

Next-Day Wildfire Spread Prediction on mNDWS

A concise proposal aligned to DS6050 Deliverable #1 by Project Group 4

Robert Clay Harris, Hannah Richardson, and Chelsey Blowe
School of Data Science, University of Virginia
{jbm2rt, zhx9yf, qck2qg}@virginia.edu

Abstract—We propose a next-day ($t+1$) burned-area prediction model using the modified Next Day Wildfire Spread (mNDWS) dataset (500 m VIIRS, CONUS-West, 2018–2023). The task is framed as binary image segmentation of pixels likely to burn tomorrow given multimodal inputs (weather, vegetation/drought, fuels, topography, impervious/water). We outline a simple, reproducible baseline (logistic regression and a compact U-Net), an ablation plan to quantify feature-family importance (wind, fuels, vegetation/drought, topography), and robustness slices (high-wind, WUI). Our literature review highlights recent deep segmentation for spread prediction and motivates short-horizon modeling with calibrated thresholds.

Index Terms—wildfire, remote sensing, geospatial AI, image segmentation, multimodal learning

I. MOTIVATION: PROBLEM STATEMENT

Wildfire spread forecasts inform evacuations, resource allocation, and risk communication. While physics-based simulators can be accurate they require detailed configuration and are costly to run at regional scale. Traditional fire danger indices rely on fixed formulas and cannot capture nonlinear interactions among fuels, wind, and drought. The goal of our study is to predict where wildfire spreads with the use of a deep learning model. We target a practical question: which 500 m pixels will burn tomorrow? Our objectives are: (i) a clear, reproducible baseline; (ii) insight into which feature families drive performance; (iii) evaluation of model robustness under challenging conditions such as high winds and wildland–urban interface areas; and (iv) calibrated decision thresholds usable by practitioners.

Our study will aim to answer the following research questions:

- RQ1: How well can a multimodal model predict next-day burned pixels at 500 m?
- RQ2: Which feature families (wind, fuels, vegetation/drought, topography) contribute most?
- RQ3: How robust is performance under high wind and in WUI (impervious) areas?
- RQ4: Can we add simple, calibrated uncertainty to support thresholding?

II. DATASET: PUBLIC URL

We will use the modified Next Day Wildfire Spread (mNDWS) dataset (1): <https://www.kaggle.com/datasets/georgehulsey/modified-next-day-wildfire-spread>

Why mNDWS:

- Resolution and coverage: VIIRS at 500 m (versus 1 km in classic NDWS), CONUS-West, 2018–2023.
- Covariates: Burn Index, CHILI, ERC, NDVI, precipitation/temperature/wind aggregates including wind direction, elevation; impervious and water masks.
- Fuels: 30 m LANDFIRE fuels compressed via a convolutional autoencoder into a 3-D fuel embedding (fuel1–3) per 500 m pixel.
- Packaging: TFRecords with train/val/test splits; provided feature statistics; class imbalance roughly 3% positives.

We follow the provided splits and treat prediction as binary segmentation (burn/no-burn) for day $t+1$.

III. LITERATURE REVIEW: PRIOR WORK AND GAPS

Applying deep learning models to classify and predict wildfires has become increasingly popular to aid in wildfire relief efforts. Large-eddy simulation (LES) has become a foundational approach for modeling wildfire–atmosphere interactions over the past two decades. Early coupled fire–atmosphere systems, including FIRETEC (2), WFDS (3), and WRF-Fire (4), demonstrated that explicitly resolving buoyancy-driven turbulence, plume dynamics, and wind–fire feedbacks improves predictions of fire spread and convective behavior compared with empirical rate-of-spread models. These physics-based systems advanced understanding of multiscale interactions between the surface fire, lower atmosphere, and ambient wind fields (5; 6). However, they were primarily implemented in legacy Fortran and C++ with MPI parallelization, which limits extensibility and compatibility with modern high-performance computing (HPC) frameworks.

LES has since been widely used to investigate fireline intermittency (7), entrainment and plume rise (8), and fire behavior over complex terrain and heterogeneous fuels (9). Previous studies suggest that grid resolutions between 0.5 m and 2 m are adequate for resolving dominant fire–atmosphere coupling processes (10); however, applying such fine resolutions across landscape scales remains computationally intensive. Field validation efforts using experiments such as FireFlux I and II (11; 12) have improved confidence in coupled fire–atmosphere modeling but also reveal persistent challenges related to unresolved combustion processes, fuel heterogeneity, and moisture dynamics.

