

Capturing and interpreting wildfire spread dynamics: attention-based spatiotemporal models using ConvLSTM networks



Arif Masrur ^a, Manzhu Yu ^{a,*}, Alan Taylor ^{a,b}

^a Department of Geography, Penn State University, University Park, PA, USA

^b Earth and Environmental Systems Institute, Penn State University, University Park, PA, USA

ARTICLE INFO

Keywords:

Spatiotemporal dependence
Self-attention
ConvLSTM
Spatiotemporal modeling
Wildfire spread

ABSTRACT

Predicting the trajectory of geographical events, such as wildfire spread, presents a formidable task due to the dynamic associations among influential biophysical factors. Geo-events like wildfires frequently display short and long-range spatial and temporal correlations. Short-range effects are the direct contact and near-contact spread of the fire front. Long-range effects are represented by processes such as spotting, where firebrands carried by the wind ignite fires distant from the flaming front, altering the shape and speed of an advancing fire front. This study addresses these modeling challenges by clearly defining and analyzing the scale-dependent spatiotemporal dynamics that influence wildfire spread, focusing on the interplay between biophysical factors and fire behavior. We propose two unique attention-based spatiotemporal models using Convolutional Long Short-Term Memory (ConvLSTM) networks. These models are designed to learn and capture a range of local to global and short and long-range spatiotemporal correlations. The proposed models were tested on two datasets: a high-resolution wildfire spread dataset produced with a semi-empirical percolation model and a satellite observed wildfire spread data in California 2012–2021. Results indicate that attention-based models accurately predict fire front movements that are consistent with known wildfire spread-biophysical dynamics. Our research suggests there is considerable potential for attention mechanisms to capture the spatiotemporal behavior of wildfire spread, with model transferability, that can guide rapid deployment of wildfire management operations. We also highlight directions for future studies that focus on how the self-attention mechanism could enhance model performance for a range of geospatial applications.

1. Introduction

Accurately predicting how geographic events unfold in space and over time is crucial for reducing the socio-ecological impacts of natural hazards. Wildfire is one of such hazard that increases risks to life, property, carbon sequestration, biodiversity, and other ecosystem services in North America and worldwide (Abatzoglou and Williams, 2016; Ellis et al., 2022). Accurate prediction of wildfire occurrences and spread dynamics is challenging due to spatially and temporally varying relationships among biophysical factors (e.g. fuel, topography, weather) that control fire activity (Finney, 2005; Hawbaker et al., 2013; Hilton et al., 2015; Ruffault and Mouillot, 2017). With the recent increase in wildfire risk worldwide (Ellis et al., 2022; Jones et al., 2022), accurate prediction and understanding of the spatiotemporal processes of wildfire spread are crucial for reducing socio-ecological risks to wildlife and for developing effective fire management strategies and practices.

Wildfire spread is a complex spatiotemporal process and a series of attempts have been made to model fire behavior, dating back to as early as the 1920s (for comprehensive reviews see Sullivan, 2009a, 2009b, 2009c). Existing wildfire spread models generally can be categorized into three classes: *physical, semi-empirical, and empirical* (Finney et al., 2021). Physical (or semi-physical that represents only physics) models are based on the fundamental physical and chemical principles of combustion and heat transfer. These models require identification of the governing processes of wildfires and then formulation of mathematical equations that describe these processes and their possible interactions at all possible spatial and temporal scales. Therefore, physical models need a great degree of information such as model parameters, initial and boundary conditions, as exemplified by the Weather Research Forecast (WRF)-FIRE model and the Coupled Atmosphere-Wildland Fire Environment modeling system (Coen et al., 2013; Mandel et al., 2011) which require comprehensive ecological and meteorological data to capture

* Corresponding author.

E-mail address: mqy5198@psu.edu (M. Yu).

the fire-atmospheric feedback and dynamics at varied scale. Semi-empirical models are those that use some form of physical framework based on chemistry and physics of combustion and heat transfer to analyze experimental observational data (as opposed to quasi-physical models that use data solely for parameterization). Generally, the goal of these models is to identify the functional relationships between dependent and independent variables. The semi-empirical model of Rothermel (1972) is a pioneering work that developed an equation to predict fire spread rate, which also formed the basis of the National Fire Danger Rating System (Deeming, 1972; Deeming et al., 1977) and has seen its application in firefighting efforts globally. However, these models also demand input parameters and thus remain less practical for everyday applications. Empirical models (see early examples in McArthur, 1966, 1967; Wagner, 1990), on the other hand, contain no physical basis, rather statistical in nature and use data from fire observation along with relevant environmental factors (fuel, topography, weather) to determine the key characteristics used to describe fire spread behavior. While empirical and semi-empirical models are simple and computationally faster (compared to physical and semi-physical models), contributed to the advancement of wildfire science and produced effective operational guidelines for fire suppression and management efforts, these approaches may struggle to generalize across different ecological and environmental conditions (Higgins et al., 2008). All empirical sciences face measurement issues (e.g., quantification errors in field vs laboratory experiments) and while large-scale field experiments with wildfires can help reduce scale-dependent measurement issues, those experiments are usually costly to develop. This limitation is particularly evident in the context of large spatial scales, where the homogeneity of crucial inputs is assumed (McKenzie et al., 1996). To address this, Li (2000) suggested using an ecological modeling approach to reconstruct natural fire regimes, which can help evaluate forest policies. However, the challenge of accurately modeling fire regimes at regional to global scales remains, with the need for comprehensive model evaluation emphasized (Hanson et al., 2016).

Recent research has highlighted the potential of artificial intelligence techniques, particularly machine and deep learning, in modeling the complex dynamics of wildfire progression (Jain et al., 2020). For example, Subramanian and Crowley (2018) introduces a novel approach using spatial reinforcement learning to build forest wildfire dynamics models from satellite images, demonstrating the potential in predicting fire spread. Burge et al. (2020) explores the effectiveness of Convolutional Long Short-Term Memory (ConvLSTM) networks (Shi et al., 2015) in capturing fire-spread dynamics over consecutive time steps. Buch et al. (2023) introduces the SMLFire1.0 model, a stochastic machine learning framework for modeling wildfire activity in the western United States at a 12 km spatial resolution. This model effectively accounts for the nonlinear, spatially heterogeneous interactions between climate, vegetation, and human factors, improving the prediction of fire frequencies and sizes. While each of these approaches offers valuable insights into wildfire modeling, they also highlight the complexity and diversity of challenges in this field. Key among these challenges is accurately accounting for both local and long-range spatiotemporal dependencies in wildfire spread (Cai et al., 2020; Quinn et al., 2019). As suggested by the buoyant flame dynamics in wildfire spread (Finney et al., 2015), biophysical factors, such as topography, heterogeneous fuel patterns, moisture of dead and live vegetation, and atmospheric dynamics contribute to the spatiotemporal propagation processes that control the fire fronts at scale. Besides those local factors, the spread of a fire front is also affected by longer distance processes (i.e., beyond the front's adjacent neighborhood) such as spotting, where downwind transport firebrands from the flaming front that can ignite new fires (Alexandridis et al., 2008). Thus, advanced modeling is in demand to accurately capture both local and long-range interactions for predicting fire spread.

Addressing this need, we propose a ConvLSTM framework to capture both the local and long-range (i.e., spotting effect) spatiotemporal

interactions to predict wildfire propagation more accurately. Complementing this framework, the incorporation of attention mechanisms in deep learning has shown promise in modeling dependencies without regard to their distance in the input or output sequences (Vaswani et al., 2017). The self-attention mechanism, in particular, allows the inputs to interact with each other (hence "self") to identify focal points within the data (therefore "attention"). While traditionally used as a complementary approach to convolution, addressing both local and global contexts, recent studies suggest that global attention may not completely replace convolution for better predictions (Bello et al., 2019; Woo et al., 2018). Hence, the focus has shifted to more systematic approaches to attention, such as pairwise and patchwise self-attention (Lin et al., 2020; Zhao et al., 2020). Pairwise self-attention seeks to capture interactions between pairs of positions in the input data, while patchwise attention considers the relationships within subregions, potentially allowing the model to better understand local patterns (Zhao et al., 2020). The integration of these attention mechanisms in deep learning models has begun to address wildfire predictions. Recent efforts have primarily focused on fire detection (Majid et al., 2022), smoke identification (Cao et al., 2019), and burned area prediction (Li et al., 2021, 2023). Notably, Wang et al. (2023) demonstrated the efficacy of self-attention mechanisms to predict fire spread, using them in specific components of an LSTM network to learn from historical fire features for subsequent combustion image sequence prediction. However, a gap remains in our understanding of which type of self-attention mechanism (e.g., pair vs. patch) is most effective for learning important biophysical information (e.g., vegetation, wind, moisture, elevation) associated with fire spread, as well as their interpretability concerning the spatiotemporal progression.

In this paper, we address the challenge of accurately predicting wildfire propagation using ConvLSTM networks augmented with attention mechanisms. Specifically, our study focuses on capturing both the local interactions at the fire front and the long-range effects such as spotting - where embers carried by the wind can ignite fires at a distance from the main fire. We select wildfire spread as our primary spatiotemporal prediction problem due to its inherent complexity involving intricate input-input and input-output relationships. This complexity demands not only improved predictive accuracy but also enhanced interpretability of the models. To achieve this, we utilize both simulated and real-world data. The simulated data is a highly detailed, semi-empirical model-generated data provided by Burge et al. (2020), which offers a higher level of spatiotemporal granularity than conventional satellite-based observations. Their dataset, simulating realistic wildfire scenarios, serves as the foundation for both prediction and interpretation in our study. The real-world data is a California wildfire spread dataset derived from the Visible Infrared Imaging Radiometer Suite (VIIRS) satellite observations using an object-based system for tracking wildfire progression from 2012 to 2020 (Chen et al., 2022). Methodologically, our research explores a range of attention formulations within the ConvLSTM framework, specifically designed to address the different aspects of wildfire spread, such as the local spread typically represented by patchwise attention and long-range spotting effects potentially captured by pairwise attention. In addition to developing these attention-enhanced ConvLSTM models, we compare their performance against a baseline non-attention ConvLSTM network.

The main objectives of our paper are twofold: 1) To design and develop spatiotemporal attention-based ConvLSTM models for the improved prediction of wildfire progression, emphasizing the dual consideration of local spread and long-range spotting effects. 2) To investigate the strengths and weaknesses of different attention mechanisms within the ConvLSTM framework for comprehensive spatiotemporal wildfire prediction. Furthermore, we aim to predict the progression of wildfire spread over multiple time steps, with a primary goal of forecasting the spread pattern over the next 10 days, based on data from previous days' wildfire progression. Our approach involves a rolling prediction strategy, where we continually update the prediction

window with new data, allowing for dynamic and iterative forecasting. This strategy enables us to predict the wildfire spread for the next day and progressively extend the forecast up to 10 days ahead. In the following sections, we introduce the formulation of the wildfire spread prediction problem, the model architecture, experiment design, and result demonstrations, followed by conclusions and future directions.

2. Formulation of the wildfire spread prediction problem

Consider a spatial region under observation for wildfire spread, represented by an $X \times Y$ grid where X signifies the number of rows, and Y denotes the number of columns. Each cell within the $X \times Y$ grid contains F features (since we only use wildfire spread as the predicting variable, the value of F is 1) that change over time. As a result, at any time slice t , the observation can be captured by a three-dimensional tensor $X \in R_{F \times X \times Y}$, where R represents the domain of the observed wildfire spread. Over a time period t , we can gather a spatiotemporal sequence of observations $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_t$. Consequently, the 1-day-ahead wildfire spread prediction problem becomes forecasting the most probable tensor in the future given the past J -days observations, which can be formulated as in Eq. (1) and depicted in Fig. 1:

$$\tilde{X}_{t+1} = \text{argmax} \{ p(X_{t+1} | \hat{x}_{t-J+1}, \hat{x}_{t-J+2}, \dots, \hat{x}_t) \} \quad (1)$$

Here, \tilde{X}_{t+1} is the predicted wildfire spread field at time $t + 1$, and p is a conditional probability.

3. The spatiotemporal deep learning model for wildfire spread modeling and prediction

3.1. Building block: ConvLSTM

Having established the context and the problem statement, we proceed to introduce the ConvLSTM network (Shi et al., 2015), a pivotal component in the architecture of our predictive model. ConvLSTM is an extension of the traditional Long Short-Term Memory (LSTM) network, a type of recurrent neural network designed to learn long-term dependencies in sequence data. The ConvLSTM improves upon the LSTM by integrating convolutional structures into the LSTM's gating mechanisms. The convolutional operation in a ConvLSTM performs two crucial functions, much like in a conventional Convolutional Neural Network (CNN): *feature aggregation* and *feature transformation*. Feature aggregation occurs through a set kernel (for instance, a 3×3 cell window) that employs pre-trained, fixed weights to linearly combine feature values from neighboring cells. This allows the ConvLSTM to gather and incorporate local spatial information from its surroundings, a key feature that sets it apart from traditional LSTM networks. Feature transformation, on the other hand, is accomplished through linear

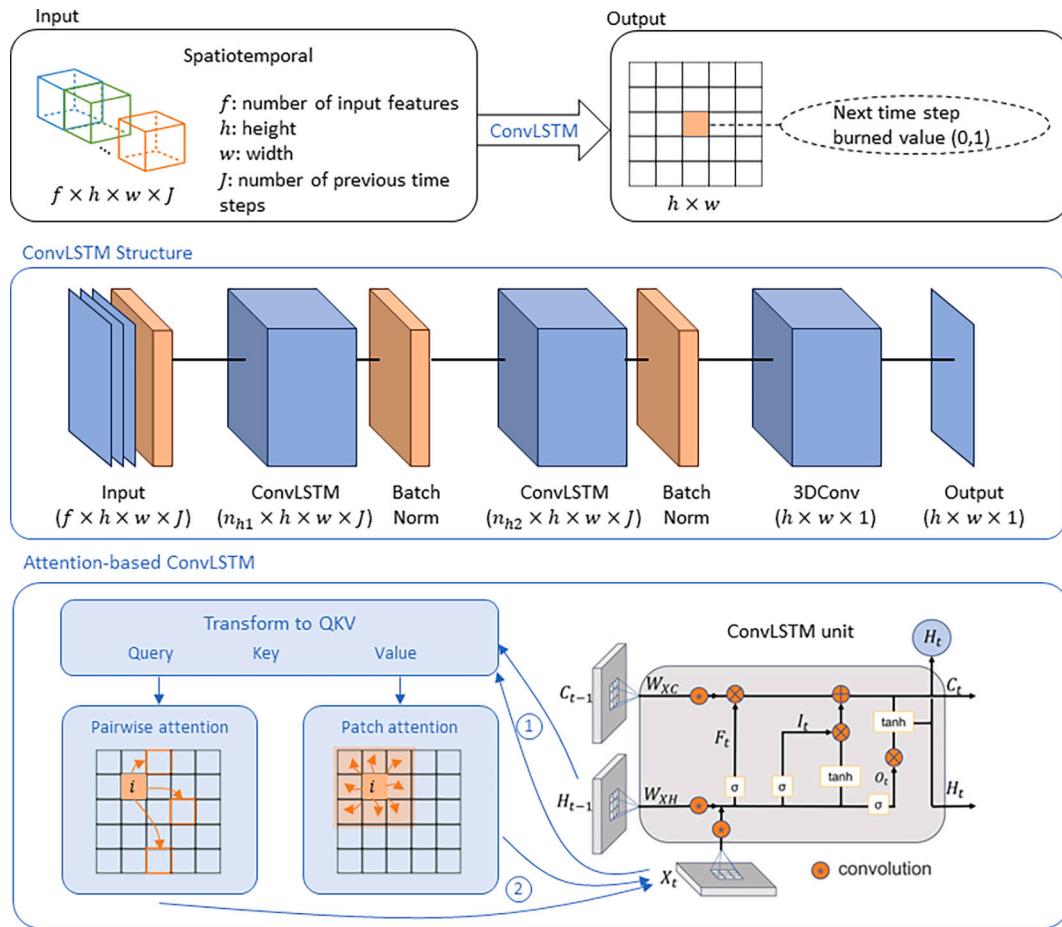


Fig. 1. Schematic representation of the ConvLSTM architecture integrated with pairwise and patchwise attention mechanisms. The input includes spatiotemporal features with dimensions $f \times h \times w \times J$, where f is the number of input features, h is the height, w is the width, and J is the number of previous time steps. The ConvLSTM structure processes these inputs through a series of ConvLSTM and Batch Normalization layers, followed by a 3D convolutional layer to predict the next time step's burned area. The Attention-based ConvLSTM unit integrates pairwise and patchwise attention mechanisms to enhance spatial feature extraction. Model inputs encompass various variables (details in Section 4.1 Data), including the location of all unburned vegetation, the location of tree vegetation that was burned, the location of all vegetation that was burned prior to this time step, the horizontal component of wind velocity, the vertical component of wind velocity, elevation, and moisture content.

mappings coupled with nonlinear scalar functions, such as sigmoid or tanh. These transformations enable the ConvLSTM to manipulate the aggregated features and generate complex, nonlinear representations, thereby facilitating more accurate predictions of future states.

