IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 12 March 2026, accepted 25 March 2026, date of publication 2 April 2026, date of current version 7 April 2026.

Digital Object Identifier 10.1109/ACCESS.2026.3679978

RESEARCH ARTICLE

Wildfire Spread Forecasting With Deep Learning

NIKOLAOS ANASTASIOU[1], SPYROS KONDYLATOS[1,2], AND IOANNIS PAPOUTSIS[1,3]
[1]Orion Laboratory, National Observatory of Athens, National Technical University of Athens
[2]Image Processing Laboratory (IPL), Universitat de València
[3]Archimedes, Athena Research Center

Corresponding author: Nikolaos Anastasiou (nikolas619065@gmail.com)

This work was supported in part by the Project: Small-Satellites implemented by the Hellenic Ministry of Digital Governance with European Space Agency (ESA) Assistance in the Management and Implementation under Grant 16855; and in part by the National Recovery and Resilience Plan 'Greece 2.0', through the Recovery and Resilience Facility (RRF), Core Program of European Union-NextGenerationEU.

ABSTRACT Accurate prediction of wildfire spread is crucial for effective risk management, emergency response, and strategic resource allocation. In this study, we present a Deep Learning (DL)-based framework for forecasting the final extent of burned areas, using data available at the time of ignition. We leverage a spatio-temporal dataset that covers the Mediterranean region from 2006 to 2022, incorporating remote sensing data, reanalysis meteorological observations, vegetation maps, land cover classifications, topography data, and thermal anomalies. To evaluate the influence of temporal context, we conduct an ablation study examining how the inclusion of pre- and post-ignition data affects model performance, benchmarking the temporal-aware DL models against a baseline trained exclusively on ignition-day inputs. Our results indicate that multi-day observational data substantially improve predictive accuracy. Particularly, the best-performing model, incorporating a temporal window of four days before to five days after ignition, improves both the F1 score and the Intersection over Union by almost 5% in comparison to the baseline on the test dataset. We publicly release our dataset and models to enhance research into data-driven approaches for wildfire modeling and response.

INDEX TERMS Deep learning, vision transformer, wildfire spread, forecasting.

I. INTRODUCTION

Wildfires are natural disturbances that play a crucial role in maintaining ecosystem health and biodiversity, offering important ecological benefits [1]. Nevertheless, they also pose significant threats to human life, infrastructure [2], and public health, due to the release of particulate matter and toxic pollutants [3]. With climate change projected to intensify both the frequency and severity of wildfires [5], the development of resilient systems for enhancing wildfire management has become increasingly urgent.

One of the most critical wildfire-related applications is the forecasting of wildfire spread, which aims to predict the future evolution of a wildfire following its ignition [6]. The implementation of fire spread models varies considerably between studies, particularly in terms of the spatial and temporal resolution. Thus, applications range from near-real-time hourly forecasts [7] to cumulative estimates of the total burned area throughout an entire fire event [8], and even to early predictions of final fire extent immediately after ignition [9].

Remote sensing imagery, when combined with weather measurements, can play a vital role in identifying regions at high risk of ignition and spread [10]. However, the accuracy of these predictions is constrained by the spatial resolution of meteorological forecasts, along with the revisit frequency and coverage of satellites [11]. As a result, unmanned aerial vehicles (UAVs) are often employed for high-resolution fire monitoring [12], whereas satellites are better suited for broader, multi-day forecasts across large geographic areas [13].

Wildfire spread is driven by complex, non-linear interactions between atmospheric conditions, fuel properties (such as vegetation type and density), and topography. Wind, in particular, has a significant effect on the direction and extent of burned areas [14]. Historically, wildfire spread prediction has been estimated using empirical systems like the Canadian Forest Fire Behavior Prediction System [15].

The associate editor coordinating the review of this manuscript and approving it for publication was Monidipa Das.

© 2026 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License.
For more information, see https://creativecommons.org/licenses/by/4.0/

More recent physics-based models, such as WRF-SFIRE [7], have enabled near real-time burned area forecasts [16], relying on high-resolution meteorological inputs.

Lately, advances in Machine Learning (ML) and Deep Learning (DL) have opened new avenues for data-driven approaches to wildfire modeling [17], showing great promise in predicting wildfire spread [18]. These methods can be trained in large-scale datasets—including satellite imagery, meteorological records, and historical fire perimeters—to learn patterns and capture the complex spatio-temporal behavior of wildfires.

Motivated by these recent advancements, this study explores DL-based approaches for predicting wildfire spread, specifically aiming to estimate the final extent of burned areas starting from an ignition point, using data available at the time of ignition and future meteorological conditions. This study utilizes ERA5 reanalysis data, which provide physically consistent, observation-based weather conditions generated by a static ECMWF IFS model [19]. Using ERA5 enables the models to learn stable relationships between fire behavior and environmental drivers, without the additional variability introduced by forecast uncertainty. While operational deployment would rely on numerical weather forecasts from evolving NWP systems, our framework is directly applicable for scenario-based simulations that support preparedness planning, resource allocation, and anticipatory mitigation strategies. Extending this framework to emergency response would require models explicitly designed to remain robust under the distribution shifts imposed by operational NWP forecasts.

The proposed approach captures the spatiotemporal context surrounding wildfires, leveraging satellite imagery, meteorological data, and topographic information. We explore both Vision Transformers (ViTs), Convolutional Neural Networks (CNNs) and ConvLSTM-based architectures using input features from the ignition and preceding days, as well as information for the post-ignition period. The main contributions of this work can be summarized as follows:

- Building on the Mesogeos datacube [20], we develop a novel, large-scale ML-ready dataset comprising approximately 9,500 fire events, containing spatio-temporal information of fire drivers and curated for wildfire spread analysis. The dataset is designed to be readily applicable for training and evaluating data-driven models.
- We train DL models that predict the final burned area, using the available state of relevant wildfire-driving variables at the time of the ignition. Additionally, we provide a quantitative and qualitative evaluation of model predictions, including visual comparisons of predicted and actual burned areas for selected events to assess the practical utility of the predictions.
- We perform an ablation study to assess the effect of varying temporal input windows on model performance. Our findings demonstrate that increasing post-ignition information context on the input variables significantly enhances predictive accuracy.

To promote reproducibility and facilitate future research, the dataset and source code for the developed models are being made publicly available at https://github.com/Orion-AI-Lab/WildFireSpread.

II. RELATED WORK

Early wildfire modeling efforts primarily relied on physics-based and mathematical models to simulate fire behavior and spread. One widely used operational approach is the Rothermel model [21], which characterizes fire spread as a function of fuel properties, terrain, and environmental conditions, and has been extensively adopted in operational fire behavior systems.

Recently, DL techniques have been increasingly applied in a range of wildfire-related applications, including wildfire danger forecasting, active fire detection, burned area mapping, and fire spread prediction [22]. Le et al [23] used a Multilayer Perceptron for predicting wildfire danger, while other studies use CNNs to model forest fire susceptibility [24], [25], [26], [27]. Kondylatos et al. [28] used models like Long Short-Term Memory (LSTM) and Convolutional LSTM (ConvLSTM) to predict short-term wildfire danger, demonstrating improved performance over traditional indices such as the Fire Weather Index (FWI) [29], and used explainable Artificial Intelligence to uncover the main drivers that influence the occurrence of fires. CNNs have also proven effective for the rapid detection of active wildfires from aerial vehicle thermal imagery [30], [31] and satellite images, which cover a broader spatial extent [32] with frequent data updates [33]. Recently, a transformer-based solution was proposed to segment active fire pixels from the VIIRS satellite, demonstrating improved performance compared to other spatial and temporal-aware models [34]. Moreover, DL models have accurately segmented burned areas from high-resolution multispectral imagery (e.g., Sentinel-2), often outperforming traditional methods like Normalized Burn Ratio (NBR) [35] thresholding, especially in handling varying burn severities [36], [37].

A. WILDFIRE SPREAD

Recent studies have explored a variety of DL approaches for modeling wildfire spread, with a strong emphasis on capturing the involved complex spatial and temporal dynamics. Shadrin et al. [11] evaluated several CNN-based architectures for wildfire spread prediction, emphasizing the effectiveness of convolutional models in learning spatial patterns from geospatial data. Large-scale datasets such as Next Day Wildfire Spread [13] and WildfireSpreadTS [38] have been introduced, enabling the daily prediction of wildfire spread and the benchmarking of DL models in this task. FireCast [18] used CNNs trained on historical fire data to identify high-risk zones for imminent spread. Burge et al. [39] applied a Convolutional Long Short-Term Memory (ConvLSTM) network to model wildfire propagation dynamics using fire spread simulation datasets.