Recent reviews highlight methodological limitations that constrain the broader application of LES in wildland fire science (6; 13). Despite growth in computing capabilities, most

existing wildfire LES frameworks do not leverage modern accelerator hardware such as GPUs and TPUs and lack integration with machine learning-oriented workflows. This hinders performance scaling and limits opportunities for data-driven parameter estimation, surrogate model development, and hybrid physics–ML approaches. Additionally, few studies have systematically examined how grid resolution affects global fire behavior metrics—such as rate of spread or heat release rate—compared to local turbulence characteristics, leaving a lack of practical guidance for efficiency versus accuracy trade-offs.

The study by Wang et al. (14) addresses several of these limitations by developing a TensorFlow-based LES wildfire simulation framework optimized for TPU acceleration, performing a resolution sensitivity analysis, and validating the system using the FireFlux II dataset. Nonetheless, important gaps remain. The framework is tested only in flat grassland environments, limiting relevance to complex landscapes with variable fuel loading and topography. Its combustion model uses a prescribed volumetric heat source rather than explicit pyrolysis, restricting physical realism and the ability to simulate fuel moisture effects. Validation is currently limited to a single experimental dataset, which prevents generalization across ecological fuel types. Furthermore, the model does not incorporate realistic atmospheric boundary layer evolution, including diurnal transitions and wind shear, which strongly influence fire behavior. Finally, although the TensorFlow implementation shows promise for scalability and future machine learning integration, it has not yet been adapted for real-time fire prediction or operational decision support.

A study by Papakis et al. (15) adds a number of factors that identify the lack of realistic atmospheric fire influence from the previous study. This study focused on correctly classifying wildfires in Greece. It merged topographic, satellite, vegetation and meteorological factors from several data sources, including normalized vegetation index (NDVI) imagery data, Visible Infrared Imaging Radiometer Suite (VIIRS) fire data, and other imagery data sources. The overall goal of the model is to explore different prediction approaches and find a method that correctly classifies wildfires most accurately.

Two different approaches were explored; the first only included numerical features (no imagery data) and the second was a multimodal approach. Both of these approaches were trained twice, once with a Daynight feature and once without. This feature was used to indicate whether the data used was from the day or night. The importance of including this feature was determined to be influential in predicting wildfires.

The non multimodal numerical features approach analyzed long short term memory (LSTM) networks, 1D convolutional neural networks (CNNs), and ensemble models during training. These three models were selected due to their “architectures [being] well-suited to time series classification” (15). The ensemble model was a combination of the LSTM and CNN models with the goal of reducing bias. The LSTM model performed best with a 90% accuracy.

The multimodal approach contained NDVI images while the previous approach did not. CNN was used for image process-

ing, and a multilayer perceptron (MLP) was used to process numerical features. “Multimodal fusion combines” (15) pre-processed images and numerical features using a combination of batch normalization, ReLU, and softmax activation. The multimodal training explored the performance of stochastic gradient descent (SGD) with step decay, AdamW with Cosine Annealing, AdamW with exponential decay, and SGD with aggressive step decay to determine the best parameters. The SGD with step decay performed the best with a 96.15% accuracy.

The main gap of this study is that the focus was on identifying areas where a fire existed or not, the focus of our study will include predicting where the fire will likely spread. Another is that the study is specific to Greece’s climate and land makeup, the aim of our study is to predict wildfire spread throughout the varying climates within the US. A study by Shadrin et al. (16) addresses these gaps by introducing another country of study as well as predictive metrics.

The study propose a deep learning approach for short-term wildfire spread prediction using multimodal remote-sensing and meteorological data. The study focuses on several fire-prone regions in Russia and frames the task as a pixel-level segmentation problem, predicting which areas will be affected by fire one to five days after ignition. Each sample is represented as a small raster window centered on an ignition point that combines static geospatial variables such as land cover, topography, and population density with daily weather fields derived from ERA5. The resulting input tensor contains multiple stacked channels, enabling spatial models to capture interactions between terrain, fuels, and meteorological drivers of spread.

Several encoder–decoder architectures are evaluated, including U-Net, U-Net++, DeepLabV3, and MA-Net, each trained using cross-entropy and Dice-based losses. MA-Net achieves the best overall performance, with F1 scores around 0.64–0.68 for one- to five-day horizons. The model demonstrates that spatial attention mechanisms can modestly improve the delineation of burned perimeters, particularly under variable wind conditions. Evaluation metrics also include mean absolute error in burned area and a spread-velocity estimate based on the distance from the ignition to the predicted fire front.