The ConvLSTM network contains four key components that work together to learn and retain relevant information from spatiotemporal data. These components are the input gate, forget gate, cell state, and output gate. The input gate decides how much new information will be stored in the cell state. It uses the current input (X_t), the previous hidden state (H_{t-1}), and the previous cell state (C_{t-1}) to compute the input gate (i_t) using a sigmoid function, which converts the values between 0 and 1.

$$i_t = \sigma(W_{xi}^* X_t + W_{hi}^* H_{t-1} + W_{ci}^* C_{t-1} + b_i) \quad (2)$$

The forget gate decides how much of the previous cell state to retain. It uses the current input, the previous hidden state, and the previous cell state to compute the forget gate (f_t) using a sigmoid function.

$$f_t = \sigma(W_{xf}^* X_t + W_{hf}^* H_{t-1} + W_{cf}^* C_{t-1} + b_f) \quad (3)$$

The new memory cell ($C_{\sim t}$) is computed using the current input and the previous hidden state. This represents the new information to store in the cell state.

$$C_{\sim t} = \tanh(W_{xc}^* X_t + W_{hc}^* H(t-1) + b_c) \quad (4)$$

The final memory cell (C_t) is updated by retaining certain information from the previous cell state (controlled by the forget gate) and adding some new information (controlled by the input gate and the new memory cell).

$$C_t = f_t^* C_{t-1} + i_t^* C_{\sim t} \quad (5)$$

The output gate decides what part of the cell state will be outputted. It uses the current input, the previous hidden state, and the updated cell state to compute the output gate (o_t) using a sigmoid function.

$$o_t = \sigma(W_{xo}^* X_t + W_{ho}^* H_{t-1} + W_{co}^* C_t + b_o) \quad (6)$$

The hidden state (H_t) is updated by applying the tanh function to the cell state and multiplying it by the output gate. This gives the final output of the ConvLSTM at the current time step.

$$H_t = o_t^* \tanh(C_t) \quad (7)$$

where $*$ denotes the convolutional operation, σ is the sigmoid function, i_t, f_t, o_t are the input, forget, and output gates, C_t is the cell state, H_t is the hidden state, X_t is the input, W and b are the weights and biases, respectively, with subscripts denoting which gate or cell they correspond to.

3.2. Attention-based methods for ConvLSTM networks

For the wildfire spread prediction, we integrate two different variants of self-attention to the ConvLSTM: pairwise and patchwise self-attentions. Our contribution is spatiotemporal attention-based sequence forecasting frameworks using the ConvLSTM network that predicts the spatial propagation of fire fronts at multi-time steps based on sequentially learning the dynamic wildfire-environmental driver relationships or dependencies over local and long-range (i.e., to capture spotting effect) spatiotemporal distances (Fig. 1). J.

The structure of the ConvLSTM network is three layers, including two ConvLSTM layers and a 3D CNN layer, with hidden dimensions of 64, 128, and 256. The first ConvLSTM layer expands the input channel dimensions by applying spatial filters to each time step of the input sequence, the second ConvLSTM later further expands the channel dimension, and the final convolutional layer then aggregates the temporal information (over 10 time steps) and reduces the channel dimension to produce a single output channel that represents fire probabilities at the next time step.

3.2.1. Pairwise self-attention ConvLSTM

The pairwise self-attention mechanism initially maps the input feature maps at time step t , denoted by H_t , into three distinct feature spaces: Query (Q_h), Key (K_h), and Value (V_h).

$$\begin{cases} Q_h = W_q H_t \in R^{c \times N} \\ K_h = W_k H_t \in R^{c \times N} \\ V_h = W_v H_t \in R^{c \times N} \end{cases} \quad (8)$$

where $\{W_q, W_k, W_v\}$ denote s set of weights for 1×1 convolutions, c and c represent the number of channels or predictors in the input feature map and in the transformed feature spaces respectively, and N is the product of the height (H) and width (W) of the feature map, i.e., the total number of cells.

The pairwise similarity scores among all the cells are computed via the matrix multiplication of the Query and Key feature spaces, creating a similarity score matrix, SS.

$$SS = Q_h^T K_h \in R^{N \times N} \quad (9)$$

This SS matrix, of dimensions $N \times N$, holds the similarity score between any two cells at the i^{th} and j^{th} locations. The similarity score between two cells is calculated as the dot product of the i^{th} cell's feature vector in the Query space ($H_{(t,i)}^T W_q^T$) and the j^{th} cell's feature vector in the Key space ($W_k H_{(t,j)}$).

$$SS_{i,j} = (H_{(t,i)}^T W_q^T)(W_k H_{(t,j)}) \quad (10)$$

where $H_{t,i}$ and $H_{t,j}$ denote feature vectors having shape $c \times 1$.

These pairwise similarity scores are then normalized across the matrix. The aggregated feature of a cell from the input feature map, denoted by f_i , is calculated as the weighted sum of the Value feature vectors across all cells, with the weights being the normalized similarity scores:

$$f_i = \sum_{j=1}^N \alpha_{ij} (W_v H_{t,i}) \quad (11)$$

where α_{ij} denotes normalized similarity scores.

The attention-based aggregated features, denoted by \hat{H}_t , are derived by a shortcut connection:

$$\hat{H}_t = w^* f + H_t \quad (12)$$

where w represents trainable weights, and f is the aggregated feature map. These final aggregated features are then fed into the LSTM gates for prediction at the next time step.

In the context of wildfire spread, the query, key, and value components capture relationships between different spatial locations within the input grid. Specifically, the model uses the query to represent the current state of a particular cell (e.g., whether it is burning, unburned, or previously burned); uses the key to include the influence of all cells on the current cell's state (e.g., proximity to burning cells); and uses the value to convey information about the fire spread potential and relevant features (e.g., wind velocity, moisture content) of all cells. By incorporating pairwise self-attention, the model can dynamically weigh the importance of interactions between different cells, capturing both local and global dependencies. This allows the model to better predict wildfire spread by considering the combined influence of various spatial relationships within the input grid.

3.2.2. Patchwise self-attention ConvLSTM

While pairwise self-attention is a one-to-one mapping between the target grid and all other grids, patchwise attention focuses on mapping one-to-many relationships. The patch includes spatially adjacent grid points in the feature map and the relationships within these points are captured by attention weights. These relationships, as calculated by the

self-attention mechanism within each patch, are used by the model to generate the output. This allows for more complex, localized relationships to be considered than in traditional convolution; for that, we propose to replace the basic convolution in ConvLSTM with the patchwise attention module for accurately predicting geo-event progression in space and time.

First, a footprint R_i is defined in the data, which refers to the region or ‘patch’ of the 2D lattice (or grid) that is being considered for each target cell or location (i). In other words, if we are trying to calculate the attention-weighted feature map for a particular grid cell (i), we would consider not just the features at that specific location, but also the features in the surrounding areas within this footprint R_i . This localized focus allows it to capture spatial patterns and contextual information more effectively within a given region. The optimal footprint size is determined empirically through experimentation.

For each patch, the self-attention mechanism computes attention scores. This process involves applying a learnable function, denoted as α , on the feature vectors within the patch ($X_{\{R_i\}}$). This results in a tensor with the same spatial dimensions as the patch.

We then determine the weights for each feature vector within the patch. These weights, denoted as $\beta(x_j)$, are not only based on the individual data point but also the collective information from all feature vectors within the patch. This distinguishes the patchwise approach from pairwise self-attention.

The patchwise self-attention process culminates in the aggregation of weighted features within the patch. This is represented by the equation:

$$Y_i = \sum_{\{j \in R_i\}} [\alpha(X_{\{R_i\}})_j \odot \beta(x_j)] \quad (13)$$

where Y_i is the output for the patch, $\alpha(X_{\{R_i\}})_j$ is the attention score for the feature vector at location j within the patch, and \odot denotes element-wise multiplication.

In the context of wildfire spread, the learnable function α and weights β capture localized relationships within specific patches. The learnable function α captures localized relationships within each patch, such as the distribution of unburned and burned vegetation patches, the influence of wind patterns within the patch, and terrain features within the patch affecting the fire’s path. The weights β captures the importance of features within the patch, including current burn status and its potential to spread, wind velocity and moisture content within the patch, historical data on burn status and other relevant features. By incorporating patchwise self-attention, the model dynamically weighs the importance of localized interactions within patches, enhancing its ability to predict wildfire spread by considering complex spatial relationships within specific regions.

4. Experiment design

4.1. Data

4.1.1. Simulated wildfire spread from fire percolation model

In our experiments, we used wildfire spread data generated by a semi-empirical simulation model known as the percolation model (Burge et al., 2020), which utilizes a mathematical analogue approach based on percolation theory (Beer and Enting, 1990), enriched with factors such as fuel density, moisture content, slope, and wind to simulate wildfire spread. This model is abstract and is not calibrated to replicate specific real-world fire events but rather to emulate general wildfire behaviors under controlled conditions. The use of this model allows us to explore the impact of various parameters on fire dynamics in a systematic manner, which is essential for training and testing our ConvLSTM models. Each data sample in this set represents a 110×110 cell grid, which corresponds to a forest patch with an active wildfire, including 1) the location of all unburned vegetation, 2) the location of tree vegetation that was burned, 3) the location of all vegetation that

was burned prior to this time step, 4) the horizontal component of wind velocity, 5) the vertical component of wind velocity, 6) elevation, and 7) moisture content. This setup allows us to explore various fire behaviors under simulated conditions without specific resolution constraints, providing flexibility in modeling different scenarios.

4.1.2. California wildfire spread from satellite observations

The second dataset we used in the experiment is the California wildfire spread dataset derived from the Visible Infrared Imaging Radiometer Suite (VIIRS) satellite observations using an object-based system for tracking wildfire progression from 2012 to 2020 (Chen et al., 2022). The data originates from the VIIRS active fire detections, which provide spatially detailed fire information with a resolution of 375 m. Fire pixels identified by VIIRS are clustered based on their spatial proximity. These clusters are either appended to existing fire events or categorized as new incidents, depending on their location relative to previously identified fires. Similar to the fire percolation model simulations, we structured the fire spread data with collocated biophysical factors into the same seven variables. The first three variables (the location of all unburned vegetation, the location of tree vegetation that was burned, and the location of all vegetation that was burned prior to this time step) are derived based on the California wildfire spread data with collocation of MODIS Normalized Difference Vegetation Index (NDVI) with a resolution of 250 m over a 16-day period. The wind components and soil moisture are downloaded from ERA5 Land at a resolution of 9 km, and we spatially interpolated the three variables into the resolution of 375 m. The elevation variable is derived from NASA-DEM with a resolution of 30 m, and we spatially resampled the data into 375 m.

To accommodate the requirements of our ConvLSTM network, we have structured the input data as follows: $[N_s, N_T, H, W, N_c]$, where N_s is the number of samples, N_T is the number of previous time steps, H and W are the height and width of the grid, and N_c is the number of predictors. Each sample is a 110×110 grid, representing a forest patch, with data from the 10 previous time steps. Most channels were normalized between 0.0 and 1.0 using min-max normalization. For the wind speed channels, we applied standard normalization to achieve zero mean and unit variance due to the presence of both positive and negative initial values.

The target data for our prediction is structured as: $[N_s, 1, H, W, \text{fire probability}]$. The fire probability for each cell in the 110×110 grid indicates the probability of a fire occurring at the next time step. These probabilities can range between 0 (no fire) and 1 (definite fire), with values in between representing varying degrees of risk. This allows the model to predict the spread of the wildfire over the entire grid for the upcoming time step, based on the previous 10 time steps and the specified predictors.

4.2. Experiment setup

Our experiment design primarily focuses on comparing the performances of 1) Baseline Model: ConvLSTM without attention, 2) ConvLSTM with pairwise self-attention, and 3) ConvLSTM with patchwise self-attention. This design allows us to evaluate the effectiveness of the proposed attention mechanisms. Our baseline model sets a performance threshold that we aim to surpass with our attention mechanisms. The second model aims to capture spatial dependencies within the wildfire spread data by incorporating the relationships between each pair of points in the feature map. The third model extends standard convolution by considering relationships across multiple grids, thereby allowing for consideration of local spatial contexts. For each model, we train the model on our training set (~60% of data samples), tune the model parameters to a validation set (~20%), and then evaluate the model’s performance on a previously unseen test set (~20%). The simulated dataset contains 3000 samples, and the satellite observed dataset contains 1929 samples. To test the optimal footprint size in the

patchwise attention mechanism, we designed a subset of experiments that systematically vary the size of the footprint - within the range of 3×3 , 5×5 , ... and 11×11 , and evaluate the performance of the model under each configuration. We did not test footprint sizes beyond 11×11 because it exceeds our computation capability while training the model.

