

## Article

# Decadal Wildfire Risk Projection in Guangxi, China, Using a Two-Stage Deep Learning Framework with Recursive Environmental Factor Forecasting

Kun Xu <sup>1,2,†</sup>, Zhentao Liang <sup>2,3,†</sup>, Xin Pan <sup>2,3,\*</sup>, Kevin Tansey <sup>4</sup>, Rufat Guluzade <sup>1</sup>, Zi Yang <sup>1,3</sup>, Yingbao Yang <sup>2,3</sup> and Akram Abdulla <sup>4</sup>

<sup>1</sup> School of Earth Sciences and Engineering, Hohai University, Nanjing 211100, China

<sup>2</sup> Key Laboratory of Soil and Water Processes in Watershed, Hohai University, Nanjing 211100, China

<sup>3</sup> College of Geography and Remote Sensing, Hohai University, Nanjing 210098, China

<sup>4</sup> School of Geography, Geology and the Environment, University of Leicester, Leicester LE1 7RH, UK

\* Correspondence: px1013@hhu.edu.cn

† These authors contributed equally to this work.

## Highlight

A two-stage deep learning framework enabled decadal wildfire prediction without future observational inputs.

- The LSTM-Transformer hybrid model achieved an AUC of 0.967 on the test set.
- 1D-CNN contributed most to model performance, followed by Skip-LSTM.
- Attention weights revealed 12-month periodicity aligned with the fire season.
- Projected risk maps (2019–2028) identified stable high-risk zones in western and northern Guangxi

## Abstract

Multi-year wildfire risk projection requires environmental drivers to be estimated beyond the observational period. This study develops a two-stage framework for Guangxi, China. FactorLSTM recursively estimates nine time-varying environmental factors, and an LSTM-Transformer combines these estimates with four projection-static factors to generate monthly wildfire probability estimates on a 1 km grid for 2019–2028. Observed-driver classification and recursive, predicted-driver evaluation are distinguished. An additional retrospective experiment using a separately trained classifier with a positive-class weight of 10 and Platt calibration yielded a pooled ROC-AUC of 0.7835 and average precision of 0.001475 for available valid observations in 2019–2025. However, its mean predicted probability was approximately 4.50 times the observed positive fraction, indicating residual overestimation. Seasonal maps summarise monthly outputs for January–April, May–September, and October–December. The ten-year outputs are conditional projections under extrapolated environmental drivers and fixed projection-stage land cover. The results identify limitations in probability calibration, rare-event detection, and observation coverage; they do not establish reliable monthly forecasting across the complete ten-year horizon.

Received: 21 July 2026

Revised: 22 September 2026

Accepted: 29 September 2026

Published: 2 October 2026

**Copyright:** © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

**Keywords:** wildfire prediction; LSTM–Transformer; deep learning; remote sensing; recursive factor prediction; Guangxi

---

## 1. Introduction

Wildfire is a major disturbance in terrestrial ecosystems, releasing vast quantities of carbon, degrading air quality, destroying property, and reshaping vegetation patterns across time scales from months to centuries [1,2]. Global fire regime analyses based on multi-source satellite records [3] have further documented the spatial and temporal heterogeneity of wildfire activity across biomes. In China alone, an average of over 3000 forest fires are reported annually, with southern provinces bearing a disproportionate share due to the combination of dense vegetation cover, complex terrain, and seasonal drought [4]. Guangxi Zhuang Autonomous Region, situated in subtropical southern China, records some of the highest fire frequencies in the country, driven by a pronounced dry season from October to March and extensive tracts of fire-prone coniferous and broadleaf forests [5].

Predicting where and when wildfires are likely to occur is central to effective fire management. Traditional approaches rely on meteorological fire danger indices—such as the Canadian Forest Fire Weather Index (FWI) [6] or the Keetch–Byram Drought Index [7]—that estimate daily fire weather conditions from a small set of weather variables. While operationally useful for short-term warnings, these indices do not account for vegetation state, fuel accumulation, or terrain effects, and they provide no mechanism for projecting risk beyond the weather forecast horizon of a few days.

Statistical and machine learning methods have expanded the scope of wildfire prediction by incorporating a wider range of environmental drivers [8]. Logistic regression, random forests, and gradient boosting machines have been applied to map fire susceptibility from combinations of topographic, climatic, vegetation, and anthropogenic variables [9–11]. These approaches typically produce static probability maps that characterise the long-term average fire environment but do not capture temporal dynamics: they cannot distinguish, for example, between a dry-season month with anomalously low fuel moisture and a normal month at the same location.

Deep learning methods, particularly recurrent neural networks and their variants, have been increasingly applied to temporal wildfire prediction because they can model sequential dependencies in multi-variate time series [12,13]. Long Short-Term Memory (LSTM) networks [14] have been used for wildfire occurrence and burned area prediction at time scales ranging from daily to monthly, owing to their ability to capture both short-term weather fluctuations and longer-term seasonal and interannual patterns [5]. More recently, Transformer architectures [15], originally developed for natural language processing, have been adapted for time-series modelling in environmental science, offering the advantage of attending to any position in the input sequence regardless of temporal distance. Despite this progress, two gaps remain. Most deep learning studies for wildfire prediction assume that all driving factors are known at the time of prediction. This assumption holds for hindcast evaluation but breaks down for multi-year projections. Additionally, while both LSTM and Transformer architectures have been applied separately, their combination in a hybrid architecture inspired by LSTNet [16] has received limited attention in wildfire science. Recent studies have demonstrated the effectiveness of deep learning for wildfire danger prediction from historical remote sensing data [17,18], further motivating the adoption of such approaches in the present study.

This study addresses both gaps by proposing a two-stage LSTM–Transformer framework for decadal wildfire probability prediction in Guangxi, China. A FactorLSTM module recursively predicts nine dynamic factors up to 120 months ahead, enabling an LSTM–Transformer hybrid model—integrating 1D-CNN, Skip-LSTM, multi-head attention, and autoregressive correction—to project pixel-level monthly wildfire probabilities from 2019 to 2028 without requiring future observations. A systematic ablation study and attention weight analysis are conducted to quantify each component’s contribution and interpret the temporal patterns learned by the model. Here, “decadal projection” denotes the 120-month output horizon conditional on extrapolated drivers, rather than a claim that monthly fire occurrence is equally predictable throughout all ten years.

(1) We design a two-stage prediction pipeline in which a FactorLSTM module recursively predicts nine dynamic environmental factors up to 120 months ahead, enabling the LSTM–Transformer to project wildfire risk from 2019 to 2028 without requiring future observations.

(2) We develop an LSTM–Transformer hybrid architecture integrating temporal convolution, recurrent encoding with a 12-month skip connection, self-attention, and autoregressive correction. The factor predictor and wildfire classifier are trained separately and linked sequentially for end-to-end prediction and evaluation.

(3) We conduct a systematic ablation study and attention weight analysis to quantify the contribution of each architectural component and to interpret the temporal patterns learned by the model.

The remainder of this paper is organised as follows. Section 2 describes the study area and data sources. Section 3 presents the methodology. Section 4 reports the results. Section 5 discusses the findings. Section 6 concludes the paper.

