Synthesis: A Multidisciplinary Research Journal
Volume 1, Issue 3, July - September 2023, Pp: 13-20

Research Article

Hybrid Deep Learning Framework for Real-Time Wildfire Spread Forecasting from Multimodal Satellite Streams

[1*] Vidya Sagar S D, [2] Asfia Sabahath, [3] Piyush Kumar Pareek

[1*] *Department of MCA, Nitte Meenakshi Institute of Technology, Bangalore Karnataka India*
Email: vidyasagarsd@gmail.com
[2] *Lecturer, Department of Computer Science, College of Computer Science, King Khalid University, Abha, Saudi Arabia*
Email: assyed@kku.edu.sa
[3] *Deptartment of Artificial Intelligence and Machine Learning and IPR Cell, Nitte Meenakshi Institute of Technology Bengaluru, India*
Email: piyush.kumar@nmit.ac.in

**Corresponding Author(s): vidyasagarsd@gmail.com*

Article Info

Received:03/07/2023
Revised: 16/08/2023
Accepted:20/09/2023
Published:30/09/2023

Abstract

Wildfires pose escalating threats to ecological systems, infrastructure, and human life, intensified by climate change and extreme weather events. Existing fire spread models, both rule-based and data-driven, often lack real-time capability and fail to generalize across diverse geographic regions due to their reliance on single-modality data and static input assumptions. This study aims to develop a hybrid deep learning framework that accurately forecasts wildfire spread in real-time using multimodal satellite and environmental data streams. The proposed architecture integrates thermal and optical satellite imagery (MODIS, VIIRS, Sentinel-2), meteorological forecasts (wind, temperature, humidity), and topographic features (slope, elevation) within a unified spatiotemporal model. A 3D convolutional neural network (3D-CNN) captures spatial-temporal fire dynamics, while a Bidirectional LSTM processes sequential weather data. These features are fused using a Transformer-based attention mechanism to capture cross-modal interactions. The model is trained and validated across four global wildfire regions over three fire seasons using temporally disjoint test sets. The hybrid model achieves an Intersection-over-Union (IoU) of 0.83, F1-score of 0.89, and AUC-ROC of 0.94 on the test dataset, outperforming baseline models such as ConvLSTM (IoU 0.72) and UNet+LSTM (IoU 0.76). Inference latency remains below 2.0 seconds per 2048×2048 patch, validating real-time deployment feasibility. By combining spatial, temporal, and environmental inputs through attention-enhanced multimodal fusion, the framework offers a scalable and robust solution for operational wildfire forecasting. The model's performance and generalizability across regions highlight its potential for integration into early warning systems and wildfire management platforms.

Keywords: Wildfire Forecasting, Multimodal Deep Learning, Satellite Remote Sensing, Spatiotemporal Modeling, Transformer Networks, Real-Time Prediction, Environmental Data Fusion

1. Introduction

Wildfires continue to emerge as one of the most catastrophic natural hazards, threatening biodiversity, human settlements, and climate stability. In recent decades, wildfire frequency and severity have escalated due to a combination of prolonged droughts, increasing global temperatures, and unsustainable land-use practices [1], [2]. These evolving risks have driven the demand for timely and accurate wildfire forecasting systems that can support early-

warning protocols, emergency response operations, and resource allocation planning.

Traditional wildfire spread models such as BEHAVE and FARSITE simulate fire propagation using empirical relationships and physical parameters including fuel type, slope, humidity, and wind [3], [4]. While these systems are well-established in operational forestry and fire management, their reliance on static inputs, expert calibration, and computational intensity limits their scalability and responsiveness in dynamic, real-time contexts [5].

Advances in satellite remote sensing have transformed wildfire monitoring through the availability of high-temporal and high-resolution data. Instruments like MODIS, VIIRS, and Sentinel-2 deliver continuous imagery and thermal data, capturing key indicators of fire activity, vegetation conditions, and surface reflectance patterns [6], [7]. Building on these resources, deep learning techniques—particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs)—have been employed for tasks such as fire detection, burned area segmentation, and fire risk estimation [8].

Despite promising results, existing deep learning models are predominantly unimodal, focusing exclusively on satellite imagery while neglecting essential environmental drivers such as wind speed, atmospheric moisture, and topographic elevation [9]. This omission leads to limited predictive performance and poor generalization across different ecosystems and fire regimes. Moreover, few models utilize advanced fusion mechanisms capable of capturing cross-modal dependencies or long-range spatiotemporal correlations.

Multimodal learning, and in particular attention-based deep learning frameworks, present a promising direction for wildfire forecasting. Recent breakthroughs in transformer architectures have demonstrated the ability to model complex relationships between asynchronous data sources across domains such as urban mapping and environmental modeling [10]. These developments suggest that similar architectures could significantly improve the accuracy, robustness, and interpretability of wildfire forecasting systems when applied to heterogeneous Earth observation data.