We evaluate the performance of our models using root mean squared error (RMSE) for fire grids and non-fire grids separately. We evaluate the accuracy of our model on predicting fire fronts, by reclassifying the output fire probability into fire and non-fire grids using a predefined fire probability threshold of 0.1 via the classification metrics of precision, recall, and F1-score.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (14)$$

$$Precision = \frac{TP}{TP + FP} \quad (15)$$

$$Recall = \frac{TP}{TP + FN} \quad (16)$$

$$F1Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (17)$$

Where N is the number of grid cells, y_i is the ground truth for the i^{th} grid cell, and \hat{y}_i is the predicted value for the i^{th} grid cell. TP (True Positives) is the count of correctly predicted fire grid cells, FP (False Positives) is the count of non-fire grid cells incorrectly predicted as fire grid cells, and FN (False Negatives) is the count of fire grid cells that were incorrectly predicted as non-fire grid cells.

We also evaluate the model's capability for predicting locations of spotting fires or fires not directly adjacent to the existing fire front. We calculate true positives (correctly predicted spot fires), false positives (predicted spot fires that are incorrect), and false negatives (actual spot fires that were not predicted) using the classification metrics of precision, recall, and F1-score on non-neighboring fire grids after reclassifying fire probabilities using a series of thresholds from 0.01 to 0.1.

4.3. Model interpretability

We employ a model interpretability technique for feature attribution, known as *Integrated Gradients* (Sundararajan et al., 2017), to illustrate the path integral of the gradients along the straight-line path from the baseline x' to the input x . The baseline is the starting point for the path integral and is often a neutral input that has no contribution to the output prediction. The integrated gradient along the i^{th} dimension for an input x and baseline x' can be defined as follows:

$$Integrated\ Gradients_i(x) = (x_i - x'_i) \times \int_{\alpha=0}^1 \frac{\partial F(x' + \alpha \times (x - x'))}{\partial x_i} d\alpha \quad (18)$$

where x_i is the i^{th} feature of the input, x'_i is the i^{th} feature of the baseline input, F is the model's prediction function, $\frac{\partial F(x)}{\partial x_i}$ is the gradient of F along the i^{th} dimension, and α scales the path from the baseline x'_i to the input x_i . The integral accumulates the gradients at all points along the straight path from the baseline to the input in the feature space. The product with $(x_i - x'_i)$ scales the integrated gradient by the difference in the feature value from the baseline, which gives the attribution of that feature. With the integrated gradients, this study investigates the efficacy of the pair-patchwise self-attention modules into the standard ConvLSTM to adequately learn the local-to-global interactions to accurately predict and explain wildfire spread behavior.

5. Results with the simulated fire spread

5.1. Prediction performance

The pairwise attention ConvLSTM (M2) outperformed the standard ConvLSTM (M1) and the patchwise attention ConvLSTM (M3) across all considered metrics, indicating that pairwise attention mechanisms may be more effective for spatial and temporal wildfire prediction tasks (Table 1). The changes in RMSE are minimal across the different footprint sizes for patchwise attention ConvLSTM (M3), indicating that patchwise attention ConvLSTM using this metric is insensitive to the changes in footprint size. While the precision stays relatively high across all footprint sizes, recall varies, reflecting changes in the model's capacity to detect all fire instances across different sizes. Despite the variations in recall, our focus on prioritizing fire location prediction accuracy leads us to select a footprint size of 7 for further analysis. This choice is grounded in the fact that, at this size, the model achieves one of the highest among the sizes tested. While this comes with a slightly lower recall, it aligns with our objective of maximizing accuracy in pinpointing fire locations, ensuring that when the model identifies an area as a fire, it does so with high reliability. Thus, footprint size 7 will be the default in our subsequent analysis with the patchwise attention ConvLSTM model.

All three models indicate most grid cells have little to no error suggesting that the model performs well in areas far from the fire (Fig. 2). The model's tendency to over-predict near the fire's edge could be due to an attempt of the model to capture potential fire spread, while under-predictions within the fire zone suggest challenges in accurately modeling the fire's core intensity or shape. The **pairwise ConvLSTM (M2) prediction** aligns more closely with the **target** than the standard ConvLSTM (M1) and patchwise ConvLSTM (M3) predictions. In predictions of Pairwise ConvLSTM (M2), the errors are more localized around the fire's perimeter rather than scattered, and the areas of under-prediction (blue cells) within the fire zone are less prominent than the other two model predictions, indicating that Pairwise ConvLSTM (M2) is better at capturing the core area of the fire. The background (non-fire areas) in the Pairwise ConvLSTM (M2) error image has fewer red spots than the ConvLSTM (M1) and Patchwise ConvLSTM (M3) model, indicating fewer false positives. This suggests that the Pairwise ConvLSTM (M2) model may be more conservative or accurate in predicting non-fire areas.

Table 2 provides a detailed comparison of three different models – the standard ConvLSTM (M1), the pairwise attention ConvLSTM (M2), and the patchwise attention ConvLSTM (M3) – based on their performance in predicting the location of spot fires. In predicting the location of spot fires, the pairwise attention ConvLSTM (M2) consistently outperforms the other two models across all thresholds. The standard ConvLSTM (M1) and the patchwise attention ConvLSTM (M3) are similar in performance to each other, but M3 tends to lag slightly behind, especially in terms of recall.

Fig. 3 illustrates the comparison of spot fire predictions from the three models. The standard ConvLSTM (M1) results show that the true positives are well scattered throughout the prediction area. However, noticeable clusters of false positives suggest overestimations, and false negatives indicate missed fires. The Pairwise ConvLSTM model (M2) shows a higher concentration of true positives, especially towards the central parts of the fire region, indicating a more accurate prediction in those areas. The false positives and false negatives are present but seem less compared to the standard ConvLSTM (M1) model, suggesting an improvement in prediction accuracy. The Patchwise ConvLSTM (M3) shows more clusters of false positives and false negatives, indicating that the model might not efficiently capture the broader, more dispersed fire events. The spread of errors, particularly the false negatives, could imply that the Patchwise model struggles with continuous prediction across the entire fire region. Overall, the Pairwise ConvLSTM (M2) shows the capability in capturing the long-range dependencies of fire spread, as

Table 1
Model performance comparison.

Model	Overall RMSE	Fire RMSE	Non-Fire RMSE	Precision	Recall	F1 Score
ConvLSTM (M1)	0.0056	0.0521	0.0031	0.9728	0.5940	0.7376
Pairwise attention (M2)	0.0051	0.0499	0.0028	0.9840	0.6193	0.7602
	3	0.0060	0.0594	0.0032	0.9604	0.5392
	5	0.0060	0.0577	0.0034	0.9569	0.5597
Patchwise attention (M3) with varying footprint size	7	0.0060	0.0595	0.0031	0.9645	0.5236
	9	0.0060	0.0590	0.0032	0.9593	0.5393
	11	0.0060	0.0587	0.0033	0.9564	0.5475
						0.6964

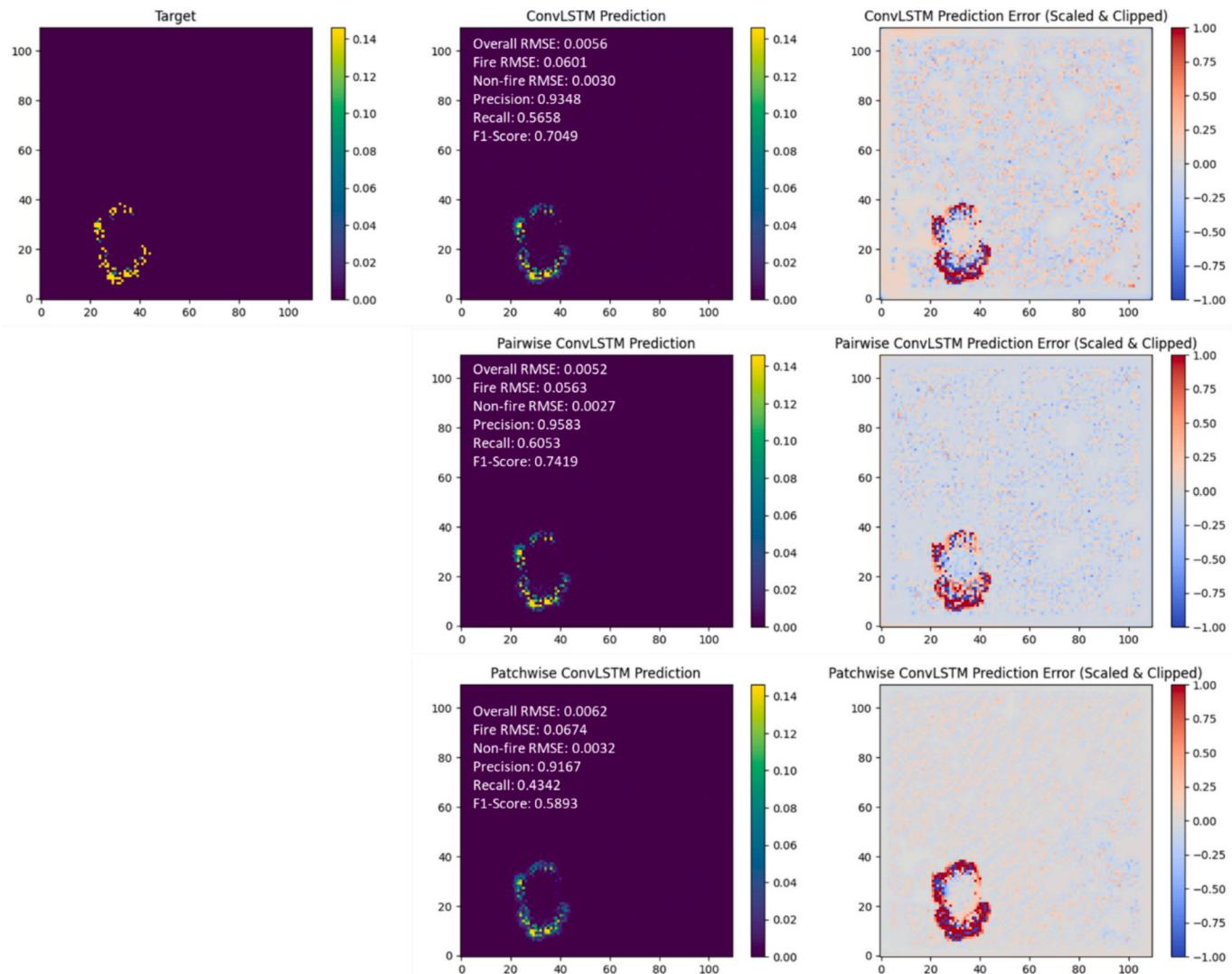


Fig. 2. Error visualization between the target and the predictions for a specific test sample. In the error image (rightmost), the color intensity represents the magnitude of the prediction error at each grid cell, with red indicating areas where the model over-predicted (predicted fire where there is none in the target) and blue indicating areas where the model under-predicted (missed predicting fire that is present in the target). The scale has been adjusted and clipped to enhance visibility. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

evidenced by the spread and distribution of true positives and the reduced error spots (false positives and false negatives). This suggests that the model is not only learning local interactions but is also effectively utilizing the additional contextual information provided by the pairwise attention to predict fire spread over a broader area.

5.2. Physical interpretations

Across the three models, unburned vegetation shows the most

dispersion across integrated gradient values, indicating the highest influence on the model's output. The wide range of values suggests that the model pays close attention to the state and condition of unburned vegetation, as it represents potential fuel for the fire. Current and previous burned area show less dispersed clusters of integrated gradient values. Moisture, elevation, and wind components are clustered around a low, negative value with slight variation, suggesting a minor environmental influence on the model's output (Fig. 4).

In the pairwise ConvLSTM model (M2), most variables show tighter

Table 2

Precision, recall, and F1-score of different models – the standard ConvLSTM (M1), the pairwise attention ConvLSTM (M2), and the patchwise attention ConvLSTM (M3) – based on their performance in predicting the location of spot fires. Threshold levels (0.01 to 0.10) represent different cutoffs for predicting a spot fire; a lower threshold implies a more sensitive model that predicts fires more readily, while a higher threshold indicates a more conservative model.

Threshold	Precision			Recall			F1_Score		
	M1	M2	M3	M1	M2	M3	M1	M2	M3
0.01	0.4329	0.4817	0.3454	0.4091	0.5189	0.2636	0.4077	0.4938	0.2975
0.02	0.5655	0.6137	0.5241	0.5515	0.6197	0.4669	0.5481	0.6114	0.4931
0.03	0.6567	0.7079	0.6324	0.6063	0.6731	0.5282	0.6205	0.6855	0.5752
0.04	0.7158	0.7779	0.7028	0.6291	0.6987	0.5547	0.6593	0.7317	0.6196
0.05	0.7499	0.8134	0.738	0.6312	0.6948	0.5426	0.6764	0.7446	0.6249
0.06	0.7644	0.8179	0.7364	0.6172	0.6707	0.5143	0.6734	0.7326	0.6048
0.07	0.7423	0.8014	0.7011	0.5704	0.623	0.467	0.634	0.6961	0.5599
0.08	0.7043	0.7621	0.6437	0.5218	0.5592	0.4123	0.5888	0.6396	0.502
0.09	0.6405	0.6939	0.5764	0.4565	0.4903	0.3621	0.5218	0.5695	0.4442
0.1	0.5768	0.6262	0.5109	0.3908	0.4221	0.3089	0.4557	0.4994	0.3842

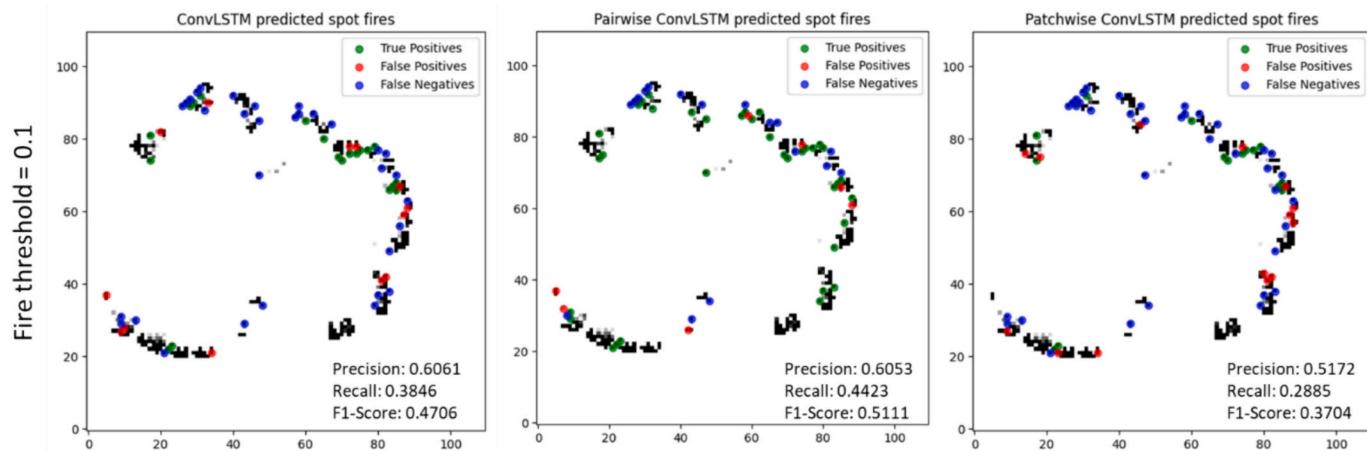


Fig. 3. Location accuracy of predicted spot fire by different models. Each sub-figure displays the location of true positives (green dots), false positives (red dots), and false negatives (blue dots) for the predicted spot fires. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