## 2. Study Area and Data

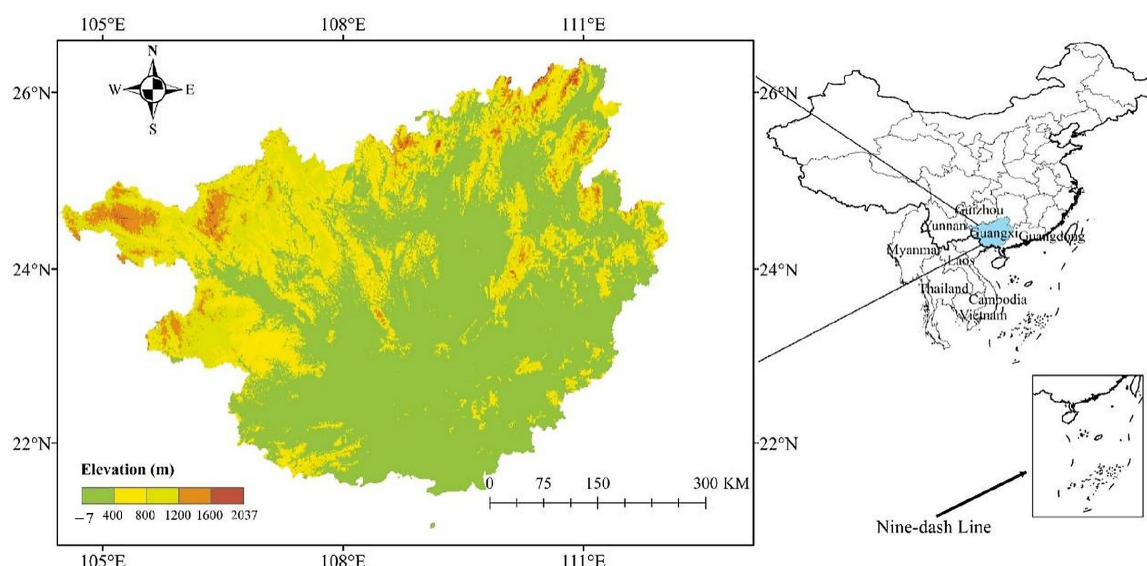
### 2.1. Study Area

Guangxi Zhuang Autonomous Region (20°54′–26°24′N, 104°26′–112°04′E) is located in southern China, bordering Vietnam to the southwest and the Gulf of Tonkin to the south (Figure 1). The region covers approximately 237,600 km<sup>2</sup> and features a complex topography dominated by mountains and hills in the northwest (elevations exceeding 2000 m in the Yuecheng and Dupang ranges) transitioning to alluvial plains and coastal lowlands in the southeast. Karst landscapes are widespread in the central and western portions.

Guangxi has a subtropical monsoon climate characterised by hot, humid summers and mild, relatively dry winters. Mean annual temperature ranges from 17 °C in the northwest highlands to 23 °C along the coast, and mean annual precipitation varies from approximately 1200 mm in sheltered valleys to over 2000 mm on windward slopes. The dry season extends from October to March, during which reduced precipitation, lower humidity, and accumulated litter create conditions favourable for wildfire. For seasonal analysis, each calendar year is divided into Period A (January–April, early dry-season period), Period B (May–September, rainy-season period), and Period C (October–December, late dry-season period). These periods contain four, five, and three months, respectively; fire counts are interpreted together with period length and valid observation coverage.

Forest cover in Guangxi exceeds 62%, comprising subtropical evergreen broadleaf forests in the south and east, mixed coniferous–broadleaf forests in the central hills, and Masson pine (*Pinus massoniana*) plantations throughout much of the region. The pine plantations are particularly fire-prone due to their resinous litter and open canopy struc-

ture. Agricultural burning, land clearing, and accidental ignition associated with rural activities are the primary fire causes. Previous work has assessed wildfire probability in Guangxi using multi-source environmental data [19], providing a foundation for the present decadal prediction framework.



**Figure 1.** Location of the study area.

## 2.2. Fire Occurrence Data

Several satellite-derived fire products are available for mapping burned areas, including the MODIS active fire product (MCD14ML), the MODIS burned area product (MCD64A1) [20], the VIIRS 375 m active fire product (VNP14IMG) [21], and the FireCCI51 product from the ESA Climate Change Initiative [5]. Active fire products detect thermal anomalies at the time of satellite overpass and are sensitive to the timing and duration of fire events; consequently, short-lived or low-intensity fires may be missed if they do not coincide with a satellite pass. The FireCCI51 product offers longer temporal coverage but has been shown to underestimate small fires in tropical and subtropical regions. We selected the MODIS Collection 6.1 MCD64A1 burned area product for three reasons: (1) it provides a consistent 500 m resolution record spanning the entire study period (2003–2018) with monthly compositing, which matches the temporal resolution of our prediction framework; (2) its burned area algorithm uses both active fire detections and surface reflectance change to identify burn scars, making it less dependent on overpass timing than active-fire-only products; (3) it has been extensively validated across diverse biomes globally [22]. MCD64A1 provides monthly burned-area observations at a nominal spatial resolution of 500 m. Its detection performance and observation availability can vary across environments and time periods, which must be considered when interpreting satellite-derived fire labels.

Monthly MCD64A1 rasters cover January 2003–December 2025. The intended 1 km aggregation rule assigns a positive label when at least one contributing valid 500 m pixel is burned. Valid unburned observations provide negative labels, whereas NoData represents an unknown observation and must not be treated as unburned. Historical labels are used for model development, and 2019–2025 labels support retrospective evaluation. Positive pixels whose centres lie inside the Guangxi boundary are exported for map overlays. These points represent burned-pixel centres, not independent ignition events. A month

with no exported point indicates no detected positive pixel among the available observations; it does not establish the absence of fire in unobserved areas.

### 2.3. Driving Factors

Thirteen environmental factors were used, retaining FMC and HPF and excluding TVDI from the fourteen-factor set used in the preceding study. Nine factors are treated as time-varying inputs: FMC, HPF, LAI, LST, NMDI, SIF, wind speed, rainfall, and humidity. Land cover, elevation, slope, and aspect form the four factors held fixed during future projection. This distinction concerns their treatment in the forecasting workflow: it does not imply that land cover is physically invariant over time. The fixed input order is FMC, HPF, LAI, Landcover, LST, NMDI, SIF, WindSpeed, Elevation, Rain, Slope, Aspect, and Humidity. All inputs are aligned to the common model grid before sequence construction.

The nine time-varying factors describe fuel and vegetation conditions (FMC, LAI, SIF, and NMDI), meteorological conditions (LST, wind speed, rainfall, and humidity), and anthropogenic pressure (HPF). Their sources follow the preceding published factor dataset summarised in Table 1. LAI is obtained from MOD15A2H; NMDI is derived from MOD09A1; SIF follows the study by Zhang et al. (2018) [23], and LST follows the study by Pan and Ding (2024) [24]. HPF denotes Human Footprint and HPF represents human pressure on the terrestrial environment. Annual HPF inputs are repeated across the twelve monthly time steps of the corresponding year. Wind speed, rainfall, and humidity are retained as separate meteorological predictors.

**Table 1.** Summary of the 13 driving factors.

| Category      | Factor    | Source/Reference         | Reported Scale | Model Role        |
|---------------|-----------|--------------------------|----------------|-------------------|
| Fuel          | FMC       | Quan et al. (2021) [25]  | 0.05°          | Dynamic           |
| Fuel          | LAI       | MOD15A2H                 | 500 m          | Dynamic           |
| Fuel          | SIF       | Zhang et al. (2018) [23] | 0.05°          | Dynamic           |
| Fuel          | Landcover | MCD12Q1                  | 500 m          | Projection-static |
| Fuel          | NMDI      | MOD09A1-derived          | 500 m          | Dynamic           |
| Climate       | WindSpeed | Jing et al. (2016) [26]  | 1000 m         | Dynamic           |
| Climate       | Rain      | Jing et al. (2016) [26]  | 1000 m         | Dynamic           |
| Climate       | Humidity  | Sun et al. (2014) [27]   | 1000 m         | Dynamic           |
| Climate       | LST       | Pan et al. (2025) [24]   | 1000 m         | Dynamic           |
| Topography    | Elevation | GMTED2010                | 90 m           | Projection-static |
| Topography    | Slope     | Derived from DEM         | 90 m           | Projection-static |
| Topography    | Aspect    | Derived from DEM         | 90 m           | Projection-static |
| Anthropogenic | HPF       | Mu et al. (2022) [28]    | 1000 m         | Dynamic           |

Elevation, slope, aspect, and land cover describe the terrain and vegetation context of each grid cell and are held fixed during projection. Elevation was obtained from GMTED2010, and slope and aspect were derived from the same elevation dataset [29]. Elevation and slope influence local microclimate, fuel moisture dynamics, and fire accessibility, while aspect determines solar exposure and consequently affects fuel drying rates [30] and influences local microclimate, fuel moisture dynamics, and fire accessibility. Land cover type, derived from the MODIS MCD12Q1 product [31], differentiates forest, shrubland, cropland, and urban areas with fundamentally different fuel loads and ignition patterns.