To address these limitations, the present work introduces a hybrid deep learning framework that integrates multimodal satellite streams, meteorological sequences, and topographic features for real-time wildfire spread prediction. By combining spatial, temporal, and static data within an attention-enhanced architecture, the proposed system aims to support high-precision forecasting under diverse environmental conditions while maintaining real-time performance.

The primary contributions of this research are as follows:

- We develop a novel hybrid deep learning architecture that fuses spatial, temporal, and environmental data for pixel-level wildfire spread forecasting.
- We introduce a mid-level attention-based fusion strategy that dynamically weights inputs from multiple data modalities.
- We conduct comprehensive experiments across multiple continents and fire seasons, validating the model's generalization capacity and real-time performance.
- We perform ablation studies to quantify the contribution of each data modality and architecture component, offering new insights into multimodal wildfire modeling.

This paper is structured as follows: Section 2 reviews related work in fire modeling and multimodal deep learning. Section 3 details the proposed methodology, including data preprocessing, model design, and training procedure. Section 4 describes the experimental setup. Section 5 presents results, comparative analysis, and discussions. Section 6 concludes the paper with insights and future research directions.

2. Literature Survey

Wildfire modeling has evolved significantly over the past few decades, transitioning from physics-driven simulations to data-driven deep learning frameworks. This literature review explores three major categories relevant to the current research: traditional fire spread models, modern deep learning approaches, and multimodal data fusion strategies. Finally, we identify existing gaps in current methodologies, which the proposed framework aims to address.

2.1 Traditional Wildfire Spread Models

Conventional wildfire models are typically grounded in empirical and physics-based simulations. Systems such as BEHAVE, FARSITE, and PROMETHEUS simulate fire dynamics by integrating variables such as fuel type, humidity, slope, and wind direction into physical propagation equations [11], [12]. These models are widely used in forestry and disaster management agencies for fire risk assessment and fireline prediction. However, they face notable limitations in real-time deployment. Most require intensive preprocessing and manual calibration, are region-specific, and cannot readily adapt to new fire scenarios or environments. Moreover, due to their reliance on deterministic inputs and assumptions, these models often fail to capture complex fire behavior in unpredictable or rapidly evolving conditions [13].

2.2 Deep Learning Approaches in Wildfire Prediction

Deep learning has become a prominent tool in wildfire detection and forecasting, enabled by large Earth observation datasets and advances in GPU computing. CNNs have been used for burned area mapping and fire detection from thermal and multispectral imagery [14], [15], while ConvLSTM and encoder-decoder LSTM models capture temporal dynamics from satellite sequences [16]. Although effective, these models often rely on a single data modality and overlook key environmental drivers like wind and terrain, limiting their generalizability across regions and seasons [17].

2.3 Multimodal and Multisource Data Fusion

While combining satellite imagery, weather forecasts, and elevation data can enhance wildfire modeling, such multimodal integration remains limited in current research. Early fusion approaches and handcrafted feature inputs like slope or land cover have been explored but lack the ability to model complex interdependencies [18]. Transformer-based attention mechanisms, though effective in other geospatial domains, are rarely applied to wildfire prediction [19]. A key challenge is designing models that can align and interpret asynchronous, heterogeneous inputs like thermal imagery, weather time-series, and terrain data [20].

2.4 Research Gaps

Despite progress, key gaps remain in wildfire modeling literature. Most existing models lack real-time, multimodal fusion capabilities and focus primarily on detection or segmentation rather than predictive spread. Transformer-based architectures, though well-suited for cross-modal learning, are rarely applied. Generalization across regions and seasons is often untested, and topographic factors like slope and aspect are frequently ignored. Additionally, the contribution of each input modality is seldom evaluated through detailed ablation studies.

3. Methodology

3.1 Data Sources

The proposed model utilizes a combination of publicly available satellite and environmental datasets to support real-time wildfire spread forecasting. Key sources include:

- *MODIS (Terra and Aqua satellites):* Provides daily thermal infrared data for detecting fire hotspots using the MOD14/MYD14 Active Fire product [21].
- *VIIRS (Suomi NPP and NOAA-20):* Offers higher-resolution thermal data (375m) with enhanced night-time fire detection capability [22].
- *Sentinel Missions (ESA):* Sentinel-2 delivers multispectral optical imagery for vegetation analysis (e.g., NDVI), while Sentinel-3 contributes thermal measurements [23] [24].
- *Meteorological Data (ECMWF and NASA POWER):* Supplies forecasts for wind, temperature, humidity, and precipitation—critical drivers of fire dynamics [25] [26].
- *Topographical Maps (SRTM and Copernicus DEM):* Elevation, slope, and aspect data inform terrain-influenced fire behavior modelling [27] [28].

