has seen similar advances in wildfire technologies, particularly in regions vulnerable to seasonal fires (Department 2024). However, South America lags significantly behind, despite its vulnerability to wildfires, particularly in the Amazon basin and the savanna regions.

The lack of dedicated platforms or widespread implementation of remote sensing and predictive technologies in South America represents a critical gap in global wildfire risk management. While EFFIS serves as a model for multinational cooperation and the effective use of satellite data, South America lacks such coordinated systems. Given the ecological impor-

tance of its forests and their role in global climate regulation, investment in wildfire monitoring technologies tailored to the region’s specific challenges is urgently needed. Expanding remote sensing-based platforms and incorporating predictive tools like those used in EFFIS or FireAid could significantly enhance South America’s capacity to monitor and mitigate wildfire risks.

The current wildfire situation in Latin America is primarily analyzed using information from the European Commission’s Joint Research Centre through the Global Wildfire Information System (GWIS). While more detailed and updated data may be available from regional institutions such as Brazil’s National Institute for Space Research (INPE) or Mexico’s National Institute of Statistics and Geography (INEGI) (Pletsch et al. 2022, United Nations Office for Disaster Risk Reduction (UNDRR) 2022), GWIS was selected to provide a comprehensive and consistent overview of wildfire activity across the entire region. This choice ensures the use of data derived from a robust and validated methodology, enabling a reliable depiction of wildfire trends in Latin America. Combined with the recognized data gaps in the region, this highlights the pressing need for more localized and specialized platforms to address wildfire monitoring and management challenges effectively.

2.5 Deep Learning in Wildfire related studies

Remote sensing technologies, including satellite and aerial imagery, are essential for disaster management by providing data to identify and monitor wildfires, floods, earthquakes, and landslides. Although processing this data is complex and requires experts, a recent study by Prapas et al. 2021 on wildfire forecasting used machine learning with ten years of Earth observation data to predict daily fire danger. Researchers created an open-access data cube with weather variables, satellite data, topography, and human-related activity data. They tested various deep learning models against a Random Forest baseline, finding that the ConvLSTM model, which uses spatiotemporal data, performed best with an AUROC of 0.926, offering more accurate national-scale daily fire danger maps than existing methods.

Researchers developed a novel deep learning model called static location-aware ConvLSTM (SLA-ConvLSTM) to predict global wildfire danger. This model incorporates both static positional elements and dynamic variables, along with global teleconnection information, enhancing predictive accuracy. By integrating spatiotemporal features and static positional data using U-Net and SKNet for feature extraction, the SLA-ConvLSTM model utilizes the SeasFire cube, a dataset for sub-seasonal to seasonal wildfire prediction. The study shows that the SLA-ConvLSTM model significantly outperforms commonly used Recurrent Neural Networks (RNN) architectures, and emphasizes the importance of including both static and dynamic variables in wildfire prediction, as well as the benefits of integrating global and teleconnection data for improved forecast accuracy (Ji et al. 2024).

The authors in Huot et al. 2020 investigate the use of deep learning models to predict wildfire danger by combining nearly a decade of remote-sensing data and historical fire records. They framed the prediction problem into three machine-learning tasks: image segmentation with daily fire masks, aggregated fire masks, and sequence segmentation. Four deep

learning models, including convolutional autoencoders and residual U-Nets, with and without convolutional LSTM layers, were evaluated. The models achieved an AUC of 83%, showcasing the potential of deep learning to identify high-risk wildfire areas using extensive data on vegetation, weather, and topography. The study demonstrates the promise of deep learning in improving wildfire management and disaster preparedness through accurate risk predictions based on comprehensive remote-sensing datasets.

Recent advancements in cloud computing and machine learning have paved the way for more efficient and automated approaches to remote sensing data analysis. OpenEO, an open-source initiative for Earth observation data processing, provides a standardized API for accessing and processing data cubes, facilitating interoperability and scalability across different platforms and datasets (Pondi et al. 2024). Deep learning algorithms, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have shown promise in extracting useful information from remote sensing imagery, enabling tasks such as image classification, object detection, and semantic segmentation.

Recent advancements have incorporated cloud-based platforms such as Google Earth Engine (GEE) and Open Data Cube (ODC), which allow for the efficient handling of vast amounts of satellite data (Giuliani et al. 2020). The OpenEO API introduces an innovative method by offering standardized access to multiple cloud-based backends, enabling adaptable and scalable data analysis workflows using popular programming languages like Python and R. This methodology has been successfully employed in various studies to merge and analyze extensive datasets for land use and cover classification (Yue et al. 2020). Previous research has demonstrated the effectiveness of various approaches in wildfire danger prediction. Kondylatos et al. 2022 utilized deep learning techniques such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to analyze spatiotemporal data for wildfire danger prediction. While Prapas et al. 2022 explored deep learning methods like convolutional autoencoders and residual U-Nets, highlighting the importance of combining remote sensing data with historical fire records. Bilidas et al. 2023 demonstrated the use of data cubes, machine learning, and OBDA systems for fire risk management, emphasizing the integration of multidimensional data for accurate predictions. These studies underscore the potential of deep learning and data cube technologies in enhancing wildfire prediction models.

2.5.1 Wildfire Risk Modelling in Germany

Germany employs a combination of advanced machine learning techniques, data-driven modeling, and collaborative evaluations to assess and predict wildfire risk and spread. Given the increasing threat of wildfires, particularly in regions like Brandenburg, researchers have applied machine learning algorithms such as Random Forest (RF) to model forest fire susceptibility (FFS). These models integrate a range of environmental and anthropogenic factors, including topography, climate conditions, human activities, soil composition, and vegetation characteristics. Recent studies have demonstrated that human influence and vegetation types play significant roles in determining wildfire susceptibility, with predictive accuracies ranging from

69% to 71%. Climate change projections suggest that wildfire risk will escalate in the coming years, necessitating more proactive and adaptive fire management strategies. (Horn et al. 2025)

Beyond machine learning, Germany has also embraced deep learning models to enhance wildfire prediction accuracy. A notable example is the implementation of Spatiotemporal Graph Neural Networks (STGNNs), which are designed to predict next-day burned areas by analyzing extensive wildfire time series datasets. These models effectively capture the complex spatial and temporal dynamics of wildfires, improving forecasting reliability and response planning. The integration of such data-driven approaches provides crucial insights for emergency services and policymakers, allowing them to allocate resources more efficiently and implement targeted mitigation measures. Moreover, the adoption of satellite-based monitoring and climate data analytics further strengthens Germany’s capability to anticipate and manage wildfire outbreaks in vulnerable regions. (Rösch et al. 2024)

In addition to technological advancements, Germany has actively engaged in international peer reviews to refine its wildfire risk management strategies. In November 2024, Brandenburg, the country’s most fire-prone state, underwent an extensive peer review through the Union Civil Protection Mechanism (UCPM). Experts from Spain, Italy, and Portugal evaluated the state’s wildfire governance, prevention, preparedness, and recovery strategies, offering valuable recommendations based on their own experiences. This collaborative effort underscores the importance of knowledge exchange in strengthening wildfire resilience at both national and European levels. As climate change continues to intensify fire risks, Germany’s multi-faceted approach—leveraging artificial intelligence, real-time data analysis, and international cooperation—positions the country to better predict, prevent, and respond to wildfires in the future. (The Knowledge Network 2024)

3 Dataset and Study Area

3.1 Study Area

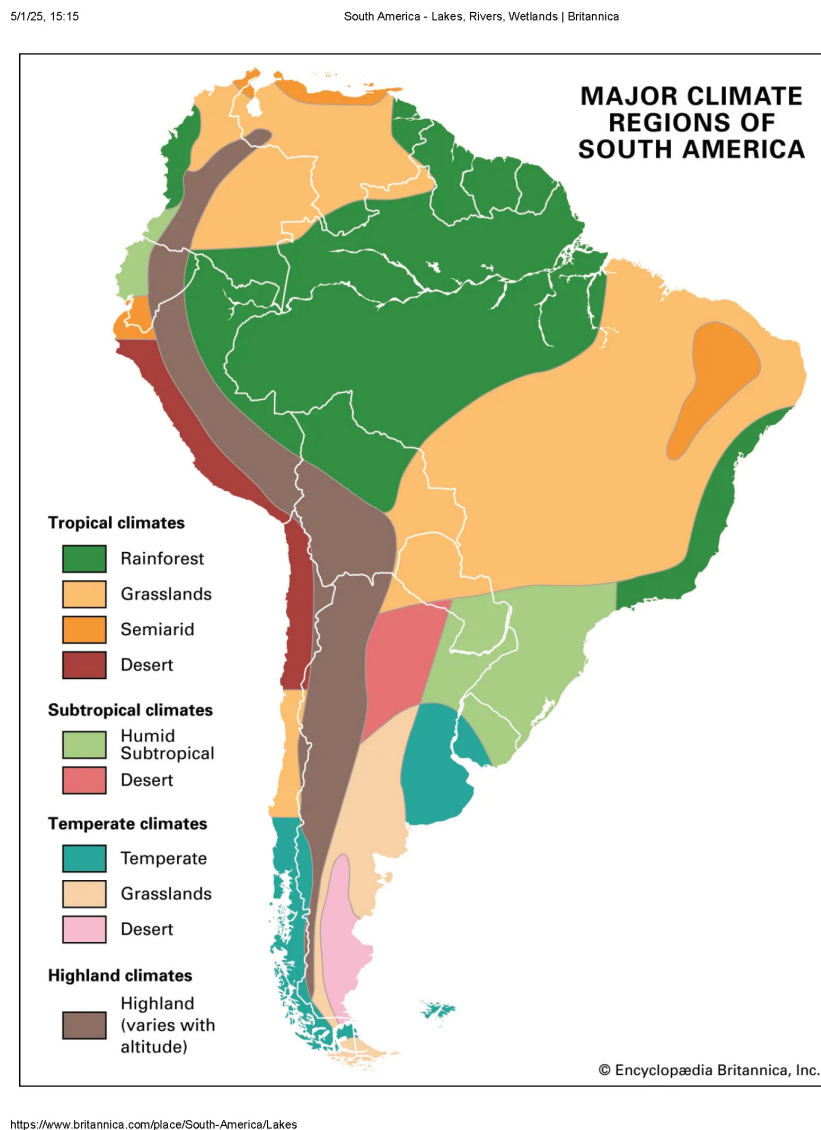
South America is a continent of extraordinary geographical diversity, shaped by vast mountain ranges, extensive river systems, and unique ecosystems. The Andes, the world’s longest mountain range, stretch along the continent’s western edge, while the Amazon River, the largest river by volume, winds through the Amazon rainforest, the planet’s largest tropical rainforest. These regions play a crucial role in global water cycles, oxygen production, and carbon storage. In stark contrast, the Atacama Desert in Chile is one of the driest places on Earth, showcasing the continent’s extreme environmental variation.

The continent’s interior consists of expansive lowlands, including the Amazon Basin, the Orinoco Basin, and the Paraná-Paraguay River system. These areas are home to tropical rainforests, savannas, and wetlands, supporting a vast array of flora and fauna. The Brazilian and Guiana Highlands, located in the eastern part of the continent, are ancient geological formations that add to the region’s ecological richness. Meanwhile, the fertile Pampas in

Argentina provide some of the world's most productive agricultural land, known for wheat cultivation and cattle ranching.

South America's physical geography also contributes to its diverse climate patterns (Figure 1), ranging from the icy glaciers of Patagonia in the south to the tropical rainforests of the Amazon. The continent's varied landscapes foster rich biodiversity, with unique species thriving in its mountains, plains, and forests. These physical features not only define the continent's natural beauty but also influence the livelihoods, cultures, and economies of its people, making South America a region of immense ecological and geographical significance (National Geographic 2023).

Figure 1: Major climate regions of South America



Major climate regions of South America, highlighting rainforests, deserts, grasslands, and highland climates across the continent

3.2 Dataset

3.2.1 Target variable: MODIS Burned Area

The Terra and Aqua MCD64A1 Version 6 Burned Area product is a monthly global dataset that provides 500m resolution information on burned areas and data quality at the pixel level. This mapping method combines 500m MODIS Surface Reflectance imagery with 1km MODIS active fire data. The algorithm utilizes a burn-sensitive vegetation index (VI) derived from MODIS shortwave infrared surface reflectance bands 5 and 7, adjusted for atmospheric conditions, along with a measure of temporal texture. Dynamic thresholds are applied to composite data to detect burned areas. The algorithm assigns a burn date to each 500m grid cell, recorded as the ordinal day of the calendar year when the burn occurred. Unburned pixels are given specific values, while additional codes are used for missing data and water-covered grid cells (Giglio et al. 2015).

3.2.2 Predictive variables

To estimate wildfire likelihood, data from various remote sensing sources available on Google Earth Engine (GEE) were aggregated. These sources were chosen for their high spatial and temporal resolution, broad geographical and historical coverage, and minimal data gaps. All selected datasets are publicly accessible on Earth Engine, making them suitable for research and disaster management purposes. Additionally, these datasets are regularly updated on the platform, ensuring access to the most recent information. The logistical advantage of using Earth Engine lies in its centralized storage, which facilitates consistent and synchronized access to datasets across time and locations. This compiled dataset is ideal for training models to predict wildfire likelihood based on recent remote sensing observations, emphasizing the importance of consistently updated and reliable data. The sources and datasets in Table 1 were selected for inclusion:

Land Surface Temperature and Emissivity (LST): The thermal infrared signal detected by satellite sensors is influenced by several factors, including surface temperature, surface emissivity and reflectivity, atmospheric emission, absorption and scattering of thermal radiation from the surface, and solar radiation during the daytime. Land Surface Temperature (LST) is derived from MODIS thermal channel data, covering the Earth’s entire land area, including evergreen and deciduous forests and shrubs, croplands, grasslands, inland water bodies, snow and ice, barren lands with exposed soil, sand, and rocks, as well as urban environments. LST serves as a key indicator of Earth’s surface energy balance and the greenhouse effect, playing a crucial role in understanding land-surface processes at regional and global scales. Reflects the interactions between the surface and atmosphere and the exchange of energy fluxes. However, the MODIS LST product, based on thermal infrared data, is only available under clear sky conditions (Zhengming 1999).

To enhance the understanding of Earth’s systems on a global scale, the Earth Observing System provides surface kinetic temperature measurements with an accuracy of 0.3°K for oceans

Table 1: Datacube input sources

Dataset	Description	Spatial Resolution	Temporal Resolution
MOD11	Land Surface Temperature and Emissivity (LST)	1km	Daily
MOD13	Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI)	1km	16-day
MOD15	Photosynthetically Active Radiation (FPAR), Leaf Area Index (LAI)	500m	8-day
MOD16	Evapotranspiration	500m	8-day
ERA5-Land	Wind Speed & direction, Precipitation, Temperature, Surface Pressure, Dew Point Temperature	9km	5 hours/Daily
MCD12	MODIS Land Cover Type	500m	Yearly
SRTM	Digital Elevation Model	90m	-
Aspect	Aspect	90m	-
Slope	Slope	90m	-
World pop	Human Modification	1km	1990-2015

Overview of datasets used for wildfire risk prediction, including MODIS, ERA5-Land, SRTM, and human modification data, with their spatial and temporal resolutions

and 1°K for land. MODIS is particularly valuable for LST retrieval due to its global coverage, high radiometric resolution, wide dynamic range for various land cover types, and precise calibration across multiple thermal infrared bands designed for LST and atmospheric property analysis. The computation of spectral radiance as a function of wavelength is given by the equation 1:

$$L(\lambda, T) = \varepsilon(\lambda)B(\lambda, T_s) \quad (1)$$

Where emitted spectral radiance L at wavelength λ from a surface at thermodynamic temperature T_s is given by multiplying the Planck function by spectral emissivity $\varepsilon(\lambda)$.

Vegetation indices (VIs): MODIS vegetation index (VI) products enable consistent spatial and temporal monitoring of global vegetation conditions, supporting analyses of plant activity, phenology, and environmental changes. These indices, generated at 16-day and monthly intervals, help track seasonal and interannual variations in vegetation.

Vegetation indices (VIs) are radiometric measurements based on the spectral properties of vegetation, particularly in the red and near-infrared (NIR) bands. They not only detect vegetation presence at a pixel level but also assess its quantity and condition. The core idea is to isolate the vegetation signal, with stronger signals indicating denser vegetation. VIs are simple to compute and require no additional data beyond the spectral measurements. As plants grow, red light reflectance decreases due to chlorophyll absorption, while NIR reflectance increases due to scattering in healthy leaves. However, direct reflectance measurements are unreliable

for global vegetation monitoring because they are influenced by factors like solar radiation, atmospheric conditions, and canopy structure. To address this, vegetation indices combine multiple spectral bands into formulas that enhance the vegetation signal while minimizing interference from non-vegetation elements (Huete et al. 1999).

The simple ratio (SR) (equation 2) was the first index to be used, formed by dividing the NIR response by the corresponding ‘red’ band output.

$$SR = \frac{X_{nir}}{X_{red}} \quad (2)$$

Here, X can represent digital counts, at the units of the measured physical property. However, in areas with dense vegetation, the amount of red light reflected becomes extremely low, causing the ratio to increase indefinitely. To address this, the normalized difference vegetation index (NDVI) was introduced, which scales this ratio between -1 and +1 by calculating the difference between the near-infrared (NIR) and red bands, then dividing by their sum as shown in equation 3.

$$NDVI = \frac{X_{nir} - X_{red}}{X_{nir} + X_{red}} \quad (3)$$

However, with MODIS-NDVI, improved atmospheric correction may unintentionally amplify canopy background issues, creating a feedback effect where reducing one type of noise increases another. To address this, Enhanced Vegetation Index (EVI) (equation 4) were created, which is similar to NDVI and can be used to quantify vegetation greenness. However, EVI corrects for some atmospheric conditions and canopy background noise and is more sensitive in areas with dense vegetation (Huete et al. 1999).

$$EVI = 2 * \frac{\rho_{nir} - \rho_{red}}{L + \rho_{nir} + C_1\rho_{red} + C_2\rho_{blue}} \quad (4)$$

where ρ is ‘apparent’ (top-of-the-atmosphere) or ‘surface’ directional reflectances, L is a canopy background adjustment term, and $C1$ and $C2$ weigh the use of the blue channel in aerosol correction of the red channel.

Leaf Area Index (LAI) And Fraction Of Photosynthetically Active Radiation Absorbed By Vegetation (FPAR): Leaf Area Index (LAI) is defined as the one-sided green leaf area per unit of ground area in broadleaf canopies and as half the total needle surface area per unit ground area in coniferous canopies. The Fraction of Photosynthetically Active Radiation (FPAR) refers to the proportion of solar radiation (400-700 nm) absorbed by green vegetation. Both LAI and FPAR are essential for modeling surface photosynthesis, evapotranspiration, and net primary production, which contribute to understanding terrestrial energy, carbon, and water cycles, as well as vegetation biogeochemistry. Algorithm improvements have enhanced retrieval accuracy and alignment with field measurements across all biomes, with a

particular emphasis on woody vegetation (Myneni et al. 2010).

The MOD15 LAI and FPAR products are available at a 1 km resolution and are provided on a daily and 8-day basis. The algorithm utilizes spectral reflectance data from up to seven MODIS bands, collected by sensors aboard the TERRA and AQUA satellites. These Level 4 MODIS products are derived from MODIS Reflectance (MR) data along with ancillary surface characteristics such as land cover type and background conditions. They are generated globally at an 8-day interval, corresponding to the MODIS Reflectance compositing period. Spatial resolution is determined by the MODIS reflectance dataset, ranging from 250m to 1km, with the 8-day product being composited based on maximum FPAR values (Knyazikhin et al. 1999).