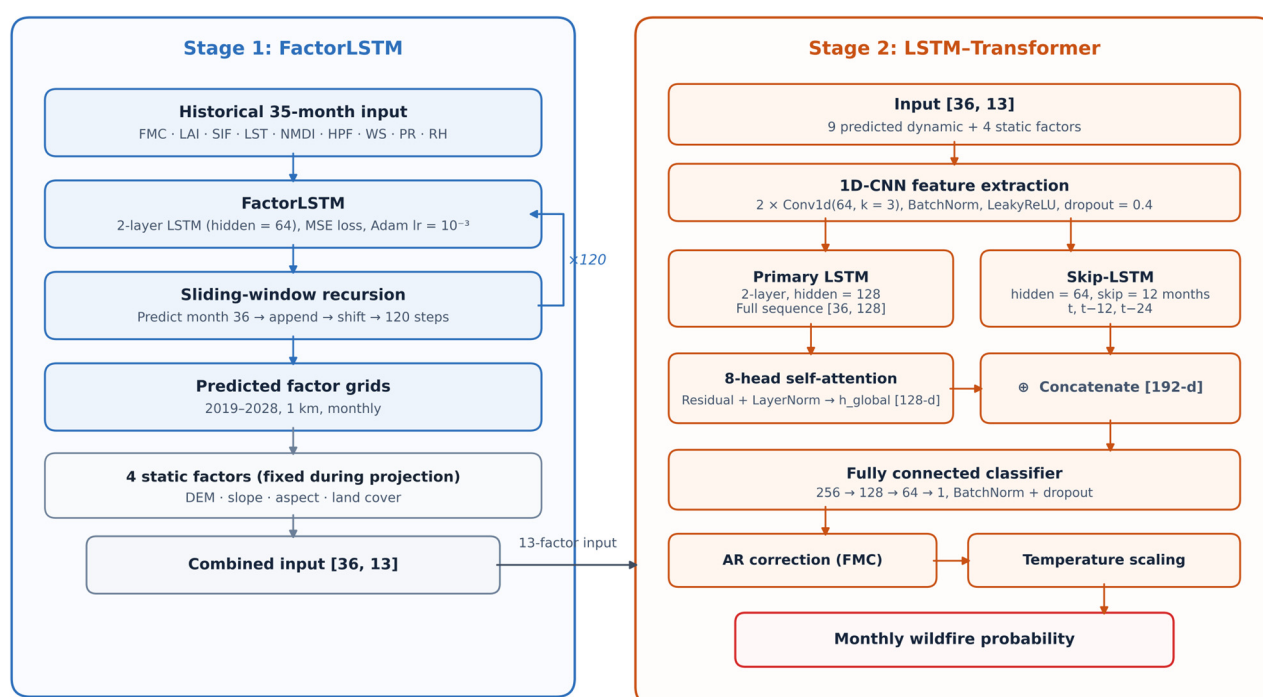
### 2.4. Data Partitioning

The dataset spans 192 months (January 2003–December 2018) and was split temporally into three non-overlapping subsets: training (2003–2016), validation (2017), and testing (2018). Temporal splitting was chosen over random splitting to prevent information leakage: because environmental conditions are strongly autocorrelated across months, a random split would allow temporally adjacent months to appear in both training and test sets, inflating performance estimates. The validation set was used for early stopping and hyperparameter selection; the test set was reserved for final evaluation and is reported throughout the Results section.

### 3. Methodology

#### 3.1. Framework Overview

Given the preceding 36-month sequence of 13 factors at each 1 km cell, the model estimates wildfire probability for the target month. The operational target is a positive satellite-derived burned-area label at the pixel-month scale. This probability does not refer to the precise ignition location or the number of independent fire events. The workflow proceeds in two stages (Figure 2).



**Figure 2.** Architecture of the two-stage wildfire prediction framework.

In Stage 1, a lightweight FactorLSTM module is trained to predict the next month's values of nine dynamic driving factors (FMC, HPF, LAI, LST, NMDI, SIF, WindSpeed, Rain, and Humidity) from the preceding 35-month sequence. Once trained, this module is applied recursively: each one-step-ahead prediction is appended to the input window as if it were an observation, and the window advances by one month. Repeating this process up to 120 times extends the dynamic factor series from 2018 to 2028 without requiring any future observational data. The four static factors (elevation, slope, aspect, and land cover type) are held constant throughout the projection period.

In Stage 2, the LSTM–Transformer processes 13-channel sequences and outputs monthly wildfire probabilities. Observed factors are used for training in 2003–2016, validation in 2017, and testing in 2018. For the 2019–2028 projection, dynamic-factor values after the December 2018 cutoff are provided recursively by Stage 1. Early projection windows can still contain pre-cutoff observations; the post-cutoff portions use predictions. Land cover, elevation, slope, and aspect are held fixed during projection.

This two-stage design decouples factor extrapolation from probability inference, so that each module can be trained, diagnosed, and improved independently.

### 3.2. Model Architecture

The FactorLSTM in Stage 1 is a two-layer LSTM network (hidden size = 64, dropout = 0.2) followed by a fully connected output layer. It takes an input sequence from the studies in [9,32] and predicts the nine factor values at month 36, trained with MSE loss using Adam ( $\text{lr} = 1 \times 10^{-3}$ , batch = 512). At inference time, the recursive loop starts from the last 35 observed months and generates 120 monthly steps for the 2019–2028 projection, using the historical scaling transformation consistently. Normalisation controls numerical scale but does not itself demonstrate control of recursive prediction error.

The Stage 2 classifier takes a 36-month sequence of 13 factors at each grid cell [14]. Two one-dimensional convolutional layers, each with 64 filters and a kernel size of 3, extract temporal features [15]. Batch normalisation and Leaky ReLU follow each convolution, and dropout is applied after the convolutional block. The resulting sequence is processed by a two-layer LSTM with 128 hidden units.

An eight-head self-attention layer operates on the LSTM outputs [16], using the same hidden-state sequence as queries, keys, and values. The attention output at the final sequence position provides a 128-dimensional representation. A parallel single-layer Skip-LSTM with 64 hidden units processes convolutional features at positions 1, 13, and 25 of the 36-month input window, corresponding to a 12-month sampling interval [33].

The attention and Skip-LSTM representations are concatenated into a 192-dimensional vector and passed through fully connected layers with 256, 128, 64, and 1 units [34]. The first three layers use batch normalisation, Leaky ReLU, and dropout rates of 0.5, 0.4, and 0.3, respectively [35]. An additional branch processes the six most recent FMC values through layers with 32 and 1 units. Its output, multiplied by 0.1, is added to the classifier logit [32].

Probability calibration is evaluated separately using fixed classifier outputs, as described in Section 3.4.