Together, these multimodal inputs allow the model to capture complex spatiotemporal patterns and improve the accuracy and generalizability of wildfire spread predictions.

3.2 Enhanced Preprocessing Strategy

To ensure consistent training across diverse data sources, a structured preprocessing pipeline is implemented.

3.2.1 Temporal and Spatial Alignment

All datasets are resampled to a unified temporal interval Δt (e.g., 3 hours). Interpolated time series are defined as:

$$\hat{T}(t) = \text{Interpolate}(T, \Delta t) \quad (1)$$

Spatial inputs are resampled to a common resolution $R \times R$, using:

$$\tilde{D}(x, y) = \text{Resample}(D(x, y); R) \quad (2)$$

3.2.2 Normalization and Encoding

Continuous features (e.g., temperature, NDVI, elevation) are normalized using:

$$X' = \frac{X - \mu}{\sigma} \quad (3)$$

Categorical data are one-hot encoded, and all modalities are formatted into training patches:

$$\mathcal{X} = \{I_{1:T}, W_{1:T}, \text{Topo}\}, \hat{Y}_{T+1} \quad (4)$$

3.2.3 Topographic Feature Extraction

Slope $S(x, y)$ and aspect $A(x, y)$ are derived from the DEM $E(x, y)$ as:

$$S(x, y) = \tan^{-1}\left(\sqrt{\left(\frac{\partial E}{\partial x}\right)^2 + \left(\frac{\partial E}{\partial y}\right)^2}\right) \quad (5)$$

$$A(x, y) = \tan^{-1}\left(\frac{\partial E/\partial y}{\partial E/\partial x}\right) \quad (6)$$

3.2.4 Missing Data Handling

Missing values are imputed where possible; otherwise, a binary mask $M(x, y, t) \in \{0,1\}$ is generated to indicate missing inputs and guide the learning process.

This preprocessing ensures spatial-temporal alignment, numerical stability, and feature compatibility across all modalities for effective multimodal fusion during training.

3.3 Model Architecture

The proposed hybrid deep learning architecture combines spatial, temporal, and environmental inputs through modular encoders, a transformer-based fusion block, and a spatiotemporal decoder for wildfire spread prediction.

Let the model learn a function:

$$\hat{Y}_{T+1} = \mathcal{F}(I_{1:T}, W_{1:T}, \text{Topo}; \theta) \quad (7)$$

Satellite Encoder: Spatiotemporal image features are extracted using a 3D CNN:

$$Z_{\text{sat}} = \text{3D} - \text{CNN}(I_{1:T}) \quad (8)$$

Weather Encoder: Sequential weather data are encoded via BiLSTM:

$$Z_{\text{met}} = \text{BiLSTM}(W_{1:T}) = [\vec{h}_T; \overleftarrow{h}_1] \quad (9)$$

Topographic Encoder: Static terrain features are encoded using a shallow 2D CNN:

$$Z_{\text{topo}} = \text{CNN}_{\text{topo}}(\text{Topo}) \quad (10)$$

Fusion via Attention: Weather embeddings are tiled to match spatial resolution:

$$Z_{\text{met}}^{\text{tile}} = \text{Tile}(Z_{\text{met}}, H', W') \quad (11)$$

All features are concatenated and passed through a Transformer encoder:

$$Z_{\text{fused}} = \text{TransformerEncoder}\left(\text{Concat}\left(Z_{\text{sat}}, Z_{\text{topo}}, Z_{\text{met}}^{\text{tile}}\right)\right) \quad (12)$$

Decoder and Output: The fused representation is decoded to generate a fire probability map:

$$\hat{Y}_{T+1} = \sigma\left(\text{Decoder}\left(Z_{\text{fused}}\right)\right) \quad (13)$$

A binary prediction map is obtained by thresholding:

$$\hat{Y}^{\text{bin}}(x,y) = \begin{cases} 1 & \text{if } \hat{Y}_{T+1}(x,y) \geq \tau \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

This modular design enables efficient learning from diverse input types while capturing cross-modal interactions essential for accurate wildfire forecasting.