Ablation experiments identify wind and land-cover variables as the most influential feature groups. Removing wind inputs leads to the largest drop in performance, reducing F1 to approximately 0.51, while the absence of land-cover layers produces a smaller but still significant effect. The authors conclude that multimodal representations combining static surface data with dynamic weather fields are essential for accurate spread modeling. They also note that generalization to new ecoregions is limited without regional retraining, underscoring the importance of spatially diverse data.

This paper provides a practical reference for our project. Its design of a multimodal tensor, short prediction horizon, and attention-based encoder–decoder architecture aligns closely with our goals for mNDWS. In our case, we will adapt this approach to the continental United States west of the

100th meridian at 500 m resolution, incorporating additional covariates available in mNDWS such as the latent fuel embeddings. Following Shadrin et al., we will include feature-group ablations to assess the marginal contribution of wind, fuels, vegetation, and topography, enabling a transparent baseline for next-day spread prediction.

A. Synthesis and Gaps

Prior work demonstrates that short-horizon wildfire spread can be effectively modeled using convolutional encoder–decoder architectures applied to multimodal geospatial inputs. Shadrin et al. (16) showed that models such as MA-Net can capture spatial fire dynamics over daily time steps, with wind and surface characteristics emerging as the most influential predictors. Other studies highlight the potential of combining diverse data sources, including fuels, vegetation indices, and topography, to improve generalization across regions.

Several gaps remain. First, the empirical value of the latent fuel embeddings introduced in mNDWS has not yet been established at 500 m resolution. Second, prior evaluations have been geographically limited, leaving uncertainty about how feature importance and model performance vary across the broader CONUS-West domain. Third, robustness under challenging conditions such as high winds and wildland–urban interface (WUI) areas has received little systematic study. Addressing these gaps will help clarify how multimodal models generalize and which features most reliably drive spread prediction in operational contexts.

IV. PROPOSED METHOD: INITIAL APPROACH

Task: next-day binary segmentation on $H \times W$ tiles (mNDWS tile size), predicting burn or no-burn.

Baselines:

- 1) Logistic regression per-pixel on tabular features (regularized).
- 2) Shallow CNN without skip connections.
- 3) Compact U-Net with a lightweight encoder (e.g., ResNet-18).

Inputs: mNDWS channels (NDVI; ERC/BI/CHILI; precipitation, temperature, and wind aggregates including direction; elevation; fuel1–3; impervious and water). Normalize using provided statistics.

Class imbalance: focal plus Dice or weighted BCE plus Dice; early stopping on validation F1.

Interpretability: feature-family ablations by removing Wind, Fuels, Vegetation and Drought, or Topography to measure changes in F1 and IoU.

Uncertainty: calibrate the score threshold on validation (precision–recall tradeoff) and report reliability curves.

V. EXPERIMENTS: EVALUATION PLAN

- 1) **Baselines:** Compare logistic regression, a shallow CNN, and a U-Net. Evaluate using F1 (micro, imagewise),

Intersection-over-Union (IoU), precision, recall, and precision–recall curves. Report parameter count and approximate training and inference times.

- 2) **Feature Ablations:** Remove one feature family at a time and record performance deltas. Hypothesis: removing wind will cause the largest degradation, while fuels provide incremental gain.
- 3) **Robustness Slices:** Stratify results by `wind_75` quartiles, impervious percentage (as a wildland–urban interface proxy), and coarse ecoregion or elevation bands if available.
- 4) **Calibration and Thresholding:** Tune classification thresholds for high-precision versus high-recall operating points, and examine reliability or calibration curves.

Reproducibility: Fix random seeds, save configurations, apply single-change ablations, and maintain basic run tracking for all experiments.

