Wildfire detection and spread prediction using machine learning models in geospatial data

Srivasa Pranathy Mutcherla, *MSCS Graduate Student, Illinois Institute of Technology,*
Shouvik Roy, *Professor, College of Computing,*

***Abstract*—The recent LA fires impacted 160,000 people, reminding the world that climate does not see class when raging its fury. It also highlighted systemic injustices in the state that were not highlighted for a while now, showing how people in power had a domino impact when things like this happened. For example, issues with insurance policy payouts highlighted how many people would not be receiving the support they need to rebuild their lives. The issues of governmental policies such as cutting the budget of the fire department and the monopoly of the ownership of the water source for the entire state were questioned. Although the state was prepared for one wildfire, they were not prepared for five. Claims that it might take as long as a decade to rebuild the region exist. Who does it really hurt, though? The taxpayers and the people whose entire lives were uprooted and destroyed by these fires. That is the motivation for choosing this problem statement, to help plan and support everyone by being able to predict the trajectory and intensity of wildfires. Even a day's worth of notice would have helped people identify what is crucial for their life moving forward - this means a lot more for people who are not on the higher bar in the income strata. This project is beneficial to teams working on emergency evacuation and rescue planning, as it helps them inform the public and plan for the staff needed to minimize damage to the best extent. This project will build a machine learning-based model that works on spatial data to help predict the trajectory of the wildfire.**

***Index Terms*—geospatial, machine learning, wind, elevation, viirs satellite data**

I. Introduction

WILDFIRES' increasing frequency, intensity, and unpredictability present a global critical environmental and public safety challenge. Particularly in fire-prone regions such as California, early detection and accurate prediction of wildfire spread are essential for timely resource deployment, evacuation planning, and mitigation of ecological damage.

Conventional wildfire monitoring methods have relied on satellite imagery, ground-based sensors, and manual reporting. While effective to some extent, these approaches often suffer from latency, limited spatial coverage, and challenges in integrating heterogeneous data sources. The advent of modern satellite constellations such as the Visible Infrared Imaging Radiometer Suite (VIIRS) has revolutionized this landscape by providing near-real-time thermal anomaly data with global coverage.

This project investigates a hybrid methodology that began with computer vision (CV) techniques aimed at deriving wildfire ground truth data from imagery, but gradually transitioned toward a geospatial machine learning pipeline based on VIIRS detections. By incorporating physical terrain features (elevation and slope) and atmospheric variables (wind vectors), the goal is to build an interpretable, scalable model to predict fire spread and render interactive visualizations for decision-making support.

II. Related Work

Wildfire monitoring and risk prediction have been long-standing challenges in remote sensing and geospatial analytics. Satellite-based thermal anomaly detection, meteorological modeling, and ecological indexing have all contributed to early warning systems and spread forecasting models.

NASA's Fire Information for Resource Management System (FIRMS) has been pivotal in providing global near real-time fire detections using data from MODIS and VIIRS instruments [1]. The VIIRS instrument, in particular, offers improved spatial resolution and detection capabilities, making it a popular choice for operational and research applications.

Recent efforts, such as the SciPy 2024 wildfire detection framework, have demonstrated methods for deriving historical fire footprints by utilizing georeferenced wildfire polygon data from FRAP and Sentinel-2 imagery, processed through Google Earth Engine and Python-based tools[2]. Similarly, projects like PyreCast [3] have advanced region-specific fire hazard modeling by integrating weather forecasts, fuel characteristics, and topographical data to provide near-term wildfire risk assessments.

The Fire Weather Index (FWI) system remains a widely used meteorological tool for estimating wildfire potential based on temperature, humidity, wind, and precipitation [4]. Other research has focused on leveraging GOES satellite systems for rapid wildfire detection at high temporal resolution [5].

Compared to these approaches, this work emphasizes the fusion of terrain (elevation, slope) and wind data with VIIRS detections in a machine learning framework, aiming for real-time interpretability through interactive geospatial visualizations.

III. Implementation

The implementation of this project followed a modular approach, involving early experimentation with computer vision, data acquisition, geospatial processing, feature engineering, and predictive modeling.