Hodges and Lattimer [40] proposed a Deep Convolutional Inverse Graphics Network to predict fire spread up to six hours in advance. Marjani et al. [8] combined convolutional layers with Bidirectional Long Short-Term Memory units (CNN-BiLSTM), effectively capturing spatio-temporal dependencies in wildfire spread data. Coffield et al. [41] employed traditional ML models, including decision trees, to estimate final fire size at the time of ignition. Li and Rad [42] introduced a combination of a spatiotemporal transformer with a goal-based decision model to capture the complex relationships between various factors influencing the fire spread and provide actionable insights for management. Zhou et al. [43] provided guidance for selecting suitable wildfire prediction model by comparing CNNs and transformer based models and enhanced model transparency through an integrative xAI. More recently, Kondylatos et al. [20] addressed the same problem using CNNs, relying solely on input features from the day of ignition.

While the majority of prior work has emphasized predicting the spatial extent of wildfires, some studies focus on estimating the final numerical scale of wildfires. For example, Liang et al. [44] employed an LSTM model with meteorological inputs to classify final fire sizes into categorical levels, underscoring the critical influence of weather conditions on fire outcomes.

Despite recent advancements in wildfire modeling, to the best of our knowledge, no study has leveraged advanced DL architectures alongside large-scale, multi-year spatio-temporal datasets to predict the final burned area of wildfires. Furthermore, there is no systematic investigation into the optimal temporal window of input data, spanning both pre- and post-ignition periods, for this specific task. This work addresses these gaps by examining the integration of ViT and CNN-based models with temporally-structured RS and meteorological data to forecast the final fire extent at the time of ignition.

III. DATASET

Training DL models for wildfire spread prediction requires high-quality, high-resolution data. For this study, a dedicated dataset was created by extracting and processing data from Mesogeos [20], a large-scale multi-purpose dataset for wildfire modeling in the Mediterranean.

A. VARIABLES

Mesogeos comprises 27 variables at a spatial resolution of 1×1 km and a daily temporal resolution spanning the period from 2006 to 2022. It includes 14 dynamic (time-varying) and 14 static (time-invariant) features, encompassing meteorological parameters, land cover, vegetation indices, and topographical characteristics. The variables selected for this study were chosen based on their established relevance to wildfire ignition and spread (Tab. 6 in Appendix).

The spatial and temporal resolutions of the dataset are determined by the Mesogeos DataCube. Given its large-scale coverage of the Mediterranean region, encompassing 28 variables over 16 years, certain limitations exist, including the spatial and temporal resolutions which are set at 1×1 km and 1 day respectively.

Weather conditions play a crucial role in determining the rates of fuel drying, directly affecting fire occurrence and spread. The dataset includes key atmospheric variables such as wind speed and direction, air temperature (at 2 meters), dew point temperature (at 2 meters), relative humidity, surface pressure, total precipitation (24-hour), and surface solar radiation downwards. Among these, wind speed and direction are particularly critical, as fire spread is highly sensitive to wind dynamics. Relative humidity provides a proxy for fuel moisture content, while temperature and solar radiation affect both fuel drying and fire intensity. High precipitation can play a mitigating role by reducing fuel flammability and limiting fire propagation.

To enhance the dataset's suitability for fire spread modeling, we derived some additional variables. In particular, wind speed was resolved into its eastward (U) and northward (V) vector components, enabling a more precise representation of wind-driven fire behavior.

$$U = W\cos(\theta)$$
$$V = W\sin(\theta)$$

where:

- W is the Wind Speed (in m/s),
- θ is the Wind Direction (in degrees).

Vegetation characteristics are key indicators of fuel type and availability. The dataset includes variables such as the Normalized Difference Vegetation Index (NDVI), Leaf Area Index (LAI), and surface soil moisture. High NDVI and LAI values signal dense vegetation and greater fuel loads, while low soil moisture increases fuel flammability, promoting easier ignition and more intense burning.

Static variables describing the land surface and terrain also play a crucial role in shaping fire spread dynamics. Land cover classification identifies dominant vegetation types and surface features, such as agricultural land, forests, grasslands, shrubland, sparse vegetation, settlements, and water bodies—each with distinct flammability and fire propagation characteristics. For instance, settlements and water bodies often act as natural barriers to the spread. Topographical features, including elevation, slope, aspect, and curvature, also play a vital role, with fires typically spreading more rapidly uphill due to convective heat and terrain-driven wind patterns.

Concerning the information related to the wildfires, two key binary variables are included: the ignition point and the final burned area mask. The ignition point is encoded as a binary image, with a value of 1 at the pixel where the fire originated and 0 elsewhere. The burned area mask serves as the prediction target, represented as a binary image in which burned pixels are labeled as 1 and unburned pixels as 0.

B. STUDY AREA

The dataset covers the Mediterranean region (Fig. 1), spanning an area of approximately $3,760,000\ km^2$. This region is of particular interest for wildfire research due to its characteristic Mediterranean climate, characterized by hot and dry summers and a high frequency of fire activity [45]. Wildfires in this region contribute significantly to carbon emissions [46], and are expected to become more frequent and severe under ongoing climate change [47]. The dataset includes fire events from multiple countries within this region, capturing a wide range of landscapes and climatic sub-regions. This geographic diversity enables the analysis of cross-border fire patterns beyond those possible with single-country datasets.

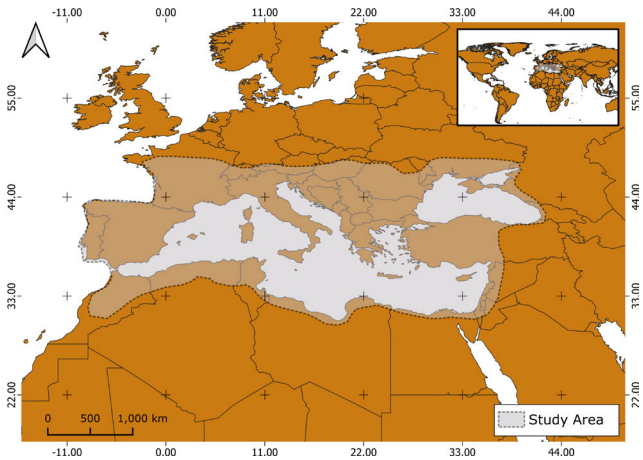

FIGURE 1. Location of the study area (Coordinate System: WGS84).

C. DATASET EXTRACTION AND PREPROCESSING

For this study, we derived an ML–ready dataset from the Mesogeos datacube. The dataset was generated by extracting spatio-temporal patches centered around historical wildfire ignition points. Specifically, for each fire event, a 64 × 64 pixel (64 × 64 km) patch was extracted, incorporating all the variables mentioned in III-A. To mitigate potential spatial bias and avoid always centering the burned area, a small random offset was applied to the patch center relative to the recorded ignition point.

To capture the temporal dynamics relevant to wildfire spread, data were extracted over a temporal window spanning from four days before the ignition day to five days after. Longer-term forecasts are constrained by the current accuracy of meteorological forecasts under operational conditions. As a result, each sample (fire event) is a spatio-temporal datacube with dimensions of (time steps × height × width) = (10 × 64 × 64).

Several post-processing steps were conducted to ensure data integrity. Missing values—primarily due to ERA5-LAND's land-only coverage affecting areas over water—were imputed with zeros. Samples with substantial coverage over water bodies were excluded to maintain focus on terrestrial wildfire behavior. The final dataset comprises 9,568 wildfire events.

D. DATASET STATISTICS

To provide insights into the dataset composition, several descriptive statistics were computed.

Figure 2 displays the annual distribution of fire events from 2006 to 2022. An upward trend is observed in recent years, particularly after 2017 (excluding 2018), where the number of recorded events exceeds 650 per year. In contrast, earlier years contain significantly fewer samples.

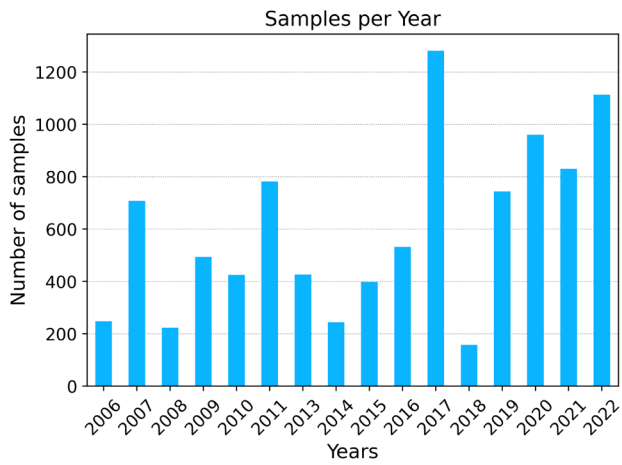

FIGURE 2. Yearly distribution of fire events in the dataset.

Figure 3 presents the cumulative monthly distribution of fire events aggregated across all years. As expected in a Mediterranean climate, the majority of fires occur during the summer months, with July and August exhibiting the highest frequency, followed by March and September, though at notably fewer occurrences.