Evapotranspiration (ET): Describes the combined water loss from evaporation and plant transpiration. In vegetated areas, it represents the amount of water consumed by plants over time. ET calculations are typically based on the conservation of energy and mass, requiring complex estimations of stomatal conductance for transpiration and evaporation from the ground surface.

The MOD16 ET algorithm runs daily, summing ET from both daytime and nighttime periods. It accounts for water vapor fluxes from soil evaporation, wet canopy evaporation, and plant transpiration at dry canopy surfaces. Remote sensing, particularly from polar-orbiting satellites, is a key tool for obtaining regional ET data, as it provides continuous spatial and temporal coverage of vegetation-related surface properties, including albedo, biome type, and Leaf Area Index (LAI). The dataset offers global 8-day (MOD16A2) and annual (MOD16A3) ET products at 0.5 km and 1 km resolutions, covering 109.03 million km² of vegetated land. The algorithm is based on the Penman-Monteith equation, which integrates daily meteorological reanalysis data with 8-day MODIS vegetation property dynamics to estimate ET. Key inputs include MODIS FPAR (Fraction of Photosynthetically Active Radiation) to determine vegetation cover fraction, MODIS albedo, daily solar radiation, air temperature, vapor pressure deficit (VPD), relative humidity, and MODIS LAI (Leaf Area Index). These variables help compute surface net radiation, soil heat flux, aerodynamic resistance, and stomatal conductance, ensuring accurate ET estimations at a global scale (Mu et al. 2013).

ERA ("ECMWF Re-Analysis"): ERA5-Land is a reanalysis dataset that provides a high-resolution view of land variables over several decades, offering an improved resolution compared to ERA5. It is generated by re-running the land component of the ECMWF ERA5 climate reanalysis, which integrates model data with global observations to create a complete and consistent dataset based on physical principles. This approach allows for the reconstruction of past climate conditions with high accuracy.

ERA5-Land relies on atmospheric variables from ERA5, such as air temperature and humidity, to regulate simulated land processes. This process, known as atmospheric forcing, ensures that the model-based estimates remain realistic. Although observations are not directly

included in ERA5-Land, they indirectly influence the dataset through the atmospheric inputs. Additionally, lapse rate correction is applied to adjust air temperature, humidity, and pressure for differences in altitude between the ERA5 forcing grid and the higher-resolution ERA5-Land grid.

This dataset provides post-processed ERA5-Land hourly data aggregated to daily time steps, excluding accumulated variables like total precipitation and runoff. Users can customize their data retrieval by selecting:

- Daily aggregation statistics (mean, max, or min)
- Sub-daily frequency sampling (1-hour, 3-hour, or 6-hour intervals)
- Local time zone adjustments (default is UTC+00:00)

In this research the daily aggregation was used for the variables described in table 2:

MODIS Land Cover Type/Dynamics: Land cover, along with both human and natural modifications, plays a crucial role in shaping global climate patterns and the Earth’s biogeochemical cycles. While the oceans primarily drive Earth’s physical climatology, the land surface significantly impacts biogeochemical processes, which in turn affect climate through greenhouse gases and reactive species. Additionally, variations in topography, albedo, vegetation cover, and other surface properties influence weather and climate by altering atmospheric circulation through surface-atmosphere energy and matter exchanges, as well as Earth’s rotational momentum (Strahler et al. 1999).

The MODIS Terra+Aqua Combined Land Cover product classifies land cover using five different classification schemes based on a supervised decision-tree approach. Its primary classification system, defined by The International Geosphere–Biosphere Programme (IGBP), consists of 17 land cover types, including 11 natural vegetation categories, three human-modified classes, and three non-vegetated categories.

The Land Cover Dynamics product tracks vegetation growth cycles by providing data on the timing of growth, maturity, and senescence. It offers biannual estimates, covering two 12-month periods (July–June and January–December), which account for seasonal differences across hemispheres and allow the detection of multiple growth cycles when applicable (Friedl et al. 2019).

Terrain-related variables: The Shuttle Radar Topography Mission (SRTM), conducted from February 11-22, 2000, aboard the space shuttle Endeavour, was a collaborative effort between NASA and the National Geospatial-Intelligence Agency (NGA) to create the first near-global digital elevation model (DEM). Using radar technology, the mission successfully captured elevation data covering approximately 80% of Earth’s land surface, between 60°N and 56°S latitude, with a spatial resolution of about 30 meters.

Table 2: ERA5-Land variables description

Name	Units	Description
Eastward Component of 10m Wind	m s-1	This measures the horizontal wind speed moving eastward at 10 meters above the surface. Local terrain and structures can influence observations, so care is needed when comparing data. It can be combined with the northward wind component to determine overall wind speed and direction.
Northward Component of 10m Wind	m s-1	This represents the horizontal wind speed moving northward at 10 meters above the surface. Since wind conditions vary in space and time, comparisons with observations should account for local influences. It is used alongside the eastward wind component to calculate full wind speed and direction.
2m Dew Point Temperature	K	This is the temperature at which air at 2 meters above the surface must be cooled to reach saturation, indicating humidity levels. It is derived through interpolation and can be used to calculate relative humidity.
2m Air Temperature	K	This represents the air temperature at 2 meters above the ground or water surfaces. It is calculated through interpolation and provided in kelvin, with conversion to Celsius possible by subtracting 273.15.
Surface Pressure	Pa	Surface pressure measures the atmospheric force exerted on the surface and varies with altitude. Mean sea level pressure is often used in mountainous areas. It is measured in Pascals (Pa), commonly reported in hPa or mb.
Precipitation	m	This includes total rainfall and snowfall reaching the surface, combining large-scale and convective precipitation. It excludes fog and dew and is measured as accumulated depth. Observational data may differ from model-based averages.

Description of key meteorological variables from ERA5-Land, including wind components, temperature, pressure, and precipitation. These variables are essential for analyzing atmospheric conditions and their impact on fire risk modeling

The radar technology used for SRTM was adapted from previous missions in 1994, where the C-band Spaceborne Imaging Radar and X-band Synthetic Aperture Radar (X-SAR) were tested. For SRTM, these radars were modified to use interferometric radar, a technique that compares two radar signals taken from slightly different angles. This was achieved through single-pass interferometry, where one radar antenna was positioned on the shuttle and another on a 60-meter-long mast. The differences between the two signals enabled the calculation of precise surface elevations. During the 11-day mission, Endeavour completed 176 orbits, collecting extensive topographic data (Earth Resources Observation and Science (EROS) Center 2018).

Derived Products from the SRTM DEM:

- **Aspect:** Indicates the compass direction a slope faces, derived from the DEM by analyzing elevation variations. It helps in understanding surface exposure to sunlight, wind, and erosion.
- **Slope:** Represents the steepness or incline of the terrain, calculated as the rate of elevation change between adjacent pixels in the DEM. Slope values are typically expressed in degrees or as percentages.
- **Spectral Analysis:** While the DEM itself does not contain spectral data, it can be combined with satellite imagery to derive spectral indices related to terrain features, vegetation, or land cover classification

Human Modification: Human activities have significantly altered the Earth’s landscapes, contributing to climate change, habitat loss, biodiversity decline, and disruption of ecosystem services. Addressing these issues is crucial for international conservation and sustainability initiatives like the Convention on Biological Diversity and the UN Sustainable Development Goals. A key challenge in tackling land use change is the lack of contemporary, high-resolution data, as most studies focus on historical patterns or use low-resolution datasets.

To address this gap, researchers developed a global dataset mapping human modification of terrestrial ecosystems from 1990 to 2015, with an update for 2017. This dataset identifies anthropogenic stressors that directly or indirectly alter natural systems by combining remote sensing data with traditional cartographic sources. Since satellite imagery alone may miss small or obscured human-made structures, integrating additional data improves accuracy (Theobald et al. 2020)

The methodology follows an established classification system that accounts for land use extent, intensity, physical impact, and uncertainty. The dataset produces a continuous human modification index (H) ranging from 0 to 1, allowing for detailed landscape assessments. The research includes finer-resolution data than previous efforts, incorporating additional stressors like reservoir impacts, air pollution, and human intrusion. This higher precision enhances applications for species conservation, land management, and policy decisions.

The index (H) helps analyze human impacts beyond broad land use categories like "urban" or "agricultural" by capturing spatial variation and modification intensity. The dataset supports conservation planning by identifying low-modification areas and ranking biomes and ecoregions based on human impact. The study also compares its findings to other global datasets to provide a comprehensive view of land transformation trends.

4 Methodology

4.1 Experimental Set-up

The experiment is structured into four main stages to achieve wildfire risk prediction using remote sensing data. The primary objective is to continuously monitor wildfire risk in near real-time across South America. To accomplish this, an Earth Observation data API is integrated into the workflow, enabling efficient data extraction. After configuring and establishing the API connection, the methodology proceeds as follows:

Active fire sampling: The first step involves retrieving active fire locations from 2021 to 2024. This includes extracting all relevant static, dynamic, and target variables associated with these locations. Active fire points serve as the foundation for identifying patterns and training the predictive models.

Non-Fire sampling: To ensure a balanced dataset, non-fire points are selected based on the land cover distribution of the active fire points. This stratified sampling ensures that negative samples (non-fire locations) share similar environmental characteristics with the positive samples (fire locations), improving model robustness.

Variable extraction: In this stage, known as variable extraction, input variables are obtained from the API based on the burn date of each previously collected data point. For each point, a tailored request is formulated considering location, time, and variable type. The retrieved data is then structured into data cubes to facilitate further processing and model training.

Experimental Design and Data Structuring: From the collected dataset, four experimental setups are designed to evaluate different modeling approaches. Each setup varies in the way spatial and temporal information is incorporated into the model. However, the patch size was selected based on the literature review. In contrast, the time range was selected based on the temporal resolution of the datasets, with the lowest one being 16 days:

- **Pixel Dataset:** For a given pixel ($1 \text{ km} \times 1 \text{ km}$ area) on a specific day, input attributes are collected along with a 17-day average for dynamic variables.
- **Temporal Dataset:** A 17-day time series of the input attributes is extracted to analyze trends over time.

- **Spatial Dataset:** Data is gathered from a $25 \text{ km} \times 25 \text{ km}$ spatial patch centered on the pixel to assess the influence of surrounding environmental factors.
- **Spatio-Temporal Dataset:** A combination of spatial and temporal features is captured by collecting data in $25 \text{ km} \times 25 \text{ km} \times 17\text{-day}$ blocks, centered on the pixel.

In all four dataset types, static variables are repeated across the time dimension, and the target variable remains consistent: the next-day burned status of the pixel (fire or no fire). This process is illustrated in Figure 2, where each dataset type is fed into a corresponding deep-learning model.

Data Balancing and Sample Selection: All positive examples, representing burned pixels, are included in the datasets. For negative examples (unburned pixels), twice as many are sampled from days without recorded fire activity. To maintain a fair comparison, negative samples are stratified to match the land cover distribution of positive examples. This stratification increases the challenge for models in differentiating between burned and unburned pixels, reducing the likelihood of the model relying on simplistic correlations or trivial patterns.

This structured experimental setup ensures a comprehensive approach to wildfire risk prediction by leveraging spatial, temporal, and spatio-temporal modeling techniques in conjunction with deep learning methodologies.

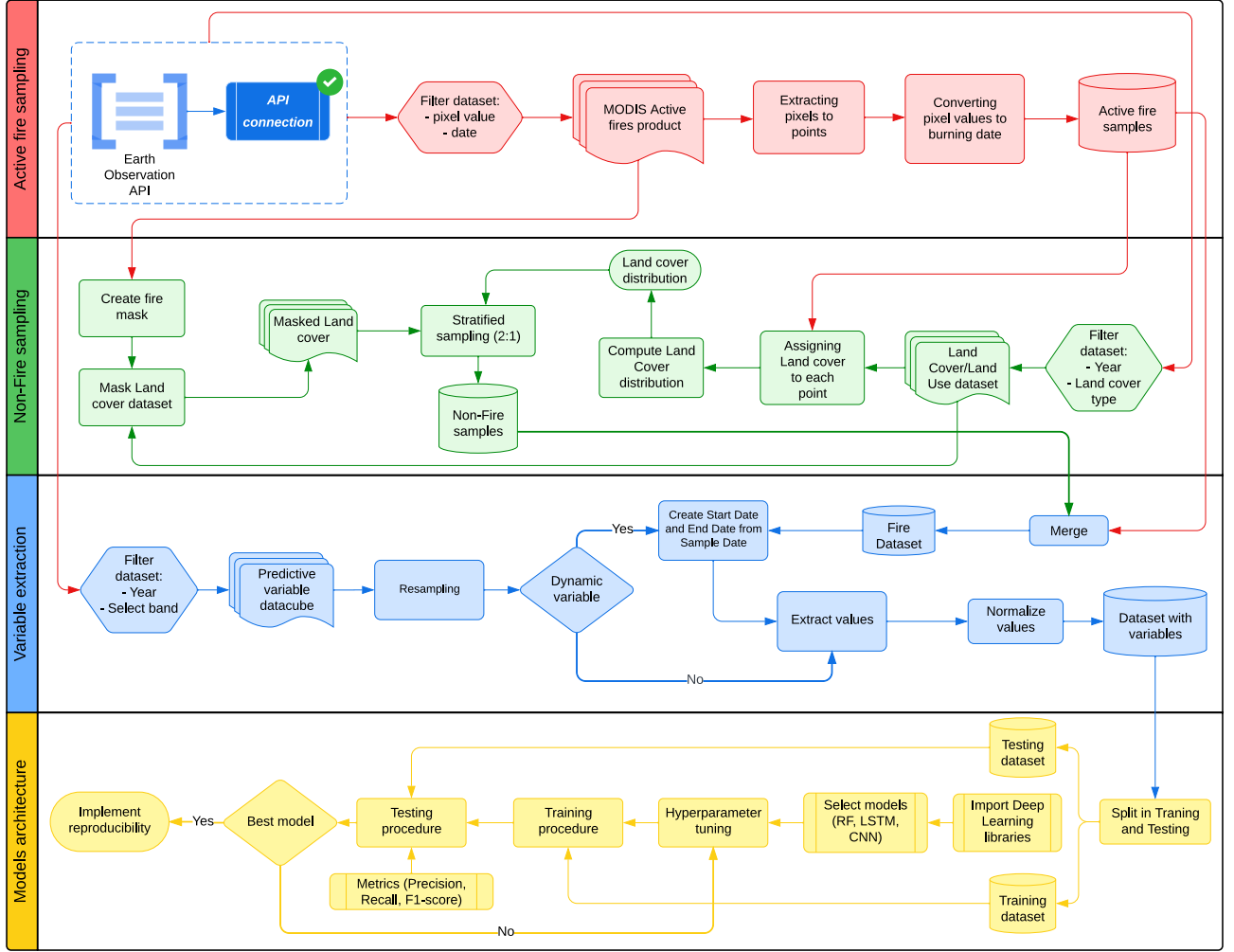
4.2 Google Earth Engine API

The Google Earth Engine (GEE) API is a cloud-based platform designed for large-scale geospatial data analysis and visualization. It provides access to a vast catalog of satellite imagery and geospatial datasets, allowing users to perform advanced analyses without requiring local data storage or extensive computational power. The API supports both JavaScript (via the Code Editor) and Python (via client libraries), enabling tasks such as data filtering, image processing, and time-series analysis. It includes tools for performing spatial and temporal computations, generating custom visualizations, and exporting results. The GEE API is widely used for applications in environmental monitoring, land cover classification, climate modeling, and disaster response, offering scalability and efficiency for global-scale geospatial projects (Gorelick et al. 2017).

4.3 Active Fire Sampling

The first step in constructing the wildfire risk prediction model involves sampling active fire locations from 2021 to 2024. This process ensures that relevant static, dynamic, and target variables are correctly extracted and structured for model training. The methodology follows a systematic approach to obtain accurate fire occurrence data and integrate it with remote sensing variables.

Figure 2: Methodology Flowchart



Workflow for wildfire risk modeling, detailing data collection, preprocessing, variable extraction, and model training. The pipeline integrates active fire sampling, non-fire sampling, predictive variable extraction, and model selection, including deep learning and traditional machine learning approaches

Data Source and Preprocessing: The active fire data is derived from MODIS (Moderate Resolution Imaging Spectroradiometer) land products, which provide fire occurrence information at multiple spatial resolutions (250 m, 500 m, 1 km, and 0.05°) and projections (Sinusoidal, Lambert Azimuthal Equal-Area, and Geographic). For consistency in spatial resolution across datasets, all fire data is resampled to a 1 km resolution and reprojected to the EPSG:4326 (WGS 84) coordinate system.

MODIS Fire Data Processing: The MODIS land product uses a tile-based grid system for higher-resolution data to maintain manageable file sizes. Most land products are produced and distributed as adjacent, non-overlapping tiles of approximately $10^\circ \times 10^\circ$. Since MODIS fire data is originally available in the Sinusoidal projection, reprojection to EPSG:4326 ensures compatibility with other geospatial datasets used in the study.

To extract active fire locations efficiently, Google Earth Engine (GEE) was leveraged for cloud-based processing. The key steps involved were:

Reprojection and Resampling:

- MODIS active fire data was resampled to 1 km resolution.
- Data was reprojected from its original Sinusoidal grid to EPSG:4326.

Fire Pixel Conversion to point geometry:

- Each active fire pixel was converted into a vector point.
- The burn date associated with each fire point was extracted to serve as the temporal reference for variable extraction.

By ensuring a standardized and systematic approach in fire data processing, this step establishes a robust foundation for subsequent wildfire risk modeling, allowing for effective integration of remote sensing variables in a near real-time predictive framework.

4.4 Non-Fire Sampling

To construct a balanced dataset for wildfire risk prediction, non-fire locations were carefully selected to ensure comparability with fire-prone areas. The MODIS land cover data was utilized to mask out non-burnable areas, such as water bodies and snow-covered regions. After filtering these areas, the land cover distribution of active fire locations was computed to guide the selection of non-fire samples.

To ensure a representative sampling, a second masking step was applied to the land cover data, removing pixels where fires had been previously recorded. This prevented bias from sampling areas recently affected by fire. Using cloud computing and API-based processing, a stratified random sampling approach was implemented, selecting twice as many random points as fire points from the remaining land cover image. Each non-fire point was assigned a random date, ensuring that it did not overlap with the fire occurrence period. This method maintained the same land cover distribution as the fire points, preventing selection in areas that are unlikely to experience wildfires. By enforcing a 1:2 fire-to-non-fire sample ratio, the dataset remains well-balanced, enhancing model performance and reducing bias in wildfire prediction.

4.5 Variable Extraction

A crucial step in wildfire risk prediction is the extraction of relevant environmental variables to construct meaningful input datasets. Earth Observation (EO) Data Cubes are used to efficiently store and analyze spatiotemporal data, enabling seamless integration of multiple variables.

4.5.1 Earth Observation Data Cubes:

EO Data Cubes are structured, multi-dimensional arrays that store time-series of remote sensing images or long lists of biophysical variable measurements. These data infrastructures facilitate

complex analyses by organizing large datasets into a grid system for efficient spatial and temporal analysis. Data Cubes comply with operational standards established by the International Organization for Standardization (ISO) and the Open Geospatial Consortium (OGC), ensuring consistency and interoperability across geospatial datasets.

An EO Data Cube typically consists of multiple dimensions, including:

- Spatial (x, y, z): Geographic coordinates and elevation.
- Temporal (t): Time-series of observations.reference for variable extraction.
- Spectral (b): Electromagnetic spectrum bands.

By structuring data in a 5-D format (x/y/z/t/b), EO Data Cubes enable comprehensive spatiotemporal analyses crucial for wildfire risk modeling.