A. Ground Truth Exploration (CV Approach)

Initially, the project attempted to replicate a methodology inspired by a SciPy 2024 paper, using Sentinel-2 imagery and FRAP wildfire polygons accessed via Google Earth Engine (GEE) to build a ground truth dataset for image-based classification. While this approach was partially successful, it was ultimately limited by the long-term goals of the project. This led to a pivot toward a structured geospatial pipeline centered around VIIRS fire detections and environmental features.

B. Data Sources

Fire detection data was obtained from NASA's FIRMS platform, which provides access to VIIRS detections via the FIRMS API. Each record includes latitude, longitude, fire radiative power (FRP), and a confidence label (low, nominal, high).

Topographic data was retrieved from the USGS 3DEP Digital Elevation Model (30m resolution). Wind data, specifically the zonal (u10) and meridional (v10) components, were downloaded from the Copernicus ERA5 reanalysis dataset using the CDS API.

C. Geospatial Feature Engineering

All datasets were reprojected into a unified coordinate reference system for spatial alignment. Slope values were derived from the DEM using the RichDEM library. For each VIIRS fire detection point, the following features were extracted:

- **FRP**: A proxy for fire intensity, measured in megawatts
- **Slope and Elevation**: Captured from DEM at each fire point
- **u10 and v10**: Hourly wind components interpolated from ERA5 grid
- **Confidence Label**: Mapped from categorical to ordinal values (0 = low, 1 = nominal, 2 = high)

D. Modeling Approach

The predictive task began as a regression problem, using a Random Forest Regressor to estimate the fire confidence value. However, this approach yielded low correlation, leading to a reformulation as a classification task.

A Random Forest Classifier was trained using FRP, slope, elevation, u10, and v10 features. This classifier showed higher performance and interpretability, providing confidence class predictions (low, nominal, high) used in downstream visualizations.

IV. Experimental Evaluation

The complete dataset was partitioned using an 80/20 train-test split. Model performance was assessed using the coefficient of determination (R^2 score) for regression, and overall classification accuracy for the categorical formulation.

Using a Random Forest Regressor to predict the numerical encoding of VIIRS confidence levels, the initial regression-based approach yielded poor performance with $R^2 \approx -0.15$. This indicated that the confidence labels were poorly suited for regression, likely due to their categorical nature, weak ordinal structure, and potential label noise. These results motivated a reformulation of the task as a classification problem, more appropriate for predicting discrete confidence classes.

To address this, the task was reframed as a multi-class classification problem using a Random Forest Classifier. The model was trained to predict one of three discrete confidence classes (low, nominal, high). This formulation improved performance, achieving an $R^2 \approx 0.86$ and providing better visual alignment with known fire activity.

Model outputs were visualized using Folium. Each fire point was rendered as a circle whose color and radius were both scaled according to Fire Radiative Power (FRP), using a continuous yellow-to-red colormap for intuitive intensity-based interpretation. Two toggleable layers were provided - one showing actual confidence labels and the other showing predicted labels - both using the same FRP-based color scheme. Wind vectors were overlaid as directional blue lines derived from the u10 and v10 wind components, and pop-ups were used to display detailed metadata including FRP, slope, and confidence.

These experiments suggest that while VIIRS confidence may not be a perfect ground truth for fire intensity, classification models enriched with environmental features yield promising predictive and interpretive results for wildfire modeling.

V. Results

The project's final output was an interactive geospatial visualization tool that overlays model predictions on topographic and atmospheric layers. Fire points were displayed as circles whose color and radius were scaled by Fire Radiative Power (FRP), using a continuous yellow-to-red colormap for intuitive intensity encoding. Two toggleable layers were provided - one for actual VIIRS confidence labels and another for model-predicted values, allowing users to compare predictions interactively. Confidence information was presented via pop-ups, preserving visual consistency.

Slope was visualized as a semi-transparent red raster derived from the DEM, while wind vectors based on ERA5 u10 and v10 components were rendered as directional blue lines showing local wind direction and strength.

