Comprehensive Fire Detection and Prediction System Utilizing Numerical Data and Satellite Imagery for Enhanced Wildfire Monitoring

Muhammad Umair Imran[1], Shahzad Waris[2], Muhammad Bilal[3], Hafiz Muhammad Hafiz[4]

Wildfires pose a significant environmental and economic threat, making accurate prediction and analysis crucial for disaster management. This study constructs a comprehensive wildfire dataset by extracting 64km × 64km patches centered on fire ignition points and analyzing various contributing factors, including atmospheric conditions, vegetation health, temperature variations, human activities, and land features. Using a U-Net segmentation model with an EfficientNet-B1 encoder, we train and evaluate fire spread prediction based on historical data. Preprocessing steps such as handling missing values, class imbalance, and feature scaling enhance model performance. The temporal analysis reveals key trends in wildfire occurrences, while visualization techniques provide insights into the spatial distribution of fire events. This work contributes to the development of more effective wildfire forecasting models, aiding in early detection and mitigation efforts.

Introduction

Wildfires are vital to ecosystems but threaten lives, infrastructure, and the environment. Climate change has increased their frequency and intensity, making early detection crucial.

Traditional models struggle with predictive accuracy due to complex environmental factors. ML and DL offer data-driven improvements but face challenges like wildfire randomness, data imbalance, and limited datasets.

This study proposes an ML-based Fire Detection and Prediction System integrating numerical data and satellite imagery to enhance monitoring and forecasting. It tackles dataset fragmentation, multi-source data integration, and the need for robust models.

A. Motivation

Despite advancements in machine learning (ML), wildfire prediction remains challenging due to fragmented data sources and the complex interplay of environmental factors. This project aims to develop a comprehensive ML-driven framework that integrates satellite imagery, numerical data, and spatiotemporal analysis to enhance fire detection and forecasting. By analyzing wildfire causes in relation to key landmarks, human activities, and meteorological conditions, our approach seeks to improve risk assessment and early warning systems, particularly in the Mediterranean region.

	aspect	burned_areas	curvature	d2m	dem	ignition_points	ndvi	rh	roads_distance	slope	...	month	day	day_of_year	week_
0	-0.023376	-0.087595	-0.152924	0.749859	-1.210648	-0.011168	-1.306742	-0.838179	-0.457263	-0.322125	...	7	29	210	
1	-0.023376	-0.087595	-0.152924	0.806068	-1.210648	-0.011168	-1.306742	-0.837083	-0.457263	-0.322125	...	7	30	211	
2	-0.023376	-0.087595	-0.152924	0.822762	-1.210648	-0.011168	-1.306742	-0.395334	-0.457263	-0.322125	...	7	31	212	
3	-0.023376	-0.087595	-0.152924	0.828469	-1.210648	-0.011168	-1.306742	-0.649035	-0.457263	-0.322125	...	8	1	213	
4	-0.023376	-0.087595	-0.152924	0.854087	-1.210648	-0.011168	-1.306742	-0.260139	-0.457263	-0.322125	...	8	2	214	

slope	...	month	day	day_of_year	week_of_year	quarter	month_sin	month_cos	day_of_year_sin	day_of_year_cos	target
.322125	...	7	29	210	30	3	-0.500000	-0.866025	-0.455907	-0.890028	1
.322125	...	7	30	211	30	3	-0.500000	-0.866025	-0.471160	-0.882048	1
.322125	...	7	31	212	31	3	-0.500000	-0.866025	-0.486273	-0.873807	1
.322125	...	8	1	213	31	3	-0.866025	-0.500000	-0.501242	-0.865307	1
.322125	...	8	2	214	31	3	-0.866025	-0.500000	-0.516062	-0.856551	1

Fig. 1. First Rows of dataset

B. Dataset Description

This study utilizes a multi-source dataset combining satellite imagery, numerical climate data, and human activity records to improve wildfire prediction. Each dataset provides unique insights into fire occurrence and propagation patterns.

1. Satellite Data. We utilize NASA satellite imagery, incorporating multi-spectral and thermal bands to analyze wildfire-prone regions. Key indices such as the Normalized Difference Vegetation Index (NDVI), Land Surface Temperature (LST), and Burned Area Index (BAI) provide essential information on vegetation stress, surface heating, and fire extent. These datasets enable real-time monitoring and historical fire pattern analysis.