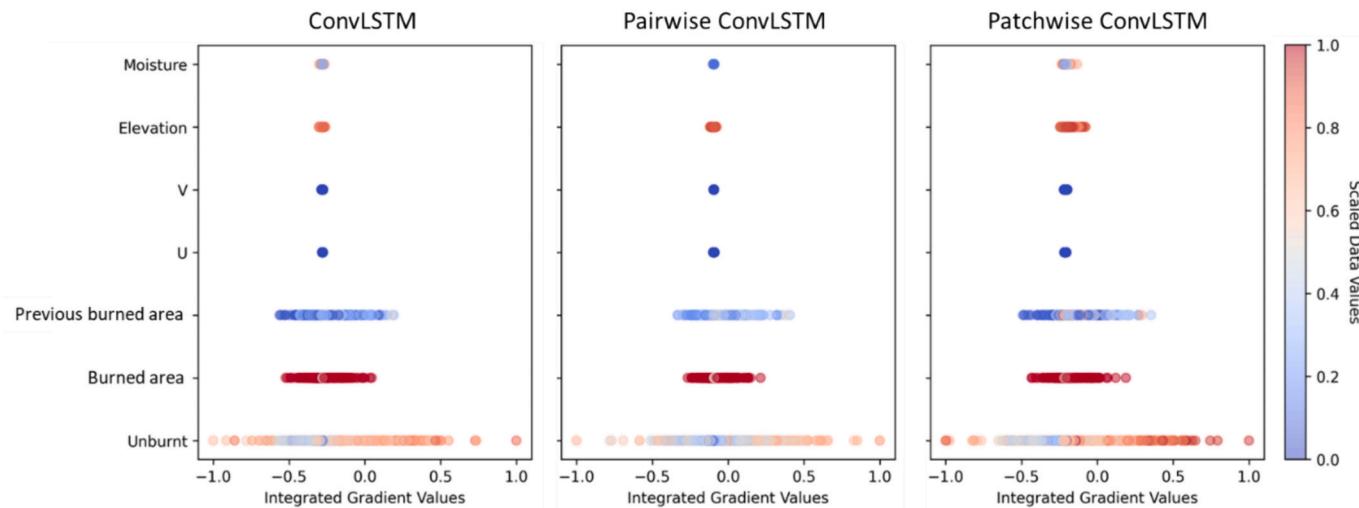


Fig. 4. Integrated gradient values for each variable. Dots represent positive examples in the test set and are colored based on the scaled value of the relevant variable.

clusters of integrated gradient values. The tight clustering indicates that the pairwise model may be capturing stable and consistent interactions between feature pairs. The patchwise ConvLSTM model (M3) displays a greater spread in integrated gradient values for elevation and moisture, implying that these features have a more varied influence on the

predictions, potentially reflecting a model that captures a wider range of conditions or a more complex relationship with the target variable. The greater spread could reflect a model attempting to learn a more complex feature space, thus not improving the predictive performance. The patchwise model might struggle to capture the key interactions

effectively, possibly because the data contains complex dependencies that are not well-represented by the patchwise approach.

5.2.1. Temporal patterns

In our three ConvLSTM models, the integrated gradient values for the unburned vegetation increase over time to a time step (T7), indicating that the fire is most likely to spread to areas that are currently unburned (Fig. 5). As the fire spreads, the number of unburned areas decreases, and the model becomes more confident that the fire will spread to these remaining areas. The influence of current and previous burned areas remains minimal, decreases to a negative integrated gradient value at T8 and increases to a larger positive value at T9. This is likely since areas have already been burned and are less likely to be re-burned. The negative integrated gradient values at T8 suggest that the model is less confident that the fire will spread to these burned areas, but the positive integrated gradient values at T9 suggest that the model is again becoming more confident that the fire could spread to these areas, potentially due to factors like wind or elevation.

In the standard ConvLSTM (M1), the horizontal wind component is initially and increasingly important for prediction until T8, but at T9 the importance drops to a negative value, suggesting a reassessment of the feature's influence. Conversely, the vertical wind component decreases until T8 and increases at T9 to a positive value, suggesting that the model is considering the possibility that vertical wind could play a role in fire spread at that particular time step. The contrast in trends for the two wind components reflects the model's adaptation to changing and distinct roles of different wind components in fire spread. Other factors like elevation and moisture content also play roles. The integrated gradients of elevation fluctuate around zero without a clear increasing or decreasing trend, suggesting that the model does not consistently attribute a strong influence on elevation. The observed fluctuations of integrated gradients for elevation around zero in our ConvLSTM model's output likely reflect a data limitation inherent in the semi-empirical fire percolation model, rather than a deficiency in the ConvLSTM model itself. This suggests that the fire percolation model may not adequately capture the complex effects of elevation on wildfire spread, particularly in varied terrain, thereby limiting the ability of our ConvLSTM model to

consistently recognize elevation as a significant influencing factor. The integrated gradient values for moisture content initially show a wide negative spread, implying that higher moisture levels are associated with a reduced likelihood or intensity of fire spread. However, as we progress to time step T6, this spread of values begins to narrow, and from around T6 to T8, it shifts towards positive values. This shift indicates a recalibration in the model's learning process, reflecting a change in how moisture content is weighed in predicting fire spread at different stages of the model's time frame.

In Pairwise ConvLSTM (M2), the integrated gradient values for elevation decrease over time in both models until T8, indicating that the fire's tendency to spread to higher elevations decreases as it spreads. However, at T9, the integrated gradient value for elevation increases to a positive value in the Pairwise ConvLSTM model, suggesting that the model is again considering the possibility that elevation could play a role in the fire's spread at that particular time step. This may be due to the Pairwise ConvLSTM model capturing the interactions between elevation and other variables, leading to a more refined understanding of elevation's influence. In real world scenarios, as elevation increases, fuel moisture typically rises and temperature decreases, which can influence fire behavior. The Pairwise ConvLSTM model's shift at T9 might reflect a recognition of these elevation-related changes in environmental conditions and their impact on the fire's progression. The overall decrease in integrated gradient values for moisture content, coupled with larger variations or uncertainties shown in the shaded areas, indicates a changing dependence on dry fuel as the fire progresses. Initially, dry conditions might be seen as critical for fire spread, but as the fire moves to different environments or as conditions evolve, the importance of moisture content appears to change, indicating the model's recalibration in response to changing fire conditions.

In Patchwise ConvLSTM (M3), all four variables (horizontal wind component, vertical wind component, elevation, and moisture) follow the same dynamical patterns, where integrated gradient values decrease to a negative value till T8 and increases at the last time step to a positive value, suggesting that the model is struggling to capture the complex relationships between variables and their influence on the fire's spread. One possible explanation for this behavior is that patchwise

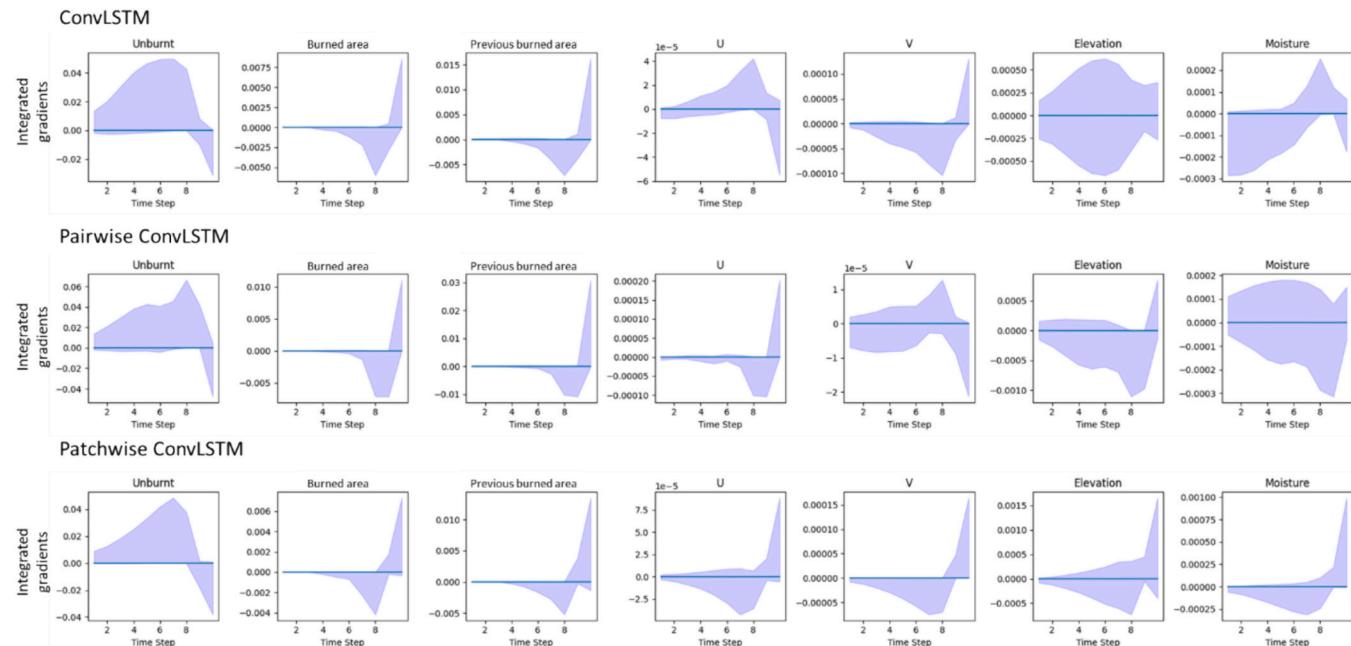


Fig. 5. Line plots with shaded areas for different features across time steps, representing the temporal dynamics of integrated gradients. Each plot corresponds to a specific feature. The x-axis denotes time steps, while the y-axis shows the integrated gradients values that indicate the importance of each feature at each time step for prediction. The shaded areas represent the range of integrated gradient values for a given feature across different scenarios or time steps. Larger shaded areas indicate greater uncertainty or variability.

ConvLSTM's focus on individual patches of fire may limit its ability to understand the broader context and interactions between variables across different patches. This could lead to the model overemphasizing certain variables in the early stages of the fire and then overcompensating in the later stages. The consistently poor performance of patchwise ConvLSTM compared to the other models further supports the notion that its limited understanding of variable interactions hinders its ability to make accurate predictions about the fire's spread.

5.2.2. Spatial patterns

The spatial patterns of integrated gradients of the Pairwise ConvLSTM model (Fig. 6) show a more concentrated distribution over predicted fire front than the ones of the standard ConvLSTM (M1). This is primarily because the ConvLSTM model (M1) processes information in a global way, considering all the pixels in the image at the same time. This can lead to a more diffuse spatial pattern of integrated gradients, as the model is not able to focus on specific areas of the image. The

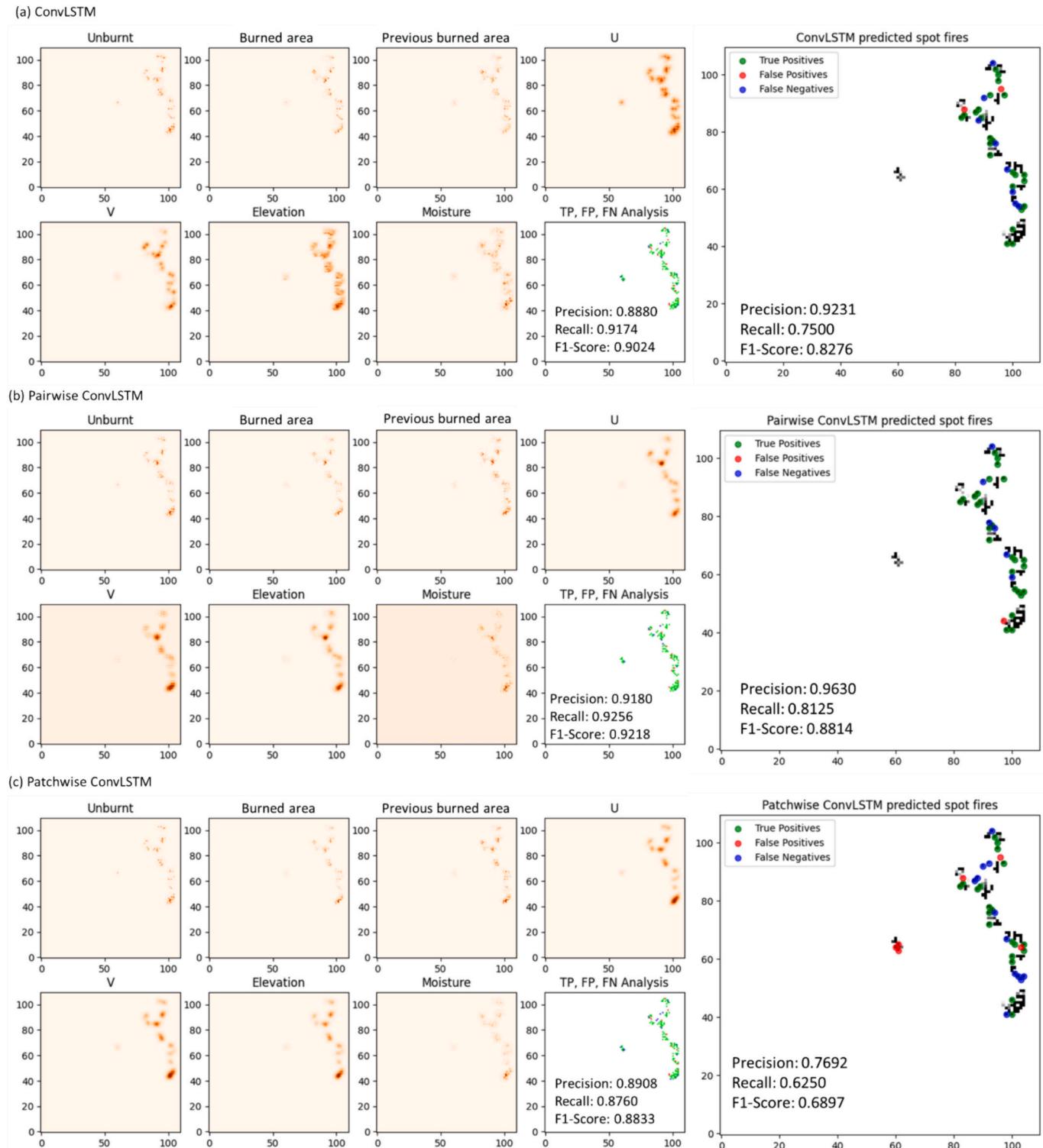


Fig. 6. Maximum Integrated Gradients of different variables over time, overall predictive performance, and spot fire predictive performance.