### 3.3. Training Configuration

The dataset was divided temporally: 2003–2016 for training (700,049 pixel–month samples), 2017 for validation (149,526 samples), and 2018 for testing (149,925 samples). Temporal splitting avoids information leakage between adjacent months that would arise from random splitting.

The model was trained by minimising the weighted binary cross-entropy loss:

$$L = -[w \cdot y \cdot \log(\sigma(z)) + (1 - y) \cdot \log(1 - \sigma(z))]$$

Here  $z$  is the model logit,  $y$  is the binary label,  $\sigma$  is the sigmoid function, and  $w$  is the positive-class weight. The original experiment used  $w = 50$ . Weighting increases the contribution of positive samples to the training objective, but does not guarantee optimal dis-

crimination or calibrated probability estimates. The additional sensitivity analysis compares  $w = 10, 30, 50, 70$ , and  $100$ ; the selection criterion is defined using validation performance [36].

The hyperparameter configuration of the LSTM-Transformer model is summarized in Table 2. All models were implemented in PyTorch 2.0 and trained on a single NVIDIA RTX 4060 GPU (NVIDIA Corporation, Santa Clara, CA, USA). Training each model variant took approximately 40–60 min. We used the Adam optimiser with an initial learning rate of  $1 \times 10^{-3}$  [37] and a cosine annealing schedule. Training was terminated early if the validation AUC did not improve for 10 consecutive epochs, and the weights corresponding to the best validation AUC were restored for testing [38].

**Table 2.** Hyperparameter configuration of the LSTM-Transformer.

| Hyperparameter               | Value                                        |
|------------------------------|----------------------------------------------|
| Sequence length              | 36 months                                    |
| Input features               | 13 (9 dynamic + 4 static)                    |
| CNN filters                  | 64, kernel size = 3                          |
| Primary LSTM hidden/layers   | 128/2                                        |
| Skip-LSTM hidden/skip period | 64/12 months                                 |
| Attention heads              | 8                                            |
| AR window                    | 6 months (FMC)                               |
| Optimiser                    | Adam ( $\beta_1 = 0.9$ , $\beta_2 = 0.999$ ) |
| Learning rate/scheduler      | $1 \times 10^{-3}$ /cosine annealing         |
| Batch size                   | 512                                          |
| Positive-class weight        | 50                                           |
| Early stopping patience      | 10 epochs                                    |

### 3.4. Evaluation Metrics

Performance is evaluated through ROC-AUC, average precision (AP), precision, recall, F1, accuracy, and confusion-matrix counts. AP complements ROC-AUC under severe class imbalance, while accuracy is interpreted relative to the positive fraction and an all-negative reference. Threshold-dependent metrics are reported with the threshold and its selection procedure. For the additional weight-10 experiment with Platt calibration, the stored validation-selected threshold is 0.0094973285 and remains fixed during retrospective evaluation. Only valid paired labels and probabilities are evaluated; valid zero labels are retained as negatives.

For the recovered-checkpoint calibration analysis, Platt scaling was fitted to validation-set logits and labels and then applied to the test set without updating the classifier weights. Brier score and ECE were calculated using all evaluated samples. ECE used 15 equal-width probability bins. Reliability curves used 10 equal-width bins; bins containing fewer than 30 samples were omitted from the displayed curves but retained in the overall metric calculations.

Probability calibration is assessed using the Brier score, expected calibration error (ECE), and reliability diagrams with bin counts. The Brier score measures mean squared probability error and is not a probability cutoff for defining high-risk areas. ECE depends on the binning scheme and event prevalence. Small numerical scores alone do not establish useful calibration for rare events; we also compare mean predicted probability with observed prevalence and use reference predictions. An all-zero forecast has a Brier score equal to the observed positive fraction.

### 3.5. Ablation and Baseline Design

Post hoc calibration is distinguished from architectural ablation. For a fixed set of logits, scaling by a positive temperature preserves ranking and therefore preserves ROC-AUC in exact arithmetic. Differences between separately retrained models cannot be attributed solely to post hoc temperature scaling. Architectural comparisons require matched data, evaluation samples, and model-selection procedures; uncertainty in differences should account for temporal and spatial dependence.

Architectural ablation considers four variants of the full model: replacing the 1D-CNN with a linear projection, removing the Skip-LSTM branch, removing self-attention, and removing the autoregressive branch. These comparisons examine the effects of temporal feature extraction, periodic sampling, attention, and autoregressive correction.

Temperature scaling is assessed separately as a post hoc calibration procedure. For fixed classifier logits, a positive temperature changes probability values without changing their ranking. Its effect is therefore evaluated using calibration metrics rather than changes in ROC-AUC.

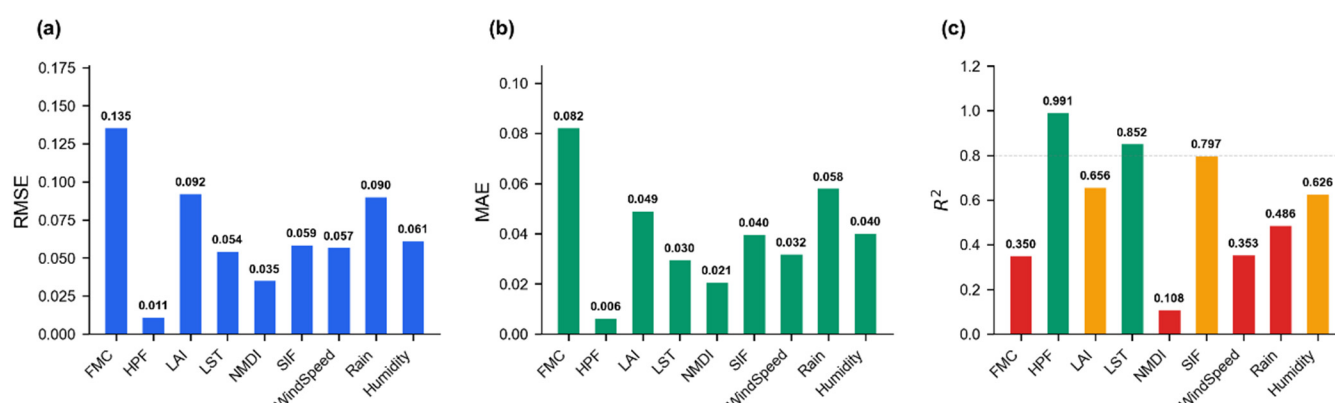
In addition, two standalone baselines were trained: an LSTM-only model (two-layer LSTM with hidden size of 128, no CNN, no attention, no Skip-LSTM, no AR) and a Transformer-only model (three-layer Transformer encoder with sinusoidal positional encoding,  $d_{\text{model}} = 128$ , 8 heads). The LSTM-only baseline uses the same four-layer classifier as the main model; the Transformer-only baseline uses a lighter two-layer classifier to avoid overfitting, since it lacks the regularisation provided by the CNN and Skip-LSTM pathways.

## 4. Results

### 4.1. Factor Prediction Quality

Before evaluating the wildfire classifier, we assessed one-step prediction of the nine dynamic factors. One-step accuracy characterises historical next-month prediction and does not by itself quantify errors accumulated over 120 recursive steps. The FactorLSTM module was trained on 2003–2017 monthly sequences (35-step input, one-step output) and evaluated on the 2018 hold-out set for all nine dynamic factors [39].