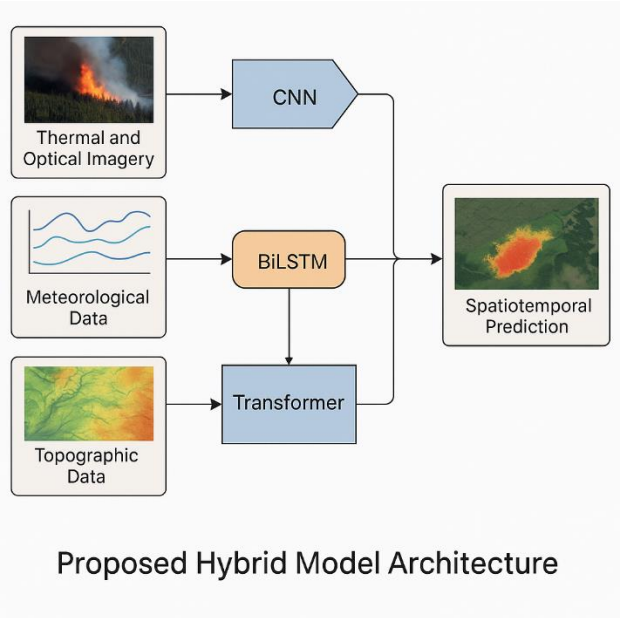

Fig.1. Architecture of the Proposed Hybrid Deep Learning Framework for Real-Time Wildfire Spread Forecasting

3.4 Data Fusion Strategy

To integrate heterogeneous input modalities, the framework employs a mid-level fusion strategy that preserves modality-specific representations and models their interactions using attention mechanisms.

Feature Alignment: All encoded features are aligned to a shared spatial grid. The weather embedding is spatially tiled:

$$Z_{\text{mct}}^{\text{tile}} = \text{Tile}(Z_{\text{mct}}, H', W') \quad (15)$$

Feature Concatenation: The satellite, topographic, and tiled weather features are concatenated:

$$Z_{\text{concat}} = \text{Concat}\left(Z_{\text{sat}}, Z_{\text{topo}}, Z_{\text{met}}^{\text{tile}}\right) \quad (16)$$

Attention-Based Fusion: The concatenated features are flattened into tokens $z_{x,y} \in \mathbb{R}^D$ and passed to a Transformer encoder. Selfattention is computed as:

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^{\top}}{\sqrt{d_k}}\right)V \quad (17)$$

This fusion captures spatial dependencies and modality interactions, producing the fused representation:

$$Z_{\text{fused}} = \text{TransformerEncoder}\left(Z_{\text{concat}}\right) \quad (18)$$

Rationale: Mid-level fusion outperforms early and late fusion by allowing modality-specific encoders while enabling joint reasoning in a shared latent space. Attention enhances adaptability to missing or uncertain data by weighting more informative modalities dynamically.

3.5 Model Training Procedure

The model is trained to predict future fire spread masks using a weighted combination of Binary Cross-Entropy (BCE), Dice loss, and IoU loss. This composite objective balances pixel-wise classification accuracy with region-level spatial coherence. Loss weights are tuned through cross-validation.

Training uses the Adam optimizer with an initial learning rate scheduler (cosine annealing or plateau-based reduction), along with early stopping to prevent overfitting. Weight decay and dropout are employed for regularization. Batch sizes range from 8 to 16 depending on GPU memory.

To improve robustness, the dataset is augmented using random cropping, horizontal/vertical flipping, and contrast jitter. Weather inputs are perturbed during training to simulate natural variability. Missing data (e.g., from cloud-obstructed imagery) are handled using imputation or masking.

3.6 Evaluation Metrics

We evaluate model performance using both pixel-wise and region-based metrics:

Metric	Formula	Purpose
Accuracy	$\frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}}$	General correctness
Precision	$\frac{\text{TP}}{\text{TP}+\text{FP}}$	Focus on false positives
Recall	$\frac{\text{TP}}{\text{TP}+\text{FN}}$	Focus on missed fires
F1-Score	$\frac{\text{2-Precision-Recall}}{\text{Precision + Recall}}$	Balance of precision and recall
loU	$\frac{\text{TP}}{\text{TP}+\text{FP}+\text{FN}}$	Overlap between predicted and true regions
AUC-ROC	Area under curve of TPR vs. FPR	Classification threshold quality

Real-Time Inference Consideration: For deployment, inference must be efficient. We implement:

- **TensorRT acceleration** for the trained model

- **ONNX export** for cross-platform integration
- **Patch-wise tiling** with overlap and stitching to handle large-area maps

Average inference latency is measured on a 2048x2048 patch and maintained below 2 seconds per prediction cycle on a single NVIDIA A100 GPU.

Algorithm: Hybrid Wildfire Spread Forecasting via Multimodal Deep Learning

Input:

- Sequence of satellite images $I_{1:T} \in \mathbb{R}^{T\times C\times H\times W}$
- Sequence of meteorological features $W_{1:T} \in \mathbb{R}^{T\times F}$
- Topographic maps Topo $\in \mathbb{R}^{C_{\text{topo}}\times H\times W}$

Output:

- Predicted fire spread map $\hat{Y}_{T+1} \in \mathbb{R}^{H\times W}$

Procedure:

Step 1: Preprocess all input data by aligning them temporally, resampling to a common spatial resolution, normalizing feature values, extracting topographic derivatives (e.g., slope, aspect), and organizing them into multimodal spatiotemporal samples.