Pairwise ConvLSTM model (M2) can capture both local and far-away spatial dependencies. This allows the model to identify both subtle patterns in the data and larger-scale patterns that are indicative of spot fires. The Patchwise ConvLSTM (M3) shows an even more focused spatial pattern compared to the Pairwise ConvLSTM (M2), indicating that Patchwise ConvLSTM model has a narrower focus on local spatial interactions. The Patchwise ConvLSTM model's small footprint size allows it to focus on specific areas of the image. This can be helpful for identifying small, localized patterns of fire activity, but it can also lead to the model missing important patterns that are spread across a larger area of the image.

6. Results with the satellite observed wildfire spread in California 2012–2020

In the satellite-observed wildfire spread experiment, the elevation and moisture variables did not converge effectively to produce reliable prediction results. Consequently, these two variables were omitted from the result assessment, focusing the evaluation on the first five variables, which provided more stable and informative insights for the predictive models. This adjustment demonstrates the complexity of real-world wildfire spread, where not all input factors may contribute to model performance. In addition, the satellite-observed wildfire spread datasets do not account for spot fire scenarios, since the fire spread dataset was derived from an object tracking algorithm, where all fires are considered as part of a continuous spread rather than isolated incidents.

Results of model performance comparisons (Table 3) show that the patchwise ConvLSTM (M3) model has the highest precision, recall, and F1-score, and the lowest fire and non-fire RMSE, followed by the pairwise attention ConvLSTM (M2). Both M2 and M3 outperform the standard ConvLSTM (M1) model across the examining metrics. Furthermore, Fig. 7 illustrates the differences in model performances using a series of box plots. For M1, the upper range, mean, and lower range values of precision, recall, and F-score were consistently lower and wider than the other two models. The F-1 Score values for different attention footprint sizes (3, 5, 7, 9, 11) of M3 were very close to each other, suggesting that those footprint sizes had minimal to no impact in the model outcome.

All three ConvLSTM models perform well in active fire areas, as manifested by a few target wildfire events (Fig. 8). The standard ConvLSTM model (M1) exhibited high precision (between 0.98 and 0.99) in fire predictions but shows a limitation in fire detection sensitivity, as shown by a lower recall (between 0.83 and 0.91) than M2 and M3, pointing to a higher number of false negatives. Moreover, the M1 model's predictions are characterized by a more dispersed spread, suggesting a tendency towards broader identification of areas as potentially affected by fire, which correlates with its relatively increased false positive rate. The pairwise ConvLSTM (M2) improves upon M1 by reducing false negatives, leading to a higher recall (between 0.88 and 0.96) without loss in precision (~0.99), resulting in a higher F1-Score (between 0.94 and 0.98) than M1 model (between 0.90 and 0.96). The patchwise ConvLSTM (M3) with footprint size 9×9 , maintains high precision (~0.99), further reduces false negatives, evidenced by slightly higher recall (between 0.90 and 0.96) than M2 and a competitive F1-

Score (between 0.95 and 0.98), indicative of a balanced performance in predicting fire and non-fire classes.

Fig. 9 demonstrates the maximum Integrated Gradients for different variables over time, illustrating the importance of each variable in the model predictions. Alongside these gradients, the figure includes true positive (TP), false positive (FP), and false negative (FN) predictions for three ConvLSTM-based models—standard, pairwise, and patchwise. In the standard ConvLSTM model, the integrated gradient values are relatively uniform across the variables, suggesting that the model relies on a more distributed contribution of features to make its predictions, which might contribute to the higher number of false negatives indicated in the TP, FP, FN analysis map. In the Pairwise ConvLSTM model, the integrated gradients are more focused, with certain areas appearing distinctly higher. This concentration of high gradient values indicates a more targeted reliance on specific features for making predictions. The corresponding classification map shows improved predictive accuracy, with a higher number of true positives and fewer false negatives. The patchwise ConvLSTM model also shows focused areas of high attribution, but the distribution differs from the pairwise model.

For the standard ConvLSTM model, the areas of FP are noticeable, representing missed prediction of active fire fronts. Compared to the standard ConvLSTM, the pairwise model shows an increased number of TPs and a reduced number of FNs, indicating an improvement in detecting burnt areas. The FPs appear slightly increased, which may point to a higher occurrence of misclassified unburnt areas, but the overall classification accuracy, as indicated by the TPs, is higher. The patchwise ConvLSTM also displays an improved performance over the standard ConvLSTM, with a dense distribution of TPs and reduced FNs. The FPs are minimal, suggesting a high level of precision in the model.

7. Discussions

We addressed the challenge of accurate, reliable spatiotemporal prediction and meaningful physical interpretations of geographic event progression with wildfire spread as an example using two datasets: 1) a high-resolution simulated dataset produced with a semi-empirical percolation model, and 2) satellite observed wildfire spread data in California 2012–2021. The primary contribution of the research is the demonstration of a set of proposed attention-based ConvLSTM models for prediction of propagation of natural hazards with wildfire as a case study. This includes a comprehensive evaluation of the different models' performance in accurately forecasting fire spread and biophysical interpretability (i.e., how these models process various environmental factors to make spatiotemporal predictions and how that coincides with the known wildfire ecology), followed by a comparison of the impact of parameters and factors like fire threshold (i.e., fire probability cut-off) and foot-print size on predictions, estimation of each model's prediction uncertainty, and explore model transferability aspects using real-world wildfire datasets. Lastly, we discuss a few challenging aspects of implementing the proposed models in real-world applications and guidelines for future research.

Table 3
Model performance comparison.

Model	RMSE			Precision	Recall	F1-Score	
	Overall	Fire	Non-fire				
ConvLSTM (M1)	0.082	0.212	0.043	0.941	0.908	0.918	
Pairwise attention ConvLSTM (M2)	0.049	0.151	0.023	0.976	0.936	0.953	
Patchwise attention ConvLSTM (M3) with varying footprint size	3 5 7 9 11	0.043 0.045 0.045 0.044 0.045	0.125 0.130 0.130 0.132 0.134	0.019 0.021 0.019 0.019 0.017	0.974 0.979 0.976 0.979 0.983	0.949 0.947 0.946 0.945 0.944	0.959 0.960 0.958 0.959 0.960

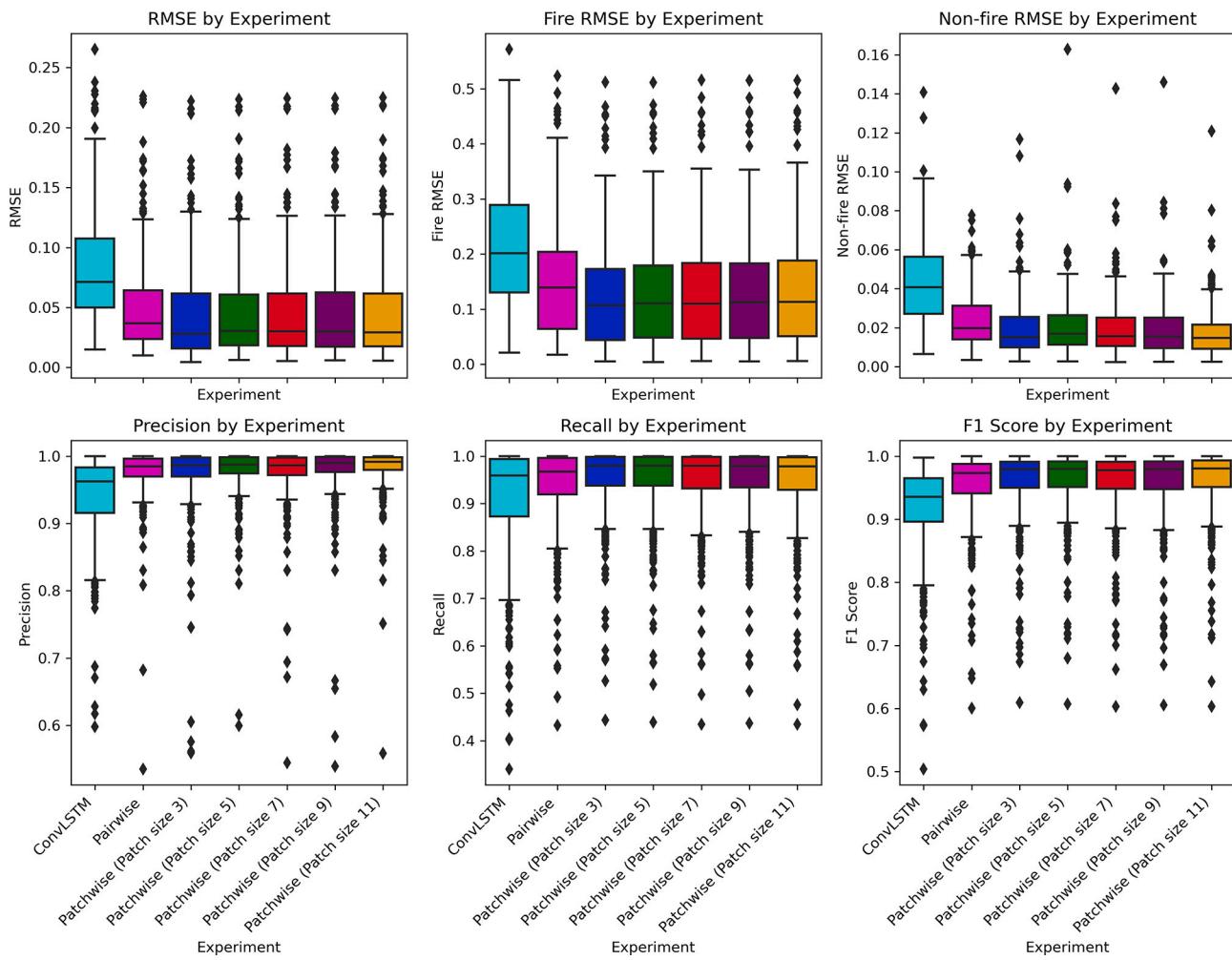


Fig. 7. Box plots showing the range of error metrics comparison by model experiments.

7.1. Predictability of attention-based ConvLSTMs

Pairwise and patchwise ConvLSTM models outperformed the standard (i.e., non-attention) ConvLSTM to accurately predict wildfire spread in real-world experiments. This outperformance can be seen across metrics, including prediction, recall, and F1-score over different fire thresholds, and the RMSEs. The pairwise attention mechanism's effectiveness is particularly notable in its enhanced ability to accurately capture and predict core areas of wildfire and the dynamics of spot fire occurrence (see section 5.1). We recognize that the simulated spot fires might not fully capture the complex, irregular patterns of real wildfire spotting due to the abstract nature of the model. The percolation model incorporates stochastic elements influenced by wind and terrain. Spot fires in the simulations are generated based on ember transport mechanisms, which are inherently stochastic (Keeley and Zedler, 2009) and influenced by the simulated wind conditions at each timestep. This approach may lead to the appearance of spot fires around the perimeter, which might seem unnatural compared to specific real-world observations where spot fires are more sporadic and less uniformly distributed (Cruz et al., 2012).

In the satellite-observed wildfire spread experiment, the pairwise attention ConvLSTM model demonstrated the highest precision and F1-score, with the lowest non-fire RMSE, indicating its superior performance in wildfire prediction tasks. The patchwise attention ConvLSTM, while exhibiting similar precision to pairwise, achieved the highest recall, indicating its strength in detecting the majority of fire instances, with a slightly increased false positive rate. Both models outperformed

the regular ConvLSTM across the examined metrics. We further explored the location accuracy of spot fire predictions, emphasizing the pairwise ConvLSTM's consistent superior performance over the other models, which supports the idea that pairwise self-attention can be powerful over the convolutional baselines (Zhao et al., 2020).

7.2. Model interpretability

The integrated gradient analysis for model interpretability presented in Section 5.2 provides insights into why the pairwise ConvLSTM model outperforms the other models for predicting fire spread and spot fires. In the pairwise ConvLSTM, the tighter clustering of integrated gradient values for most variables indicates this model can capture stable and consistent interactions between feature pairs, which is critical for understanding the complex dynamic in wildfire spread. This is further evidenced by the concentrated spatial patterns of integrated gradients in the pairwise ConvLSTM model, which show a more concentrated pattern over the predicted fire front compared to the standard ConvLSTM. The pairwise model's capacity to capture both local and distant spatial dependencies enables it to identify subtle as well as larger-scale patterns such as spot fires. These capabilities are key to its superior performance in wildfire prediction, as they allow for a more comprehensive and focused analysis of fire behavior, making the pairwise ConvLSTM model particularly effective in these predictive tasks.