As shown in Figure 3, the model reproduced HPF almost perfectly ( $R^2 = 0.991$ , RMSE = 0.011) and also achieved high fidelity for LAI ( $R^2 = 0.852$ , RMSE = 0.092) and NMDI ( $R^2 = 0.852$ , RMSE = 0.035). LST ( $R^2 = 0.656$ ) and SIF ( $R^2 = 0.797$ ) fell in the moderate range. The weakest predictions were obtained for FMC ( $R^2 = 0.350$ ), WindSpeed ( $R^2 = 0.353$ ), and Rain ( $R^2 = 0.486$ ), all of which exhibit strong stochastic variability at the monthly scale. These three factors are driven by synoptic-scale weather events whose timing and magnitude are inherently difficult to predict beyond a few weeks. Errors on normalised scales must be interpreted alongside  $R^2$  and simple reference forecasts. Reproducing typical seasonal variation does not establish skill in individual monthly anomalies or guarantee reliable multi-year recursion. The consequences for wildfire probability must be assessed by feeding predicted drivers into the classifier and comparing its outputs with valid observed labels.



**Figure 3.** Factor prediction performance of the FactorLSTM on the 2018 test set: (a) RMSE, (b) MAE, and (c) R<sup>2</sup>.

#### 4.2. Model Performance and Component Analysis

Table 3 reports the 2018 observed-driver comparison. The reported AUC difference between the hybrid and Transformer-only models is 0.004, which requires uncertainty assessment before being interpreted as a consistent advantage. Accuracy alone provides limited information under the low positive rate; precision–recall performance and the number of detected and missed positives are also required. These observed-driver results do not represent the accuracy of the entire two-stage system over the 2019–2028 projection.

**Table 3.** Test-set performance of three model architectures. Best values in bold.

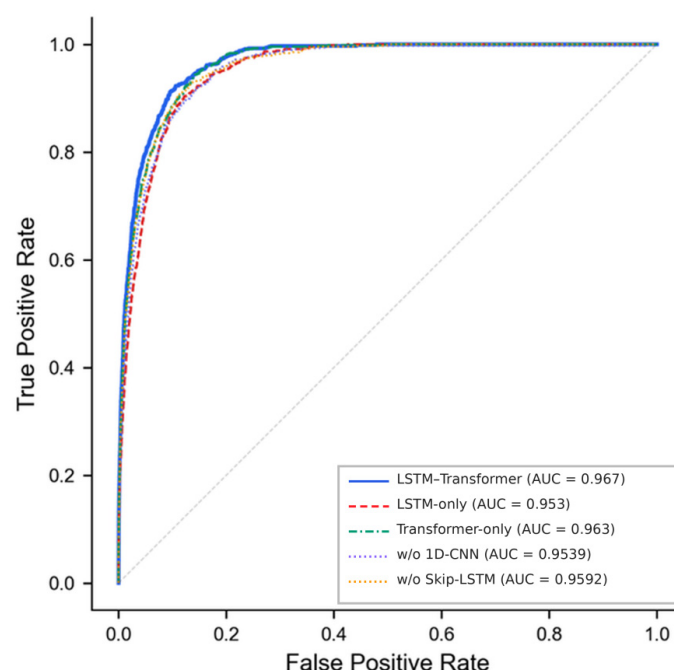
| Model            | AUC   | Precision | Recall | F1    | ACC   |
|------------------|-------|-----------|--------|-------|-------|
| LSTM–Transformer | 0.967 | 0.093     | 0.649  | 0.163 | 0.970 |
| Transformer-only | 0.963 | 0.093     | 0.649  | 0.163 | 0.970 |
| LSTM-only        | 0.953 | 0.098     | 0.588  | 0.152 | 0.977 |

The ROC curves (Figure 4) confirm this ranking. The LSTM–Transformer and Transformer-only curves nearly overlap across most of the operating ranges, both dominating the LSTM-only curve particularly in the low false-positive-rate region (FPR < 0.15), which is the operationally relevant regime for a rare-event prediction task [40]. The 0.014 AUC gap between the LSTM–Transformer and LSTM-only models is modest in absolute terms but is observed across the evaluated thresholds. This result indicates that the complete hybrid architecture achieved better discrimination than the LSTM-only baseline on the 2018 test set. However, the baseline comparison does not isolate the contribution of the attention module, and the subsequent ablation experiment does not demonstrate an independent performance gain from attention.

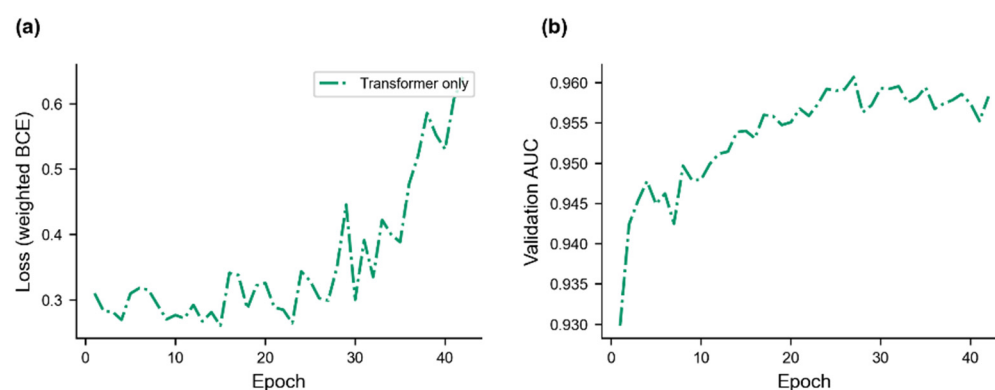
Figure 5 shows the training loss and validation AUC of the Transformer-only baseline. Validation AUC increased during the early epochs and fluctuated thereafter.

To examine the contribution of individual model components, we removed one component at a time and retrained each variant under the same experimental settings. As shown in Table 4 and Figure 6, removing the 1D-CNN reduced ROC-AUC from 0.9668 to 0.9539 ( $\Delta\text{AUC} = -0.0129$ ), representing the largest performance decrease. Removing the Skip-LSTM reduced ROC-AUC to 0.9592 ( $\Delta\text{AUC} = -0.0076$ ), indicating that explicitly representing the annual cycle contributed to discrimination. Removing the autoregressive component produced only a small decrease to 0.9661 ( $\Delta\text{AUC} = -0.0007$ ). In contrast, the model without attention achieved an ROC-AUC of 0.9695, which was 0.0027 higher than that of the full model. Therefore, the current experiment does not provide evidence that

the attention module improved discriminative performance. Additional repeated experiments would be required to determine whether this small difference is systematic or caused by training variability. Temperature scaling is a post hoc calibration procedure rather than an architectural component. Its effect is therefore evaluated primarily through calibration metrics rather than ROC-AUC.



**Figure 4.** ROC curves for the LSTM-Transformer, two baseline models, and two ablated variants.



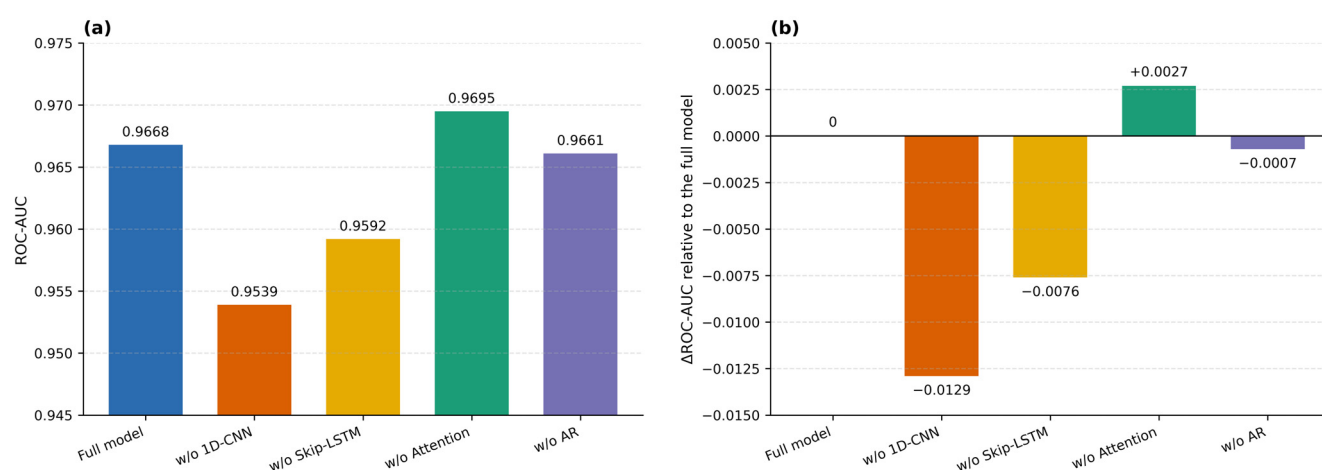
**Figure 5.** Training curves of the Transformer-only baseline: (a) weighted binary cross-entropy loss and (b) validation AUC.