2. Numerical Climate Data. The dataset integrates satellite imagery, numerical meteorological data, and land cover information to provide a comprehensive view of wildfire dynamics. High-resolution satellite data, including MODIS-derived NDVI, LAI, and land surface temperature, help assess vegetation health and fire susceptibility, while EFFIS-enhanced burned area maps and ignition points improve fire event tracking. Numerical data from ERA5-Land captures key meteorological variables such as temperature, humidity, wind speed, surface pressure, and precipitation, alongside soil moisture and solar radiation, which influence fire ignition and spread. Additionally, land cover data from Copernicus CCS details the distribution of vegetation, settlements, and water bodies, while static variables such as elevation, slope, aspect, curvature, and road distances provide critical geographical context. Together, these datasets enable accurate wildfire prediction and risk assessment by integrating diverse environmental and anthropogenic factors.

3. Temporal Dataset for fire detection. Building on the extracted 64km × 64km patches centered on fire ignition points, a refined dataset is created for training a segmentation model. During training, these patches are randomly cropped

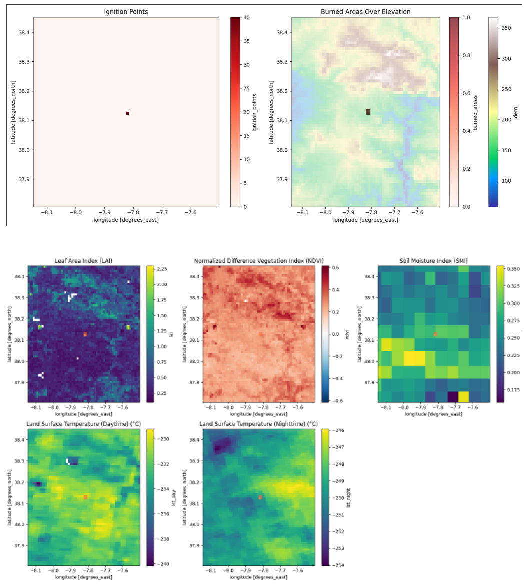

Fig. 2. Some Sample Images from dataset for one particular longitude latitude.

to 32 × 32 to maintain the ignition point within the frame while minimizing bias toward fire expansion. The dataset is processed using a U-Net architecture with an EfficientNet-B1 encoder, where multiple input variables are stacked as separate channels. Models are trained using cross-entropy loss, with an additional baseline model utilizing only ignition points for comparison. A temporal split ensures no data leakage, with samples from 2006–2019 used for training, 2020 for validation, and 2021–2022 for testing. Input variables are normalized within the range [0,1], and model evaluation is conducted using cross-entropy loss and AUPRC metrics. This dataset is crucial for improving segmentation-based wildfire prediction, leveraging historical fire data to enhance forecasting accuracy.

Binary Classification Fire Prediction

This section discusses the preprocessing and exploratory data analysis (EDA) steps involved in preparing the dataset for binary classification fire prediction. The preprocessing steps include handling missing values, addressing class imbalance, formatting date-time variables, and performing feature scaling to ensure a well-processed dataset for model training.

Preprocessing Steps. **1. Handling Missing Values:** All null values in the dataset are removed to prevent inconsistencies during model training. The distribution of missing values before removal is shown in the figure below.

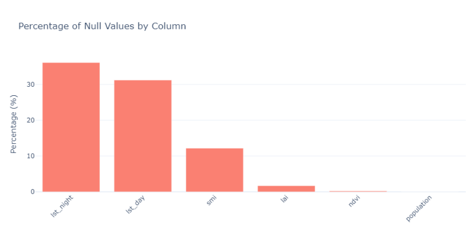

Fig. 3. Distribution of missing values in the dataset

2. Addressing Class Imbalance: Since wildfire occurrence data is often imbalanced, techniques such as oversampling, undersampling, or synthetic data generation (e.g., SMOTE) are applied to balance the class distribution. The distribution of class labels before applying balancing techniques is illustrated below.

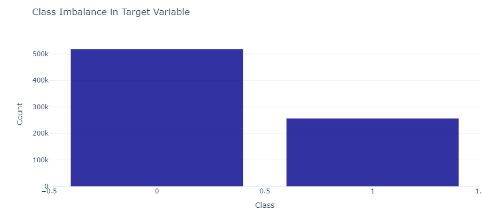

Fig. 4. Class label distribution in the dataset

3. Date-Time Formatting: The dataset contains timestamped fire event records, which are standardized to a uniform date-time format for consistency in time-series modeling.