Data Extraction Using Xarray: To facilitate efficient variable extraction, Xarray (Hoyer et al. 2017) was utilized to retrieve and structure remote sensing data from Google Earth Engine (GEE) into EO Data Cubes. Xarray’s capability to handle multi-dimensional arrays made it an ideal tool for organizing geospatial data into structured, analysis-ready formats.

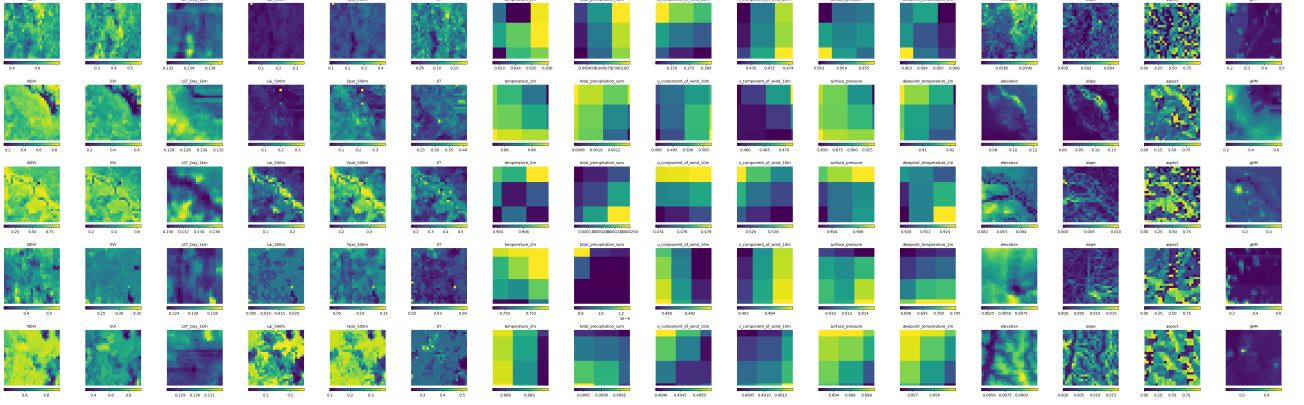
Once the points dataset were extracted, whether fire or non-fire, a tailored request was made for each point. First, for each point, a buffer was made to extract the images and therefore ensure that the variables were retrieved within 25×25 km patches. In this sense, each API request was done for polygons instead of points. However, given the extensive latitudinal variation in the study area, a dynamic scaling formula (Equation 5) was applied to maintain consistency in spatial resolution:

$$scale = \frac{1000}{111320 * \cos(\lambda)} \quad (5)$$

This adjustment ensured that, at each location, 25×25 pixels were properly extracted, regardless of its position on the globe as is shown in Figure 3, where 5 random locations and their variable patches were plotted. Additionally, a time window was applied to filter the dataset based on the burn date attribute associated with each fire point, indicating 17 days before the burn date as the starting date and one day before as end date. This allowed the retrieval of relevant environmental conditions for each observation period. The request parameters included the time window, location, and variables to extract, ensuring precise and contextually relevant data acquisition. Finally, the Data Cube created for each point was systematically stored in Google Drive, allowing for efficient access and further processing in the modeling pipeline.

This structured methodology ensures that all spatial, temporal, and spectral dimensions are effectively captured, optimizing the dataset for deep learning-based wildfire prediction models.

Figure 3: Patch Images



Workflow for wildfire risk modeling, detailing data collection, preprocessing, variable extraction, and model training. The pipeline integrates active fire sampling, non-fire sampling, predictive variable extraction, and model selection, including deep learning and traditional machine learning approaches

4.5.2 Data Preprocessing:

The temporal resolution of the extracted variables varies, with some available on a daily, 8-day, or 16-day basis. To ensure data availability for all samples, a 17-day time window was selected, guaranteeing that at least one observation per variable is present. Prior to model training, extensive preprocessing was performed to structure and refine the dataset.

Static variables were duplicated along the time axis to align with the temporal models, ensuring consistency across time-series inputs. The time coordinates of each dataset were carefully synchronized with the time window of each sample. Additionally, an initial data quality assessment was conducted by evaluating the percentage of missing values per sample, variable, and timestamp. If any patch exhibited more than 40% missing data in any variable for at least one timestamp, it was removed from the dataset.

To address missing values, a two-step interpolation approach was implemented. First, backward and forward interpolation was applied along the time dimension, which was particularly effective for Vegetation Indices, where data is only available for a few days within the time window. For variables such as Land Surface Temperature, which have a daily resolution but are often affected by atmospheric conditions, spatial interpolation was performed across the latitude and longitude dimensions using linear interpolation techniques. This method effectively mitigated data gaps while preserving spatial coherence.

Finally, all variables were normalized using min-max scaling, leveraging the predefined minimum and maximum values provided by the API. This normalization process enhanced the model's learning rate by ensuring that all features remained within a comparable numerical range, thereby improving convergence and predictive performance.

4.6 Models Architecture

4.6.1 Random Forest (RF)

The Random Forest (RF) algorithm is a versatile machine-learning technique that relies on decision trees as its fundamental component. By combining multiple trees through ensemble learning, RF effectively reduces noise and is resistant to both overfitting and underfitting. Given a dataset with n forest-fire records and m influencing factors, the RF algorithm generates n_{tree} sample sets using bootstrap sampling. It then constructs n_{tree} classification trees, where m_{try} features are randomly chosen at each node of each tree. The most influential variable is selected for branching, and the final classification result is determined by majority voting among the individual trees (weak classifiers), forming a strong classifier. The parameter m_{try} is set to $\sqrt{m}$, while n_{tree} is chosen sufficiently large to ensure the overall error rate remains stable.

Assume that the train set $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$. The final classification decision is given in equation 6:

$$H(x) = \operatorname{argmax} \sum_{t=1}^{n_{tree}} I(h_t(x) = y) \quad (6)$$

where $h_i(x)$ is the weak classifier, and $I()$ is the indicator function. This model serves as the baseline for the rest of the architectures presented in the following sections.

4.6.2 Long Short-Term Memory (LSTM)

Recurrent Neural Networks (RNNs) are highly effective for processing sequential data, as they can extract temporal patterns and semantic information while retaining past data to influence current outputs. In this model, both the recurrent and recurrent-skip components take input from the convolutional layer. Long Short-Term Memory (LSTM) units are utilized to mitigate gradient vanishing and enable association extraction before a single-step recurrence (Lin et al. 2023). The architecture of these LSTM cells is depicted in Figure 4.

The model structure of the hidden state of the cyclic unit at the time t can be expressed as follows.

$$r_t = \sigma(x_t W_{xr} + h_{t-1} W_{hr} + b_r) \quad (7)$$

$$u_t = \sigma(x_t W_{xu} + h_{t-1} W_{hu} + b_u) \quad (8)$$

$$c_t = \operatorname{ReLU}[x_t W_{xc} + r_t * (h_{t-1} W_{hc}) + b_c] \quad (9)$$

$$h_t = (1 - u_t) * h_{t-1} + u_t \odot c_t \quad (10)$$

where:

- r_t is the degree of ignoring the state information of the previous moment.
- σ is the sigmoid activation function.

Figure 4: Long short-term memory (LSTM)

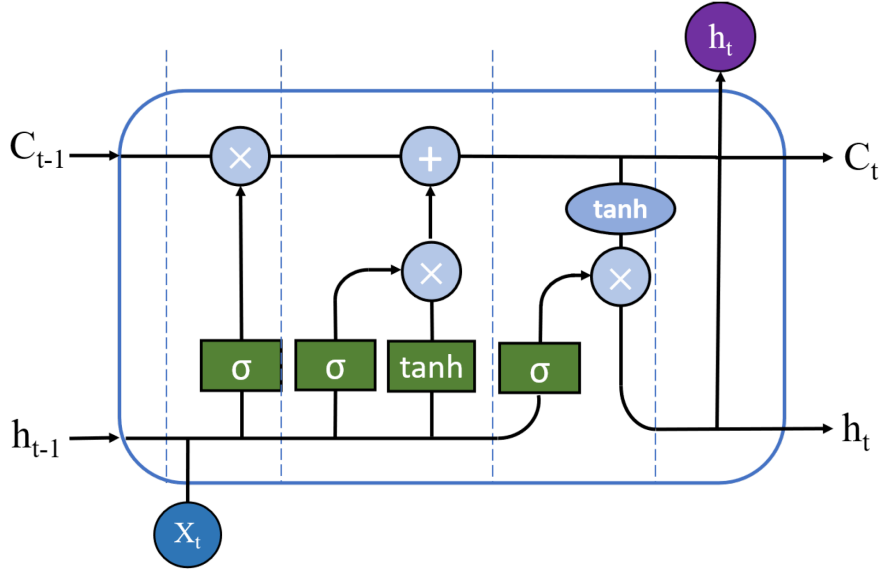


Illustration of a Long Short-Term Memory (LSTM) cell architecture, showing the flow of information through the input, forget, and output gates. The diagram highlights activation functions (sigmoid and tanh) that regulate memory updates and information retention, enabling the network to capture long-term dependencies in sequential data

- x_t is the input of the layer at time step t .
- $*$ represents element-wise multiplication.
- $W_{hr}, W_{xu}, W_{hu}, W_{xc}, W_{hc}$ are the connection weights between different states.
- h_t is the hidden state at each time step.
- b_r, b_u, b_c are the biases for different states.
- u_t represents the degree of influence of the previous state on the current state.

4.6.3 Convolutional Neural Network (CNN)

The convolutional neural network (CNN) was initially developed to address speech recognition challenges. By utilizing hidden layers, it can detect simple patterns in raw data and progressively build more complex patterns in deeper layers. Essentially, a CNN functions as an input-to-output mapping system, comprising input, hidden, and output layers. Despite the absence of a precise mathematical definition for the hidden layer, the network effectively learns various mapping relationships. Hidden layers typically include Convolutional Layers, Pooling Layers, and Fully Connected Layers. The convolutional layer extracts features from different regions of the input, employing techniques such as local receptive fields and weight sharing to optimize memory usage in deep networks. Pooling layers retain key features while reducing data size. Fully connected layers distribute high-level features from convolutional layers, similar to traditional classification neural networks, to generate the final output (You et al. 2023).

Various processing layers serve distinct functions within a CNN. Convolutional layers apply linear convolution operations between the input tensor and a set of filters, producing feature maps as output. Typically, each feature map is then processed by a nonlinear activation function. The most widely used activation function is the rectified linear unit (ReLU), which transforms the feature map generated by the convolutional layer and introduces nonlinearity into the network. The purpose of the convolution operation is to extract different input layer features and achieve weight sharing. The input and output of each stage are sets of arrays called feature maps. For example, if the input is a 2-dimensional image x , the input is first decomposed into a sequential array $x = \{x_1, x_2, \dots, x_N\}$. The convolutional layer is defined as:

$$y_j = f \left(b_j + \sum_i k_{ij} * x_i \right) \quad (11)$$

where y_j denotes the j th output of the convolutional layer, and x_i represents each input feature map. k_{ij} denotes the convolutional kernel associated with the i th input map x_i . The symbol $*$ represents the discrete convolution operator, b_j is a trainable bias, and f is the nonlinear activation function.

4.6.4 Convolutional Long Short-Term Memory (ConvLSTM)

The fundamental concept of a ConvLSTM closely resembles that of a standard LSTM. The key distinction lies in how the learnable weights in the forget, input, and output gates, as well as in the cell state, are computed—ConvLSTM uses convolution operations instead of matrix multiplication. As a result, the input vector, forget vector, cell state, and output vector transition from being 2D tensors (vectors over time) to 3D tensors (matrices over time), where the first dimension represents time and the remaining two represent spatial dimensions. (Ji et al. 2024)

The ConvLSTM model predicts the future cell state based not only on its own input and past cell states but also on those of its neighboring cells. This enables the model to preserve spatial structure while simultaneously capturing temporal dependencies, following the standard LSTM mechanism of updating and utilizing the cell state. The modified LSTM equations are now as follows:

$$f(t) = \sigma(b_f + W_f^x * X(t) + W_f^h * H(t-1)) \quad (12)$$

$$g(t) = \sigma(b_g + W_g^x * X(t) + W_g^h * H(t-1)) \quad (13)$$

$$\tilde{C}(t) = \tanh(b_c + W_c^x * X(t) + W_c^h * H(t-1)) \quad (14)$$

$$C(t) = f(t) \odot C(t-1) + g(t) \odot \tilde{C}(t) \quad (15)$$

$$o(t) = \sigma(b_o + W_o^x * X(t) + W_o^h * H(t-1)) \quad (16)$$

$$h(t) = o(t) \odot \tanh(C(t)) \quad (17)$$

In these equations, the star symbol represents the convolution operator, while the filled-in dot signifies the Hadamard product. The Hadamard product involves element-wise multiplication of two matrices, meaning each element in the first matrix is multiplied by the corresponding element in the second matrix at the same position. To ensure this operation is valid, both matrices must have the same shape. Therefore, padding is applied before the convolution process to maintain consistency in the dimensions of the cell state and input matrices.

In all these equations, W represents a set of trainable weights, analogous to a single kernel in a CNN. The total number of filters—each containing the same number of kernels as the number of data channels—is a hyperparameter that can be tuned similarly to a CNN. Figure 5 illustrates the operations input into the ConvLSTM model.

Figure 5: Convolutional Long short-term memory (ConvLSTM)

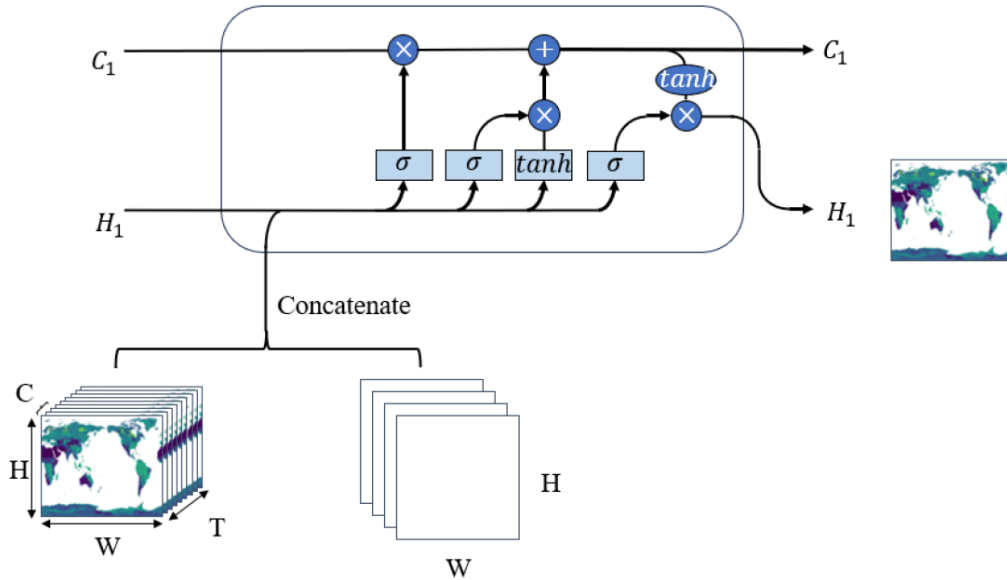


Illustration of a ConvLSTM-based spatiotemporal processing framework. The diagram depicts a sequence of multi-channel geospatial data over time as input, processed through a ConvLSTM cell, which captures spatial and temporal dependencies. The LSTM gates (sigmoid and $\tanh$) regulate information flow, and the final output preserves spatial structure for fire risk prediction.

4.6.5 Model evaluation

The primary research question involves a binary classification problem, meaning all models trained for these tasks produce categorical outputs. Consequently, the evaluation metrics used in this thesis include accuracy, precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUROC).

- Accuracy is calculated as the number of correct predictions divided by the total number of predictions. In a balanced dataset, it effectively reflects model performance, as both true positives (TP) and true negatives (TN) are equally considered.

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

- Precision measures the proportion of correctly identified positive predictions. It is computed as the number of true positives divided by the sum of true positives and false positives (FP). In imbalanced datasets, precision is often a more reliable indicator of performance than accuracy. It is particularly crucial in applications where minimizing false positives is essential, such as fraud detection.

$$\text{Precision} = \frac{TP}{TP + FP}$$

- Recall, on the other hand, assesses the proportion of actual positives that the model correctly identifies. It is determined by dividing true positives by the sum of true positives and false negatives (FN). Recall is especially useful in scenarios where avoiding false negatives is critical, such as disease detection in medical applications.

$$\text{Recall} = \frac{TP}{TP + FN}$$

- Since precision and recall often trade off against each other, improving one can reduce the other, it can be challenging to compare models that excel in different areas. To address this, the F1-score is used, which is the harmonic mean of precision and recall. A high F1-score signifies that the model effectively identifies true positives while minimizing false positives.

$$\text{F1-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

- Finally, AUROC measures the model's ability to distinguish between classes. It is derived from the area under the receiver operating characteristic curve, which plots the true positive rate against the false positive rate across various decision thresholds. An AUROC score of 0.8 indicates that 80% of the time, a randomly chosen positive instance will be assigned a higher probability than a randomly chosen negative instance. The best possible AUROC score is 1, while 0.5 represents random guessing—scores below 0.5 indicate that

the model’s predictions could be inverted for better performance.

$$\text{AUROC} = \int_0^1 \text{True Positive Rate(TPR)} d(\text{False Positive Rate(FPR)})$$

4.6.6 Proposed models architecture

For the pixel dataset, a Random Forest (RF) baseline model is considered, as it has been extensively used in related work. Its hyperparameters are selected based on the validation Area Under the Receiver Operating Characteristic (AUROC). The final RF consists of 200 trees with a maximum depth of 50, the maximum number of features considered at each split set to the square root of the total number of features, and a random state of 42 for reproducibility.

For the temporal dataset, a Long-Short Term Memory (LSTM) architecture is employed to effectively capture temporal dynamics. The model comprises two stacked LSTM layers with 128 and 64 neurons, respectively. The second LSTM layer incorporates a dropout of 0.3 for both input and recurrent connections to mitigate overfitting. This is followed by three fully connected linear hidden layers with 128, 64, and 32 neurons, respectively, each succeeded by a dropout layer with probability $p = 0.5$ and ReLU activation. The output layer is a 2-class softmax layer. The LSTM model is compiled with the Adam optimizer using a learning rate of 0.001 and trained with sparse categorical cross-entropy loss. Early stopping based on validation loss with a patience of 5 epochs and a batch size of 32 is employed during training.

For the spatial dataset, a Convolutional Neural Network (CNN) architecture is utilized, which is known to effectively capture spatial features. The architecture begins with a convolutional layer with 64 filters of size 2x2, padding set to "same", and ReLU activation, followed by a 2x2 max-pooling layer. A second convolutional layer follows with 32 filters of size 2x2, padding set to "same", and ReLU activation, followed by another 2x2 max-pooling layer. The feature maps are then flattened and passed through two fully connected layers with 64 and 32 neurons, respectively, each followed by dropout with $p = 0.5$ and ReLU activation. The final output is a 2-class softmax layer. The CNN is trained with the Adam optimizer using a learning rate of 0.001 and an L2 weight decay of 0.01. The model employs sparse categorical cross-entropy as the loss function.