The biophysical interpretability of the proposed models is aligned with the exiting literature in wildfire ecology (see Section 5.2). Overall, across all models, the unburned vegetation or availability of fuel

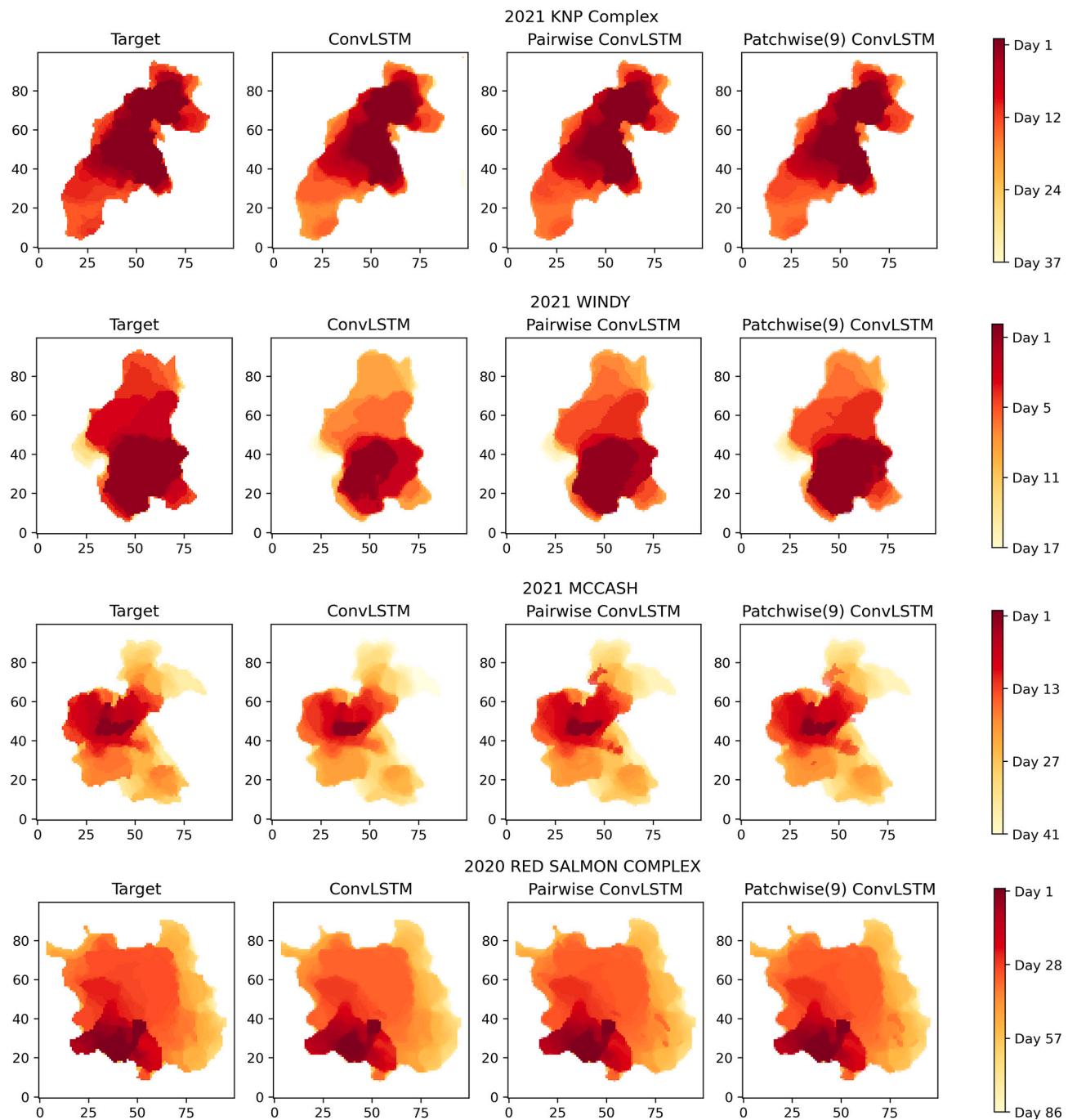


Fig. 8. Visualization of the target and the predictions for specific test fire samples from California. The color intensity represents the day of the target and predicted fire event at each grid cell, with red indicating early-stage fires and yellow indicating areas of late-stage fires. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

appeared to be the most influential on the model's output. Fuel quantity and characteristics are integral to fire both the fire severity (Taylor et al., 2021) and spread (Oliveira et al., 2021). In the standard ConvLSTM, a contrast in trends for the two wind components (horizontal and vertical) was observed, which reflects the model's adaptation to varied roles of different wind components in fire spread behavior (Rossa and Fernandes, 2018). As described in section 5.2.1, the pairwise ConvLSTM model's integrated gradient values for elevation highlighted on fire's tendency to spread to higher elevations decreases as it spreads – a pattern that coincides with the scale-dependent relationship between fire severity and spread behavior (Povak et al., 2018). We also found that the dependence fire spread on dry fuel changes as the fire

progresses, suggesting initially dry conditions might be critical for fire spread, but the importance of moisture content change as the fire moves to different environments (e.g., different type and arrangement of fuels) or as conditions evolve (Holsinger et al., 2016). These examples of the influences of fuel, wind, elevation, and moisture on wildfire spread suggest that the trained model captured the fire-environmental dynamics that were introduced in our simulated data by the percolation model.

7.3. Effects of fire threshold values and footprint size

One of the important parameters in wildfire spread prediction is the

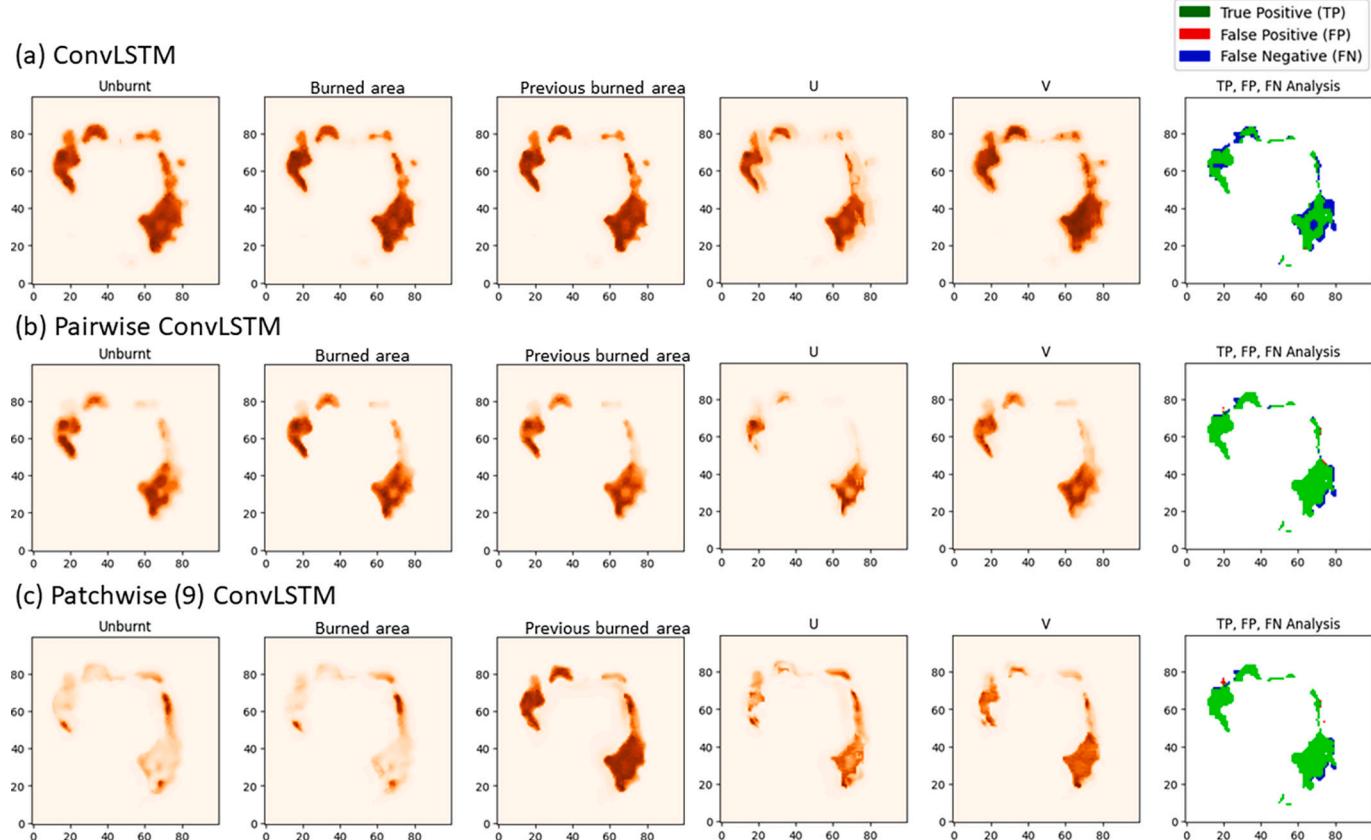


Fig. 9. Maximum Integrated Gradients of different variables and true positive, false positive, and false negative predictions.

fire threshold value – i.e., a cut-off value for fire probability determining whether or not each grid cell has been burned (Hodges and Lattimer, 2019), which directly impacts the model performance metrics (e.g., precision, recall, F-1 score). To demonstrate such sensitivity, we computed these metrics for the real-world dataset using a range between 0.1 and 0.6 for fire threshold values (thresholds chosen after experiments). Fig. 10 shows values for all three metrics are affected by the threshold values, helping us to determine the optimum threshold value for each to accurately predict fire front. For instance, a fire threshold of 0.2 resulted in the highest F1-Score for all three models. However, for the non-attention ConvLSTM, F-Score decreases rapidly with increasing

fire threshold value. Patchwise and pairwise attention models were competitive across the fire threshold range for precision, recall, and F1-Score.

Our study also illustrates the distinct behavior of the patchwise attention ConvLSTM in relation to footprint size. Despite varying footprint sizes, the changes in RMSE were minimal, suggesting that the patchwise model's performance was not significantly influenced by the size of the attention footprint. This finding is critical, as it points to the model's stability across different spatial resolutions. However, as mentioned previously in Section 5.1, the patchwise ConvLSTM model achieved one of the highest performances among the footprint sizes

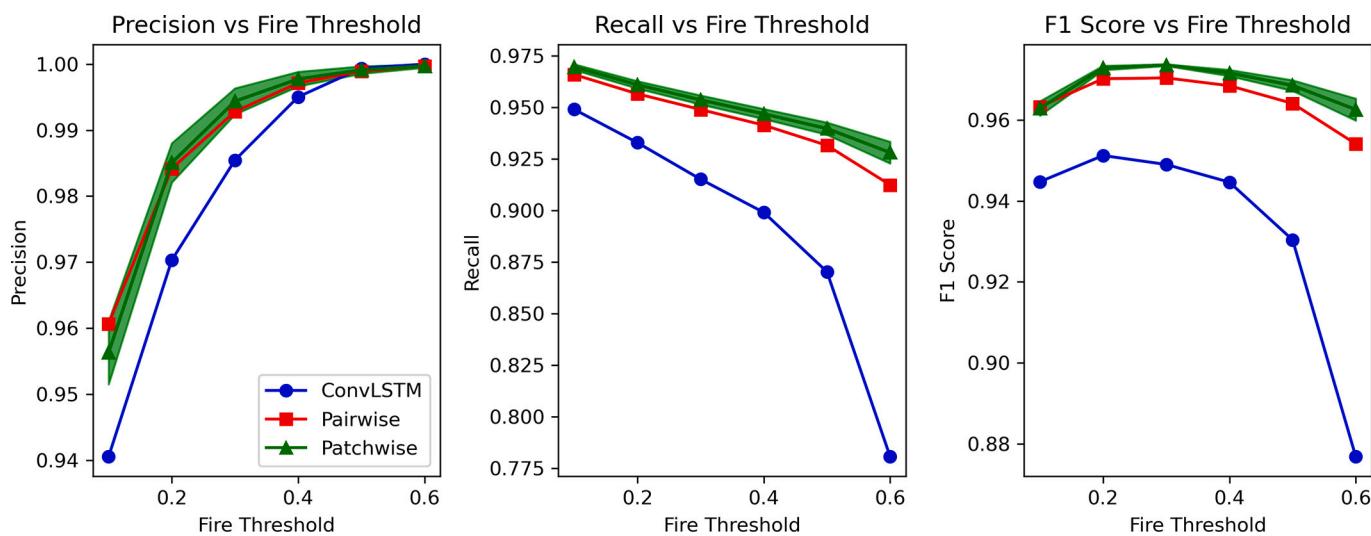


Fig. 10. Effects of fire thresholds (fire probability cut-off) on error metrics for overall fire predictions in the test dataset.

tested, similar to Marjani et al. (2024) that showed a 7×7 footprint size in a CNN-based approach outperformed other models.

7.4. Transferability of models

We compared the transferability of each model that were trained over a region to accurately predict wildfire spread behavior in another region, following the approaches mentioned in (Applestein et al., 2021; Marjani et al., 2023). Data for model transferability tests include several wildfires in northern California (Fig. 11). The trained non-attention ConvLSTM and pairwise and patchwise ConvLSTM models were employed to predict wildfire spread in both the ‘main’ (over the training regions) and ‘test’ (over the transfer regions) examples at varied fire thresholds ranging between 0.1 and 0.6. Fig. 12 shows the performance comparison for all models between the main and test fires based on precision, recall and F1-Score metrics. Overall, the patchwise attention (with footprint size 5) mechanism’s precision score exhibited the least gap while predicting fire spread behavior in the main and test regions – a pattern which was also consistent over different fire thresholds. Comparing the recall scores showed similar patterns for all three models, however, the ConvLSTM model had the lowest recall at fire threshold 0.6, indicating its struggle to predict true fire fronts than the other two models. The F1 score also suggests that the patchwise attention is slightly superior in transferability than the other two models. It is evident that the pairwise attention struggles in terms of transferability, performing almost like the non-attention model, which suggests that

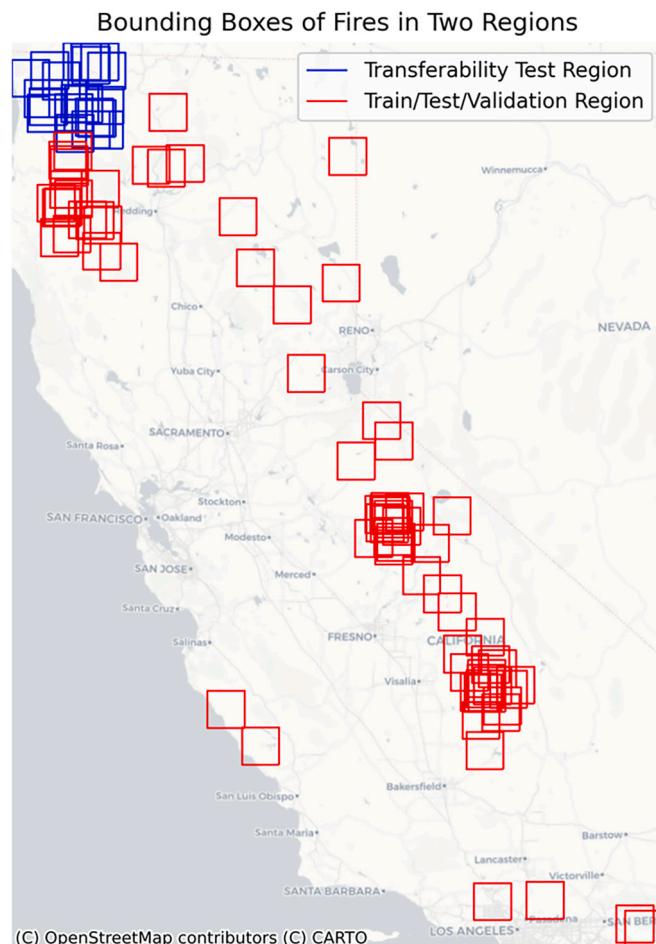


Fig. 11. Bounding boxes of fires in two regions in California: blue and red boxes include fires for transferability test and model train/test/validation, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

pairwise attention mechanism might be better suited for capturing individual fire spread dynamics in a region but not applicable to other regions. In contrast, patchwise attention mechanisms may manifest generalizability and transferability across regions.

7.5. Uncertainty estimation using Monte Carlo dropout

Estimating model uncertainty is crucial in deep learning (Lemay et al., 2022) and specially in predicting inherently stochastic nature of wildfire spread behavior that are influenced by a complex interaction among data variables and oftentimes incomplete information. Using Monte Carlo dropout approach – an approximate Bayesian inference technique (Gal and Ghahramani, 2016) – we computed mean and standard deviation of all three model uncertainties in fire spread predictions at different fire thresholds (0.1 to 0.6). For that, we used a dropout rate of 0.2 and 50 stochastic predictions for each sample. Our held-out validation dataset contained 15 fire events in California over the 2020–2021 period. Fig. 13 box plots show that the non-attention ConvLSTM had the highest mean and standard deviation of uncertainty across fire thresholds. The least uncertainty is represented by the patchwise attention model.