REFERENCES

- [1] G. Hulsey, “Modified Next Day Wildfire Spread (mNDWS),” Kaggle Dataset, Sep. 2024. <https://www.kaggle.com/datasets/georgehulsey/modified-next-day-wildfire-spread>.
- [2] R. R. Linn, J. M. Reisner, J. J. Colman, and J. Wintertkamp, “FIRETEC: A physics-based model for coupled fire–atmosphere interaction,” *International Journal of Wildland Fire*, vol. 11, no. 4, pp. 233–246, 2002. DOI: <https://doi.org/10.1071/WF02006>.
- [3] W. Mell, S. L. Manzello, A. Maranghides, D. Butry, and R. G. Rehm, “The wildland–urban interface fire problem: Current approaches and research needs,” *International Journal of Wildland Fire*, vol. 19, no. 2, pp. 238–251, 2009. DOI: <https://doi.org/10.1071/WF07131>.
- [4] J. L. Coen, M. Cameron, J. Michalak, E. G. Patton, P. J. Riggan, and K. Yedinak, “WRF-Fire: Coupled weather–wildland fire modeling with the Weather Research and Forecasting model,” *Journal of Applied Meteorology and Climatology*, vol. 52, no. 1, pp. 16–38, 2013. DOI: <https://doi.org/10.1175/JAMC-D-12-023.1>.
- [5] A. L. Sullivan, “Wildland fire behaviour modelling: Review of models, evaluation, and applications,” *International Journal of Wildland Fire*, vol. 18, no. 4, pp. 369–386, 2009. DOI: <https://doi.org/10.1071/WF06143>.
- [6] A. Bakhshaii and E. A. Johnson, “A review of wildland fire spread modelling, 1990–2019,” *International Journal of Wildland Fire*, vol. 28, no. 11, pp. 827–842, 2019. DOI: <https://doi.org/10.1071/WF19002>.
- [7] F. Pimont et al., “Fireline intensity of spreading wildfires: Variability and multiplicative structure,” *International Journal of Wildland Fire*, vol. 30, no. 1, pp. 1–14, 2021. DOI: <https://doi.org/10.1071/WF19133>.
- [8] A. K. Kochanski, M. A. Jenkins, J. Mandel, J. D. Beezley, C. B. Clements, and S. Krueger, “Evaluation of WRF-Sfire performance with field observations from the FireFlux experiment,” *Geoscientific Model Development*,

vol. 6, no. 3, pp. 1109–1126, 2013. DOI: <https://doi.org/10.5194/gmd-6-1109-2013>.

- [9] E. V. Mueller, R. R. Linn, J. Winterkamp, and C. H. Sieg, “Modeling crown fire behavior in complex terrain using coupled fire–atmosphere models,” *Forest Science*, vol. 60, no. 4, pp. 781–792, 2014. DOI: <https://doi.org/10.5849/forsci.13-100>.
- [10] J. K. Hiers, J. J. O’Brien, J. M. Varner *et al.*, “Interactions between fire and atmosphere in wildland fire spread,” *Forest Ecology and Management*, vol. 475, p. 118373, 2020. DOI: <https://doi.org/10.1016/j.foreco.2020.118373>.
- [11] C. B. Clements, S. Zhong, S. Goodrick *et al.*, “Observing the dynamics of wildland grass fires: FireFlux—A field validation experiment,” *Bulletin of the American Meteorological Society*, vol. 88, no. 9, pp. 1369–1382, 2007. DOI: <https://doi.org/10.1175/BAMS-88-9-1369>.
- [12] J. M. Canfield, C. B. Clements, S. Zhong, S. Goodrick, and S. Li, “FireFlux II: An evaluation of plume and fire behavior in grass fires,” *International Journal of Wildland Fire*, vol. 23, no. 9, pp. 1163–1174, 2014. DOI: <https://doi.org/10.1071/WF12190>.
- [13] R. R. Linn, J. M. Canfield, F. Pimont *et al.*, “A conceptual framework for coupled fire–atmosphere modeling,” *International Journal of Wildland Fire*, vol. 30, no. 8, pp. 593–608, 2021. DOI: <https://doi.org/10.1071/WF20132>.
- [14] Z. Wang, C. M. Hoffman, R. Kooper, R. M. Hadden, A. Simeoni, and W. T. Scherer, “A high-resolution large-eddy simulation framework for wildland fire prediction using TensorFlow,” *International Journal of Wildland Fire*, vol. 32, no. 12, pp. 1711–1725, 2023. DOI: <https://doi.org/10.1071/WF23084>.
- [15] I. Papakis, V. Linardos, and M. Drakaki, “A Multimodal Ensemble Deep Learning Model for Wildfire Prediction in Greece Using Satellite Imagery and Multi-Source Remote Sensing Data,” *Remote Sensing*, 17, no. 19: 3310, 2025. DOI: <https://doi.org/10.3390/rs17193310>.
- [16] D. Shadrin, S. Illarionova, F. Gubanov, K. Evteeva, M. Mironenko, I. Levchunets, R. Belousov, and E. Burnaev, “Wildfire spreading prediction using multimodal data and deep neural network approach,” *Scientific Reports*, vol. 14, no. 2606, 2024. DOI: <https://doi.org/10.1038/s41598-024-52821-x>.