**Table 4.** Ablation study results.  $\Delta$ AUC is measured relative to the full model.

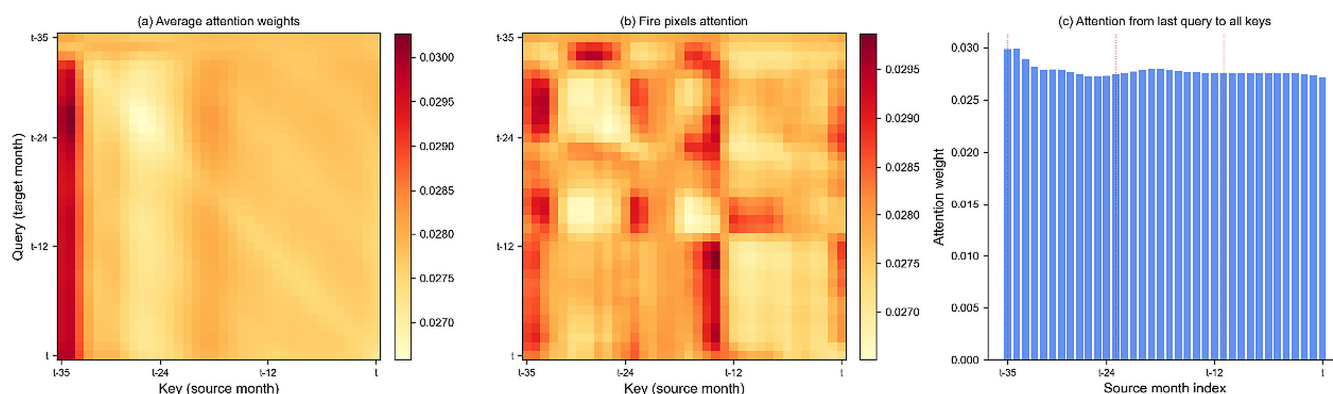
| Variant       | AUC    | $\Delta$ AUC | F1    | ACC   |
|---------------|--------|--------------|-------|-------|
| Full model    | 0.9668 | —            | 0.163 | 0.970 |
| w/o 1D-CNN    | 0.9539 | −0.0129      | 0.152 | 0.977 |
| w/o Skip-LSTM | 0.9592 | −0.0076      | 0.155 | 0.975 |
| w/o Attention | 0.9695 | +0.0027      | 0.161 | 0.971 |
| w/o AR        | 0.9661 | −0.0007      | 0.162 | 0.971 |

Note: Architectural ablation results: (a) ROC-AUC values reported in Table 4; (b) changes relative to the full model.

The self-attention weight analysis (Figure 7) reveals that the model learns distinct temporal patterns for fire-positive and fire-negative pixels. In the average-case heatmap (Figure 7a), each query position attends most strongly to the earliest months in the 36-month window (source months  $t - 35$  to  $t - 30$ ), producing a warm vertical band along the left edge. This recency-independent pattern suggests the model has learned to anchor its representations to a baseline seasonal state observed roughly three years prior. The fire-pixel heatmap (Figure 7b) shows a notably different structure. Attention weights are re-distributed into a block-diagonal pattern with elevated activity at approximately 12-month intervals, consistent with annual periodicity. Fire-positive pixels also exhibit stronger attention to recent months ( $t - 12$  to  $t$ ), particularly around the same calendar season as the target month. This indicates that, for samples that actually experienced fire, the model dynamically shifts its attention toward recent and seasonally aligned time steps—a pattern that aligns with the established influence of antecedent drought and vegetation drying on fire occurrence. The bar chart (Figure 7c) further decomposes the attention from the final query ( $t$ ) to all 36 source positions. The distribution is roughly uniform, with a slight elevation for the earliest months and a gentle decline toward the most recent steps. The near-uniform profile reflects the multi-head averaging effect across eight attention heads, each of which may attend to different temporal windows. Attention weights describe internal model behaviour and do not establish causal effects of any particular months or independently validate a physical wildfire mechanism.

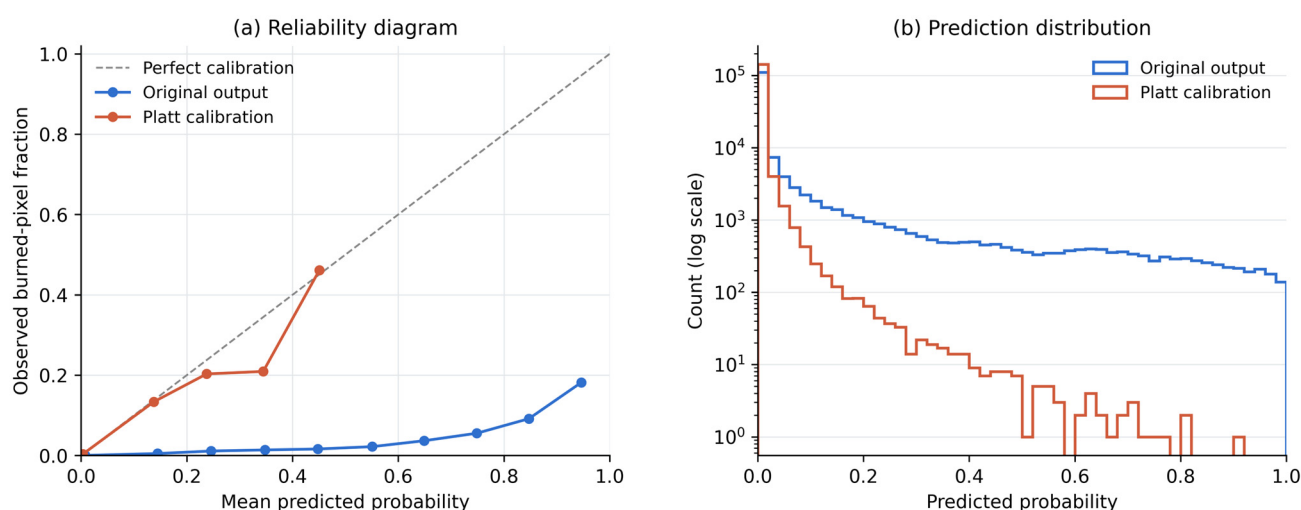


**Figure 6.** Architectural ablation results: (a) ROC-AUC values reported in Table 4; (b) changes in ROC-AUC relative to the full model.



**Figure 7.** Self-attention weight patterns: (a) averaged over all test-set samples, (b) averaged over fire-positive pixels only, and (c) attention distribution from the final query position to all 36 source months. Dashed vertical lines in (c) mark 12-month intervals.