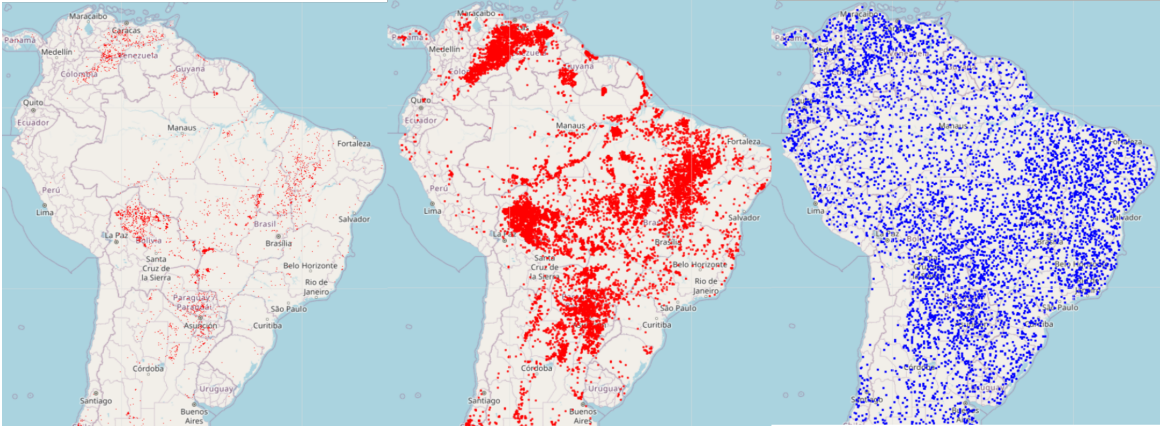
For the spatio-temporal dataset, a Convolutional Long-Short Term Memory (ConvLSTM) architecture is employed, known for its ability to capture spatio-temporal features effectively. The model starts with a ConvLSTM layer with 64 filters of size 3x3, padding set to "same", and ReLU activation. This is followed by a Conv2D layer with 32 filters of size 3x3, padding set to "same", stride of 1, and ReLU activation, along with a subsequent 2x2 max-pooling layer. The output is then flattened and passed through two fully connected layers with 64 and 32 neurons, respectively, each followed by dropout with $p = 0.5$ and ReLU activation. The final output is a 2-class softmax layer. The ConvLSTM model is trained with the Adam optimizer using a learning rate of 0.0001 and an L2 weight decay of 0.01, employing sparse categorical cross-entropy as the loss function.

5 Results and Discussion

5.1 Sampling distribution

A total of 137,940 samples were extracted for model training, maintaining a fire-to-non-fire ratio of 1:2. These samples were spatially distributed across South America for the years 2021, 2022, and 2023, with 42,712, 48,150, and 47,078 samples, respectively. For model validation, an independent dataset comprising 52,623 samples was extracted from 2024, ensuring a temporal separation between training and validation data (Figure 6). To preserve ecological represen-

Figure 6: Samples spatial distribution



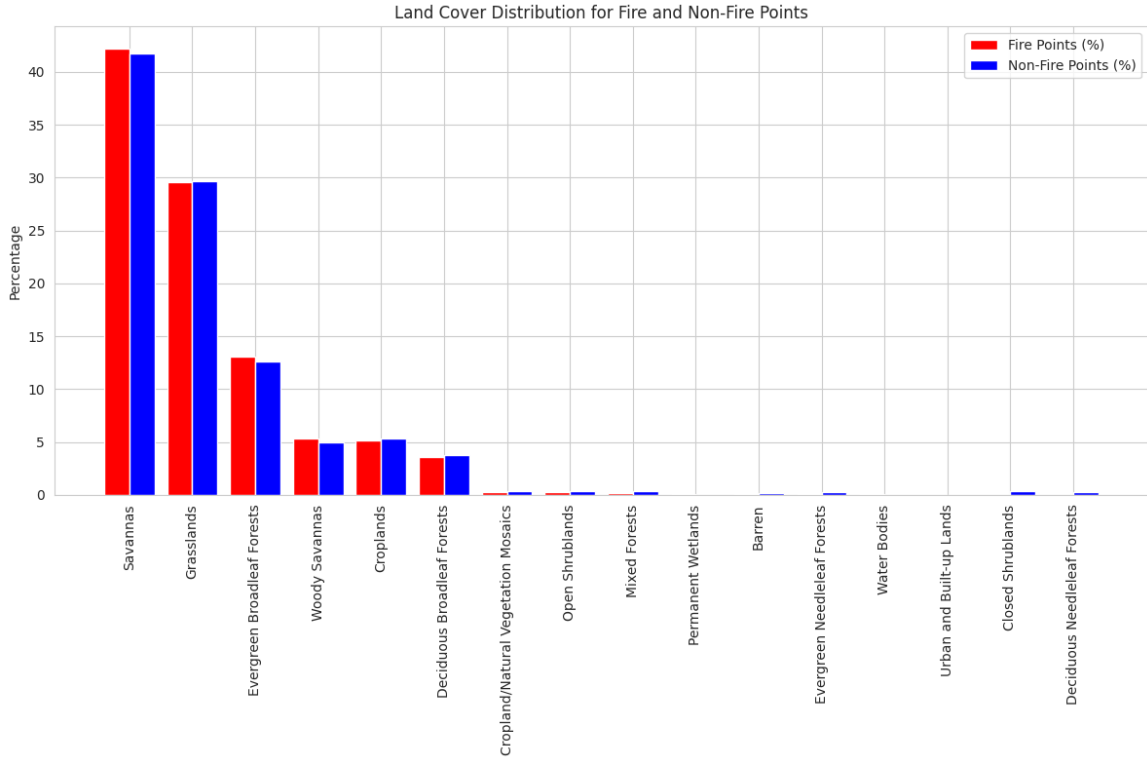
Visualization of fire and non-fire sampling points across South America. The first two maps display active fire locations (red points) detected by satellite data, while the third map shows non-fire samples (blue points) used for model testing

tativeness, the land cover distribution was maintained across both fire and non-fire samples. As illustrated in Figure 7, the distribution of land cover types remains consistent between fire-affected and non-fire areas, ensuring that the model is trained on a balanced representation of different ecosystems.

Notably, the land cover categories most frequently impacted by wildfires include Savannas, Grasslands, Evergreen Broadleaf Forests, and Croplands, highlighting the vulnerability of these ecosystems to fire disturbances. The prevalence of fires in these land cover types aligns with ecological and climatic patterns, as Savannas and Grasslands are characterized by highly flammable vegetation, seasonal dry periods, and frequent fire-adaptive dynamics. Evergreen Broadleaf Forests, although typically more humid, can experience severe wildfires under prolonged drought conditions, exacerbated by deforestation and land-use changes. Meanwhile, Croplands are often exposed to fire events due to agricultural burning practices, land clearing, and unintentional spread from nearby vegetation fires.

The overall land cover distribution within the dataset reflects the diverse fire regimes across South America, capturing variations in climate, vegetation structure, and human influence. By maintaining this distribution in both training and validation datasets, the model is better equipped to generalize across different ecological contexts and predict wildfire risk across various landscapes.

Figure 7: LULC distribution for fire and non fire points



Land cover distribution for fire and non-fire points. The bar chart compares the percentage of fire (red) and non-fire (blue) samples across different land cover types. Savannas and grasslands dominate both categories, while forests, croplands, and shrublands contribute to varying extents. This distribution helps assess the relationship between land cover and wildfire occurrence.

5.2 Environmental and Climatic Differentiation

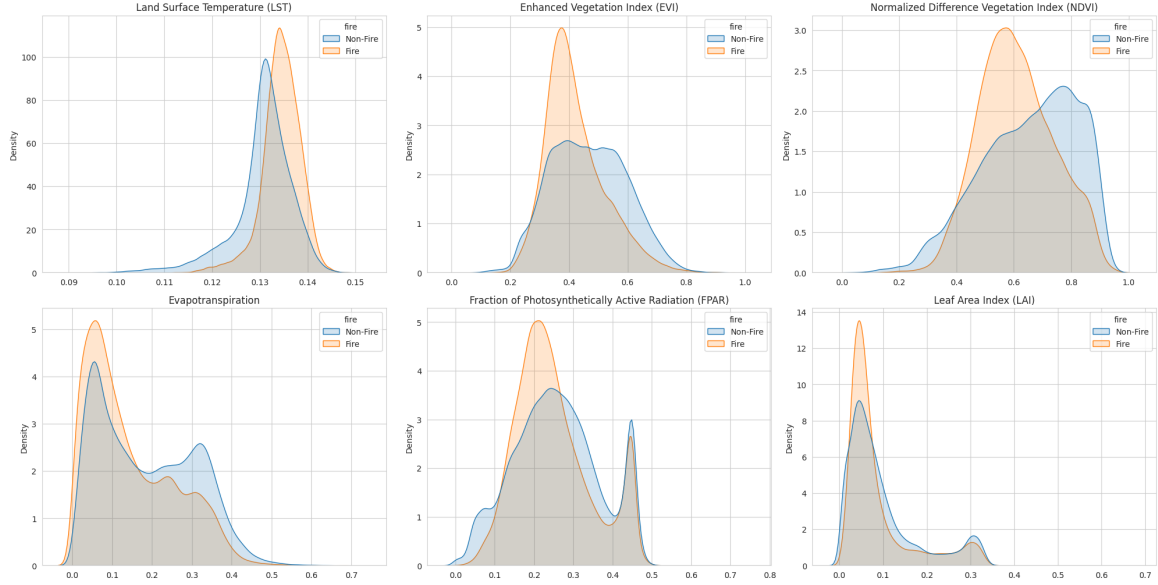
To further understand the environmental and climatic conditions influencing wildfire occurrence, the distribution of dynamic (MODIS and ERA5-Land) and static (topography and human modification) variables across fire and non-fire samples are analyzed. These distributions reveal key differences that align with known fire dynamics across South America's ecosystems.

5.2.1 Surface and Vegetation Conditions

Fire-affected areas generally exhibit higher daytime land surface temperatures compared to non-fire areas. This slight but consistent increase in surface temperature highlights the role of heat accumulation in fire-prone regions, particularly in Savannas and Grasslands, which experience intense solar radiation and dry conditions. Similarly, vegetation indices such as EVI and NDVI are notably lower in fire pixels as shown in Figure 8, indicating reduced vegetation vigor and biomass availability. This pattern suggests that fire-prone areas tend to have more open vegetation with lower greenness, characteristic of fire-adapted ecosystems like Savannas and Grasslands.

Leaf Area Index (LAI) and Fraction of Photosynthetically Active Radiation (Fpar) also show lower values in fire-affected pixels, reflecting reduced canopy cover and primary pro-

Figure 8: Distribution of Surface and Vegetation variables



Kernel density estimation (KDE) plots comparing fire and non-fire points across key dynamic variables. The distributions show differences in Land Surface Temperature (LST), vegetation indices (EVI, NDVI), evapotranspiration, Fraction of Photosynthetically Active Radiation (FPAR), and Leaf Area Index (LAI). These variations highlight the environmental conditions associated with fire occurrences.

ductivity. These characteristics are particularly evident in Croplands and degraded Evergreen Broadleaf Forests, where periodic burning or deforestation can significantly alter vegetation density.

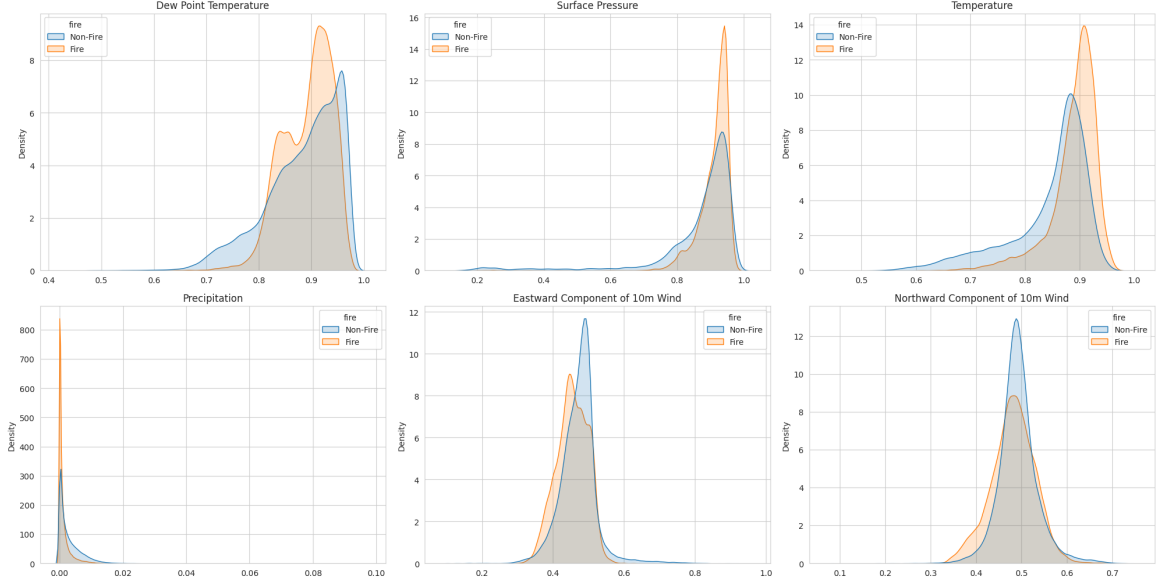
Evapotranspiration (ET), which is significantly lower in fire pixels compared to non-fire pixels. This suggests that areas experiencing wildfires tend to have lower moisture loss through vegetation and soil, reinforcing the notion that fire-prone regions are often drier. Lower ET values indicate reduced water availability, which, in combination with high temperatures, increases fire susceptibility. This trend is particularly pronounced in Grasslands and Croplands, where seasonal drought and land management practices limit soil moisture retention.

5.2.2 Climatic Factors and Fire Susceptibility

Meteorological variables further distinguish fire and non-fire regions in Figure 9. The dew point temperature distribution indicates a slight shift towards higher values in non-fire points compared to fire points, suggesting that areas experiencing wildfires tend to have lower atmospheric moisture, which can contribute to drier fuel conditions and increased fire risk.

For surface pressure, the distribution curves exhibit similar shapes for both fire and non-fire samples, making direct differentiation based on mean values challenging. However, the fire curve appears taller, indicating a higher density of values around the central distribution. This pattern suggests that while fire-prone areas do not necessarily have distinct surface pressure levels, there is a more consistent clustering of values, possibly reflecting stable atmospheric conditions conducive to fire spread.

Figure 9: Distribution of climatic variables



Kernel density estimation (KDE) plots comparing fire and non-fire points across key climatic variables. The distributions illustrate differences in dew point temperature, surface pressure, temperature, precipitation, and wind components. These variations highlight the atmospheric conditions associated with fire occurrences and their potential influence on wildfire risk.

Similarly, precipitation shows a distribution pattern akin to surface pressure, where both fire and non-fire regions share a comparable shape, but the fire distribution tends to have a more concentrated peak. While the overall precipitation levels remain low across both classes, this indicates that fire-prone regions may experience more uniform dry conditions compared to non-fire regions, where precipitation variability is slightly greater.

Temperature remains consistently higher in fire pixels than in non-fire pixels, reinforcing the correlation between higher temperatures and wildfire occurrence. This is particularly relevant in Savannas, where seasonal temperature peaks coincide with prolonged dry periods, intensifying fire susceptibility.

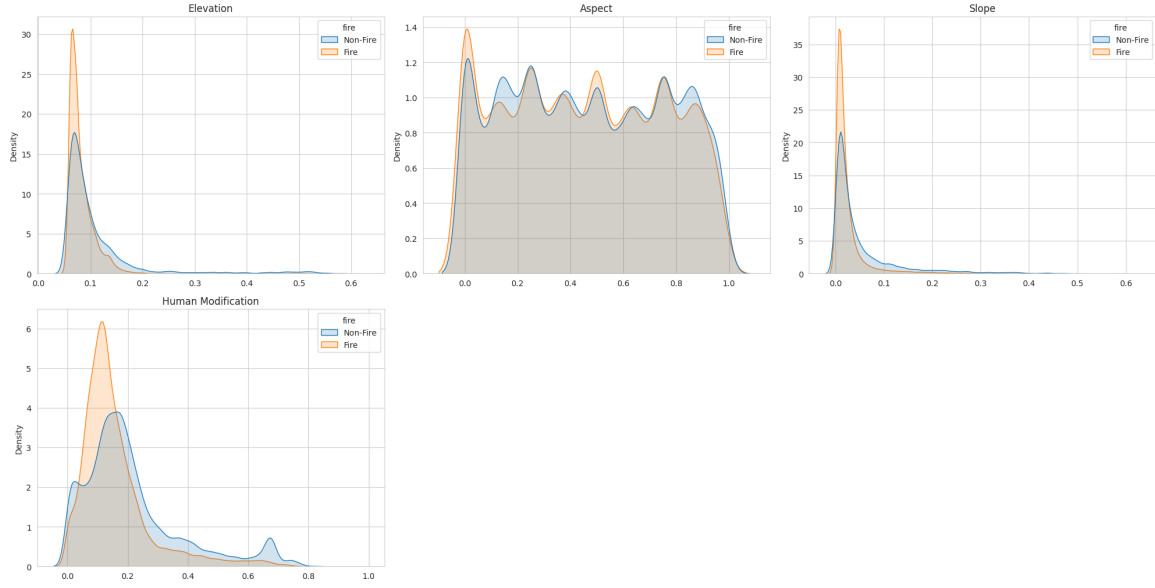
Wind components reveal that fire-prone areas experience slightly lower but still significant wind speeds. These wind dynamics are crucial in shaping fire behavior, influencing flame spread direction and intensity, particularly in open landscapes such as Grasslands and Croplands.

5.2.3 Topographical and Anthropogenic Influences

Topographic features, including elevation and slope distributed as shown in 10, provide additional context for fire occurrence patterns. While the mean values for both variables are slightly higher in non-fire regions, the overall distributions show similar shapes. However, the fire curve is significantly taller around the mean, indicating that fire-prone areas tend to cluster more tightly around a specific range of elevation and slope values. This suggests that while elevation and slope alone do not distinctly separate fire and non-fire areas, fires are more consistently

occurring within a narrower range of terrain conditions.

Figure 10: Distribution of Topographical and Anthropogenic variables



Kernel density estimation (KDE) plots comparing fire and non-fire points across key static variables. The distributions illustrate differences in elevation, slope, aspect, and human modification index, highlighting how topographical and anthropogenic factors influence fire occurrence.

Fire events tend to occur in relatively flat terrains, where fuel accumulation and wind-driven spread are more efficient. This is consistent with the dominance of wildfires in lowland Savannas and Grasslands, as well as in agricultural areas where land slopes are generally low. Conversely, non-fire regions exhibit slightly higher variation in both elevation and slope, potentially reflecting the presence of more fire-resistant ecosystems in mountainous or rugged landscapes.

Aspect distributions are similar between fire and non-fire samples, suggesting that while slope orientation may influence microclimatic conditions, it does not significantly differentiate fire-prone regions from non-fire regions at a broad scale.