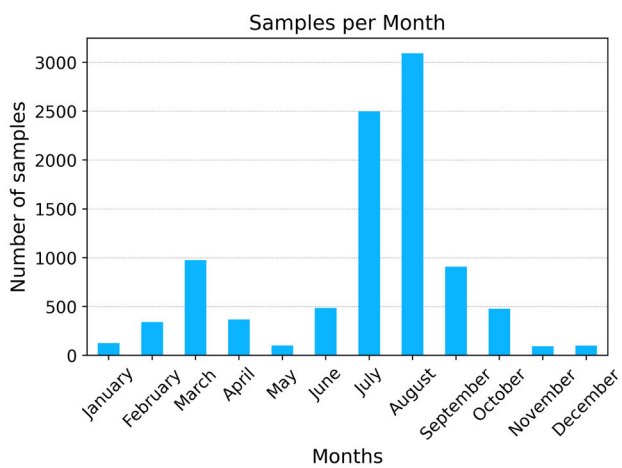

FIGURE 3. Monthly distribution of fire events across all years in the dataset.

To assess the variability in wildfire sizes, fire events were grouped using k-means clustering [48] based on their final burned area. An optimal cluster count of $k = 10$ was selected to adequately capture the distribution of fire sizes. Figure 4 presents the number of samples in each cluster, with ranges defined by the minimum and maximum burned area (in hectares) per cluster. The results reveal a significant class imbalance, with the majority of fire events concentrated in the smallest size cluster (<1500 ha), while in contrast, larger fires are underrepresented. This class imbalance is a known challenge in data-driven wildfire modeling [49].

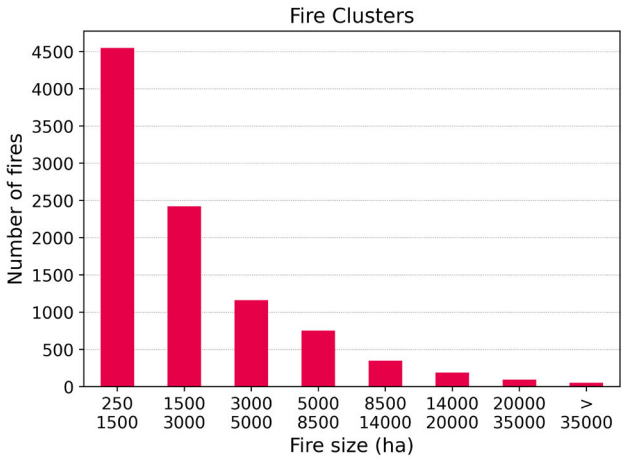

FIGURE 4. Number of samples (fire events) per fire cluster in the dataset, represented in hectares (ha).

While common imbalance techniques like oversampling (e.g., SMOTE) [50] are widely used, they were intentionally avoided in this study. For a physical process like wildfire spread, artificially generating additional large-fire samples may produce events that are not physically plausible. Training on such synthetic instances could distort model learning and ultimately degrade generalization to real-world conditions [51]. To mitigate the imbalance without compromising physical realism, we employ a BCEDice loss, which suppresses the dominance of unburned pixels and encourages learning of burned-area patterns across all fire sizes.

The distribution of fire sizes in the training and test sets is presented in Figure 5. Red bars denote the proportion of fires in each cluster within the training set, while blue bars represent the test set. Fires ranging from 250 to 1500 ha account for 39% of the training set and approximately 41% of the test set. This consistent pattern extends across all fire size categories, indicating that both the training and test sets contain a well-balanced distribution of fire samples across size classes.

Fire samples were categorized according to climate zones following the classification of Kottek et al. [52]. All 9,500 samples originate from the Mediterranean region, which encompasses diverse climate types. Table 1 reports the percentage of fire events within each zone. The majority of

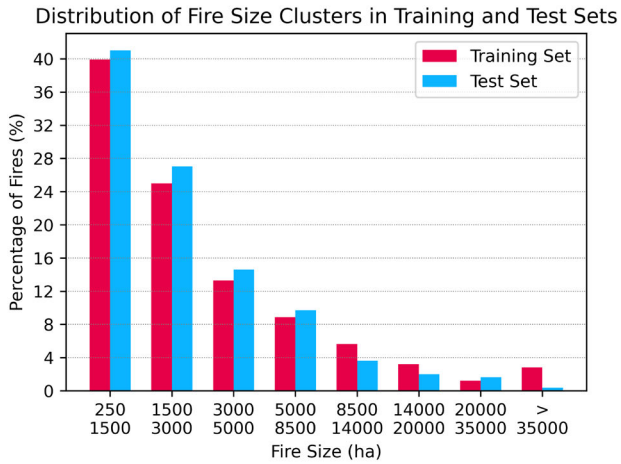

FIGURE 5. Comparison of fire size cluster distributions between training and test sets.

TABLE 1. Percentage of fire events across climate zones within the study area (see Figure 1).

Climate Zones	Percentage %	Description KöppenGeiger Classification
Csa	45.2	Warm Temperate, Dry & Hot Summer
Cfb	21.6	Warm Temperate, Fully Humid, Warm Summer
Csb	19.6	Warm Temperate, Dry & Warm Summer
Cfa	7.1	Warm Temperate, Fully Humid, Hot Summer
BSk	4.6	Arid, Steppe, Cold Arid
Other	1.9	

fires (45.2%) occurred in the Csa climate zone, characterized by warm temperatures and hot, dry summers and its widely distributed across countries such as Spain, Portugal, Italy, Greece, Turkey, and parts of the Balkans. Both the Cfb and Csb zones, which account for approximately 20% of the fires each, are marked by warm temperatures and mild summers; however, the Csb zone is distinguished by dry summers, whereas the Cfb zone is fully humid. In contrast, the Cfa and BSk zones each contributed less than 10% of the fire events, exerting relatively limited influence on the dataset. Overall, these results indicate that most fire events originate in regions with warm climates and dry to warm summers.

IV. METHODOLOGY

This section first details the experimental setup, followed by a description of the models developed, and finally outlines the evaluation procedure followed in this study.

A. EXPERIMENTAL SETUP

Each data sample comprises 27 variables, including both static and dynamic features. The final model input consists of 14 dynamic variables (excluding burned areas), each recorded over 10 time steps, and 12 static variables loaded once per sample. The ignition point is incorporated only on the day of ignition, while for all other days it is represented by a

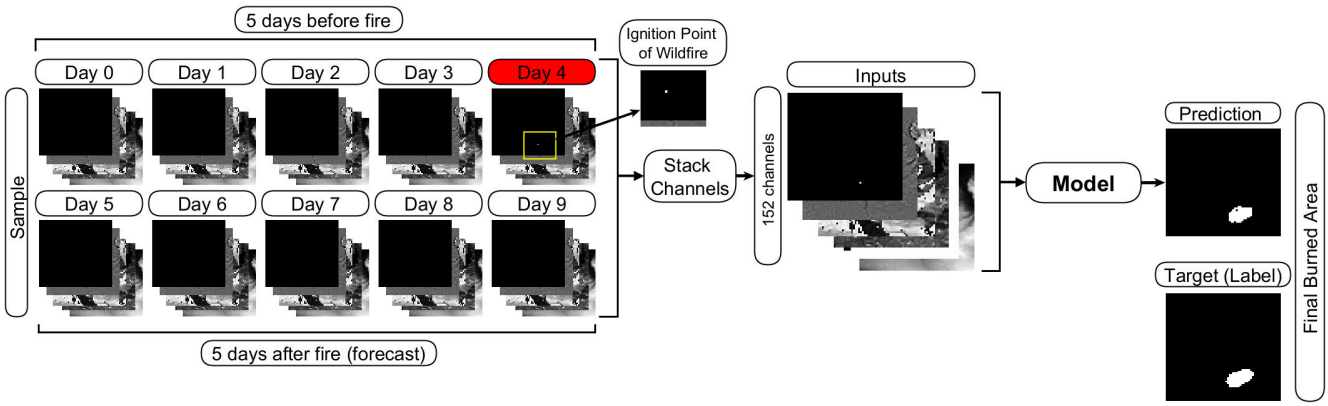

FIGURE 6. Fire event (sample) structure shown in left, containing 4 days before fire ignition day, day 4 (ignition day), and five days after ignition day. Each day incorporates 27 variables, including both dynamic and static variables. The model takes as input 152 channels (including the ignition point) representing a 10-day temporal window, and produces a 64 × 64 pixels binary masks, where burned pixels are labeled as 1 and unburned pixels as 0.

zero-valued array. The 10 time steps cover a temporal window from 4 days before to 5 days following the fire ignition, forming a time series. Left side of Figure 6 illustrates the structure of the data for a single fire event (sample). All variable data are normalized to the range [0, 1].

The final burned area variable serves as the prediction target (label) for all models. The model output is a binary segmentation mask with dimensions identical to the input channels (64 × 64 pixels), where burned pixels are labeled as 1 and unburned pixels as 0. This mask represents the predicted final burned area resulting from a wildfire event, as shown in the right side of Figure 6. The dataset exhibits significant class imbalance between burned and unburned pixels; within each 64 × 64 patch (4096 pixels in total), on average, only about 100 pixels correspond to burned areas, with the remainder being unburned.