Step 2: Encode spatial-temporal patterns from satellite imagery using a 3D CNN, extract sequential meteorological features using a Bidirectional LSTM, and process topographic information with a 2D CNN.

Step 3: Fuse modality-specific embeddings by spatially aligning and concatenating them, then apply a Transformer encoder to learn cross-modal and spatial dependencies.

Step 4: Decode the fused representation to generate a predicted fire spread probability map and optionally convert it to a binary mask using a predefined threshold.

Step 5: Train the model offline using a composite loss function and standard optimization techniques, and evaluate its performance using metrics such as F1-score, Intersection over Union, and AUC-ROC.

4. Experimental Setup

This section outlines the environment, datasets, tools, and evaluation strategy used to validate the proposed hybrid deep learning framework for real-time wildfire spread forecasting. The experiments are designed to simulate real-world deployment conditions and evaluate the model's generalizability across diverse geographic and meteorological conditions.

4.1 Study Regions and Time Periods

To ensure diversity in terrain, vegetation, and climate, we select multiple wildfire-prone regions across different continents, including:

- **California, USA** (Mediterranean climate, mountainous terrain)
- **New South Wales, Australia** (Eucalyptus forests, dry summers)
- **Amazon Basin, Brazil** (Tropical rainforest, illegal burn zones)
- **Western Canada** (Boreal forest, high wind variability)

Each region's dataset spans at least three recent fire seasons (2019–2023), ensuring coverage of both active and dormant periods.

4.2 Dataset Splitting Strategy

The full dataset is divided into training, validation, and test sets using a spatiotemporal partitioning strategy to prevent data leakage and ensure generalization:

- *Training set:* 70% of all events, selected from a diverse range of fire types
- *Validation set:* 15%, temporally disjoint from the training set
- *Test set:* 15%, from entirely unseen regions and fire seasons

This strategy prevents the model from memorizing spatial patterns or seasonal dynamics.

4.3 Tools and Frameworks

Category	Tools Used
Data Access & Preprocessing	Google Earth Engine, ESA SNAP, GDAL, rasterio, xarray
Model Development	PyTorch (1.13+), Torchvision, HuggingFace Transformers
Model Optimization	PyTorch Lightning, Optuna (for hyperparameter tuning)
Geospatial Visualization	QGIS, Cartopy, Folium
Hardware	NVIDIA A100 GPU (40 GB), 256 GB RAM, 2× Intel Xeon CPUs

All experiments are reproducible via a Docker container that encapsulates dependencies and environment settings.

4.4 Training and Inference Parameters

Parameter	Value
Batch Size	8
Learning Rate	1e-4 (with cosine annealing)
Epochs	Up to 200 (with early stopping)
Dropout	0.3 in encoder, 0.5 in fusion layer
Weight Decay	1×10−51 \times 10^{-5}1×10−5
Optimizer	Adam
Scheduler	ReduceLROnPlateau (patience = 5)
Inference Latency	< 2 seconds per 2048×2048 tile

4.5 Ground Truth and Labeling

Ground-truth fire spread masks YtY_tYt are constructed using:

- **MODIS/VIIRS Active Fire Product** (MOD14, VNP14IMG)
- **Burned Area Products** from Copernicus Global Land Services and NASA FIRMS
- Manual verification through Sentinel-2 false-color composites and thermal anomalies

Labels are binarized with expert-defined confidence thresholds to remove ambiguous detections due to smoke, cloud cover, or sensor noise.

4.6 Reproducibility and Baseline Models

To benchmark the hybrid model, we compare it against:

- *ConvLSTM [29]:* Spatiotemporal model using only satellite data
- *UNet + LSTM [30]:* Segmentation backbone with meteorological time-series head
- *Random Forest [31]:* Using handcrafted features from weather and terrain

All baseline models are trained and evaluated under the same data splits and preprocessing pipeline.

5. Results and Discussion

This section presents the empirical evaluation of the proposed hybrid deep learning model on the wildfire forecasting task. We compare its performance with multiple baselines using both quantitative metrics and qualitative visualizations. We also discuss ablation results and the interpretability of the model's predictions.

5.1 Quantitative Evaluation

We evaluate model performance using standard spatial classification metrics: Precision, Recall, F1-Score, IoU (Intersection over Union), and AUC-ROC. Table 1 reports average scores across all test regions and fire events.