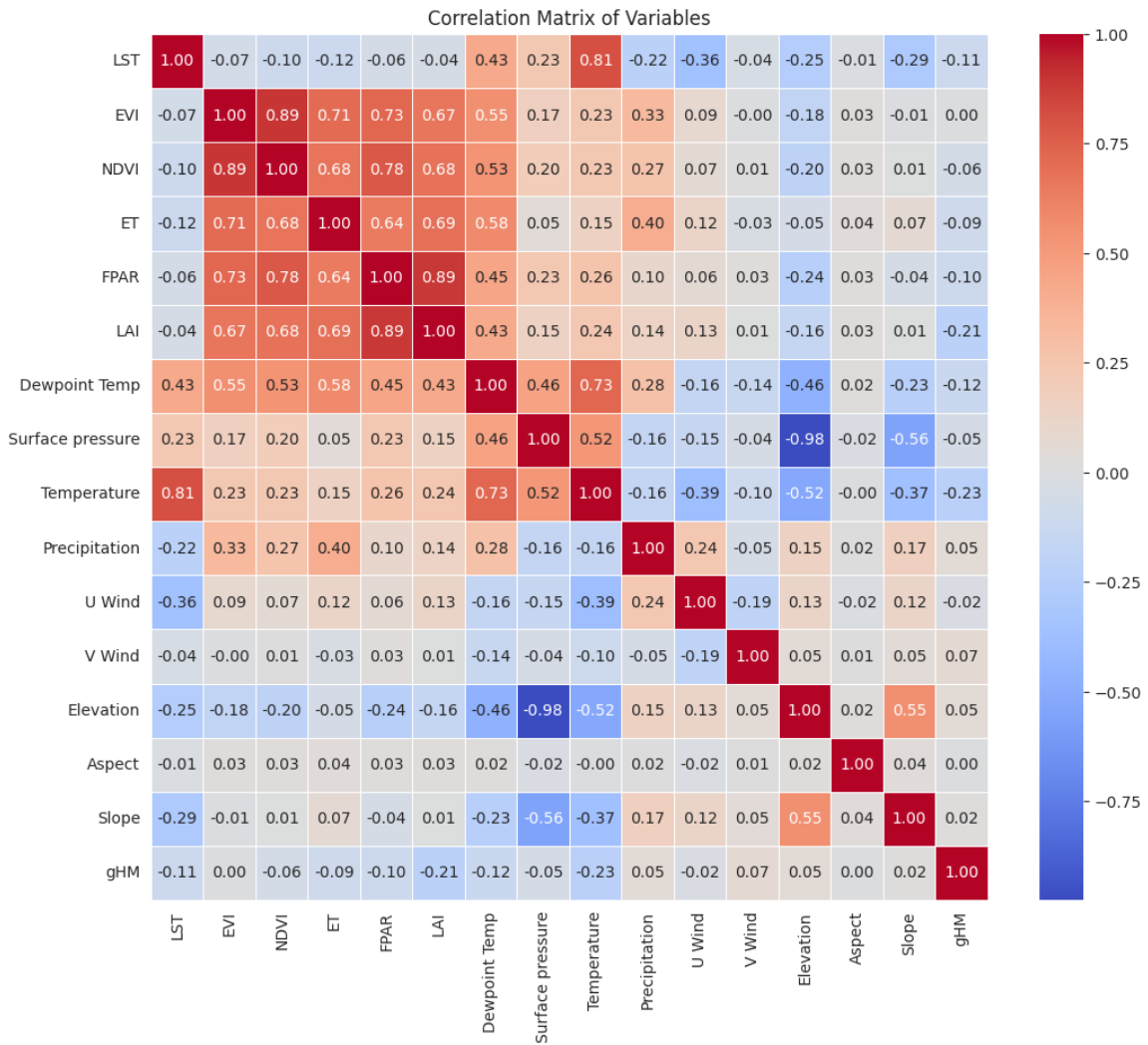
Human modification values highlight the impact of anthropogenic activities on fire occurrence. Fire pixels exhibit lower human modification scores compared to nonfire pixels, suggesting that while fire-prone areas include some modified landscapes such as Croplands, a significant portion remains in semi-natural conditions like Savannas and Grasslands. This observation aligns with fire management practices, where human influence in croplands often leads to controlled burning, while natural wildfires are more frequent in less disturbed ecosystems.

5.2.4 Analysis of Variable Correlations

The correlation analysis highlights the key environmental and climatic factors that influence fire occurrence (Figure 11). Higher land surface temperature and air temperature are strongly associated with fire-prone areas, while variables related to vegetation productivity, such as NDVI, EVI and evapotranspiration (ET), show negative correlations with temperature. This suggests

that fire-prone regions tend to be hotter, drier, and have less vigorous vegetation. Precipitation and ET are positively correlated, reinforcing that wetter areas retain more moisture, reducing fire susceptibility. Furthermore, the temperature of the dew point is positively correlated with temperature and ET, indicating that even in warm areas, the actual moisture available in the air may not be sufficient to prevent fuel drying. The components of wind speed show weak correlations with most variables, but a slight negative correlation with temperature suggests that in the hotter regions, wind speeds tend to be lower, which could influence the dynamics of fire spread.

Figure 11: Variables correlation



Correlation matrix of environmental variables used for wildfire risk modeling. The heatmap shows the relationships between dynamic and static variables, with red indicating strong positive correlations and blue representing negative correlations. Key insights include high correlations between vegetation indices (EVI, NDVI, FPAR, LAI) and evapotranspiration, as well as strong associations between land surface temperature, dew point temperature, and atmospheric pressure.

Topography also plays a role, as higher elevations and steeper slopes are associated with lower temperatures, lower LST, and reduced vegetation productivity (NDVI, EVI, and Fpar), making high-altitude regions less prone to fires. Surface pressure decreases with elevation, a natural atmospheric effect, but does not strongly differentiate fire-prone areas. Human

modification (gHM) exhibits negative correlations with NDVI, ET and temperature, indicating that human-altered landscapes generally have lower vegetation productivity and slightly cooler conditions, possibly due to land-use practices such as irrigation or deforestation. In general, fire-prone regions are characterized by hotter and drier conditions, less vegetation health, and reduced moisture retention, making them more vulnerable to wildfires, particularly in lowland savannas, grasslands, and cultivated areas.

5.3 Wildfire Risk Prediction

In this study, four classification algorithms—Random Forest (RF), Long Short-Term Memory (LSTM), Convolutional Neural Network (CNN), and Convolutional LSTM (ConvLSTM)—were trained and tested to predict wildfire probability. The problem was approached as a forecasting task, meaning that the training data strictly preceded the testing data chronologically. Unlike standard k-fold cross-validation, which is often used to reduce variance in model performance, time-series split cross-validation was applied to ensure that training and testing data maintained a temporal sequence, preventing data leakage and better simulating real-world forecasting conditions.

The models were trained on wildfire data from 2021 to 2023 and validated using 2024 data, maintaining a temporal separation between the training and testing datasets. The evaluation was conducted using five key metrics: accuracy, precision, recall, F1-score, and AUROC, which provide a thorough assessment of the models' ability to distinguish between fire and non-fire events. The results presented in Table 2 demonstrate the performance of each model when trained on historical wildfire data and tested on the most recent fire occurrences. By analyzing these models, this study aims to determine the most effective approach for forecasting wildfire risk across South America.

Table 3 presents the performance of different wildfire prediction models. The Random Forest (RF) model achieved the highest accuracy (84%) and precision (84%), indicating a strong ability to correctly classify non-fire samples and minimize false positives. Additionally, it scored the highest AUROC (91%), suggesting it is the best at distinguishing between fire and non-fire samples. However, RF exhibited the lowest recall (64%), meaning it struggled to identify all actual wildfire occurrences. This suggests that while RF is highly reliable when predicting a fire event, it may miss a significant number of true fire cases.

The LSTM model demonstrated a more balanced trade-off between precision (75%) and recall (70%), leading to an F1-score of 72%. Its AUROC (89%) was slightly lower than RF's, but still indicates strong performance in separating fire from non-fire regions. The CNN model, on the other hand, had the highest recall (84%), meaning it was the best at detecting fire events. However, its precision (67%) was the lowest among all models, leading to a higher false positive rate. This suggests that while CNN excels in capturing fire occurrences, it is more prone to misclassifying non-fire regions as fire. Meanwhile, the ConvLSTM model, which integrates spatial and temporal dependencies, performed slightly worse than the CNN and LSTM models in terms of accuracy (80%), recall (68%), and AUROC (87%). Its F1-score (70%) was also the

Table 3: Metrics for wildfire risk prediction models

Model	Accuracy	Precision	Recall	F1-Score	AUROC
RF	0.84	0.84	0.64	0.73	0.91
LSTM	0.82	0.75	0.70	0.72	0.89
CNN	0.81	0.67	0.84	0.75	0.90
ConvLSTM	0.80	0.72	0.68	0.70	0.87

Performance comparison of wildfire risk prediction models. The table presents accuracy, precision, recall, F1-score, and AUROC for Random Forest (RF), LSTM, CNN, and ConvLSTM models. RF achieves the highest accuracy and AUROC, while CNN demonstrates the best recall, highlighting different model strengths in fire prediction.

lowest among the neural network models, suggesting it did not significantly outperform either the CNN or LSTM individually.

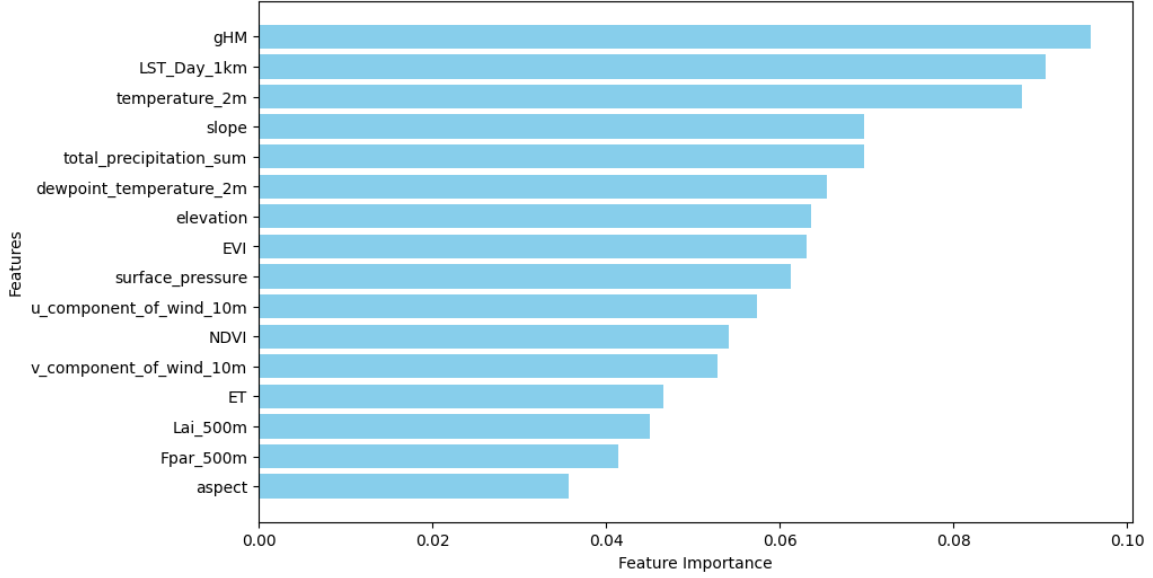
These results lead to several key observations. First, RF remains the most precise model, meaning it is highly reliable when it predicts a fire event, though it sacrifices recall. Second, CNN achieves the highest recall, making it the best at capturing fire-prone areas, but at the cost of a higher false positive rate. The LSTM model provides the most balanced trade-off between recall and precision, making it an effective choice for wildfire forecasting. Lastly, ConvLSTM does not significantly outperform other neural network models, suggesting that additional tuning or modifications may be needed to better capture spatiotemporal fire dynamics. Despite the differences in performance, the overall accuracy of these models (ranging from 80% to 84%) demonstrates that deep learning and machine learning approaches can effectively predict wildfire risk, even when trained on historical wildfire occurrences from previous years.

5.3.1 Influence of Dynamic and Static variables

The feature importance analysis reveals in Figure 12 the distinct roles of fuel, weather, topographic, and anthropogenic variables in shaping wildfire risk predictions. Among these categories, anthropogenic factors, particularly human modification (gHM), had the highest influence on the model, reinforcing the strong link between land-use change, deforestation, and fire activity. While gHM ranked as the most important predictor, its distribution across fire and non-fire samples suggests significant overlap, indicating that human influence alone is not sufficient to determine fire risk. Instead, it interacts with climatic and landscape features that modulate fire conditions.

Weather variables also played a dominant role, with land surface temperature (LST), air temperature, precipitation, and dew point temperature ranking among the most influential predictors. The strong influence of temperature-related features suggests that short-term thermal conditions had a direct impact on fire risk predictions, with higher temperatures contributing to drier fuel and increased ignition probability. Precipitation and dew point temperature, while slightly lower in importance, were crucial in differentiating between fire-prone and less fire-prone areas by capturing variations in atmospheric moisture.

Figure 12: Impact of variables in Wildfire Risk prediction



Feature importance ranking for wildfire risk prediction. The bar chart shows the relative contribution of different environmental variables to the model’s decision-making process. Human modification index (gHM), land surface temperature (LST), air temperature, slope, and precipitation are among the most influential factors in predicting fire occurrence.

Fuel-related variables, such as NDVI, EVI, Fpar, and Lai, ranked lower in importance compared to weather and anthropogenic features. While these variables help characterize vegetation density and fuel availability, their influence appears secondary to climatic and human-driven factors. The distributions of NDVI and EVI confirm that while fire-prone areas tend to have lower vegetation indices, the difference is not substantial, indicating that fuel presence alone is not a decisive factor in fire risk when compared to meteorological and human influences.

Lastly, topographic features, including slope and elevation, also contributed to fire risk assessments, albeit with lower influence compared to weather and human factors. Their importance suggests that terrain characteristics influence fire behavior by shaping wind patterns, water retention, and fuel accumulation. The relatively lower ranking of topography-related variables indicates that while slope and elevation modify fire spread dynamics, they are not the primary drivers of ignition probability.

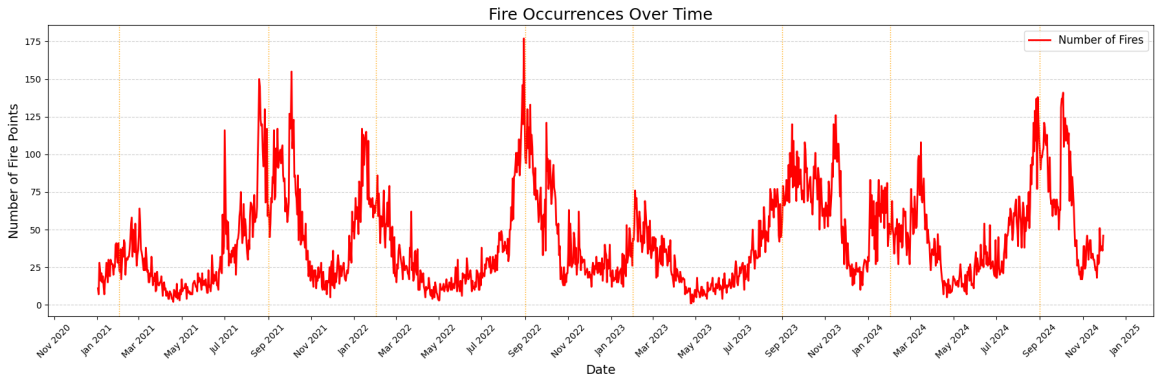
In general, the balance between static and dynamic factors in the feature importance ranking highlights the model’s ability to generalize across different regions and seasons. While static variables such as land cover and topography establish long-term fire susceptibility, dynamic weather conditions dictate short-term fire activity, emphasizing the need for real-time climate data to refine predictions. The model’s reliance on human-driven and meteorological factors suggests that wildfire risk is not solely determined by vegetation and fuel availability but is the result of a complex interaction between land-use patterns, atmospheric conditions, and fire-prone landscapes.

5.4 Real-World Application of Wildfire Risk Models

To assess the practical applicability of the wildfire risk prediction models, a real-world evaluation was conducted by analyzing their predictions in relation to actual fire occurrences. This analysis aimed to validate the models' ability to differentiate between fire-prone and non-fire-prone regions and to examine how their spatial predictions align with observed wildfire events.

To achieve this, a time series of fire activity was first plotted (Figure 13) to identify the periods with the highest fire occurrence. Based on this analysis, two study areas, each covering approximately 900,000 km², were selected for further investigation—one in February and another in September of 2024. These areas were chosen to capture seasonal variations in wildfire patterns and assess the spatial distribution of predicted fire risk under different climatic conditions.

Figure 13: Fire activity time-series



Time series of fire occurrences from 2020 to 2024. The plot shows seasonal patterns in fire activity, with peaks occurring annually around August–September, as indicated by the orange dotted lines. These recurring peaks are likely influenced by climatic and environmental factors, emphasizing the need for seasonal wildfire risk assessment and mitigation strategies.

5.5 Wildfire risk prediction

For each selected region and time period, four images were generated, with each image containing four wildfire risk maps corresponding to the predictions made by the RF, LSTM, CNN, and ConvLSTM models. Additionally, each image includes an observed fire occurrence map, displaying the actual fires that occurred after the prediction date. This setup allows for a direct visual comparison between the predicted fire risk and the real wildfire events, providing insights into the strengths and limitations of each model in capturing spatial fire dynamics.

By evaluating these spatial predictions, this section aims to highlight the differences between the models, assess their reliability in real-world conditions, and determine which approaches best capture fire-prone areas. This analysis provides valuable information for improving wildfire forecasting and guiding decision-making in fire management and prevention efforts.

5.5.1 Zone 1 - Dry season (february)

The first zone under analysis encompasses a region in northern South America, specifically between latitudes 2.2°N and 8.7°N and longitudes 71.6°W and 60.4°W. This area includes parts of southeastern Colombia and southern Venezuela, characterized predominantly by tropical rainforest ecosystems. The climate in this region is typically hot and humid, with consistent temperatures throughout the year and significant annual rainfall.

In February, the region experiences its dry season, leading to reduced precipitation and lower humidity levels. These conditions can increase the susceptibility to wildfires, as drier vegetation provides more fuel. During this period, variables such as elevated temperatures, decreased soil moisture, and lower relative humidity are experienced, all of which contribute to a heightened fire risk. Historical fire activity data indicates a surge in wildfire incidents during February, correlating with these climatic conditions. Understanding the behavior of these variables is crucial for accurate wildfire risk prediction and effective mitigation strategies in this ecologically sensitive area. (Jolly et al. 2015)

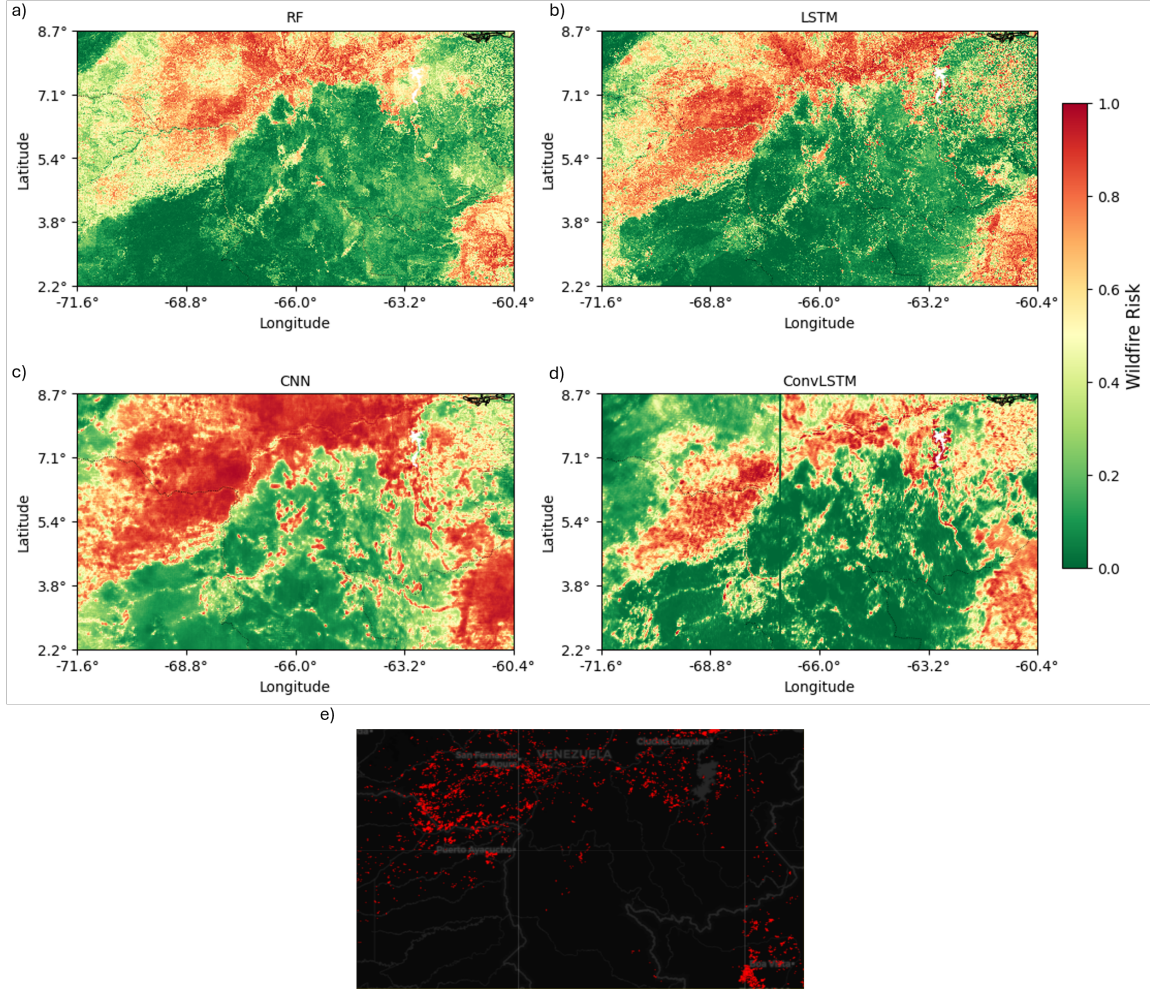
The wildfire risk predictions generated by the four models in February exhibit notable differences in their spatial distributions, as shown in Figure 14. These predictions are compared against the actual fire occurrences.