For training and evaluation, we adopted a temporal split. Fire events from 2006 to 2020 were allocated to the training set, the year 2021 was used for validation during model training to monitor performance and tune hyperparameters, and the year 2022 was reserved as the test set. This resulted in 7619 fire events for training, 841 for validation, and 1101 for testing.

B. ALGORITHMS

CNNs are well-suited algorithms for extracting hierarchical spatial features, making them a natural choice for modeling wildfire spread patterns. However, standard 2D CNNs process each variable at each time step independently, limiting their ability to directly capture temporal dependencies inherent in fire spread dynamics. To address this limitation, each input sample was structured as a multi-channel image, where each channel corresponds to a specific variable at a given time step. This design effectively makes use of temporal information as additional input dimensions for 2D CNN models. The CNN models in this study were based on the U-Net [53] architecture.

A principal 2D U-Net model was trained using the full 10-day temporal window. This input comprises 152 channels (14 dynamic variables × 10 time steps + 12 static variables) stacked from each sample as illustrated in the middle of Figure 6. The inclusion of pre-ignition data allows the model to learn pre-fire environmental conditions indicative of fire spread risk, while post-ignition data capture critical changes in dynamic factors, such as wind patterns, that directly influence the final burned area extent and direction.

A modified 3D U-Net [54] architecture was trained using the same spatio-temporal dataset. Unlike the 2D U-Net, the 3D U-Net employs 3D convolutional filters that simultaneously process spatial and temporal dimensions. This architecture inherently models temporal dependencies more directly and effectively than the 2D channel-stacking approach, enabling improved feature learning across consecutive time steps.

These two models were compared to a baseline model that was implemented using a modified 2D U-Net architecture trained exclusively on data from the ignition day. This model utilized 26 input channels (excluding the burned area label) and aimed to establish the predictive capability achievable with minimal temporal context.

Figure 7 presents the U-Net architecture employed in this study, which was modified from the original U-Net design to better accommodate the characteristics of the input data. Specifically, two encoder and decoder blocks were removed, resulting in a shallower network architecture, while an extra convolutional layer was added within each convolutional block to enhance local feature representation, as shown in Fig. 8. This architectural adjustment was motivated by the relatively low spatial resolution of the input samples (64 × 64 pixels), which does not require the extensive hierarchical feature extraction provided by a full U-Net with numerous pooling layers. Instead, the primary complexity in this task arises in capturing patterns between the numerous temporal and variable channels.

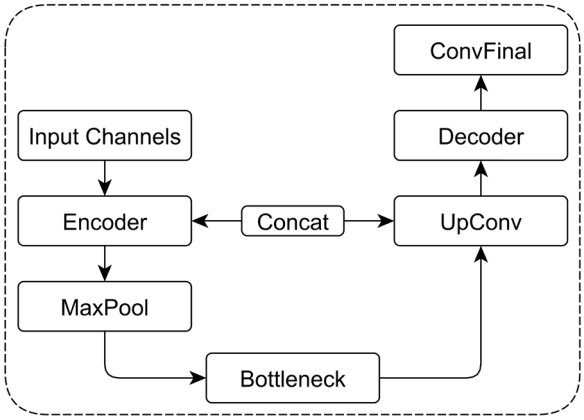

FIGURE 7. The modified architecture for both U-Net2D and U-Net3D consists of an Encoder Block followed by a MaxPooling Layer and a Bottleneck. This is followed by an UpConvolution Layer, which is concatenated with the corresponding Encoder Block, and a Decoder Block, culminating in a final Convolution Layer. The Encoder, Bottleneck, UpConvolution, and Decoder Blocks each include a Convolution Block, which is illustrated in Fig. 8.

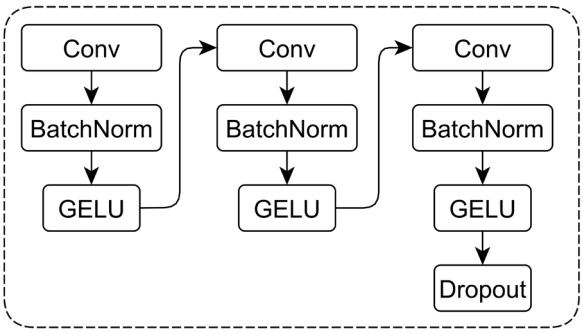

FIGURE 8. The modified Convolution Block used in both U-Net2D and U-Net3D models consists of a Convolution Layer followed by Batch Normalization and a GELU (Gaussian Error Linear Unit) Activation function. This sequence is repeated twice, with a Dropout Layer added after the second repetition.

Another key architectural modification was the replacement of the standard Rectified Linear Unit (ReLU) [55] activation function with the Gaussian Error Linear Unit (GELU) [56], which was applied in both the 2D and 3D U-Net models. Additionally, the models were trained using the BCEDice Loss function [57], which combines Binary Cross-Entropy and the Dice Loss, allowing the models to classify pixels as burned or unburned while simultaneously optimizing for spatial context.

In addition to the U-Net, we also experimented with a ViT model to explore the potential of attention-based mechanisms for this task, and a ConvLSTM model to investigate the benefits of spatiotemporal feature learning.

C. EVALUATION METRICS

Model performance was evaluated using standard evaluation metrics commonly applied in binary segmentation tasks: Precision, Recall, Dice Score, F1-Score, and Intersection over Union (IoU). Notably, in the context of binary 2D segmentation, the F1-score is mathematically equivalent to the Dice Score. As such, only the Dice Score is included in the results tables for clarity. Among these scores, the Dice Score was selected as the primary criterion for hyperparameter tuning and model selection during training, due to its effectiveness in quantifying the spatial overlap between predicted and reference burned areas. While IoU is also reported, it serves as a more conservative metric, placing greater emphasis on penalizing both false positives and false negatives. The mathematical definitions of all evaluation metrics are provided below:

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

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

$$\text{F1Score} = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

$$\text{Dice Score} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} = \frac{2|A \cap B|}{|A| + |B|}$$

$$\text{IoU} = \frac{TP}{TP + FP + FN} = \frac{|A \cap B|}{|A \cup B|}$$

where:

- TP (True Positives) refers to instances correctly identified as positive.
- FP (False Positives) refers to instances incorrectly identified as positive.
- FN (False Negatives) refers to instances incorrectly identified as negative (i.e., missed positives).
- $|A \cap B|$ represents the area of intersection between the predicted region (A) and the ground truth label (B).
- $|A \cup B|$ represents the area of union between the predicted region (A) and the ground truth label (B).

V. RESULTS

This section presents the results of the DL models to highlight both overall effectiveness and the influence of key factors on model performance. First, we report the aggregate performance metrics, calculated over all 1,101 test fire events, to provide a general assessment of model accuracy. However, these metrics alone may not fully capture performance variations related to the different fire sizes. To provide deeper insight into model performance, we also report results by fire size clusters, distinguishing between small, medium, and large events. Furthermore, we present results from a series of models trained with varying temporal windows. This analysis aims to identify the optimal temporal configuration and assess the extent to which incorporating additional temporal context improves the model's ability to predict wildfire spread. Finally, we provide representative examples of the DL model predictions to enable a qualitative assessment of model performance.

A. MODELS COMPARISON

Table 2 summarizes the overall performance of the five evaluated models on the test set: the baseline ignition-day

U-Net2D, the 10-day U-Net2D, the 10-day U-Net3D, the 10-day ViT, and the 10-day ConvLSTM. The results demonstrate that the 10-day U-Net3D model achieves the highest performance, outperforming both the baseline and the 10-day U-Net2D variants. In contrast, the ViT model exhibits the lowest performance, which may be attributed to the limited dataset size relative to the data-intensive nature of transformer-based architectures. The ConvLSTM model demonstrates performance comparable to that of the U-Net2D model. Moreover, it achieves the second-best performance behind U-Net3D, highlighting that models that explicitly account for temporal components yield improved results.

TABLE 2. Evaluation results of the models on the test set. Higher performance metrics are indicated in bold.

Metric %	U-Net2D Baseline	U-Net2D 10 days	U-Net3D 10 days	ViT 10 days	ConvLSTM 10 days
Dice Score	48.3	51.7	**53.6**	43.7	52.7
IoU	31.9	34.8	**36.6**	28.0	35.7
Precision	50.2	58.6	**59.6**	59.4	59.4
Recall	46.6	46.2	**48.8**	34.6	47.3

The top-performing model achieves a Dice Score of 53.6% on the test set. The corresponding IoU is 36.6%, reflecting the stringent nature of the IoU metric in penalizing segmentation errors. Analyzing the components of the F1-Score, the model demonstrates a Precision of 59.6%, suggesting a relatively low rate of false positives and a conservative prediction of the burned area extent as evident from the low Recall rate of 48.8%. Such conservatism contributes to reducing the risk of overestimating the final burned area, while potentially underrepresenting some affected regions.

The strong performance of the DL models highlights the significant potential of data-driven approaches for modeling wildfire spread. This effectiveness can be attributed to the ability of DL models to learn from the complex, nonlinear interactions among the multiple fire drivers—such as vegetation types, land cover, topographic features, and meteorological observations. By jointly modeling these diverse inputs, DL methods offer a robust framework for capturing the multifaceted dynamics that govern fire behavior.