4. Feature Scaling: To ensure all numerical features contribute proportionally to the model, feature scaling is applied using min-max normalization or standardization, depending on the model requirements.

These preprocessing steps enhance data quality and ensure better predictive performance for the binary classification task.

Visualization Mediterrain

This section presents key visualizations to analyze fire occurrence trends, spatial distribution, and intensity patterns. These insights help understand the temporal and geographical characteristics of wildfires.

Temporal Trends (Burned Areas Over Time). The figure below illustrates the yearly trend of burned areas, providing insights into how wildfire occurrences have evolved over time.

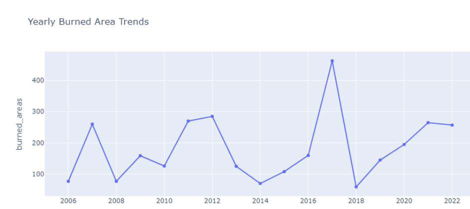

Fig. 5. Yearly trend of burned areas in the dataset

Correlation Heatmap. A heatmap is used to visualize correlations between various features in the dataset. This helps identify key variables that influence wildfire occurrences.

Fire Intensity Heatmap. The following heatmap represents the spatial intensity of fires, highlighting areas with higher wildfire activity.

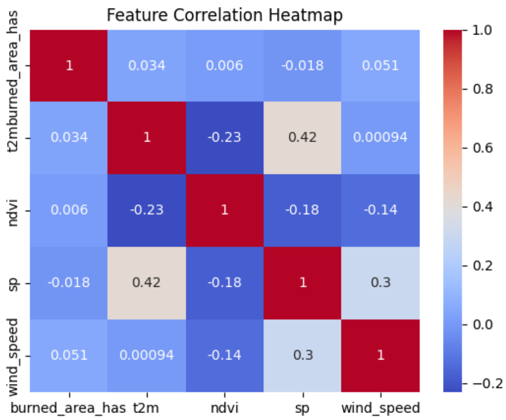

Fig. 6. Feature correlation heatmap

Fig. 7. Fire intensity heatmap

Monthly Fire Trends. This visualization captures the monthly distribution of wildfire occurrences, revealing seasonal patterns in fire activity.

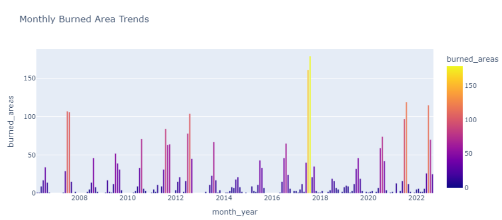

Fig. 8. Monthly trend of wildfire occurrences

Wildfire Locations. The following map displays the spatial distribution of wildfires, showing where fire events are concentrated.

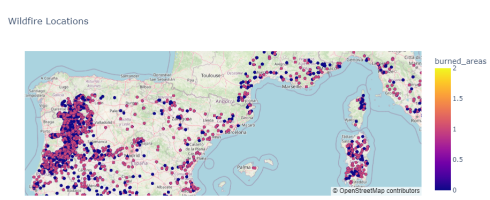

Fig. 9. Geographical distribution of wildfire locations

Visualization from Temporal Data for Fire Spread Analysis

This section presents key visualizations derived from temporal data to analyze fire spread patterns. Various factors influencing wildfire behavior, including atmospheric conditions, vegetation health, temperature, human activities, and land characteristics, are examined.

Atmospheric Conditions at Ignition. Wildfire ignition is significantly influenced by atmospheric factors such as wind speed, humidity, and air pressure. The following figure visualizes the prevailing atmospheric conditions at the time of fire ignition.

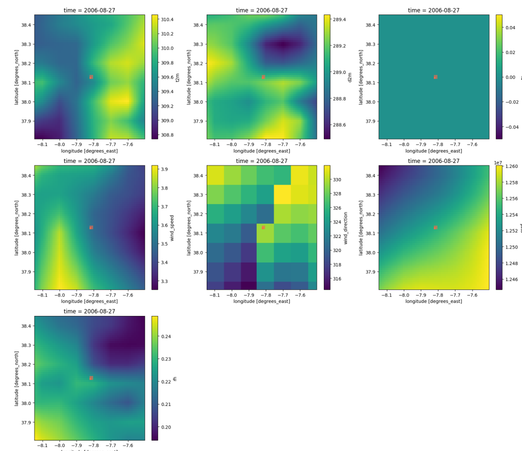

Fig. 10. Atmospheric conditions at wildfire ignition

Vegetation Health and Moisture. Vegetation plays a critical role in wildfire spread. Healthier, moisture-rich vegetation reduces fire risk, while dry, stressed vegetation accelerates fire propagation. The figure below illustrates vegetation indices related to wildfire occurrences.