An additional calibration analysis was conducted using the recovered LSTM–Transformer checkpoint and its associated validation and test samples. The test sample contained 149,925 observations, including 675 positives. Platt calibration was fitted on the validation set and applied to the fixed test logits. The Brier score decreased from 0.03313 to 0.00411, and ECE decreased from 0.06383 to 0.00027, while ROC-AUC remained approximately 0.9571. Figure 8a compares the reliability curves, and Figure 8b displays the corresponding probability distributions using a logarithmic count axis. These results demonstrate improved aggregate calibration in this evaluation sample, but do not establish reliable estimates in sparsely populated high-probability bins or under recursively predicted environmental inputs. This recovered-checkpoint analysis is distinct from the originally reported comparison in Table 3, whose reported hybrid AUC was 0.967.



**Figure 8.** Calibration assessment of the LSTM–Transformer model on the test set: (a) reliability curves of the original and Platt-calibrated outputs; and (b) corresponding prediction probability distributions. Platt parameters were fitted using the validation set only. Probability bins containing fewer than 30 samples are omitted from the reliability curves, while all samples are retained in the reported calibration metrics.

#### 4.3. End-to-End Retrospective Evaluation for 2019–2025

An additional end-to-end retrospective experiment used December 2018 as the environmental-input cutoff and recursively generated dynamic inputs for January 2019–December 2025. These inputs were supplied to a separately trained wildfire classifier with a positive-class weight of 10 and Platt calibration. This experiment is distinct from the original observed-driver model comparison and the recovered-checkpoint calibration analysis. Evaluation used available valid label–prediction pairs and retained valid unburned cells as negatives.

Across the 84-month interval, pooled ROC-AUC was 0.783492 and AP was 0.001475. Precision, recall, and F1 were 0.002138, 0.205154, and 0.004233, respectively. Mean predicted probability was 0.00198021, compared with an observed positive fraction of 0.00044042. The Brier score was 0.00046219, exceeding the all-zero reference score of

0.00044042. The end-to-end retrospective performance across cumulative forecast windows from January 2019 is summarized in Table 5. Consequently, the experiment identifies ranking information but also substantial limitations in rare-event detection and probability calibration.

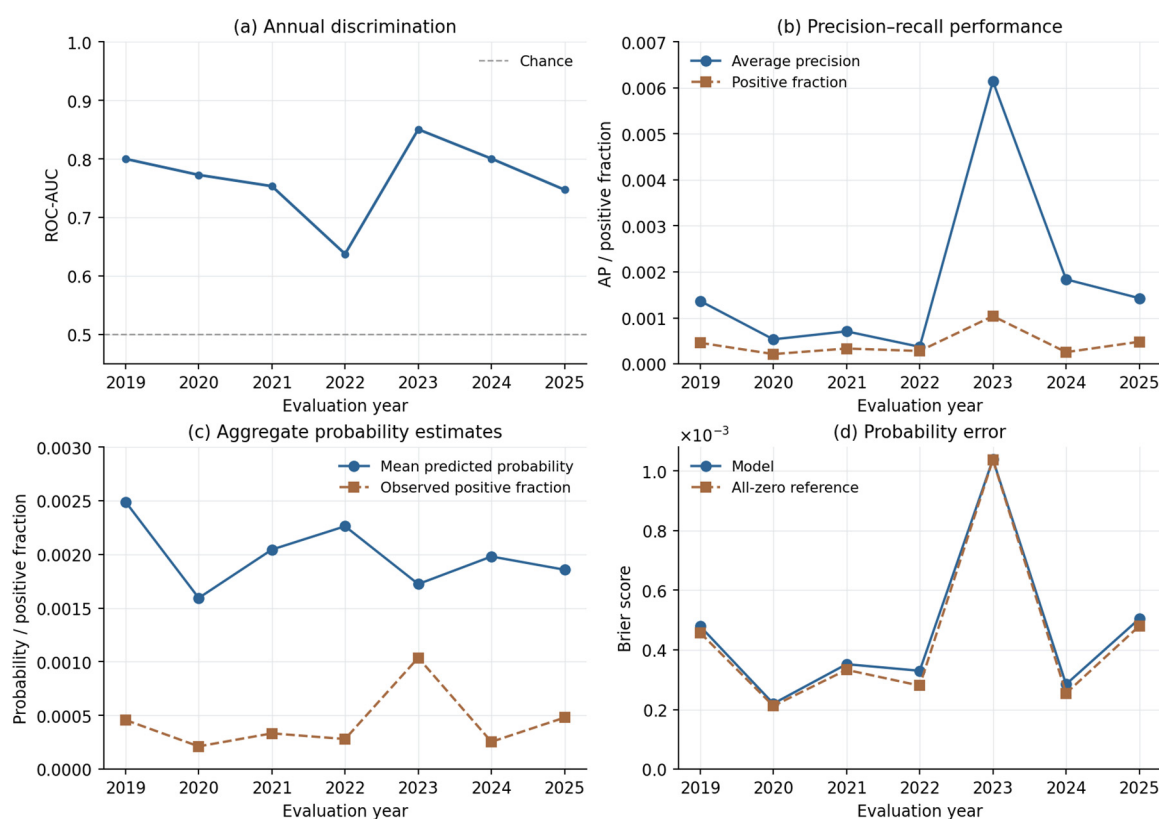
**Table 5.** End-to-end retrospective performance over cumulative forecast windows from January 2019.

| Cumulative Forecast Months | ROC-AUC  | AP       | F1       | Brier Score | ECE        |
|----------------------------|----------|----------|----------|-------------|------------|
| 1–12                       | 0.800174 | 0.001367 | 0.003462 | 0.00047915  | 0.00203333 |
| 1–24                       | 0.799576 | 0.001025 | 0.002844 | 0.00033798  | 0.00167923 |
| 1–36                       | 0.782655 | 0.000885 | 0.002451 | 0.00034284  | 0.00169072 |
| 1–60                       | 0.789608 | 0.001526 | 0.004441 | 0.00048789  | 0.00153782 |
| 1–84                       | 0.783492 | 0.001475 | 0.004233 | 0.00046219  | 0.00153979 |

AP denotes average precision. ECE uses 10 equal-width bins. Metrics are calculated from pooled valid pixel-month pairs within each cumulative window. Results refer to the weight-10 classifier with Platt calibration and a validation-selected decision threshold of 0.0094973285.

Cumulative-window metrics combine forecast lead time with calendar-year variation and observation coverage. A single forecast origin does not constitute rolling-origin validation, and these results do not establish the accuracy of the separate original-model maps or validate the complete 120-month projection horizon.

Annual performance varied substantially (Figure 9). ROC-AUC ranged from 0.638 in 2022 to 0.851 in 2023. Mean predicted probability exceeded the observed positive fraction in every year. These results show that discrimination and probability calibration must be assessed separately.