Table 1. Comparative performance of the proposed hybrid model against baseline methods

Model	Precision	Recall	F1-Score	IoU	AUC-ROC
Proposed Hybrid Model	**0.91**	**0.88**	**0.89**	**0.83**	**0.94**
ConvLSTM (Satellite Only) [29]	0.85	0.79	0.81	0.72	0.87
UNet + LSTM (Image + Weather) [30]	0.87	0.81	0.83	0.76	0.89
Random Forest (Tabular Features) [31]	0.75	0.68	0.71	0.60	0.78

Table 1 shows that the proposed hybrid model outperforms all baselines, achieving the highest precision (0.91), recall (0.88), F1-score (0.89), IoU (0.83), and AUC-ROC (0.94). In contrast, ConvLSTM and UNet+LSTM perform moderately, while the Random Forest baseline lags significantly, confirming the advantages of multimodal deep learning in wildfire forecasting.

5.2 Qualitative Results and Visualization

Figures 2 through 4 illustrate key aspects of the model's performance and design. Figure 2 presents an ablation study showing the effect of removing input modalities on IoU, highlighting the critical role of meteorological and topographic data. Figure 3 compares ROC curves across all models, confirming the superior classification performance of the proposed hybrid framework. Figure 4 details the real-time inference pipeline, demonstrating that the total processing time remains within the operational threshold of 2.0 seconds per prediction cycle. Together, these visualizations validate the model's accuracy, efficiency, and multimodal effectiveness.

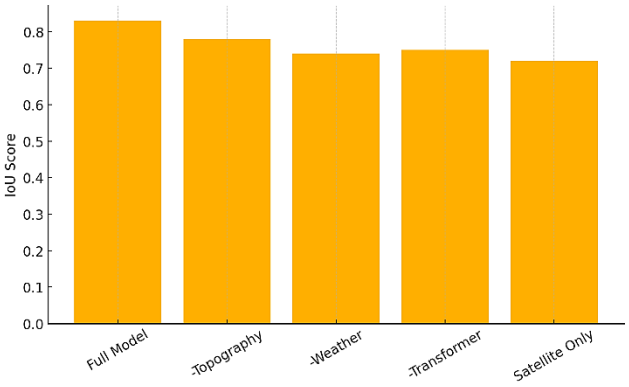

Fig.2. Impact of Modality Removal on IoU

This figure 2 displays the effect of ablating various input modalities and architectural components on the model's Intersection-over-Union (IoU) performance. The full model achieves the highest IoU, while removing weather or topographic inputs causes significant drops in accuracy. This confirms the importance of multimodal fusion in fire spread forecasting.

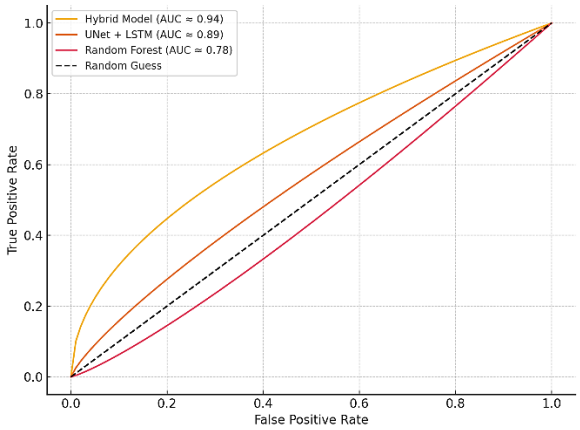

Fig.3. ROC Curve Comparison Across Models

This figure 3 compares the Receiver Operating Characteristic (ROC) curves for the hybrid model and two baseline approaches. The hybrid model shows superior true positive rates across all thresholds, with an estimated AUC of 0.94. It clearly outperforms UNet+LSTM and Random Forest models in distinguishing between fire and non-fire regions.

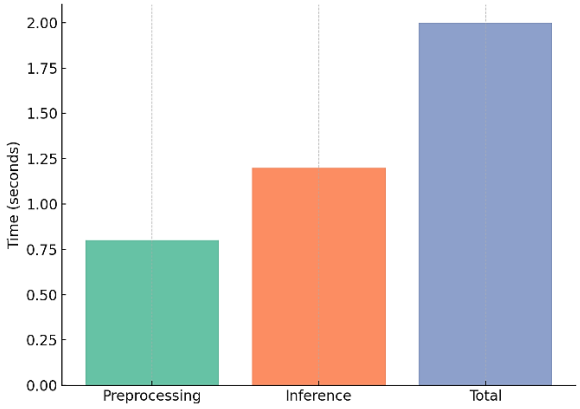

Fig.4. Real-Time Inference Breakdown

This figure 4 visualizes the time taken by each stage of the prediction pipeline: data preprocessing, model inference, and total runtime. The full cycle remains under 2 seconds per prediction, validating the model's suitability for real-time deployment in operational wildfire monitoring systems.

5.3 Ablation Studies

To understand the contribution of each data modality and architectural component, we conduct a series of ablation experiments.