Model Performance Analysis

- **RF:** The RF model provides a relatively conservative prediction, with a mix of moderate and high-risk areas concentrated in the northern and western parts of the region. It captures some of the observed fire activity, but it also leaves many actual fire locations classified as low risk. The model seems to follow vegetation patterns but underestimates the extent of high-risk zones compared to observed fire occurrences.
- **LSTM:** The LSTM model predicts a broader extent of wildfire-prone areas, particularly in the northern and western regions, aligning more closely with the actual fire distribution. The model generally identifies high-risk areas correctly but includes a larger number of false positives compared to RF, predicting elevated risk in areas where no fires were observed.
- **CNN:** The CNN model exhibits the most extensive high-risk classification, covering large portions of the region. While this results in higher recall, ensuring that most of the actual fire locations are identified, it also increases the number of false positives, as a significant portion of non-fire areas are classified as high risk. This suggests that the CNN model is more sensitive to fire-prone conditions but sacrifices precision.
- **ConvLSTM:** The ConvLSTM model provides a more spatially fragmented prediction, with clusters of high-risk areas spread across the northern and western portions of the region. While it captures several areas with actual fire activity, the fragmented nature of the predictions might reduce its effectiveness in distinguishing continuous fire-prone

zones. The model performs reasonably well but does not offer a significant improvement over the standalone LSTM or CNN models.

Figure 14: Fire Risk prediction zone 1 - Dry season



Comparison of wildfire risk predictions across different models. The maps display wildfire risk probabilities for the first zone in February, with predictions from (a) Random Forest (RF), (b) LSTM, (c) CNN, (d) ConvLSTM, while (e) shows MODIS Active Fires. Warmer colors indicate higher fire risk (a - d) and the red pixels are the active fires (e). Variations among models highlight differences in how each approach captures spatial fire risk patterns.

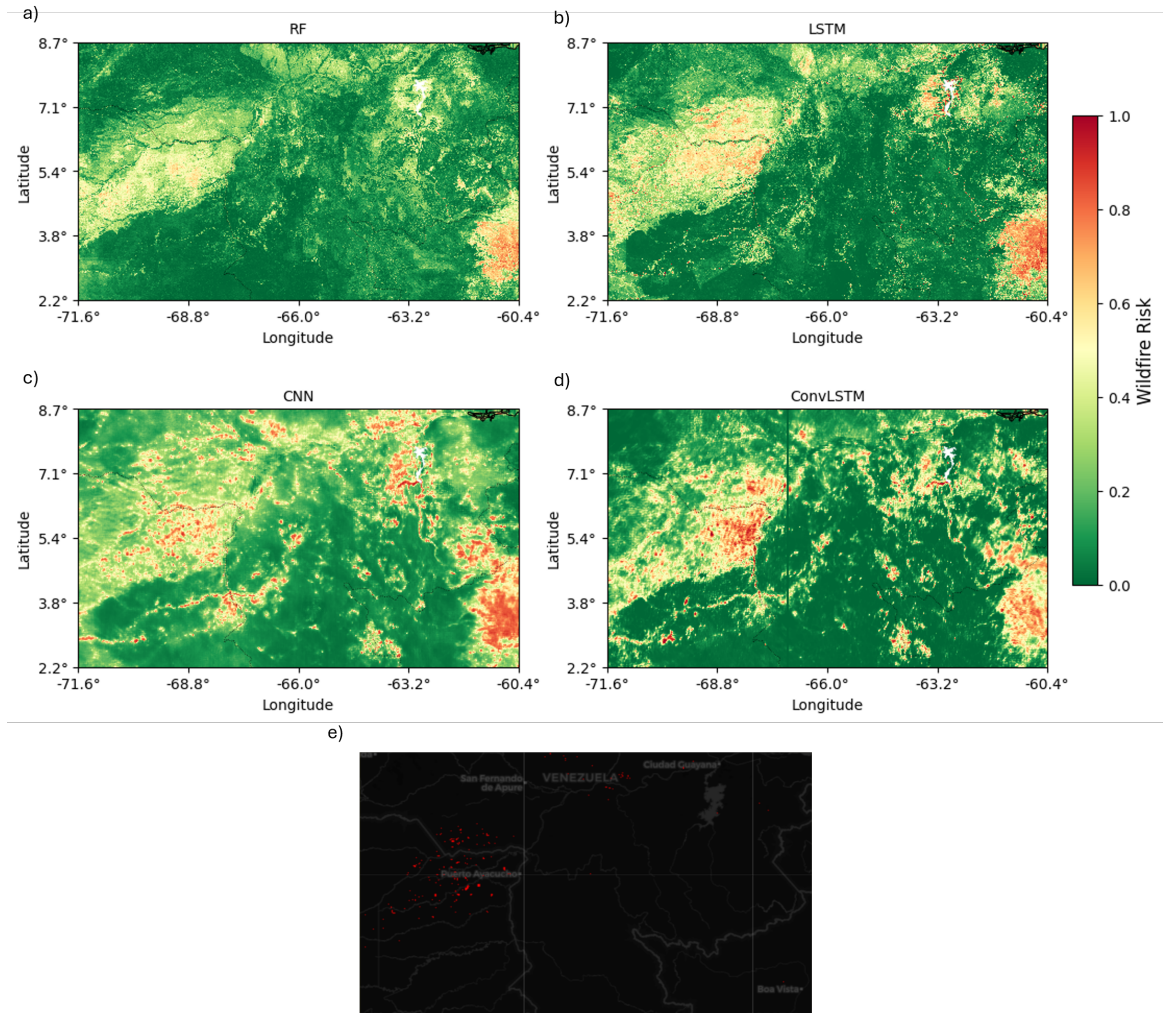
The Figure 14 (e), displaying actual fire occurrences, confirms that the highest concentration of fires occurred in the north, northwest and southeast of the region, aligning with the general patterns predicted by the models. However, CNN and LSTM provide a more extensive identification of fire-prone areas, while RF and ConvLSTM appear more conservative, missing some observed fire locations.

This analysis demonstrates how different models prioritize different aspects of wildfire risk prediction, with CNN and LSTM favoring recall (identifying more fires) and RF favoring precision (minimizing false positives). The choice of model ultimately depends on whether the application prioritizes accurate fire detection or reducing false alarms in wildfire management and mitigation efforts.

5.5.2 Zone 1 - Rain season (September)

The wildfire risk predictions for September (Figure 15) demonstrate how each model responds to seasonal shifts in fire dynamics. Unlike February, when fire-prone areas were widespread across the region, the predictions for September show a more concentrated distribution of risk, particularly in northwestern Venezuela, near Puerto Ayacucho. The RF model detects very few high-risk areas, leading to an underestimation of fire occurrences, whereas the LSTM model captures more potential wildfire zones, especially in the north and east, but still does not fully align with the observed fires. The CNN model highlights an expansive region as susceptible to fires, successfully identifying many of the fire locations but also marking several areas as at risk where no fire occurred. Meanwhile, the ConvLSTM model produces scattered predictions, marking some fire-prone zones but without forming a well-defined spatial pattern.

Figure 15: Fire Risk prediction zone 1 - Rain season



Wildfire risk predictions for the zone 1 in September using different models. The maps compare outputs from (a) Random Forest (RF), (b) LSTM, (c) CNN, (d) ConvLSTM and (e) MODIS Active Fires, where warmer colors indicate higher fire risk. Differences in predictions reflect the models' varying sensitivity to environmental and climatic conditions during this period.

A striking detail in all model outputs is the repeated classification of a specific area in the southeast as high risk, despite no recorded fire activity there. This suggests that some

models rely heavily on static environmental conditions, such as vegetation type, topography, or historical trends, even when seasonal variations might reduce fire probability. The seasonal contrast between February and September reveals that some models, particularly CNN and LSTM, are less adaptable to shifting fire patterns, continuing to highlight extensive high-risk areas even when fire occurrences are more localized. Conversely, RF and ConvLSTM appear to react more strongly to seasonal changes, but at the cost of missing actual fire events. The results underscore the importance of ensuring that wildfire prediction models effectively incorporate seasonal influences, rather than relying predominantly on static landscape features, to improve forecasting accuracy across different climatic conditions.

5.5.3 Zone 2 - Dry season (September)

The second zone, located between latitudes -13° to -6° and longitudes -59° to -48.4° , encompasses a significant portion of the southeastern Amazon rainforest, primarily within the Brazilian states of Mato Grosso and Pará. This region is characterized by a tropical climate with distinct wet and dry seasons. During the dry season, typically from June to September, reduced precipitation and lower humidity levels prevail, creating conditions conducive to fire outbreaks. In recent years, this area has experienced severe droughts, intensifying the susceptibility of forests to wildfires. The interplay of climatic factors, such as prolonged dry spells, and human activities, including deforestation and land-use changes, has led to an increase in fire incidents. Notably, the 2024 fire season was particularly devastating, with a significant surge in both the frequency and intensity of fires, exacerbated by extreme drought conditions. (Haynes [2024](#))

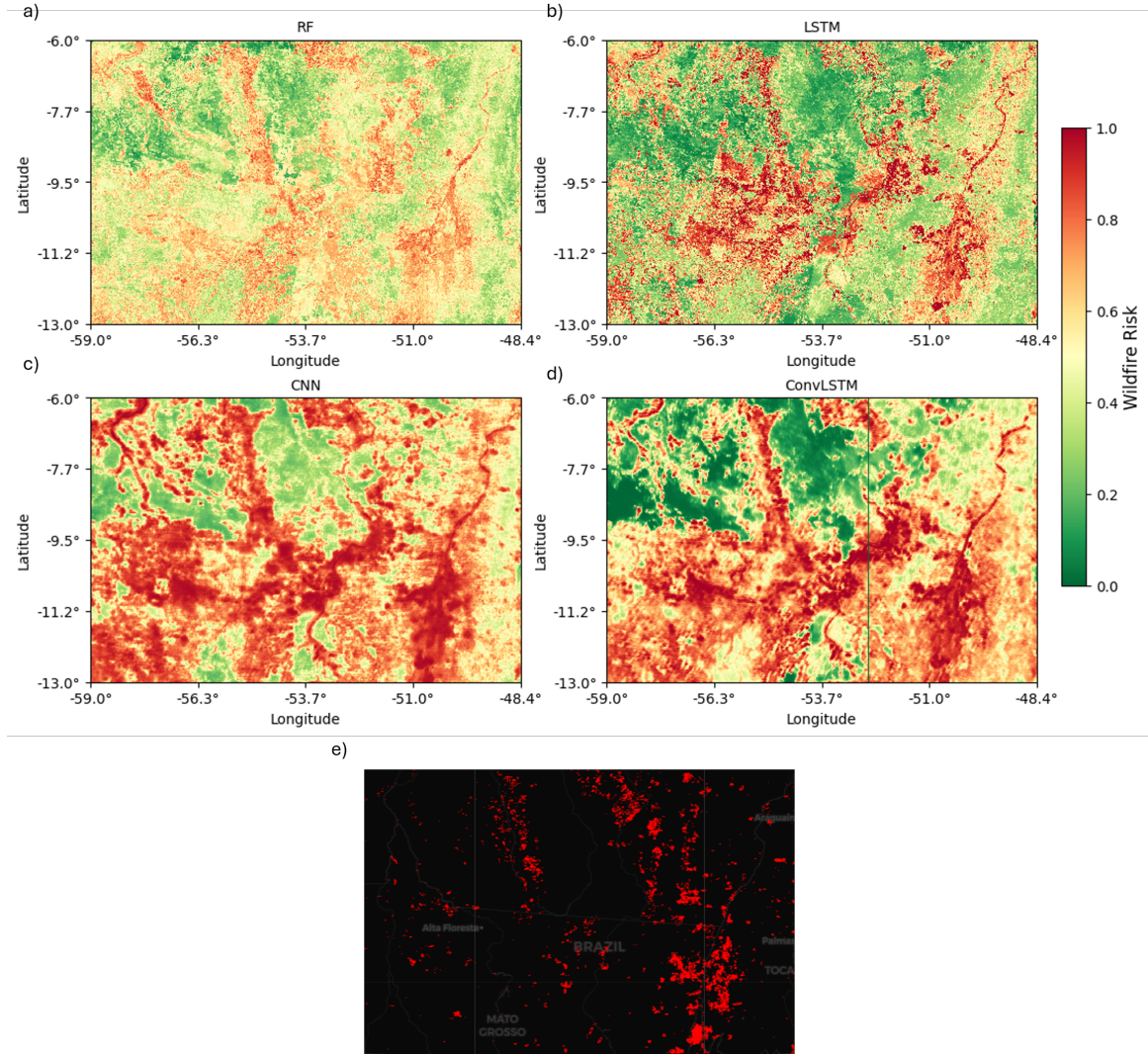
The environmental variables in this region during September reflect the peak of the dry season. Elevated temperatures, often surpassing 30°C , combined with minimal rainfall, result in low soil moisture and desiccated vegetation. These arid conditions are further amplified by deforestation activities, which not only reduce canopy cover but also disrupt local hydrological cycles, leading to even drier microclimates. The combination of natural climatic variability and anthropogenic influences has rendered the southeastern Amazon increasingly vulnerable to fires, with significant implications for biodiversity, carbon emissions, and local communities. (NASA Earth Observatory [2023](#))

The wildfire risk predictions for the dry season in this region (Figure [16](#)), covering parts of Mato Grosso and Pará, Brazil, demonstrate how each model responds to extreme fire-prone conditions. The RF model presents a relatively moderate risk distribution, with a mix of low to medium wildfire probability across most of the area, except for some high-risk patches in deforested and transitional zones. While this model captures some fire-prone locations, it appears to underestimate the extent of wildfire occurrences compared to the actual fire data. In contrast, the LSTM model predicts a more widespread presence of high-risk areas, particularly in the central and eastern parts of the region, aligning better with observed fire activity but also introducing some false positives.

The CNN model produces the most extensive fire risk predictions, marking large portions of the landscape as highly susceptible to wildfires. This prediction closely follows the

distribution of actual fire events, suggesting a high sensitivity to fire-prone conditions, though it likely overestimates fire risk in areas that did not burn. The ConvLSTM model also indicates a broad extent of fire-prone regions, but its predictions appear more fragmented, with high-risk patches interspersed with low-risk zones. Compared to the observed fire occurrences, both CNN and ConvLSTM correctly identify most of the major fire hotspots, particularly in areas affected by deforestation and land-use changes. However, all models classify some locations as high risk where no fires were detected, suggesting that static environmental conditions, such as land cover type or historical fire occurrence, heavily influence their outputs.

Figure 16: Fire Risk prediction zone 2 - Dry season



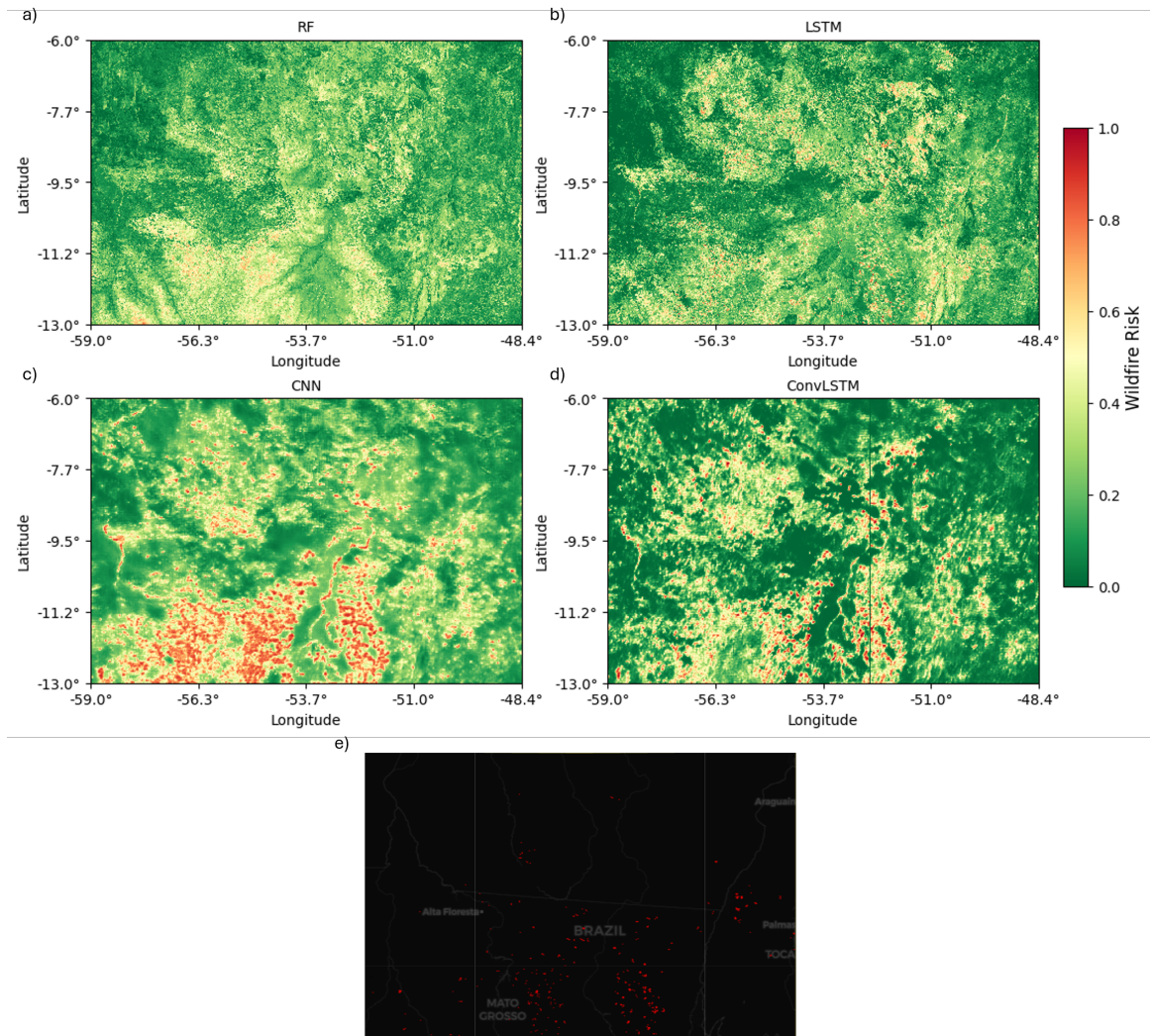
Wildfire risk predictions for the second region in September using different models. The maps compare outputs from (a) Random Forest (RF), (b) LSTM, (c) CNN, and (d) ConvLSTM, with warmer colors indicating higher fire risk, and (e) indicate the active fires from MODIS

5.5.4 Zone 2 - Rain season (February)

The wildfire risk predictions for February (Figure 17) reflect the seasonal shift in fire conditions, with all models generally predicting lower fire probabilities compared to the dry season. The RF model assigns mostly low-risk values across the region, with only small, scattered patches

of moderate risk in areas prone to fire. The LSTM model provides a slightly more nuanced risk assessment, detecting some moderate-risk zones but still predicting fewer fire-prone areas than in September. The CNN model, while still capturing fire-prone zones more aggressively, shows reduced fire probabilities in comparison to its dry season predictions, indicating sensitivity to seasonal variations. The ConvLSTM model continues to predict fragmented high-risk areas, though with a clear reduction in fire probability across most of the region. When comparing these predictions to the observed fire activity, the actual fire occurrences in February are significantly lower than in September, confirming the seasonal reduction in fire incidents. Most of the active fires detected are concentrated in deforested areas, rather than in primary forest zones.