B. INFLUENCE OF THE TEMPORAL WINDOW

One main part of this study is to assess the impact of the temporal window size, particularly the extent of post-ignition data. For this, additional experiments are conducted by training a series of models using progressively shorter forecasting windows (Tab. 3), using the U-Net 3D as backbone architecture. Although the full dataset includes temporal sequences spanning ten days (four days before to five days after ignition), these supplementary models are limited to post-ignition data up to one, two, three, four, or five days after the fire started. Each model retained the same architecture as the primary 10-day model, only differing in the amount of post-ignition information available during training. All models were tested on the same independent test set used for the main performance comparison to ensure consistency and comparability across temporal configurations.

TABLE 3. Evaluation of model performance across different forecasting horizons of 5, 4, 3, 2, and 1 days. Results illustrate how predictive accuracy varies with the length of the forecasting window, with higher performance metrics highlighted in bold.

Metric %	U-Net3D 10 days	U-Net3D 9 days	U-Net3D 8 days	U-Net3D 7 days	U-Net3D 6 days
Dice Score	**53.6**	51.2	49.4	48.7	47.5
IoU	**36.6**	34.4	32.8	32.2	31.1
Precision	**59.6**	56.6	55.0	53.6	53.6
Recall	**48.8**	46.7	44.9	44.6	42.5

A consistent trend observed across these experiments is that model performance declines as the number of post-ignition days included in the input decreases. The model utilizing the full 10-day temporal context—including data from five days post-ignition—achieves up to a 6.1% higher Dice coefficient compared to variants using fewer forecasting days. This highlights the substantial benefit of incorporating post-ignition temporal data to enhance wildfire spread predictions.

These findings are consistent with previous research and emphasize the critical role of temporal dynamics in modeling fire behavior. Wildfires typically evolve over multiple days, particularly in larger and more complex incidents, where factors such as shifting wind conditions and interactions with topography play a significant role in determining the fire's progression. Limiting the temporal context, particularly by excluding key post-ignition periods, restricts the model's ability to capture these evolving dynamics, ultimately reducing predictive accuracy. Although this study utilizes historical post-ignition meteorological observations, it is important to note that operational deployment would depend on weather forecasts, which inherently carry uncertainties that may affect prediction accuracy. Nonetheless, our results provide compelling evidence for the substantial advantage of integrating post-fire information to improve the performance of predictive DL models.

C. RESULTS PER FIRE CLUSTER

To gain a deeper insight into how model performance varies with fire size, evaluation results are reported per each separate fire size cluster, as presented in Tab. 4, for the best-performing model (10-day U-Net 3D). The results show that the model performs well for smaller-scale events, achieving a Dice Score of approximately 59% within the first three clusters, which represent fires up to 5000 hectares. A slight drop in performance of around 2.6%, is observed in the fourth cluster, covering fires up to 8500 hectares, while for fire events exceeding this threshold, model accuracy steadily declines.

This downward trend correlates with the increasingly limited availability of training samples in the larger fire clusters, as evident from the number of samples per cluster

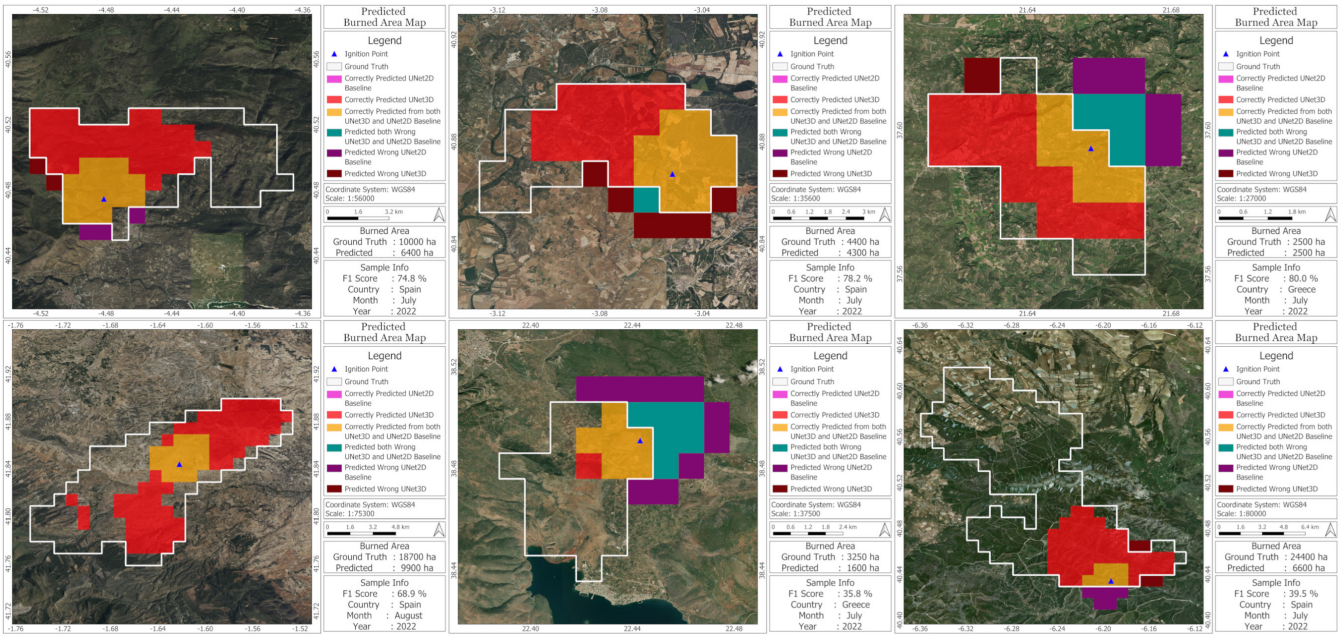

FIGURE 9. Visualizations of predicted burned area maps from the test set, comparing the best-performing 10-day U-Net3D model with the ground truth and the baseline U-Net2D model. The results demonstrate that the 10-day U-Net3D model consistently outperforms the baseline, accurately capturing the extent and spread of burned areas. It effectively predicts the majority of affected regions in both large and small fire events, with well-contained and correctly directed spread patterns.

TABLE 4. Average performance metrics per fire cluster using the 10-day U-Net 3D model, computed from test samples within each cluster.

Metric %	250 1500	1500 3000	3000 5000	5000 8500	8500 14000	14000 20000	20000 35000	> 35000
			Number	of	Samples			
Number	451	298	161	107	40	22	18	4
Dice Score	58.6	58.4	58.0	55.4	50.9	57.9	30.7	15.5
IoU	41.5	41.3	40.9	38.4	34.1	40.8	18.1	8.40
Precision	47.6	53.7	61.4	63.9	69.9	82.6	84.1	95.5
Recall	76.3	64.1	55.0	48.9	40.0	44.6	18.8	8.43

in Tab. 4. The scarcity of examples for large-scale fires likely restricts the model's ability to generalize effectively in these cases, contributing to the observed decrease in segmentation accuracy. These results highlight once again the challenge posed by class imbalance in wildfire-related applications.

D. QUALITATIVE EVALUATION

To facilitate a qualitative evaluation of the models, Figure 9 presents representative prediction maps from the test set, showcasing the outputs of the baseline model (ignition-day U-Net2D) and the best-performing model (10-day U-Net3D), alongside the ground truth fire extent. In these visualizations, the baseline model's correctly predicted burned areas is depicted in pink, the 10-day U-Net3D model's correctly predicted areas in red, and the ground truth fire boundaries in white. Regions correctly predicted, as burned, by both models are shown in orange, while those predicted wrong from both models in dark blue. Misclassifications unique to the 10-day U-Net3D are marked in dark red, whereas those unique to the baseline model are shown in dark pink.

The 10-day U-Net3D model demonstrates superior performance over the baseline for both large and small fire events. The baseline model, trained solely on ignition-day data, predicts a more uniform radial spread around the ignition, failing to accurately capture the anisotropic propagation of the fires. In contrast, the 10-day U-Net3D model, with access to multi-day temporal data—including the five days following ignition—is better equipped to capture the evolving influence of dynamic environmental conditions, such as changes in wind direction and speed.

Furthermore, the 10-day U-Net3D model shows strong capability in accurately delineating the ground truth fire extent. Across most test samples, the predicted burned area is well contained, showing only small overestimation or underestimation of the fire boundaries. The spatial progression of the fire spread is consistently aligned with the observed direction of the ground truth events, indicating once again that the model effectively captures the dynamic behavior of fire propagation.

The last two images (second row, second and third column) present a challenging scenario where both models are unable to accurately predict the final burned area, failing to reliably forecast either the overall size or the precise spatial distribution of the fire. Nonetheless, even in this case, the 10-day U-Net3D model exhibits improved performance by successfully capturing the general direction of fire spread, once again producing the most favorable results among the models evaluated.