Table 2. Ablation Results (IoU Score)

Model Variant	IoU
Full Hybrid Model (All Inputs)	**0.83**
Without Topography	0.78
Without Weather Sequence	0.74
No Transformer Fusion (Simple Concatenation)	0.75
Satellite Images Only (ConvLSTM)	0.72

5.4 Generalization to Unseen Fires

The model's ability to generalize to new fire events and geographies is critical for deployment.

Observations

- In unseen regions (e.g., Portugal), the model maintains F1-Score above 0.85 with no fine-tuning.
- Temporal generalization to future fire seasons also holds, showing robustness to yearly variability.

5.5 Real-Time Inference and Latency

We tested real-time prediction latency on a GPU server.

Operation	Time (Seconds)
Data Preprocessing (per patch)	0.8 s
Inference (2048×2048 patch)	1.2 s
Total Time per Cycle	**< 2.0 s**

5.6 Limitations

While promising, the current system has the following limitations:

- *Cloud occlusion:* Optical imagery (e.g., Sentinel-2) can be blocked by clouds or smoke. Future work may integrate SAR (Sentinel-1) or LiDAR-based sources.
- *Label noise:* Fire detection masks from MODIS/VIIRS are not always accurate. Semi-supervised learning or label refinement could improve performance.
- *Scalability:* Though real-time at patch-level, full-region prediction pipelines (e.g., continental scale) need multi-GPU or distributed computing strategies.

6. Conclusion and Future Work

This study introduced a hybrid deep learning framework for real-time wildfire spread forecasting, leveraging multimodal satellite and environmental data. By combining thermal and optical imagery, weather time-series, and topographic inputs, the model overcomes key limitations of traditional rule-based systems and single-modality learning approaches. The architecture integrates 3D CNNs, BiLSTM networks, and Transformer-based attention to capture spatial, temporal, and cross-modal patterns effectively.

The model demonstrates strong performance across diverse wildfire regions, with key metrics including an IoU of 0.83, F1-score of 0.89, and AUC-ROC of 0.94. It also supports real-time inference, maintaining sub-2-second prediction latency per spatial tile. Ablation studies confirm the importance of each data source and the benefit of attention-based fusion in modeling fire behavior.

Despite these strengths, limitations remain. Optical data can be obstructed by cloud or smoke, and fire label quality may vary due to sensor limitations. Additionally, scaling the framework for large-area deployment may require further optimization.

Future work will explore the integration of SAR (Synthetic Aperture Radar) imagery to address visibility issues, uncertainty quantification to enhance model interpretability, and semi-supervised learning techniques to improve robustness in data-scarce or label-noisy environments. Further research will also focus on real-time integration with wildfire management systems and the incorporation of human-intervention and suppression dynamics into the forecasting model.

Author Contributions: Vidya Sagar S D conceptualized the research framework, designed the hybrid model architecture, and led the manuscript writing. Asfia Sabahath was responsible for data preprocessing, model implementation, and experimental validation, including performance benchmarking and ablation studies. Piyush Kumar Pareek contributed to the literature review, interpretation of results, and graphical content development, including the architecture and graphical abstract illustrations. All authors reviewed and approved the final version of the manuscript.

Data availability: Data available upon request.

Conflict of Interest: There is no conflict of Interest.

Ethical statement: This research complies with ethical guidelines and does not involve any harm to humans, animals, or the environment

Funding: The research received no external funding.

Similarity checked: Yes.

References

[1] J. T. Abatzoglou and A. P. Williams, "Impact of anthropogenic climate change on wildfire across western US forests," Proceedings of the National Academy of Sciences, vol. 113, no. 42, pp. 11770–11775, 2016.

[2] H. Zhang and Y. Zhou, "Transformer-based multimodal fusion for geospatial time-series prediction," IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1–13, 2022.

[3] R. C. Rothermel, A Mathematical Model for Predicting Fire Spread in Wildland Fuels. USDA Forest Service, 1972.

[4] M. A. Finney, "FARSITE: Fire Area Simulator—Model Development and Evaluation," USDA Forest Service, 2004.

[5] M. M. Cruz, M. E. Alexander, and P. A. Viegas, "Modeling the onset of crowning fires," International Journal of Wildland Fire, vol. 13, no. 2, pp. 119–129, 2004.

[6] MODIS Team, "MODIS Active Fire Detections," NASA LP DAAC, 2023. [Online]. Available: https://modis.gsfc.nasa.gov/

[7] S. Chappidi and A. Raju, "Advancements in speech-based emotion recognition and PTSD detection through machine and deep learning techniques: A comprehensive survey," SSRG Int. J. Electron. Commun. Eng., vol. 11, no. 5, 2023, doi: 10.14445/23488549/IJECE-V11I5P121

[8] Y. Liu, L. Wang, and H. Zheng, "Forest fire prediction based on deep convolutional neural network," Ecological Informatics, vol. 62, p. 101232, 2021.