Figure 17: Fire Risk prediction zone 2 - Rain season



Wildfire risk predictions for the second region in February using different models. The maps compare outputs from (a) Random Forest (RF), (b) LSTM, (c) CNN, and (d) ConvLSTM, where warmer colors indicate higher fire risk. Panel (e) displays active fire detections from MODIS, serving as a reference for model performance evaluation

The aim of analyzing the four wildfire risk prediction scenarios was to demonstrate the models' ability to capture accurate spatiotemporal patterns of wildfire risk, regardless of

season or location. By evaluating predictions in two distinct regions—one in northern South America and another in the southeastern Amazon—under both dry and rainy season conditions, the analysis highlights how the models adapt to seasonal variations in fire occurrence. The results show that while all models adjust their predictions based on broader climatic patterns, their capacity to differentiate high-risk areas from non-fire-prone zones varies. This evaluation provides insight into how different modeling approaches respond to environmental dynamics, land cover changes, and seasonal fire patterns, which are crucial for effective wildfire forecasting.

Among the models, LSTM demonstrated a better distinction between fire-prone and low-risk areas, as it maintained a balance between true positives and false negatives. While some models tended to overestimate risk, classifying large regions as high probability, others were more cautious, potentially underestimating fire occurrences. LSTM, by contrast, consistently captured fire-prone zones while avoiding excessive false alarms, making it more adaptable across different climatic conditions and ecological regions. This ability to generalize while preserving accuracy is essential for wildfire risk assessment, as it ensures that predictions remain reliable across diverse landscapes and seasons, supporting better-informed fire management strategies.

5.6 Web Application

To facilitate near real-time wildfire risk prediction, a web-based application was developed using the Google Earth Engine (GEE) API, Xarray data cubes, Python, and Streamlit. The application enables users to interactively select a date, draw a polygon on an interactive map, and generate wildfire risk predictions for the specified region and time frame.

5.6.1 Web Application

The web application integrates multiple geospatial technologies to provide an efficient and user-friendly wildfire risk assessment workflow:

- **Date Selection:** A calendar widget allows users to specify a date for wildfire risk prediction, supporting both historical and near real-time analysis (Figure 18).
- **Interactive Map Interface:** Built with **Leaflet** and Streamlit, the application enables users to draw a polygon to define the region of interest for analysis (Figure 19).
- **Model Execution:** Once a date and region are selected, the application retrieves the necessary remote sensing data and processes it through a wildfire risk prediction model (Figure 20).

5.6.2 Model Implementation

The current version of the web application employs a **Long Short-Term Memory (LSTM) model** to predict wildfire risk based on multi-temporal remote sensing variables. This model was trained using an extensive dataset stored as **Xarray data cubes**, processed with Python libraries such as **xarray**, **pandas**, and **NumPy**. While the web application currently utilizes

Figure 18: Webapp - date selector

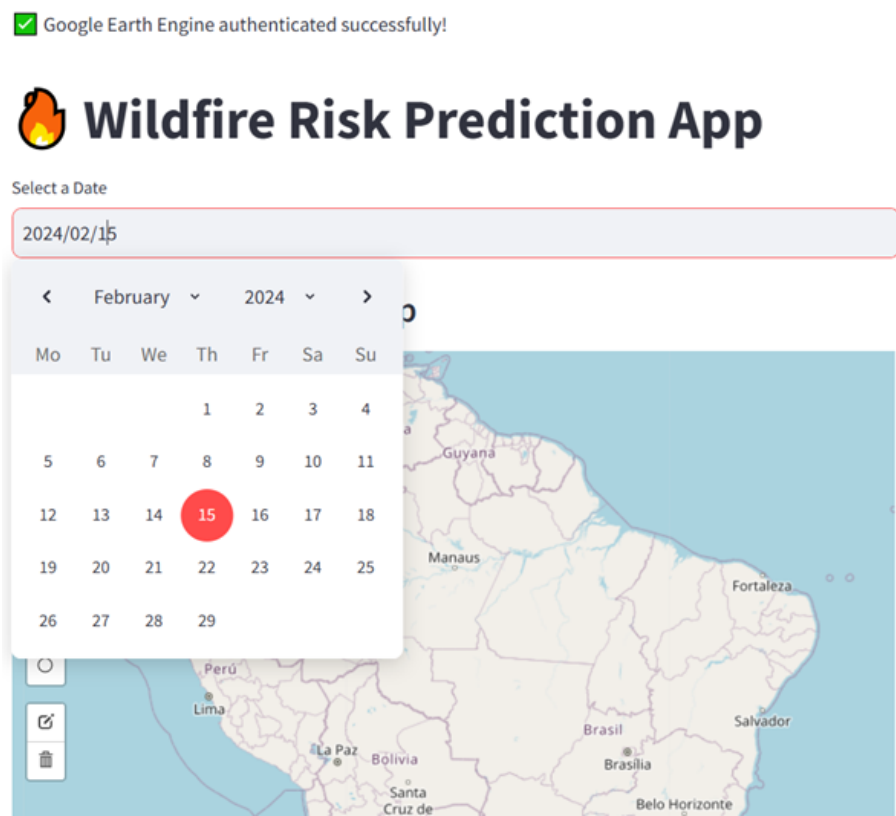
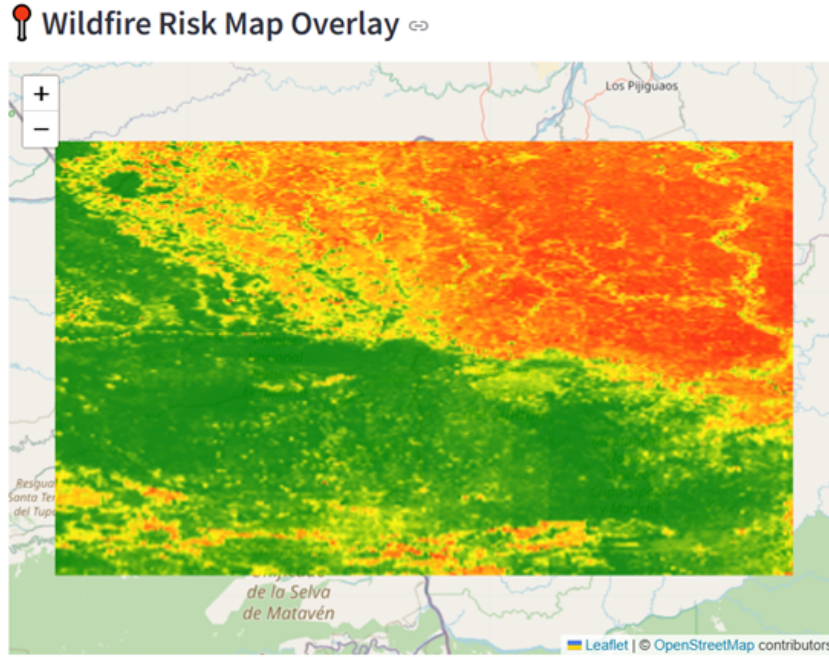


Figure 19: Webapp - draw Area of Interest



only the LSTM model, additional trained models—including convolutional and hybrid approaches—are available in the project’s **GitHub repository**. The notebooks used for model training, as well as the raw datasets, are accessible in **Google Colab**. Links to these resources

Figure 20: Webapp - calculate and visualize Wildfire Risk



can be found in the **Appendix A.2**.

5.6.3 Performance and Scalability

By leveraging **Google Earth Engine**, the application efficiently retrieves Earth observation data without the need for extensive local storage. The use of **Streamlit** ensures a lightweight and scalable deployment, providing an accessible tool for wildfire risk assessment. Future developments will focus on integrating additional models and enhancing visualization capabilities to improve predictive accuracy and usability.

5.7 Discussion

The results of this study align with and expand upon findings from previous wildfire risk prediction research. The evaluated models, including Random Forest (RF), Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), and Convolutional LSTM (ConvLSTM), demonstrated performance metrics consistent with other deep learning applications in wildfire Risk forecasting. In particular, the LSTM model showed a balanced trade-off between sensitivity and specificity with a Precision of 0.75, Recall of 0.70, and F1-Score of 0.72, which is comparable to results from Kondylatos et al. 2022, where LSTM achieved a Precision of 0.741, Recall of 0.762, and an F1-score of 0.751, indicating a strong balance between false positives and false negatives.

In contrast, Ji et al. 2024, introduced SLA-ConvLSTM, a model that integrates static location-aware features and global teleconnection information using U-Net and SKNet. Their approach significantly improved wildfire danger forecasting by enhancing spatial awareness in deep learning models. Their SLA-ConvLSTM model achieved an AUROC of 0.92 in South

America, slightly outperforming the ConvLSTM model in this study (0.87 AUROC). This suggests that incorporating static positional features and global climate patterns, as done in SLA-ConvLSTM, may enhance predictive accuracy, particularly in capturing regional fire risk variations.

Another key difference is the handling of spatial and temporal dependencies. Huot et al. 2020 framed wildfire prediction as an image segmentation task, using daily fire masks and aggregated historical fire records. Their use of residual U-Nets, convolutional autoencoders, and ConvLSTM-based models highlights the effectiveness of computer vision techniques in wildfire mapping. While the study achieved an AUC of 83%, this is lower than the AUROC scores (87%–91%) reported in this study, suggesting that integrating spatiotemporal features directly into model architectures, such as with LSTM and ConvLSTM, can yield higher predictive performance.

Compared to Prapas et al. 2021, who utilized a decade-long Earth observation dataset for daily fire danger prediction, this study employs a shorter historical window. The ConvLSTM model they proposed achieved an AUROC of 0.926, outperforming traditional machine learning methods like RF. In contrast, the ConvLSTM in this study obtained an AUROC of 0.87, which, while strong, falls slightly short of their research results. This discrepancy could be attributed to differences in dataset length, model architectures, and the inclusion of spatial-temporal interactions. Notably, Prapas et al. 2021 used an open-access data cube incorporating weather, satellite imagery, topography, and human-related variables, similar to this study’s approach, reinforcing the effectiveness of EO data cubes in wildfire prediction.

In terms of methodology, the approach taken in this study—training models on data from 2021 to 2023 and validating on 2024—ensured that temporal generalization was assessed. This method is similar to the temporal split approach used in Kondylatos et al. 2022, where models trained on past years were tested on extreme fire seasons. However, Prapas et al. 2021, employed U-Net++ architectures for wildfire segmentation, incorporated additional spatial constraints by using weekly mean seasonal cycles as baselines. A potential improvement in future work would be the integration of similar baseline comparisons to contextualize seasonal fire variations more effectively.

An important distinction in this study compared to others is the inclusion of static and dynamic variables, with feature importance analysis highlighting land surface temperature (LST), temperature at 2m, and human modification (gHM) as key predictors. Previous studies, such as Ji et al. 2024, found that the use of vision transformers (ViT) and the incorporation of static and dynamic data improve the performance of the models. However, in this study, the distinction between anthropogenic, meteorological, and fuel-related variables played a significant role in explaining model predictions. Future research could explore the integration of additional static environmental variables, such as soil moisture indices, to further refine the models’ ability to capture fire risk in diverse ecosystems.

In general, this study demonstrates that deep learning and traditional machine learn-

ing approaches are effective in wildfire prediction, but model selection depends on the trade-off between precision and recall. The findings reinforce previous research emphasizing the importance of spatiotemporal data, but also highlight areas where additional refinements—such as ensemble models or explainable AI techniques—could enhance predictive reliability.

6 Conclusion

This study explored the potential of Earth Observation (EO) data cubes for wildfire risk prediction by implementing deep learning models and integrating both static and dynamic variables. The findings highlight the efficiency of EO data cubes in handling large-scale spatiotemporal datasets, the suitability of various deep learning architectures for wildfire prediction, and the importance of combining static and dynamic predictors to enhance model performance.

The utilization of EO data cubes proved to be an effective approach for processing and analyzing wildfire-related data. The ability to access and manipulate vast amounts of Earth observation data in a structured manner significantly streamlined the extraction of variables across space and time. By integrating Google Earth Engine (GEE) APIs, the study was able to efficiently process historical and real-time data, avoiding the computational and storage limitations commonly associated with traditional methods. This approach not only facilitated data acquisition but also enabled the creation of a scalable and automated workflow for wildfire risk prediction. The results demonstrate that EO data cubes offer a practical solution for managing high-dimensional datasets and performing near real-time fire risk assessments.

In terms of algorithm selection, the study evaluated Random Forest (RF), Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), and Convolutional LSTM (ConvLSTM) models based on their ability to predict wildfire risk. The LSTM model exhibited a balanced performance across all metrics, indicating its strength in capturing temporal dependencies in wildfire occurrence. CNNs showed higher sensitivity but were prone to false positives, while ConvLSTM did not significantly outperform LSTM, likely due to limitations in patch size and the absence of explicit spatial features. These results align with findings in the literature, where LSTM and ConvLSTM models have been recognized for their ability to capture spatiotemporal patterns. The study confirms that the choice of model is influenced by the nature of the input data—whether it emphasizes temporal sequences, spatial context, or both.

The integration of static and dynamic variables played a crucial role in improving prediction accuracy. Feature importance revealed that human modification (gHM), land surface temperature (LST), and temperature at 2m were the most influential predictors, demonstrating the combined effect of anthropogenic and environmental drivers on wildfire risk. The analysis of variable distributions further supported these findings, showing that fire-prone areas had distinct temperature, vegetation, and precipitation characteristics compared to non-fire regions. The results indicate that dynamic variables, such as temperature and precipitation, are essential for capturing seasonal variability in fire risk, while static variables, such as elevation and human

modification, provide contextual information that enhances model generalization.

By applying the developed models across different regions and seasons in South America, the study demonstrated the ability of deep learning algorithms to adapt to diverse climatic conditions. The spatial predictions successfully distinguished fire-prone areas from low-risk zones, with variations in model performance reflecting differences in environmental factors. The comparison of fire risk maps for different seasons highlighted the influence of precipitation and vegetation conditions on model sensitivity, reinforcing the importance of incorporating both static and dynamic inputs.

Overall, this research contributes to wildfire risk prediction by integrating EO data cubes with deep learning models in a way that balances computational efficiency and predictive accuracy. The findings underscore the potential of spatiotemporal data-driven approaches for improving early warning systems and informing fire management strategies. Future work could explore the inclusion of additional static predictors, the refinement of spatiotemporal model architectures, and the application of explainable AI techniques to enhance model interpretability. By advancing the integration of remote sensing and deep learning, this study lays the foundation for more effective and scalable wildfire danger prediction systems.

References

- Ager, A. A. et al. (July 2014). “Wildfire risk estimation in the Mediterranean Area”. In: *Environmetrics* 25.6, pp. 384–396. DOI: [10.1002/env.2269](https://doi.org/10.1002/env.2269).
- Alto, Palo (June 2024). *Wildfire Tool Inventory and Evaluation*. URL: <https://apps.epri.com/wildfire-tool-inventory/en/citation.html>.
- Bilidas, Dimitris et al. (Nov. 2023). “Fire Risk Management using Data Cubes, Machine Learning and OBDA systems”. In: *GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems*. Association for Computing Machinery. ISBN: 9798400701689. DOI: [10.1145/3589132.3625615](https://doi.org/10.1145/3589132.3625615).
- Chen, Weijie et al. (Jan. 2021). “Wildfire risk assessment of transmission-line corridors based on naïve bayes network and remote sensing data”. In: *Sensors (Switzerland)* 21 (2), pp. 1–16. ISSN: 14248220. DOI: [10.3390/s21020634](https://doi.org/10.3390/s21020634).
- Chen, Yang, Douglas C. Morton, and James T. Randerson (June 2024). *Remote sensing for wildfire monitoring: Insights into burned area, emissions, and fire dynamics*. DOI: [10.1016/j.oneear.2024.05.014](https://doi.org/10.1016/j.oneear.2024.05.014).
- Collins, L. et al. (Oct. 2023). “Fuel reduction burning reduces wildfire severity during extreme fire events in south-eastern Australia”. In: *Journal of Environmental Management* 343. ISSN: 10958630. DOI: [10.1016/j.jenvman.2023.118171](https://doi.org/10.1016/j.jenvman.2023.118171).
- Copernicus (2024). *Effis - Welcome to EFFIS*. URL: <https://forest-fire.emergency.copernicus.eu/>.
- Department, Malaysian Meteorological (2024). *Laman Web Rasmi Jabatan meteorologist Malaysia*. URL: <https://www.met.gov.my/en/iklim/fdrs-asean/>.
- Dios, Víctor Resco de et al. (Apr. 2015). “A semi-mechanistic model for predicting the moisture content of fine litter”. In: *Agricultural and Forest Meteorology* 203, pp. 64–73. ISSN: 01681923. DOI: [10.1016/j.agrformet.2015.01.002](https://doi.org/10.1016/j.agrformet.2015.01.002).
- Earth Resources Observation and Science (EROS) Center (2018). *USGS eros archive - digital elevation - shuttle radar topography mission (SRTM) 1 ARC-second global*. URL: https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-shuttle-radar-topography-mission-srtm-1?qt-science_center_objects=0#qt-science_center_objects.
- EFI’s Mediterranean Facility (EFIMED) (2024). *Wildfire prevention*. URL: <https://efi.int/wildfire-prevention>.
- Escuin, S., R. Navarro, and P. Fernández (Feb. 2008). “Fire severity assessment by using NBR (Normalized Burn Ratio) and NDVI (Normalized Difference Vegetation Index) derived from LANDSAT TM/ETM images”. In: *International Journal of Remote Sensing* 29 (4), pp. 1053–1073. ISSN: 13665901. DOI: [10.1080/01431160701281072](https://doi.org/10.1080/01431160701281072).
- Forum, World Economic (n.d.). *The power of AI in wildfire prediction and prevention*. URL: <https://www.weforum.org/impact/artificial-intelligence-wildfire-prediction-and-prevention/>.