The predictions presented in this section are deterministic and do not include explicit measures of uncertainty. As such, they should be interpreted as plausible scenario outcomes

rather than definitive forecasts. For practical use, these outputs are best viewed as indicative patterns of fire spread under given conditions, intended to support qualitative comparison across scenarios rather than precise point estimates.

VI. WILDFIRE MANAGEMENT

Given their demonstrated reliability, these models can be integrated into wildfire management systems to support emergency planning decision-making. Systems such as AEGIS [58] exemplify this approach by combining Web Geographic Information Systems (Web GIS) with geospatial data to deliver interactive maps. These maps provide critical information on variables such as fire risk, meteorological conditions, road network, etc., thereby incorporating a wildfire spread forecast model would enhance the decision making process, and optimise allocation of firefighting resources.

In such Web GIS system, once the ignition point is detected, predictions of the final fire extent can be generated based on meteorological forecasts from the subsequent five days, retrieved from weather forecast models. The forecasts can provide estimates of both the most probable direction of the wildfire spread and its potential extent (e.g., 1000 ha, 2500 ha), which can be visualized as spatial layers in the Web GIS interface with a resolution of up to 1×1 km.

This information is critical for enabling nearby communities to implement timely protective measures. Fire response managers can further employ these forecasts for risk communication, including the prioritization of evacuation plans and the dissemination of information to affected populations. Additionally, fire brigades can use such data to optimize resource allocation and adopt proactive strategies, such as reinforcing specific fire fronts or applying prescribed burning along projected pathways. Collectively, this predictive capability substantially strengthens wildfire response and decision-making.

In addition to emergency response operations, such models can also be employed for scenario-based planning applications. By simulating fires from multiple ignition points under varying meteorological and environmental conditions, decision-makers are able to evaluate a range of "what-if" scenarios prior to the onset of fire seasons. These simulations provide a basis for strategic actions such as the pre-positioning of firefighting resources, the identification of high-risk zones requiring fuel management interventions (e.g., prescribed burning or vegetation clearance), and the optimization of evacuation routes.

VII. LIMITATIONS

This study offers valuable insights into data-driven modeling of wildfire spread; however, several limitations should be acknowledged when interpreting the results.

A key temporal limitation concerns the forecasting window, which extends up to five days post-ignition. While this window aligns with the typical temporal horizon of reliable meteorological forecasts, it does not fully capture the variability in fire duration. Smaller fires often extinguish within one to two days, whereas very large fires may continue burning well beyond five days.

Another temporal constraint arises from the use of meteorological reanalysis data rather than operational forecast products. While this choice supports accurate retrospective analysis under observation-constrained conditions, it differs from real-world operational conditions in which models rely on numerical weather prediction forecasts that contain lead-time–dependent uncertainty and bias. This discrepancy introduces a distribution shift, and predictive performance obtained using reanalysis inputs should therefore be interpreted as an upper bound on expected performance under operational forecasting conditions. Consequently, a reduction in accuracy is anticipated when transitioning to forecast-driven inputs, particularly at longer lead times.

Our approach lacks explicit uncertainty quantification. Wildfire behavior prediction is inherently uncertain due to variability in environmental conditions and model assumptions. Without uncertainty estimates, the model cannot convey confidence levels, which limits its suitability for operational decision-making contexts.

Finally, the spatial resolution of 1 km $\times$ 1 km, as provided by Mesogeos, restricts the model's capacity to capture finer-scale fire dynamics that often occur at sub-kilometer scales. Although this resolution is adequate for many operational applications, the incorporation of higher-resolution data for certain input variables, such as topography, may further enhance model accuracy and spatial detail.

VIII. CONCLUSION

Wildfires continue to pose a significant and growing threat to both human communities and natural ecosystems, emphasizing the critical need for reliable predictive tools to support mitigation efforts. This study demonstrated the feasibility of forecasting the final burned extent of wildfire events from the day of ignition using DL methods.

We introduced a carefully curated, ML-ready dataset and developed DL models capable of capturing the complex spatio-temporal dynamics that govern fire behavior, demonstrating that a purely data-driven DL approach can successfully model wildfire spread. Furthermore, we showed that incorporating an extended temporal context, specifically data spanning up to five days post-ignition, substantially improved predictive performance, highlighting the vital role of temporal information in fire modeling. Moreover, the proposed models exhibited robust performance across fires of varying sizes, showing both quantitative and qualitative skill in predicting the progression of larger wildfire events. These findings highlight the potential of DL-based methods as valuable tools for operational wildfire forecasting and risk management.

Our proposed framework is directly applicable to scenario-based planning, where meteorological conditions and ignition locations are prescribed rather than forecast in real time. By generating fire spread predictions under hypothetical ignition points and environmental conditions derived from

reanalysis data, the model can support preparedness planning, strategic resource allocation, and anticipatory mitigation efforts. This planning-oriented use case enables communities and decision-makers to evaluate potential fire evolution under plausible scenarios, providing actionable insight without relying on operational weather forecasts.

Future efforts could consider integrating additional wildfire-related variables—particularly detailed vegetation maps or land cover variables— to improve model understanding of fuel availability and flammability. Moreover, uncertainty aware deep learning techniques can be considered as have been shown to slightly improve predictive performance and the expected calibration error [59]. Beyond performance gains, uncertainty quantification is essential for operational wildfire decision support, where predictions directly inform critical actions such as evacuations and resource deployment. Probabilistic outputs and confidence bounds enable decision-makers to assess risk, compare alternative response strategies, and avoid over-reliance on single deterministic predictions. Additionally, incorporating explainable AI methods can help identify the underlying factors contributing to reduced predictive accuracy in certain fire-spread cases. Finally, exploring advanced architectures such as spatiotemporal transformers, which are designed to jointly capture spatial and temporal dependencies, may lead to more accurate modeling of fire spread.

APPENDIX A

The hyperparameters employed for the proposed model are summarized in the first row of Table 5. All experiments were conducted with identical batch size, dropout rate, learning rate, and number of filters. The batch size was fixed at 32, constrained by the 10 GB VRAM of the GPU used for training. A dropout rate of 50%, a commonly adopted value for CNN's, was applied. The initial learning rate was set to 0.001, with the ReduceLROnPlateau scheduler employed to optimize models. Instead of using a fixed epoch count, training proceeded until the learning rate was reduced to a point where no significant improvement was observed, at which time the best-performing model was selected. The number of filters was fixed at 32, as increasing this value beyond 32 did not yield performance improvements.

The kernel size of (3,3,3), together with a pooling size of (1,4,4), was found to be the most effective hyperparameter configuration for the U-Net3D model. A convolutional window of (3,3,3) enables the network to capture fine-grained spatial features more effectively than larger kernels, such as (5,5,5), which may oversmooth local details. Furthermore, employing a slightly larger pooling size of (4,4), rather than the commonly used (2,2) in CNN architectures, reduces the spatial resolution more aggressively, thereby helping to mitigate overfitting.

An architectural modification involved replacing the ReLU activation function with the Gaussian Error Linear Unit (GELU) activation function [56]. Compared to ReLU, GELU provides smoother gradients and addresses the dead neuron problem by allowing a non-zero response for small negative input values.

All experiments were conducted on a workstation equipped with an NVIDIA RTX 3080 GPU (10 GB VRAM), an Intel i5-14600K CPU, and 64 GB of RAM, running Ubuntu 22.04.5 LTS. Python version used was 3.8 within an Anaconda distribution and Pytorch version used was 1.13.3+cu117.

Using this hardware configuration, the training times for the U-Net2D Baseline (1-day), U-Net2D 10-day, U-Net3D 10-day (proposed), Vision Transformer (ViT), and ConvLSTM models were approximately 4.5 min, 10 min, 20 min, 10 min, and 40 min, respectively. Inference time

TABLE 5. Performance of the proposed model across different hyperparameter settings. The kernel size is defined as a triplet (time, height, width), where the time window denotes the number of days considered before and after the target date, and the height and width correspond to the spatial dimensions of the kernel. The pool size indicates the degree of downsampling applied to the feature maps after each Convolution.

Model	Kernel Size	Pool Size	Dice / F1 Score (%)
UNet3D Proposed	(3, 3, 3)	(1, 4, 4)	53.6
UNet3D Test I	(3, 5, 5)	(1, 2, 2)	50.6
UNet3D Test II	(3, 7, 7)	(1, 2, 2)	47.7

TABLE 6. Variables used in Study. The variables are categorized into four sections, meteorological variables, vegetation maps, land cover variables and thermal anomalies.