The maps in Fig. 14 show spatial distribution of the standard deviation of uncertainties for four different fire events. The higher values on the uncertainty maps represent the primary sources of errors in the model predictions. It is evident that the non-attention ConvLSTM had larger coverage of uncertain areas and larger std. values contributing to model predictions, whereas in pair and patchwise attention models these uncertainties were mostly presented in the border areas. In particular, the pairwise attention model had the smallest coverage of uncertainty and mostly around the edges of the fire perimeter. This kind of uncertainties on the fire borders areas were also detected by (Marjani et al., 2023).

7.6. Limitations and future research

Wildfire spread prediction using deep learning approaches face various challenges including limited good-quality empirical wildfire dataset (e.g., coarse spatiotemporal resolution of both the fire perimeter and contributing factors), modeling complex stochastic processes, model transferability and generalizability, and high computational requirements. Although we aimed to address a few of these challenges by integrating spatiotemporal attention mechanisms within a sequence-to-sequence modeling framework and applied to semi-empirical and empirical datasets, certain limitations were met with respect to data quality both for the percolation model-derived and real-world datasets. The percolation model of Burge et al. (2020) was not calibrated to reproduce real wildfire behavior, thus it was only useful for training our fire spread models to emulate the real-world behavior, therefore our prediction and interpretability results should be interpreted cautiously. For instance, although topography is one of the most important factors driving wildfire activity (Harris and Taylor, 2017; Masrur et al., 2022), our ConvLSTM models were not able to consistently recognize elevation as a significant influencing factor for fire spread behavior, potentially due to the percolation model not adequately capturing the complex effects of elevation on wildfire spread, particularly in varied terrain. Consequently, to further examine model performances we conducted a series of experiments with real-world fire spread datasets from California, which showed promise for transferability and model uncertainty. However, we were not able to include two important biophysical factors such as moisture and topography into the model since they were not available within the spatiotemporal scale of other datasets.

In future research, we plan to investigate how well both the non-attention ConvLSTM and attention-based ConvLSTM models perform in retrospective reconstruction of real-world wildfire progression. This approach will enable us to analyze past wildfire events in detail at various spatiotemporal scales, utilizing granular datasets including land

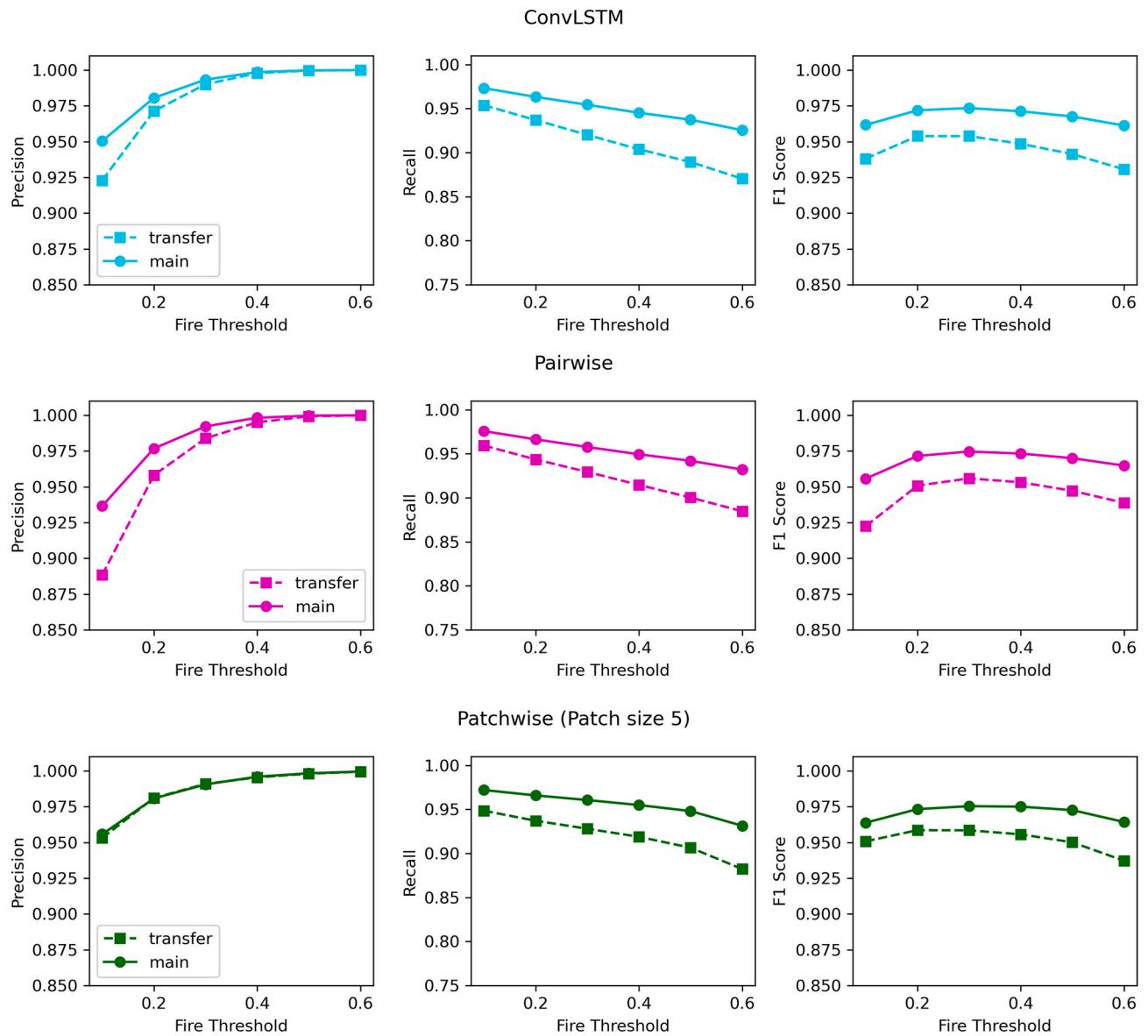


Fig. 12. Comparison of model's transferability over different geographic areas.

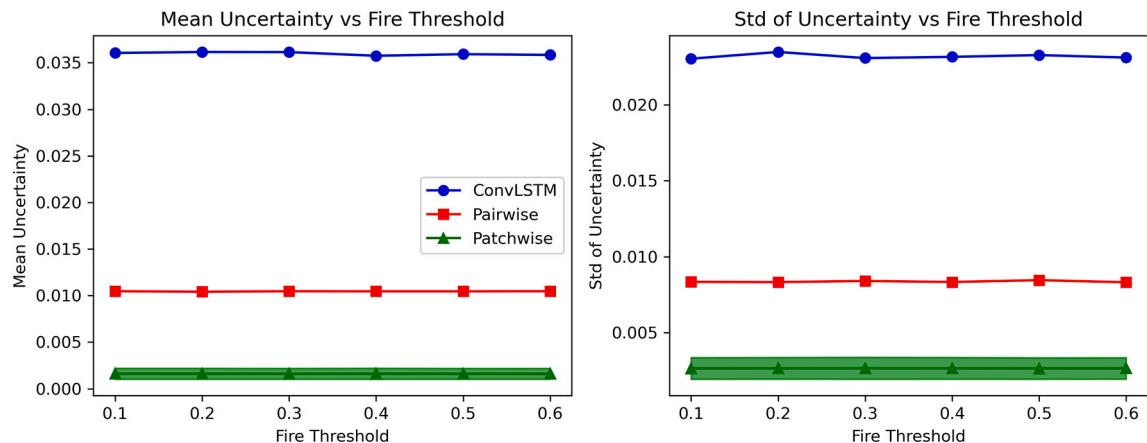


Fig. 13. Box plots of uncertainty metrices (mean and standard deviation) by model experiments.

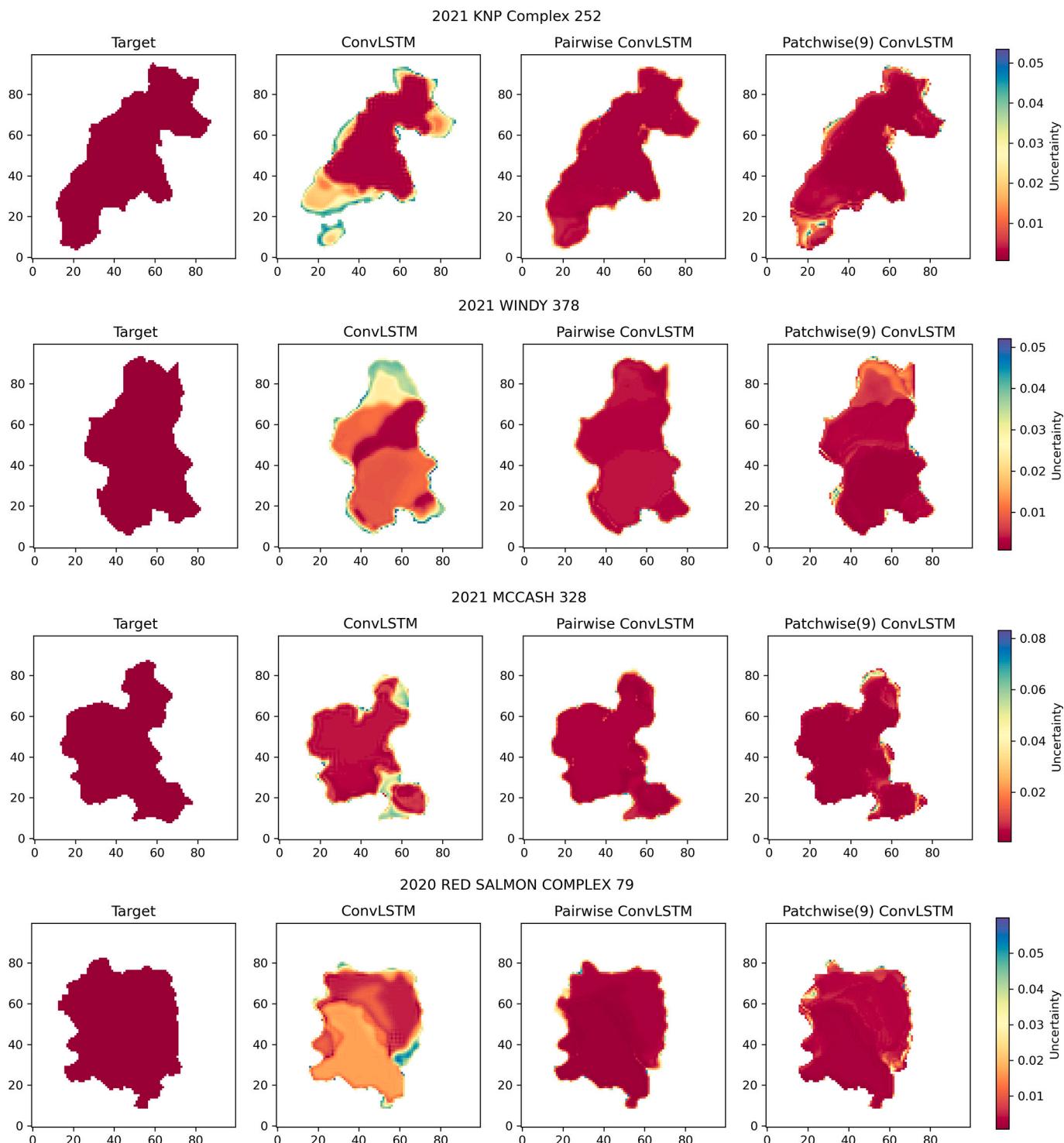


Fig. 14. Maps showing comparison of the std. of prediction uncertainties by different model experiments for four different fire events.

skin temperature, soil moisture, and historical weather conditions. By reconstructing past fire events, we aim to gain a deeper understanding of the complex interactions between various environmental factors and fire behavior. The retrospective analysis will also be instrumental in identifying patterns and trends that could improve the accuracy and effectiveness of future predictive models. Understanding fire behavior through the retrospective analysis through the model's interpretative insights can also inform conservation strategies particularly in fire-adapted ecosystems.

In addition, we intend to investigate the integration of machine or

deep learning techniques and fire percolation models to create hybrid systems for more robust and computationally efficient fire prediction models. We will validate such models under different geographic, climatic, and topographic environments to ensure the generalizability and effectiveness in different fire-prone regions. We hope to extend our research to practical applications of forest fire management and preservation of ecological environments. The enhanced predictive accuracy of the pairwise ConvLSTM model, especially in identifying fire fronts and spot fires, can potentially improve the early-warning system and the efficiency of firefighting efforts.

8. Conclusions

In this paper, we have applied attention-based spatiotemporal models using Convolutional Long Short-Term Memory (ConvLSTM) networks to address the challenge of accurately predicting and interpreting a dynamic spatiotemporal event such as wildfire spread. Wildfires frequently display short- and long-range spatial and temporal correlations: short-range effects are the direct contact and near-contact spread of the fire front, whereas long-range effects are represented by processes such as *spotting*, where firebrands carried by the wind ignite fires distant from the flaming front. We define this scale-dependent spatiotemporal dynamics of wildfire spread via local to global interactions among biophysical factors (e.g., fuel, wind, topography, and moisture) and propose two unique attention-based spatiotemporal models using Convolutional Long Short-Term Memory (ConvLSTM) networks. These networks are designed to learn and capture a range of local to global and short and long-range spatiotemporal correlations using the proposed pair and patchwise self-attention mechanisms. Our experiments on both simulated (based on a semi-empirical percolation model) and real-world fire progression (satellite observed wildfire spread data in California 2012–2021) datasets show that self-attention mechanisms within ConvLSTM networks can accurately predict fire front movements. We have also highlighted the transferability, model uncertainty, and biophysical interpretability of the proposed networks that showed consistency with the known knowledge of wildfire spread–biophysical dynamics. Our study indicates that self-attention attention mechanisms have potential to address the challenge of modeling wildfire spread behavior, thus can be useful to develop early-warning systems for firefighting and wildfire management efforts. We also highlight a set of directions for future studies to focus on how the self-attention mechanisms within sequence-sequence modeling frameworks could enhance model performance for a range of geospatial applications.

CRediT authorship contribution statement

Arif Masrur: Writing – review & editing, Writing – original draft, Methodology, Data curation, Conceptualization. **Manzhu Yu:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Project administration, Methodology, Investigation, Conceptualization. **Alan Taylor:** Writing – review & editing.

Data availability

Data will be made available on request.