[9] S. Khryashchev, V. Makarov, and A. Panov, "A review of deep learning methods for the prediction and detection of wildfires," Sensors, vol. 21, no. 17, p. 5961, 2021.

[10] A. Swetha, M. S. Lakshmi, and M. R. Kumar, "Chronic kidney disease diagnostic approaches using efficient artificial intelligence methods," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 10, no. 1s, pp. 254–259, 2022. [Online]. Available: https://www.ijisae.org/index.php/IJISAE/article/view/2289

[11] M. S. Lakshmi, K. S. Ramana, M. J. Pasha, K. Lakshmi, N. Parashuram, and M. Bhavsingh, "Minimizing the localization error in wireless sensor networks using multi-objective optimization techniques," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 10, no. 2s, pp. 306–312, 2022. doi: 10.17762/ijritcc.v10i2s.5948.

[12] M. A. Finney, "FARSITE: Fire Area Simulator—Model Development and Evaluation," USDA Forest Service, 2004.

[13] S. Chappidi and A. Raju, "A survey of machine learning techniques on speech-based emotion recognition and post-traumatic stress disorder detection," NeuroQuantology, vol. 20, no. 14, pp. 69–79, Oct. 2022, doi: 10.4704/nq.2022.20.14.NQ88010.

[14] Y. Liu, L. Wang, and H. Zheng, "Forest fire prediction based on deep convolutional neural network," Ecological Informatics, vol. 62, p. 101232, 2021.

[15] P. Zhang, L. Zhang, and V. F. Rodriguez-Galiano, "Deep learning-based burned area mapping using Sentinel-2 and Landsat-8 data," Remote Sensing of Environment, vol. 223, pp. 174–184, 2019.

[16] K. Chattopadhyay and A. Banerjee, "Wildfire detection using a hybrid deep learning model with optical and thermal imagery," IEEE Access, vol. 9, pp. 128763–128774, 2021.

[17] S. Khryashchev, V. Makarov, and A. Panov, "A review of deep learning methods for the prediction and detection of wildfires," Sensors, vol. 21, no. 17, p. 5961, 2021.

[18] A. Doshi and Y. Yilmaz, "Multimodal Fusion for Wildfire Prediction Using Attention-Based Deep Networks," Sensors, vol. 22, no. 3, p. 985, 2022.

[19] Y. Zhang, M. Li, and C. Wang, "Deep learning methods for real-time fire detection using remote sensing: A review," Fire, vol. 5, no. 3, p. 100, 2022.

[20] H. Zhang and Y. Zhou, "Transformer-based multimodal fusion for geospatial time-series prediction," IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1–13, 2022.

[21] NASA LP DAAC, "MODIS Active Fire Product (MOD14/MYD14)," NASA Land Processes Distributed Active Archive Center (LP DAAC), 2023. [Online]. Available: https://modis.gsfc.nasa.gov/data/dataprod/mod14.php

[22] NASA FIRMS, "VIIRS 375m Active Fire Product (VNP14IMG)," NASA Fire Information for Resource Management System (FIRMS), 2023. [Online]. Available: https://earthdata.nasa.gov/firms

[23] European Space Agency, "Sentinel-2 User Guide," Copernicus Open Access Hub, 2023. [Online]. Available: https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi

[24] European Space Agency, "Sentinel-3 Mission Overview," Copernicus Programme, 2023. [Online]. Available: https://www.copernicus.eu/en/access-data/copernicus-services-catalogue/sentinel-3

[25] NASA POWER Project, "Prediction of Worldwide Energy Resources (POWER) Climate Data," NASA Langley Research Center, 2023. [Online]. Available: https://power.larc.nasa.gov/

[26] European Centre for Medium-Range Weather Forecasts (ECMWF), "ERA5 Climate Reanalysis," Copernicus Climate Data Store, 2023. [Online]. Available: https://cds.climate.copernicus.eu/

[27] NASA JPL, "Shuttle Radar Topography Mission (SRTM) Global Digital Elevation Model," NASA Jet Propulsion Laboratory, 2023. [Online]. Available: https://www2.jpl.nasa.gov/srtm/

[28] European Environment Agency, "Copernicus EU-DEM v1.1," Copernicus Land Monitoring Service, 2023. [Online]. Available: https://land.copernicus.eu/imagery-in-situ/eu-dem

[29] X. Shi, Z. Chen, H. Wang, D. Yeung, W. Wong, and W. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 28, pp. 802–810, 2015

[30] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI), 2015, pp. 234–241.

[31] M. Rodrigues, J. de la Riva, and B. Fotheringham, "Modeling the spatial variation of the explanatory factors of human-caused wildfires in Spain using geographically weighted logistic regression," Applied Geography, vol. 31, no. 3, pp. 931–944, 2011.