Variable	Spatial Resolution	Temporal Resolution
Meteorological		
Max Temperature	9x9 km	1 day
U Component of Wind	9x9 km	1 day
V Component of Wind	9x9 km	1 day
Max Dewpoint Temperature	9x9 km	1 day
Max Surface Pressure	9x9 km	1 day
Min Relative Humidity	9x9 km	1 day
Total Precipitation	9x9 km	1 day
Mean Surface Solar Radiation	9x9 km	1 day
Day Land Surface Temperature	1x1 km	1 day
Night Land Surface Temperature	1x1 km	1 day
Vegetation Maps		
Vegetation Index (NDVI)	0.5x0.5 km	16 days
Leaf Area Index (LAI)	0.5x0.5 km	8 days
Soil moisture	5x5 km	10 days
Land Cover		
Fraction of Agriculture	0.3x0.3 km	-
Fraction of Forest	0.3x0.3 km	-
Fraction of Grassland	0.3x0.3 km	-
Fraction of Settlements	0.3x0.3 km	-
Fraction of Shrubland	0.3x0.3 km	-
Fraction of Sparse Vegetation	0.3x0.3 km	-
Fraction of Water Bodies	0.3x0.3 km	-
Fraction of Wetland	0.3x0.3 km	-
Topography		
Elevation	0.03x0.03 km	-
Slope	0.03x0.03 km	-
Aspect	0.03x0.03 km	-
Curvature	0.03x0.03 km	-
Thermal Anomalies		
Burned Areas	1x1 km	1 day
Ignition Points	1x1 km	1 day

was comparable across all models, with a single-sample prediction requiring approximately 1 sec.

APPENDIX B

Table 6 lists all the variables used, along with their initial spatial and temporal resolutions, categorized accordingly.

REFERENCES

[1] J. G. Pausas and J. E. Keeley, "Wildfires as an ecosystem service," *Frontiers Ecology Environ.*, vol. 17, no. 5, pp. 289–295, Jun. 2019. [Online]. Available: https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1002/fee.2044

[2] H. A. Kramer, M. H. Mockrin, P. M. Alexandre, and V. C. Radeloff, "High wildfire damage in interface communities in California," *Int. J. Wildland Fire*, vol. 28, no. 9, pp. 641–650, Sep. 2019, doi: 10.1071/wf18108.

[3] S. E. Finlay, A. Moffat, R. Gazzard, D. Baker, and V. Murray, "Health impacts of wildfires," *PLoS Currents*, p. 4, 2012, doi: 10.1371/4f959951cce2c.

[4] F. Bautista Vicente, N. Carbajal, and L. F. Pineda Martínez, "Estimation of total yearly CO_2 emissions by wildfires in Mexico during the period 1999–2010," *Adv. Meteorol.*, vol. 2014, no. 1, 2014, Art. no. 958457. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1155/2014/958457

[5] M. W. Jones, A. Smith, R. Betts, J. G. Canadell, I. C. Prentice, and C. Le Quéré, "Climate change increases the risk of wildfires," Tech. Rep., 2020.

[6] M. M. Valero, O. Rios, C. Mata, E. Pastor, and E. Planas, "An integrated approach for tactical monitoring and data-driven spread forecasting of wildfires," *Fire Saf. J.*, vol. 91, pp. 835–844, 2017.

[7] J. Mandel, S. Amram, J. Beezley, G. Kelman, A. K. Kochanski, V. Y. Kondratenko, B. Lynn, B. Regev, and M. Vejmelka, "Recent advances and applications of WRF–SFIRE," *Natural Hazards Earth Syst. Sci.*, vol. 14, no. 10, pp. 2829–2845, 2014.

[8] M. Marjani, M. Mahdianpari, and F. Mohammadimanesh, "CNN-BiLSTM: A novel deep learning model for near-real-time daily wildfire spread prediction," *Remote Sens.*, vol. 16, no. 8, p. 1467, Apr. 2024.

[9] N. Alizadeh, M. Mahdianpari, E. Hemmati, and M. Marjani, "Fusion-FireNet: A CNN-LSTM model for short-term wildfire hotspot prediction utilizing spatio-temporal datasets," *Available SSRN*, 2024.

[10] D. Richardson, A. S. Black, D. Irving, R. J. Matear, D. P. Monselesan, J. S. Risbey, D. T. Squire, and C. R. Tozer, "Global increase in wildfire potential from compound fire weather and drought," *npj Climate Atmos. Sci.*, vol. 5, no. 1, p. 23, Mar. 2022.

[11] 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," *Sci. Rep.*, vol. 14, no. 1, p. 2606, Jan. 2024.

[12] F. Afghah, A. Razi, J. Chakareski, and J. Ashdown, "Wildfire monitoring in remote areas using autonomous unmanned aerial vehicles," in *Proc. IEEE INFOCOM Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Apr. 2019, pp. 835–840.

[13] F. Huot, R. L. Hu, N. Goyal, T. Sankar, M. Ihme, and Y.-F. Chen, "Next day wildfire spread: A machine learning dataset to predict wildfire spreading from remote-sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022.

[14] J. G. Pausas and J. E. Keeley, "Wildfires and global change," *Frontiers Ecol. Environ.*, vol. 19, no. 7, pp. 387–395, 2021.

[15] K. G. Hirsch, *Canadian Forest Fire Behavior Prediction (FBP) System: User's Guide*, 1996.

[16] M. P. Kale et al., "Operational forest-fire spread forecasting using the WRF-SFIRE model," *Remote Sens.*, vol. 16, no. 13, p. 2480, Jul. 2024.

[17] M. Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, and Prabhat, "Deep learning and process understanding for data-driven Earth system science," *Nature*, vol. 566, no. 7743, pp. 195–204, Feb. 2019.

[18] D. Radke, A. Hessler, and D. Ellsworth, "FireCast: Leveraging deep learning to predict wildfire spread," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Aug. 2019, pp. 4575–4581. [Online]. Available: https://www.ijcai.org/proceedings/2019/636

[19] C. D. Roberts, R. Senan, F. Molteni, S. Boussetta, M. Mayer, and S. P. E. Keeley, "Climate model configurations of the ECMWF integrated forecasting system (ECMWF-IFS cycle 43r1) for HighResMIP," *Geoscientific Model Develop.*, vol. 11, no. 9, pp. 3681–3712, Sep. 2018.

[20] S. Kondylatos, I. Prapas, G. Camps-Valls, and I. Papoutsis, "Mesogeos: A multi-purpose dataset for data-driven wildfire modeling in the Mediterranean," in *Proc. Adv. Neural Inf. Process. Syst.*, 2023, pp. 50661–50676.

[21] R. C. Rothermel, *A Mathematical Model for Predicting Fire Spread in Wildland Fuels*, vol. 115, 1972.

[22] P. Jain, S. C. P. Coogan, S. G. Subramanian, M. Crowley, S. Taylor, and M. D. Flannigan, "A review of machine learning applications in wildfire science and management," *Environ. Rev.*, vol. 28, no. 4, pp. 478–505, Dec. 2020.

[23] H. V. Le, D. A. Hoang, C. T. Tran, P. Q. Nguyen, V. H. T. Tran, N. D. Hoang, M. Amiri, T. P. T. Ngo, H. V. Nhu, T. V. Hoang, and D. T. Bui, "A new approach of deep neural computing for spatial prediction of wildfire danger at tropical climate areas," *Ecol. Informat.*, vol. 63, Jul. 2021, Art. no. 101300. [Online]. Available: https:/www.sciencedirect.com/science/article/pii/S1574954121000911

[24] G. Zhang, M. Wang, and K. Liu, "Forest fire susceptibility modeling using a convolutional neural network for Yunnan province of China," *Int. J. Disaster Risk Sci.*, vol. 10, no. 3, pp. 386–403, Sep. 2019. [Online]. Available: http://link.springer.com/10.1007/s13753-019-00233-1

[25] G. Zhang, M. Wang, and K. Liu, "Deep neural networks for global wildfire susceptibility modelling," *Ecol. Indicators*, vol. 127, Aug. 2021, Art. no. 107735. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1470160X21004003

[26] J. R. Bergado, C. Persello, K. Reinke, and A. Stein, "Predicting wildfire burns from big geodata using deep learning," *Saf. Sci.*, vol. 140, Aug. 2021, Art. no. 105276. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925753521001211

[27] A. Bjånes, R. De La Fuente, and P. Mena, "A deep learning ensemble model for wildfire susceptibility mapping," *Ecol. Informat.*, vol. 65, Nov. 2021, Art. no. 101397. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1574954121001886

[28] S. Kondylatos, I. Prapas, M. Ronco, I. Papoutsis, G. Camps-Valls, M. Piles, M.-Á. Fernández-Torres, and N. Carvalhais, "Wildfire danger prediction and understanding with deep learning," *Geophys. Res. Lett.*, vol. 49, no. 17, p. 2022, Sep. 2022.

[29] C. E. Van Wagner, *Structure of the Canadian Forest Fire Weather Index*, vol. 1333. Ottawa, ON, Canada: Environment Canada, Forestry Service, 1974.

[30] A. Bouguettaya, H. Zarzour, A. M. Taberkit, and A. Kechida, "A review on early wildfire detection from unmanned aerial vehicles using deep learning-based computer vision algorithms," *Signal Process.*, vol. 190, Jan. 2022, Art. no. 108309.