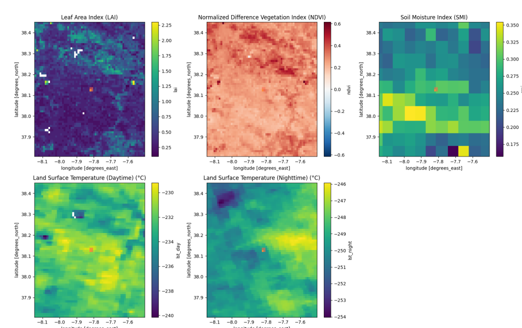

Fig. 11. Vegetation health and moisture levels in fire-prone areas

Temperature Influence. High temperatures increase wildfire risk by drying out vegetation and reducing soil moisture. The following figure represents temperature variations across wildfire-affected regions.

Human Influence Factors. Human activities such as land use changes, deforestation, and infrastructure development can significantly impact wildfire occurrences. This figure highlights human-related factors contributing to fire risks.

Land Features and Fire Susceptibility. Topographical and land cover features, including elevation, slope, and proximity to roads, affect wildfire behavior. The visualization

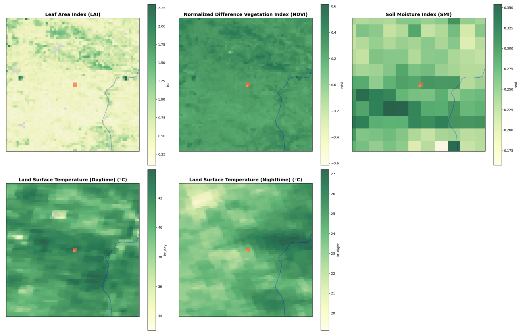

Fig. 12. Temperature variations in wildfire-prone regions

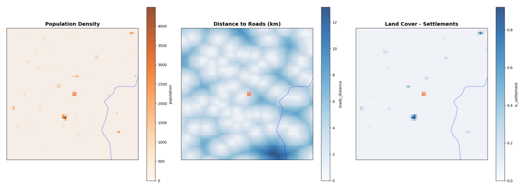

Fig. 13. Impact of human activities on wildfire occurrences

below represents key land characteristics influencing fire spread.

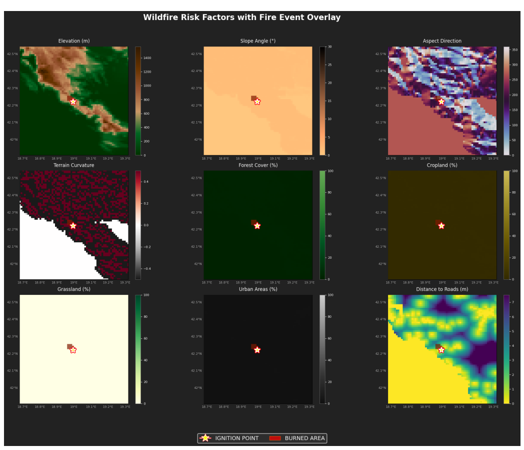

Fig. 14. Topographical and land cover characteristics affecting wildfires

Conclusions

This study presents a comprehensive analysis of wildfire spread using temporal data, integrating atmospheric conditions, vegetation health, temperature variations, human activity, and land features. Through data preprocessing, feature engineering, and visualization, we identified key factors influencing wildfire ignition and propagation.

The analysis highlights that high temperatures, dry vegetation, and human activities significantly contribute to wildfire occurrences. Additionally, topographical factors such as elevation, slope, and proximity to roads play a crucial role in fire behavior. The developed dataset and visualizations provide valuable insights for fire prediction models, aiding in more accurate forecasting and risk assessment.

Future work will focus on enhancing predictive modeling using deep learning approaches and incorporating real-time satellite imagery. Improved segmentation models and advanced feature extraction techniques can further refine fire spread predictions, contributing to more effective wildfire management and mitigation strategies.

References

[1] S. Kondylatos, I. Prapas, G. Camps-Valls, and I. Papoutsis, "Mesogeos: A multi-purpose dataset for data-driven wildfire modeling in the Mediterranean," ArXiv, 2023. Available: https://arxiv.org/pdf/2306.05144