- Friedl, M. and Sulla-Menashe (2019). “MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V006”. In: *NASA EOSDIS Land Processes Distributed Active Archive Center*. DOI: <https://doi.org/10.5067/MODIS/MCD12Q1.006>.
- Gharakhanlou, Navid Mahdizadeh and Navid Hooshangi (June 2021). “Dynamic simulation of fire propagation in forests and Rangelands using a GIS-based cellular automata model”. In: *International Journal of Wildland Fire* 30.9, pp. 652–663. DOI: [10.1071/wf20098](https://doi.org/10.1071/wf20098).
- Gheshlaghi, Hassan Abedi, Bakhtiar Feizizadeh, and Thomas Blaschke (Feb. 2020). “GIS-based forest fire risk mapping using the analytical network process and fuzzy logic”. In: *Journal of Environmental Planning and Management* 63 (3), pp. 481–499. ISSN: 13600559. DOI: [10.1080/09640568.2019.1594726](https://doi.org/10.1080/09640568.2019.1594726).
- Giglio, L et al. (2015). *MCD64A1 MODIS/Terra+Aqua Burned Area Monthly L3 Global 500m SIN Grid V006*. URL: <https://doi.org/10.5067/MODIS/MCD64A1.006>.
- Giuliani, Gregory et al. (2020). *Earth observation data cubes*. MDPI - Multidisciplinary Digital Publishing Institute.
- Goldarag, Y. Jafari, Ali Mohammadzadeh, and A. S. Ardakani (Dec. 2016). “Fire Risk Assessment Using Neural Network and Logistic Regression”. In: *Journal of the Indian Society of Remote Sensing* 44 (6), pp. 885–894. ISSN: 09743006. DOI: [10.1007/s12524-016-0557-6](https://doi.org/10.1007/s12524-016-0557-6).
- Gorelick, Noel et al. (2017). “Google Earth Engine: Planetary-scale geospatial analysis for everyone”. In: *Remote Sensing of Environment* 202. Big Remotely Sensed Data: tools, applications and experiences, pp. 18–27. ISSN: 0034-4257. DOI: <https://doi.org/10.1016/j.rse.2017.06.031>. URL: <https://www.sciencedirect.com/science/article/pii/S0034425717302900>.
- Hausfather, Zeke (July 2024). *State of the climate: 2024 now very likely to be warmest year on record*. URL: <https://www.carbonbrief.org/state-of-the-climate-2024-now-very-likely-to-be-warmest-year-on-record/>.
- Haynes, Brad (2024). *Brazil fire brigades fight Amazon blazes off to torrid start in 2024 — reuters*. URL: <https://www.reuters.com/business/environment/brazil-fire-brigades-fight-amazon-blazes-off-torrid-start-2024-2024-08-09/>.
- Horn, K. H. et al. (2025). “Modelling current and future forest fire susceptibility in north-eastern Germany”. In: *Natural Hazards and Earth System Sciences* 25.1, pp. 383–401. DOI: [10.5194/nhess-25-383-2025](https://doi.org/10.5194/nhess-25-383-2025). URL: <https://nhess.copernicus.org/articles/25/383/2025/>.
- Hoyer, S. and J. Hamman (2017). “xarray: N-D labeled arrays and datasets in Python”. In: *Journal of Open Research Software* 5.1. DOI: [10.5334/jors.148](https://doi.org/10.5334/jors.148). URL: <https://doi.org/10.5334/jors.148>.
- Huete, Alfredo, Chris Justice, and Wim van Leeuwen (1999). *MODIS VEGETATION INDEX (MOD 13): ALGORITHM THEORETICAL BASIS DOCUMENT*. URL: https://lpdaac.usgs.gov/documents/104/MOD13_ATBD.pdf.

- Huot, Fantine et al. (2020). “Deep Learning Models for Predicting Wildfires from Historical Remote-Sensing Data”. In: *34th Conference on Neural Information Processing Systems*. NeurIPS 2020. Vancouver, Canada.
- Jaafari, Abolfazl, Davood Mafi Gholami, and Eric K. Zenner (May 2017). “A Bayesian modeling of wildfire probability in the Zagros Mountains, Iran”. In: *Ecological Informatics* 39, pp. 32–44. ISSN: 15749541. DOI: [10.1016/j.ecoinf.2017.03.003](https://doi.org/10.1016/j.ecoinf.2017.03.003).
- Jaafari, Abolfazl et al. (June 2022). “Anthropogenic activities amplify wildfire occurrence in the Zagros eco-region of Western Iran”. In: *Natural Hazards* 114.1, pp. 457–473. DOI: [10.1007/s11069-022-05397-6](https://doi.org/10.1007/s11069-022-05397-6).
- Janiec, Piotr and Sébastien Gadal (Dec. 2020). “A comparison of two machine learning classification methods for remote sensing predictive modeling of the forest fire in the north-eastern Siberia”. In: *Remote Sensing* 12 (24), pp. 1–20. ISSN: 20724292. DOI: [10.3390/rs12244157](https://doi.org/10.3390/rs12244157).
- Ji, Yuheng et al. (Jan. 2024). “Global Wildfire Danger Predictions Based on Deep Learning Taking into Account Static and Dynamic Variables”. In: *Forests* 15 (1). ISSN: 19994907. DOI: [10.3390/f15010216](https://doi.org/10.3390/f15010216).
- Jolly, W. Matt et al. (July 2015). “Climate-induced variations in global wildfire danger from 1979 to 2013”. In: *Nature Communications* 6.1. ISSN: 2041-1723. DOI: [10.1038/ncomms8537](https://doi.org/10.1038/ncomms8537). URL: <http://dx.doi.org/10.1038/ncomms8537>.
- Knyazikhin, Y. et al. (1999). *MODIS Leaf Area Index (LAI) And Fraction Of Photosynthetically Active Radiation Absorbed By Vegetation (FPAR) Product*. URL: <http://eospsso.gsfc.nasa.gov/atbd/modistables.html>.
- Kondylatos, Spyros et al. (Sept. 2022). “Wildfire Danger Prediction and Understanding With Deep Learning”. In: *Geophysical Research Letters* 49 (17). ISSN: 19448007. DOI: [10.1029/2022GL099368](https://doi.org/10.1029/2022GL099368).
- Lin, Xufeng et al. (Apr. 2023). “Forest fire prediction based on long- and short-term time-series network”. In: *Forests* 14.4, p. 778. DOI: [10.3390/f14040778](https://doi.org/10.3390/f14040778).
- Mandala Partners (2024). *A Global Sampling of Wildfire Economic Impact: The Benefits of A Dedicated Satellite Constellation to Detect and Monitor Wildfires*. Tech. rep. Earth Fire Alliance.
- Miller, Jay D. and Andrea E. Thode (July 2007). “Quantifying burn severity in a heterogeneous landscape with a relative version of the delta Normalized Burn Ratio (dNBR)”. In: *Remote Sensing of Environment* 109 (1), pp. 66–80. ISSN: 00344257. DOI: [10.1016/j.rse.2006.12.006](https://doi.org/10.1016/j.rse.2006.12.006).
- Mu, Qiaozhen, Maosheng Zhao, and Steven W. Running (2013). *MODIS Global Terrestrial Evapotranspiration (ET) Product(NASA MOD16A2/A3)*. URL: https://lpdaac.usgs.gov/documents/93/MOD16_ATBD.pdf.
- Myneni, Ranga, Yuri Knyazikhin, and Nicolay Shabanov (2010). “Leaf area index and fraction of absorbed par products from Terra and Aqua Modis sensors: Analysis, validation,

- and refinement”. In: *Remote Sensing and Digital Image Processing*, pp. 603–633. DOI: [10.1007/978-1-4419-6749-7_27](https://doi.org/10.1007/978-1-4419-6749-7_27).
- NASA Earth Observatory (2023). *Drought fuels wildfires in the Amazon*. URL: https://earthobservatory.nasa.gov/images/151965/drought-fuels-wildfires-in-the-amazon?utm_source=chatgpt.com.
- National Geographic (Oct. 2023). URL: <https://education.nationalgeographic.org/resource/south-america-physical-geography/>.
- Natural Resources Canada (n.d.). *Canadian wildland fire information system: Data sources and methods for daily maps*. URL: <https://cwfis.cfs.nrcan.gc.ca/background/dsm/fm3>.
- Nikhil, S et al. (2021). “Application of GIS and AHP Method in Forest Fire Risk Zone Mapping: a Study of the Parambikulam Tiger Reserve, Kerala, India”. In: *Elsevier*. DOI: [10.1007/s41651-021-00082-x](https://doi.org/10.1007/s41651-021-00082-x). URL: <https://doi.org/10.1007/s41651-021-00082-x>.
- Ntinopoulos, Nikolaos et al. (Aug. 2023). “Fusion of Remotely-Sensed Fire-Related Indices for Wildfire Prediction through the Contribution of Artificial Intelligence”. In: *Sustainability (Switzerland)* 15 (15). ISSN: 20711050. DOI: [10.3390/su15151527](https://doi.org/10.3390/su15151527).
- Oliveira, Sandra, Jorge Rocha, and Ana Sá (2021). “Wildfire risk modeling”. In: *Current Opinion in Environmental Science Health* 23, p. 100274. ISSN: 2468-5844. DOI: <https://doi.org/10.1016/j.coesh.2021.100274>. URL: <https://www.sciencedirect.com/science/article/pii/S2468584421000465>.
- OroraTech (June 2023). *Wildfire solution*. URL: <https://ororatech.com/wildfire-solution/>.
- Pelletier, N., K. Millard, and S. Darling (Oct. 2023). “Wildfire likelihood in Canadian treed peatlands based on remote-sensing time-series of surface conditions”. In: *Remote Sensing of Environment* 296. ISSN: 00344257. DOI: [10.1016/j.rse.2023.113747](https://doi.org/10.1016/j.rse.2023.113747).
- Piralilou, Sepideh Tavakkoli et al. (Feb. 2022). “A Google Earth Engine Approach for Wildfire Susceptibility Prediction Fusion with Remote Sensing Data of Different Spatial Resolutions”. In: *Remote Sensing* 14 (3). ISSN: 20724292. DOI: [10.3390/rs14030672](https://doi.org/10.3390/rs14030672).
- Pletsch, Mikhaela A.J.S. et al. (July 2022). “Near Real-Time Fire Detection and Monitoring in the MATOPIBA Region, Brazil”. In: *Remote Sensing* 14 (13). ISSN: 20724292. DOI: [10.3390/rs14133141](https://doi.org/10.3390/rs14133141).
- Pondi, Brian, Marius Appel, and Edzer Pebesma (Apr. 2024). “OpenEOcubes: an open-source and lightweight R-based RESTful web service for analyzing earth observation data cubes”. In: *Earth Science Informatics* 17 (2), pp. 1809–1818. ISSN: 18650481. DOI: [10.1007/s12145-024-01249-y](https://doi.org/10.1007/s12145-024-01249-y).
- Pourtaghi, Zohre Sadat et al. (May 2016). “Investigation of general indicators influencing on forest fire and its susceptibility modeling using different data mining techniques”. In:

- Ecological Indicators* 64, pp. 72–84. ISSN: 1470160X. DOI: [10.1016/j.ecolind.2015.12.030](https://doi.org/10.1016/j.ecolind.2015.12.030).
- Prapas, Ioannis et al. (2021). “Deep Learning Methods for Daily Wildfire Danger Forecasting”. In: *35th Conference on Neural Information Processing Systems*. NeurIPS 2021. Sydney, Australia.
- Prapas, Ioannis et al. (2022). “Deep Learning for Global Wildfire Forecasting”. In: *35th Conference on Neural Information Processing Systems*. NeurIPS 2022.
- Prasad, V. Krishna, K. V.S. Badarinath, and Anuradha Eaturu (2008). “Biophysical and anthropogenic controls of forest fires in the Deccan Plateau, India”. In: *Journal of Environmental Management* 86 (1), pp. 1–13. ISSN: 03014797. DOI: [10.1016/j.jenvman.2006.11.017](https://doi.org/10.1016/j.jenvman.2006.11.017).
- Quan, Xingwen et al. (2021). “Application of Landsat ETM+ and OLI Data for Foliage Fuel Load Monitoring Using Radiative Transfer Model and Machine Learning Method”. In: *IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING* 14. DOI: [10.1109/JSTARS](https://doi.org/10.1109/JSTARS.2021.3095219). URL: [https://doi.org/10.1109/JSTARS](https://doi.org/10.1109/JSTARS.2021.3095219).
- Resco de Dios, Víctor (2020). *Plant-fire interactions: Applying ecophysiology to wildfire management*. Springer.
- Rezaie, Fatemeh et al. (2023). “Development of novel optimized deep learning algorithms for wildfire modeling: A case study of Maui, Hawai‘i”. In: *Engineering Applications of Artificial Intelligence* 125, p. 106699. ISSN: 0952-1976. DOI: <https://doi.org/10.1016/j.engappai.2023.106699>. URL: <https://www.sciencedirect.com/science/article/pii/S0952197623008837>.
- Root, Tik (Sept. 2024). *No one should be surprised that South America is burning*. URL: <https://grist.org/wildfires/no-one-should-be-surprised-that-south-america-is-burning/>.
- Rösch, Moritz et al. (2024). “Data-Driven Wildfire Spread Modeling of European Wildfires Using a Spatiotemporal Graph Neural Network”. In: *Fire* 7.6. ISSN: 2571-6255. DOI: [10.3390/fire7060207](https://doi.org/10.3390/fire7060207). URL: <https://www.mdpi.com/2571-6255/7/6/207>.
- Roy, David P, Luigi Boschetti, and Chris Justice (2006). *Global mapping of fire-affected areas using multitemporal MODIS data: the MCD45 product*.
- Schroeder, Wilfrid et al. (Nov. 2016). “Active fire detection using Landsat-8/OLI data”. In: *Remote Sensing of Environment* 185, pp. 210–220. ISSN: 00344257. DOI: [10.1016/j.rse.2015.08.032](https://doi.org/10.1016/j.rse.2015.08.032).
- Scull, Caitlin (July 2023). *Alertcalifornia launches to provide essential tools to understand and adapt to wildfires and natural disasters in the Golden State*. URL: <https://alertcalifornia.org/alertcalifornia-launches-to-provide-essential-tools-to-understand-and-adapt-to-wildfires-and-natural-disasters-in-the-golden-state/>.

- Shao, Yakui et al. (June 2022). “Mapping China’s Forest Fire Risks with Machine Learning”. In: *Forests* 13 (6). ISSN: 19994907. DOI: [10.3390/f13060856](https://doi.org/10.3390/f13060856).
- Strahler, Alan et al. (1999). *MODIS Land Cover Product Algorithm Theoretical Basis Document (ATBD)*. URL: https://lpdaac.usgs.gov/documents/86/MCD12_ATBD.pdf.
- Szpakowski, David M. and Jennifer L.R. Jensen (Nov. 2019). *A review of the applications of remote sensing in fire ecology*. DOI: [10.3390/rs11222638](https://doi.org/10.3390/rs11222638).
- Talukdar, Nazimur Rahman et al. (June 2024). “Forest fire estimation and risk prediction using multispectral satellite images: Case study”. In: *Natural Hazards Research* 4 (2), pp. 304–319. ISSN: 26665921. DOI: [10.1016/j.nhres.2024.01.007](https://doi.org/10.1016/j.nhres.2024.01.007).
- The Knowledge Network (2024). *Germany’s most fire-prone region undergoes a UCPM Wildfire Peer Review*. URL: <https://civil-protection-knowledge-network.europa.eu/stories/germanys-most-fire-prone-region-undergoes-ucpm-wildfire-peer-review>.
- Theobald, David M. et al. (Sept. 2020). “Earth transformed: Detailed mapping of Global Human Modification from 1990 to 2017”. In: *Earth System Science Data* 12.3, pp. 1953–1972. DOI: [10.5194/essd-12-1953-2020](https://doi.org/10.5194/essd-12-1953-2020).
- Thomas, Grangeon et al. (July 2021). “Modelling forest fire and firebreak scenarios in a mediterranean mountainous catchment: Impacts on sediment loads”. In: *Journal of Environmental Management* 289. ISSN: 10958630. DOI: [10.1016/j.jenvman.2021.112497](https://doi.org/10.1016/j.jenvman.2021.112497).
- United Nations Environment Programme (2022). *Spreading like Wildfire – The Rising Threat of Extraordinary Landscape Fires*. Tech. rep. UNEP.
- United Nations Office for Disaster Risk Reduction (UNDRR) (June 2022). URL: <https://www.undrr.org/publication/wildfires-latin-america-preliminary-analysis-messages-and-resources>.
- World Resources Institute (2024). *Global Forest Watch*. URL: <https://www.wri.org/initiatives/global-forest-watch>.
- Xu, Xiyan et al. (2020). “Climate regime shift and forest loss amplify fire in Amazonian forests”. In: *Global Change Biology* 26.10, pp. 5874–5885. DOI: <https://doi.org/10.1111/gcb.15279>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/gcb.15279>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/gcb.15279>.
- Xu, Zhengsen et al. (May 2024). “Wildfire Risk Prediction: A Review”. In: *Elsevier*. URL: <https://arxiv.org/abs/2405.01607>.
- Yahia, Oualid, Mohammed Ghabi, and Moussa Sofiane Karoui (2023). “The Prediction of Regional Wildfire Risk Using High-Resolution Remotely Sensed Soil Moisture Content Estimation, Case Study: Sidi Douma Forest, Saida, Algeria”. In: *International Geoscience and Remote Sensing Symposium (IGARSS)*. Vol. 2023-July. Institute of Electrical and Electronics Engineers Inc., pp. 3387–3390. ISBN: 9798350320107. DOI: [10.1109/IGARSS52108.2023.10281986](https://doi.org/10.1109/IGARSS52108.2023.10281986).

- Yang, Jian, Hong S. He, and Eric J. Gustafson (Dec. 2004). “A hierarchical fire frequency model to simulate temporal patterns of fire regimes in LANDIS”. In: *Ecological Modelling* 180 (1), pp. 119–133. ISSN: 03043800. DOI: [10.1016/j.ecolmodel.2004.03.017](https://doi.org/10.1016/j.ecolmodel.2004.03.017).
- You, Xingyue et al. (Dec. 2023). “A PSO-CNN-based deep learning model for predicting forest fire risk on a national scale”. In: *Forests* 15.1, p. 86. DOI: [10.3390/f15010086](https://doi.org/10.3390/f15010086).
- Yue, Peng et al. (Sept. 2020). “Geocube: Towards the Multi-Source Geospatial Data Cube in Big Data Era”. In: *International Geoscience and Remote Sensing Symposium (IGARSS)*. Institute of Electrical and Electronics Engineers Inc., pp. 3127–3130. ISBN: 9781728163741. DOI: [10.1109/IGARSS39084.2020.9324632](https://doi.org/10.1109/IGARSS39084.2020.9324632).
- Zhang, Guoli, Ming Wang, and Kai Liu (Sept. 2019). “Forest Fire Susceptibility Modeling Using a Convolutional Neural Network for Yunnan Province of China”. In: *International Journal of Disaster Risk Science* 10 (3), pp. 386–403. ISSN: 21926395. DOI: [10.1007/s13753-019-00233-1](https://doi.org/10.1007/s13753-019-00233-1).
- Zhengming, Wan (1999). *MODIS Land-Surface Temperature Algorithm Theoretical Basis Document (LST ATBD)*. URL: <https://www.scirp.org/reference/referencespapers?referenceid=1146314>.

A Appendix

A.1 AI Use Declaration

To ensure transparency and uphold academic integrity, the following table outlines the artificial intelligence (AI) tools utilized during the preparation of this report.

Table 4: AI Use Declaration

AI-Based Tool	Use Case	Scope	Remarks
ChatGPT 4.0 Last access: 19.02.2024	Paraphrasing, correct lan- guage, find synonyms	Entire work	Code debugging

A.2 Web Application documentation

- Github Repository: <https://github.com/jhonyasv/Wildfire-Risk-App.git>
- Google Colab storage: <https://drive.google.com/drive/folders/1b642Ns21oNN174mA4SWGLawvEp21B1?usp=sharing>
- Hosted Webapp: <https://wildfire-risk-app.streamlit.app/>