[31] R. Ghali, M. A. Akhloufi, and W. S. Mseddi, "Deep learning and transformer approaches for UAV-based wildfire detection and segmentation," *Sensors*, vol. 22, no. 5, p. 1977, Mar. 2022.

[32] N. T. Toan, P. Thanh Cong, N. Q. Viet Hung, and J. Jo, "A deep learning approach for early wildfire detection from hyperspectral satellite images," in *Proc. 7th Int. Conf. Robot Intell. Technol. Appl. (RiTA)*, Nov. 2019, pp. 38–45.

[33] D. Aminou, "MSG's SEVIRI instrument," *ESA Bull.*, no. 111, pp. 15–17, 2002.

[34] Y. Zhao, Y. Ban, and J. Sullivan, "Tokenized time-series in satellite image segmentation with transformer network for active fire detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10155171/

[35] S. Escuin, R. Navarro, and P. Fernández, "Fire severity assessment by using NBR (Normalized burn Ratio) and NDVI (Normalized difference vegetation Index) derived from LANDSAT TM/ETM images," *Int. J. Remote Sens.*, vol. 29, no. 4, pp. 1053–1073, Feb. 2008.

[36] L. Knopp, M. Wieland, M. Rättich, and S. Martinis, "A deep learning approach for burned area segmentation with sentinel-2 data," *Remote Sens.*, vol. 12, no. 15, p. 2422, Jul. 2020.

[37] M. Sdraka, A. Dimakos, A. Malounis, Z. Ntasiou, K. Karantzalos, D. Michail, and I. Papoutsis, "FLOGA: A machine-learning-ready dataset, a benchmark, and a novel deep learning model for burnt area mapping with Sentinel-2," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 7801–7824, 2024.

[38] S. Gerard, J. Sullivan, and Y. Zhao, "WildfireSpreadTS: A dataset of multimodal time series for wildfire spread prediction," in *Proc. Adv. Neural Inf. Process. Syst.*, 2023, pp. 74515–74529.

[39] J. Burge, M. Bonanni, M. Ihme, and L. Hu, "Convolutional LSTM neural networks for modeling wildland fire dynamics," 2020, *arXiv:2012.06679*.

[40] J. L. Hodges and B. Y. Lattimer, "Wildland fire spread modeling using convolutional neural networks," *Fire Technol.*, vol. 55, no. 6, pp. 2115–2142, Nov. 2019, doi: 10.1007/s10694-019-00846-4.

[41] S. R. Coffield, C. A. Graff, Y. Chen, P. Smyth, E. Foufoula-Georgiou, and J. T. Randerson, "Machine learning to predict final fire size at the time of ignition," *Int. J. Wildland Fire*, vol. 28, no. 11, pp. 861–873, Nov. 2019.

[42] B. S. Li and R. Rad, "Wildfire spread prediction in North America using satellite imagery and vision transformer," in *Proc. IEEE Conf. Artif. Intell. (CAI)*, Jun. 2024, pp. 1536–1541.

[43] Y. Zhou, R. Kong, Z. Xu, L. Xu, and S. Cheng, "Comparative and interpretative analysis of CNN and transformer models in predicting wildfire spread using remote sensing data," *J. Geophys. Res., Mach. Learn. Comput.*, vol. 2, no. 2, p. 2024, Jun. 2025.

[44] H. Liang, M. Zhang, and H. Wang, "A neural network model for wildfire scale prediction using meteorological factors," *IEEE Access*, vol. 7, pp. 176746–176755, 2019.

[45] P. Nastos, "The Mediterranean climate," in *Soil Protection in Sloping Mediterranean Agri-Environments: Lectures and Exercises*, 2009, p. 7.

[46] E. Garcia-Hurtado, J. Pey, M. J. Baeza, A. Carrara, J. Llovet, X. Querol, A. Alastuey, and V. R. Vallejo, "Carbon emissions in Mediterranean shrubland wildfires: An experimental approach," *Atmos. Environ.*, vol. 69, pp. 86–93, Apr. 2013.

[47] J. Ruffault, T. Curt, V. Moron, R. M. Trigo, F. Mouillot, N. Koutsias, F. Pimont, N. Martin-Stpaul, R. Barbero, J.-L. Dupuy, A. Russo, and C. Belhadj-Khedher, "Increased likelihood of heat-induced large wildfires in the Mediterranean basin," *Sci. Rep.*, vol. 10, no. 1, p. 13790, Aug. 2020.

[48] A. Likas, N. Vlassis, and J. J. Verbeek, "The global k-means clustering algorithm," *Pattern Recognit.*, vol. 36, no. 2, pp. 451–461, Feb. 2003.

[49] I. Prapas, S. Kondylatos, I. Papoutsis, G. Camps-Valls, M. Ronco, M.-Á. Fernández-Torres, M. P. Guillem, and N. Carvalhais, "Deep learning methods for daily wildfire danger forecasting," 2021, *arXiv:2111.02736*.

[50] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002.

[51] A. S. Tarawneh, A. B. Hassanat, G. A. Altarawneh, and A. Almuhaimeed, "Stop oversampling for class imbalance learning: A review," *IEEE Access*, vol. 10, pp. 47643–47660, 2022.

[52] M. Kottek, J. Grieser, C. Beck, B. Rudolf, and F. Rubel, "World map of the Köppen–Geiger climate classification updated," Tech. Rep., 2006.

[53] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," 2015, *arXiv:1505.04597*.

[54] C. Wang, "A review on 3D convolutional neural network," in *Proc. IEEE 3rd Int. Conf. Power, Electron. Comput. Appl. (ICPECA)*, Jan. 2023, pp. 1204–1208.

[55] J. He, L. Li, J. Xu, and C. Zheng, "ReLU deep neural networks and linear finite elements," 2018, *arXiv:1807.03973*.

[56] D. Hendrycks and K. Gimpel, "Gaussian error linear units (GELUs)," 2016, *arXiv:1606.08415*.

[57] A. Galdran, G. Carneiro, and M. A. G. Ballester, "On the optimal combination of cross-entropy and soft dice losses for lesion segmentation with out-of-distribution robustness," in *Diabetic Foot Ulcers Grand Challenge*. Cham, Switzerland: Springer, 2022, pp. 40–51.

[58] K. Kalabokidis, A. Ager, M. Finney, N. Athanasis, P. Palaiologou, and C. Vasilakos, "AEGIS: A wildfire prevention and management information system," *Natural Hazards Earth Syst. Sci.*, vol. 16, no. 3, pp. 643–661, Mar. 2016.

[59] S. Kondylatos, G. Camps-Valls, and I. Papoutsis, "Uncertainty-aware deep learning for wildfire danger forecasting," 2025, *arXiv:2509.25017*.

NIKOLAOS ANASTASIOU was born in Athens, Greece, in 2002. He received the Diploma degree in rural, surveying and geoinformatics engineering from the National Technical University of Athens, in 2025. His Diploma thesis focused on the development and evaluation of deep learning models for wildfire spread prediction using remote sensing data. He completed a three-month internship with the National Observatory of Athens, focusing on earth observation and wildfire mapping. He is currently with Athena Research Center, where he is involved in training foundation models for fire detection from satellite data. His primary research interests include remote sensing and deep learning, with an emphasis on wildfire related applications.

SPYROS KONDYLATOS was born in Athens, Greece, in 1991. He received the B.Sc. degree in mathematics and the M.Sc. degree in applied mathematics from the National and Kapodistrian University of Athens, in 2013 and 2016, respectively, and the M.Sc. degree in data science from Athens University of Economics and Business, in 2017. He is currently pursuing the Ph.D. degree with the University of Valencia. From 2018 to 2020, he was a Data Scientist with the industry. Since 2020, he has been a Research Associate with the Orion Laboratory, National Observatory of Athens, National Technical University of Athens. He has contributed to H2020 projects DeepCube and TREEADS, as well as the ESA-funded SeasFire Project. His research focuses on Earth Observation (EO) with an emphasis on short-term wildfire danger forecasting using deep learning. His work has been published in high-esteemed conferences and journals, such as NeurIPS and *Geophysical Research Letters*. He has helped develop an online GIS system for predicting and mapping daily fire danger in Greece. His primary scientific interest lies in applying uncertainty-aware deep learning, such as Bayesian neural networks to EO applications.

IOANNIS PAPOUTSIS received the Diploma degree in electrical and computer engineering from the National Technical University of Athens (NTUA), the Master of Science degree in telecommunications from University College London, the M.B.A. degree from Alba Business School, and the Ph.D. degree in satellite remote sensing from NTUA. He is currently an Assistant Professor of remote sensing and artificial intelligence with the National Technical University of Athens and an Adjunct Researcher with the National Observatory of Athens and the Archimedes/Athena RC. He works on big satellite data analytics and machine learning for earth observation applications, with emphasis on natural disasters management and climate change impact monitoring. His research interests encompass the development of foundational models in remote sensing, focusing on self-supervised techniques for multi-modal EO data. He is also invested in advancing vision-language models for the analysis and interpretation of remote sensing data, as well as earth system deep learning for spatiotemporal forecasting.