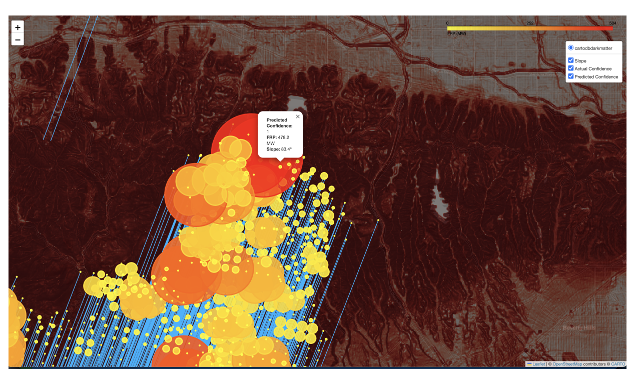

Fig. 1. Map showing FRP-scaled fire points, slope overlay (red), and wind vectors (blue).

Figure 1 illustrates the integrated visualization, showing all key features. This layered map enables users to spatially interpret the role of terrain and wind in wildfire behavior and supports the assessment of model accuracy.

These results demonstrate that even with a limited feature set-FRP, slope, and wind-wildfire dynamics can be interpreted through spatial visualizations. However, model limitations remain due to noise in the confidence labels and the absence of ecological variables such as vegetation, fuel load, or soil moisture.

VI. Conclusion

This project presents a practical exploration into wildfire detection and modeling using openly available geospatial and satellite data. Through an iterative development process that began with computer vision techniques and transitioned into machine learning-based geospatial analysis, the potential of combining VIIRS fire detections with slope, elevation, and wind data for confidence-level prediction and map-based visualization was demonstrated.

Initial attempts to model fire confidence using regression were ineffective, in part due to noise and uncertainty in the confidence labels, but more fundamentally because confidence is a categorical target with limited ordinal structure. Reformulating the task as a classification problem led to significant improvements in predictive performance and interpretability.

Future extensions of this work include integration of vegetation and fuel load data, time series modeling of fire spread, and real-time deployment via APIs or dashboards. The overall workflow and tools presented here form a robust foundation for wildfire informatics in academic and applied settings.

Appendix

Supplementary Resources

The following platforms, tools, and datasets were explored throughout this project but are not directly cited in the main text:

- **Chloris Earth Platform**: https://www.chloris.earth/
- **Chloris Documentation**: https://app.chloris.earth/docs/1.8.2/index.html
- **Microsoft AI Wildfire Demo**: https://satelliteimagerydemostg.z5.web.core.windows.net/damage-assessment/los_angeles_palisades_fire_1_8_2025.html
- **Terra.do Climate Learning Platform**: https://www.terra.do/
- **YouTube - Ankur Shah Wildfire Talk**: https://www.youtube.com/live/iMbfxlNZRSk
- **Ankur Shah GitHub - FIRMS Ingestion**: https://github.com/ankshah131/localsolve-open/blob/main/wildfires/la_wildfires/Ingest_FIRMS.ipynb
- **Google Earth Engine Wildfire Boundaries**: https://medium.com/google-earth/how-to-generate-wildfire-boundary-maps-with-earth-engine-b38eadc97a38
- **NOAA HMS Hazard Mapping System**: https://www.ospo.noaa.gov/products/land/hms.html#data
- **Kaggle Wildfire Dataset**: https://www.kaggle.com/datasets/abdelghaniaaba/wildfire-prediction-dataset/code

References

[1] NASA FIRMS, "Fire information for resource management system (firms)," https://earthdata.nasa.gov/earth-observation-data/near-real-time/firms, 2007.

[2] V. Martin, D. Morgan, and K. B. Venable, "Python-based geoimagery dataset development for deep learning-driven forest wildfire detection," in *Proceedings of the 23rd Python in Science Conference (SciPy 2024)*, 2024. [Online]. Available: https://proceedings.scipy.org/articles/YADT7194

[3] Spatial Informatics Group, "Pyrecast wildfire forecasting platform," https://sig-gis.com/pyrecast, 2021.

[4] Natural Resources Canada, "Canadian forest fire weather index (fwi) system," https://www.nwcg.gov/publications/pms437/cffdrs/fire-weather-index-fwi-system, 1987.

[5] National Oceanic and Atmospheric Administration (NOAA), "Goes fire detection and monitoring," https://www.star.nesdis.noaa.gov/goes/, 2016.