References

- Abatzoglou, J.T., Williams, A.P., 2016. Impact of anthropogenic climate change on wildfire across western US forests. Proc. Natl. Acad. Sci. USA 113, 11770–11775. <https://doi.org/10.1073/pnas.1607171113>.
- Alexandridis, A., Vakalis, D., Siettos, C.I., Bafas, G.V., 2008. A cellular automata model for forest fire spread prediction: the case of the wildfire that swept through Spetses Island in 1990. Appl. Math. Comput. 204, 191–201. <https://doi.org/10.1016/j.amc.2008.06.046>.
- Applestein, C., Cauglin, T.T., Germino, M.J., 2021. Weather affects post-fire recovery of sagebrush-steppe communities and model transferability among sites. Ecosphere 12, e03446. <https://doi.org/10.1002/ecs2.3446>.
- Beer, T., Enting, I.G., 1990. Fire spread and percolation modelling. Math. Comput. Model. 13, 77–96. [https://doi.org/10.1016/0895-7177\(90\)90065-U](https://doi.org/10.1016/0895-7177(90)90065-U).
- Bello, I., Zoph, B., Le, Q., Vaswani, A., Shlens, J., 2019. Attention augmented convolutional networks. In: Proc. IEEE Int. Conf. Comput. Vis. 2019-Octob, pp. 3285–3294. <https://doi.org/10.1109/ICCV.2019.00338>.
- Buch, J., Williams, A.P., Juang, C.S., Hansen, W.D., Gentine, P., 2023. SMLFire1.0: a stochastic machine learning (SML) model for wildfire activity in the western United States. Geosci. Model Dev. 16, 3407–3433. <https://doi.org/10.5194/gmd-16-3407-2023>.
- Burge, J., Bonanni, M., Ihme, M., Hu, L., 2020. Convolutional LSTM Neural Networks for Modeling Wildland Fire Dynamics.
- Cai, L., Janowicz, K., Mai, G., Yan, B., Zhu, R., 2020. Traffic transformer: capturing the continuity and periodicity of time series for traffic forecasting. Trans. GIS 24, 736–755. <https://doi.org/10.1111/TGIS.12644>.
- Cao, Y., Yang, F., Tang, Q., Lu, X., 2019. An attention enhanced bidirectional LSTM for early forest fire smoke recognition. IEEE Access 7, 154732–154742. <https://doi.org/10.1109/ACCESS.2019.2946712>.
- Chen, Y., Hantson, S., Andela, N., Coffield, S.R., Graff, C.A., Morton, D.C., Ott, L.E., Foufoula-Georgiou, E., Smyth, P., Goulden, M.L., Randerson, J.T., 2022. California wildfire spread derived using VIIRS satellite observations and an object-based tracking system. Sci. Data 9, 249. <https://doi.org/10.1038/s41597-022-01343-0>.
- Coen, J.L., Cameron, M., Michalakes, J., Patton, E.G., Riggan, P.J., Yedinak, K.M., 2013. WRF-fire: coupled weather–wildland fire modeling with the weather research and forecasting model. J. Appl. Meteorol. Climatol. 52, 16–38. <https://doi.org/10.1175/JAMC-D-12-023.1>.
- Cruz, M.G., Sullivan, A.L., Gould, J.S., Sims, N.C., Bannister, A.J., Hollis, J.J., Hurley, R.J., 2012. Anatomy of a catastrophic wildfire: the Black Saturday Kilmore East fire in Victoria, Australia. For. Ecol. Manag. 284, 269–285. <https://doi.org/10.1016/j.foreco.2012.02.035>.
- Deeming, J.E., 1972. National Fire-Danger Rating System. Rocky Mountain Forest and Range Experiment Station, Forest Service, US.
- Deeming, J.E., Burgan, R.E., Cohen, J.D., 1977. The National Fire-Danger Rating System, 1978. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station.
- Ellis, T.M., Bowman, D.M.J.S., Jain, P., Flannigan, M.D., Williamson, G.J., 2022. Global increase in wildfire risk due to climate-driven declines in fuel moisture. Glob. Chang. Biol. 28, 1544–1559. <https://doi.org/10.1111/gcb.16006>.
- Finney, M.A., 2005. The challenge of quantitative risk analysis for wildland fire. For. Ecol. Manag. 211, 97–108. <https://doi.org/10.1016/j.foreco.2005.02.010>.
- Finney, M.A., Cohen, J.D., Forthofer, J.M., McAllister, S.S., Gollner, M.J., Gorham, D.J., Saito, K., Akafuah, N.K., Adam, B.A., English, J.D., Dickinson, R.E., 2015. Role of buoyant flame dynamics in wildfire spread. Proc. Natl. Acad. Sci. USA 112, 9833–9838. <https://doi.org/10.1073/PNAS.1504498112>.
- Finney, M.A., McAllister, S.S., Grumstrup, T.P., Forthofer, J.M., 2021. Wildland Fire Behaviour.
- Gal, Y., Ghahramani, Z., 2016. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In: Balcan, M.F., Weinberger, K.Q. (Eds.), Proceedings of the 33rd International Conference on Machine Learning. Proceedings of Machine Learning Research. PMLR, New York, New York, USA, pp. 1050–1059.
- Hantson, S., Arnett, A., Harrison, S.P., Kelley, D.I., Prentice, I.C., Rabin, S.S., Archibald, S., Mouillot, F., Arnold, S.R., Artaxo, P., Bachelet, D., Ciais, P., Forrest, M., Friedlingstein, P., Hickler, T., Kaplan, J.O., Kloster, S., Knorr, W., Lasslop, G., Li, F., Mangeon, S., Melton, J.R., Meyn, A., Sitch, S., Spessa, A., van der Werf, G.R., Voulgarakis, A., Yue, C., 2016. The status and challenge of global fire modelling. Biogeosciences 13, 3359–3375. <https://doi.org/10.5194/bg-13-3359-2016>.
- Harris, L., Taylor, A.H., 2017. Previous burns and topography limit and reinforce fire severity in a large wildfire. Ecosphere 8, e02019. <https://doi.org/10.1002/ecs2.2019>.
- Hawbaker, T.J., Radeloff, V.C., Stewart, S.I., Hammer, R.B., Keuler, N.S., Clayton, M.K., 2013. Human and biophysical influences on fire occurrence in the United States. Ecol. Appl. 23, 565–582. <https://doi.org/10.1890/12-1816.1>.
- Higgins, S.I., Bond, W.J., Trollope, W.S.W., Williams, R.J., 2008. Physically motivated empirical models for the spread and intensity of grass fires. Int. J. Wildland Fire 17, 595–601. <https://doi.org/10.1071/WF06037>.
- Hilton, J.E., Miller, C., Sullivan, A.L., Rucinski, C., 2015. Effects of spatial and temporal variation in environmental conditions on simulation of wildfire spread. Environ. Model Softw. 67, 118–127. <https://doi.org/10.1016/j.envsoft.2015.01.015>.
- Hodges, J.L., Lattimer, B.Y., 2019. Wildland fire spread modeling using convolutional neural networks. Fire. Technol 55, 2115–2142. <https://doi.org/10.1007/s10694-019-00846-4>.
- Holsinger, L., Parks, S.A., Miller, C., 2016. Weather, fuels, and topography impede wildland fire spread in western US landscapes. For. Ecol. Manag. 380, 59–69. <https://doi.org/10.1016/j.foreco.2016.08.035>.
- Jain, P., Coogan, S.C.P., Subramanian, S.G., Crowley, M., Taylor, S., Flannigan, M.D., 2020. A review of machine learning applications in wildfire science and management. Environ. Rev. 28, 478–505. <https://doi.org/10.1139/er-2020-0019>.
- Jones, M.W., Abatzoglou, J.T., Veraverbeke, S., Andela, N., Lasslop, G., Forkel, M., Smith, A.J.P., Burton, C., Betts, R.A., van der Werf, G.R., Sitch, S., Canadell, J.G., Santin, C., Kolden, C., Doerr, S.H., Le Quéré, C., 2022. Global and regional trends and drivers of fire under climate change. Rev. Geophys. 60 <https://doi.org/10.1029/2020RG000726> e2020RG000726.
- Keeley, J.E., Zedler, P.H., 2009. Large, high-intensity fire events in southern California shrublands: debunking the fine-grain age patch model. Ecol. Appl. 19, 69–94. <https://doi.org/10.1890/08-0281.1>.
- Lemay, A., Hoebel, K., Bridge, C.P., Befano, B., De Sanjose, S., Egemen, D., Rodriguez, A.C., Schiffman, M., Campbell, J.P., Kalpathy-Cramer, J., 2022. Improving the repeatability of deep learning models with Monte Carlo dropout. Npj Digit. Med. 5, 1–11. <https://doi.org/10.1038/s41746-022-00709-3>.
- Li, C., 2000. Reconstruction of natural fire regimes through ecological modelling. Ecol. Model. 134, 129–144. [https://doi.org/10.1016/S0304-3800\(00\)00290-8](https://doi.org/10.1016/S0304-3800(00)00290-8).
- Li, Z., Huang, Y., Li, X., Xu, L., 2021. Wildland fire burned areas prediction using long short-term memory neural network with attention mechanism. Fire. Technol 57, 1–23. <https://doi.org/10.1007/s10694-020-01028-3>.
- Li, F., Zhu, Q., Riley, W.J., Zhao, L., Xu, L., Yuan, K., Chen, M., Wu, H., Gui, Z., Gong, J., Randerson, J.T., 2023. AttentionFire_v1.0: interpretable machine learning fire model for burned-area predictions over tropics. Geosci. Model Dev. 16, 869–884. <https://doi.org/10.5194/gmd-16-869-2023>.

- Lin, Z., Li, M., Zheng, Z., Cheng, Y., Yuan, C., 2020. Self-attention ConvLSTM for spatiotemporal prediction. Proc. AAAI Conf. Artif. Intell. 34, 11531–11538. <https://doi.org/10.1609/aaai.v34i07.6819>.
- Majid, S., Alenezi, F., Masood, S., Ahmad, M., Gündüz, E.S., Polat, K., 2022. Attention based CNN model for fire detection and localization in real-world images. Expert Syst. Appl. 189, 116114 <https://doi.org/10.1016/j.eswa.2021.116114>.
- Mandel, J., Beezley, J.D., Kochanski, A.K., 2011. Coupled atmosphere-wildland fire modeling with WRF 3.3 and SFIRE 2011. Geosci. Model Dev. 4, 591–610. <https://doi.org/10.5194/gmd-4-591-2011>.
- Marjani, M., Ahmadi, S.A., Mahdianpari, M., 2023. FirePred: a hybrid multi-temporal convolutional neural network model for wildfire spread prediction. Ecol. Inform. 78, 102282 <https://doi.org/10.1016/j.ecoinf.2023.102282>.
- Marjani, M., Mahdianpari, M., Ali Ahmadi, S., Hemmati, E., Mohammadimanesh, F., Saadi Mesgari, M., 2024. Application of explainable artificial intelligence in predicting wildfire spread: an ASPP-enabled CNN approach. IEEE Geosci. Remote Sens. Lett. 21, 1–5. <https://doi.org/10.1109/LGRS.2024.3417624>.
- Masrur, A., Yu, M., Mitra, P., Pequet, D., Taylor, A., 2022. Interpretable machine learning for analysing heterogeneous drivers of geographic events in space-time. Int. J. Geogr. Inf. Sci. 36, 692–719. <https://doi.org/10.1080/13658816.2021.1965608>.
- McArthur, A.G., 1966. Weather and grassland fire behaviour, p. 23.
- McArthur, A.G., 1967. Fire behaviour in Eucalypt forests, p. 35.
- Mckenzie, D., Peterson, D.L., Alvarado, E., 1996. Extrapolation problems in modeling fire effects at large spatial scales: a review. Int. J. Wildland Fire 6, 165–176. <https://doi.org/10.1071/wf9960165>.
- Oliveira, U., Soares-Filho, B., de Souza Costa, W.L., Gomes, L., Bustamante, M., Miranda, H., 2021. Modeling fuel loads dynamics and fire spread probability in the Brazilian Cerrado. For. Ecol. Manag. 482, 118889 <https://doi.org/10.1016/j.foreco.2020.118889>.
- Povak, N.A., Hessburg, P.F., Salter, R.B., 2018. Evidence for scale-dependent topographic controls on wildfire spread. Ecosphere 9, e02443. <https://doi.org/10.1002/ecs2.2443>.
- Quinn, N., Bates, P.D., Neal, J., Smith, A., Wing, O., Sampson, C., Smith, J., Heffernan, J., 2019. The spatial dependence of flood hazard and risk in the United States. Water Resour. Res. 55, 1890–1911. <https://doi.org/10.1029/2018WR024205>.
- Rossa, C.G., Fernandes, P.M., 2018. An empirical model for the effect of wind on fire spread rate. Fire 1, 31. <https://doi.org/10.3390/fire1020031>.
- Rothermel, R.C., 1972. A Mathematical Model for Predicting Fire Spread in Wildland Fuels. Intermountain Forest & Range Experiment Station. Forest Service, U.S. Department of Agriculture.
- Ruffault, J., Mouillot, F., 2017. Contribution of human and biophysical factors to the spatial distribution of forest fire ignitions and large wildfires in a French Mediterranean region. Int. J. Wildland Fire 26, 498–508. <https://doi.org/10.1071/WF16181>.
- Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W., Woo, W., 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In: Advances in Neural Information Processing Systems. Curran Associates, Inc.
- Subramanian, S.G., Crowley, M., 2018. Using spatial reinforcement learning to build forest wildfire dynamics models from satellite images. Front. ICT 5.
- Sullivan, A.L., 2009a. Wildland surface fire spread modelling, 1990–2007. 2: empirical and quasi-empirical models. Int. J. Wildland Fire 18, 369–386. <https://doi.org/10.1071/WF06142>.
- Sullivan, A.L., 2009b. Wildland surface fire spread modelling, 1990–2007. 1: physical and quasi-physical models. Int. J. Wildland Fire 18, 349–368. <https://doi.org/10.1071/WF06143>.
- Sullivan, A.L., 2009c. Wildland surface fire spread modelling, 1990–2007. 3: simulation and mathematical analogue models. Int. J. Wildland Fire 18, 387–403. <https://doi.org/10.1071/WF06144>.
- Sundararajan, M., Taly, A., Yan, Q., 2017. Axiomatic Attribution for Deep Networks.
- Taylor, A.H., Harris, L.B., Drury, S.A., 2021. Drivers of fire severity shift as landscapes transition to an active fire regime, Klamath Mountains, USA. Ecosphere 12. <https://doi.org/10.1002/ecs2.3734>.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. In: Adv. Neural Inf. Process. Syst. 2017-Decem, pp. 5999–6009.
- Wagner, C.V., 1990. Six decades of forest fire science in Canada. For. Chron. 66, 133–137.
- Wang, Xinyu, Wang, Xinquan, Zhang, M., Tang, C., Li, X., Sun, S., Wang, Y., Li, D., Li, S., 2023. Predicting the continuous spatiotemporal state of ground fire based on the expended LSTM model with self-attention mechanisms. Fire 6. <https://doi.org/10.3390/fire6060237>.
- Woo, S., Park, J., Lee, J.Y., Kweon, I.S., 2018. CBAM: Convolutional block attention module. In: Lect. Notes Comput. Sci. Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinforma. 11211 LNCS, pp. 3–19. https://doi.org/10.1007/978-3-030-01234-2_1.
- Zhao, H., Jia, J., Koltun, V., 2020. Exploring self-attention for image recognition. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 10073–10082. <https://doi.org/10.1109/CVPR42600.2020.01009>.