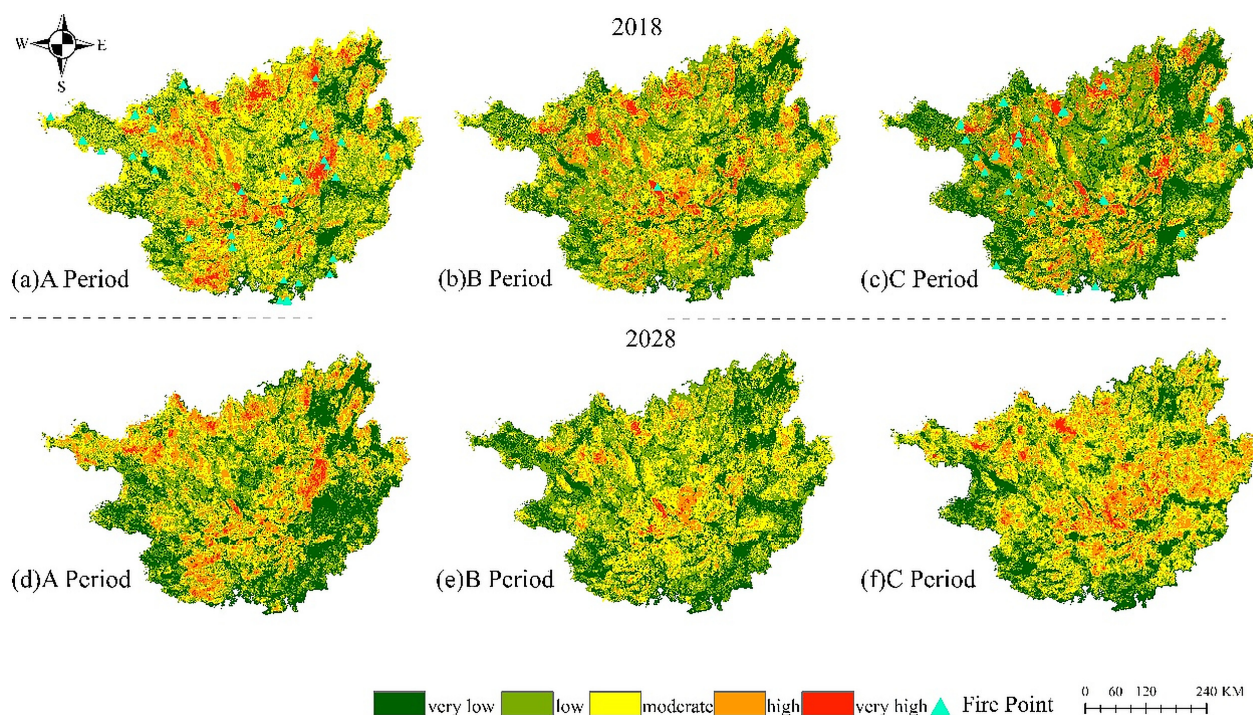
**Figure 9.** Annual results of the 2019–2025 retrospective evaluation: (a) ROC-AUC; (b) average precision and observed positive fraction; (c) mean predicted probability and observed positive fraction; and (d) Brier score compared with an all-zero reference. Results are based on available valid pixel-month pairs for the weight-10 classifier with Platt calibration, using December 2018 as the forecast origin.

#### 4.4. Projected Wildfire Risk Maps

Figure 10 presents the 2018 test-year maps and the 2028 maps at the end of the ten-year projection. Each calendar year is divided into Period A (January–April), Period B (May–September), and Period C (October–December). The model prediction unit remains monthly; the three periods provide seasonal summaries for presentation. The aggregation rule and probability-class boundaries are stated with the maps. Intermediate projections for 2019–2027 are presented in Appendix A.

Observed burned-pixel centres are overlaid on the maps to illustrate spatial correspondence. The high and very-high classes identify relatively elevated model estimates within the specified classification scheme; membership in these classes does not mean that a fire is assumed to have occurred. Quantitative validation must also include valid unburned cells to measure false positives. Seasonal comparisons require explicit consideration of the four-, five-, and three-month period lengths, as well as differences in valid observation coverage.

The 2028 maps retain the ten-year projection horizon from the December 2018 origin. Spatial similarity to earlier maps may reflect fixed terrain and land cover, persistent seasonal structure, or convergence of extrapolated drivers toward typical values. Such similarity cannot establish that recursive error accumulation is negligible. The 2019–2025 retrospective assessment provides empirical evidence over the observed interval, while 2026–2028 remains outside that validation period.



**Figure 10.** Seasonal summaries of monthly wildfire probability for 2018 and 2028: (a) 2018 Period A, (b) 2018 Period B, (c) 2018 Period C, (d) 2028 Period A, (e) 2028 Period B, and (f) 2028 Period C. Corresponding maps for 2019–2027 are provided in Appendix A.

## 5. Discussion

### 5.1. Architecture Design and Component Contributions

The central objective of this study is not only to evaluate a novel hybrid architecture, but to use its structural properties and output risk maps to improve our understanding of wildfire distribution in subtropical Guangxi. The spatial and temporal patterns revealed by the model, including the pronounced concentration of high-risk zones in the western and northern highlands during Period A and the near-complete suppression of risk during Period C, are consistent with field-based observations and reinforce the dominant role of the monsoon cycle and terrain in shaping the regional fire regime. The architectural analysis below both validates these patterns and explains why specific model components are necessary to capture them.

The LSTM–Transformer outperformed both standalone baselines, but the margin over the Transformer-only model was narrow (AUC of 0.967 vs. 0.963). The advantage is structural rather than purely numerical. The training curves (Figure 5) show that the Transformer-only model exhibited substantially more oscillation in validation loss and AUC, while the LSTM–Transformer maintained a smooth trajectory. This stability matters in practice: environmental modelling requires reproducible results across retraining runs [41]. The LSTM layers provide an ordered, locally smooth feature sequence to the attention mechanism, rather than forcing the Transformer to learn positional relationships from sinusoidal encodings alone. The LSTM-only model, while stable, plateaued at a lower AUC (0.953), as its 36-step sequence approaches the limit over which standard gates can maintain informative gradients [42].

The ablation study revealed a clear hierarchy: the 1D-CNN contributed most ( $\Delta\text{AUC} = -0.013$ ), followed by the Skip-LSTM ( $\Delta\text{AUC} = -0.008$ ). The CNN result is consistent with the autocorrelated structure of the input data [43]; adjacent months are not independent, and the convolution acts as temporal smoothing that would otherwise require hand-crafted lag features. The Skip-LSTM's 12-month interval directly encodes the annual fire-season cycle [44,45], which is the dominant feature of wildfire dynamics in subtropical Guangxi. From a wildfire science perspective, the LSTM and the periodic skip pathway provide plausible representations of gradual and seasonal dependencies. Although the attention weights describe how the model distributes information across the input sequence, the current ablation result does not establish an independent improvement in discriminative performance from the attention module. The attention patterns are therefore interpreted as descriptions of internal model behaviour rather than evidence of a causal or performance-enhancing contribution.

### 5.2. Factor Prediction and Error Propagation

Errors in recursively predicted environmental drivers can propagate into wildfire estimates and alter their distribution relative to the observed-driver training data. One-step performance does not quantify this effect over a 120-month horizon. High one-step accuracy for an individual factor also does not guarantee stable multi-step predictions. Spatial persistence in the projected maps may arise from fixed terrain and land cover or convergence toward typical seasonal conditions, and therefore is not independent evidence of forecast accuracy. The additional 2019–2025 retrospective experiment provides a partial evaluation of a predicted-driver configuration, but its low AP and residual probability overestimation limit the conclusions. Earlier forecast origins, observed-driver comparisons, and persistence or seasonal-reference baselines are still needed to distinguish predictive skill from recurring spatial and seasonal structures [46,47].

### 5.3. Calibration, Limitations, and Future Directions

Improved calibration of a recovered classifier on historical test samples does not automatically transfer to spatially post-processed maps or recursively generated future inputs. We retain the term wildfire probability for the model-estimated probability of the defined monthly burned-area target. This terminology does not imply that the estimates are perfectly calibrated. The observed overestimation must be evaluated using reliability diagrams, Brier score, ECE, and reference forecasts. Risk-class maps summarise the output field, whereas calibrated probabilities are required when numerical event likelihoods enter a decision calculation. High and very-high classes are therefore not interpreted as confirmed fire occurrences. Operational use requires validation at the intended spatial and temporal scales and consideration of the costs of false alarms and missed events.

Several limitations affect the interpretation of the ten-year projections. MCD64A1 labels represent detected burned area and may omit fires or contain invalid observations [20]. NoData must remain distinct from valid unburned cells, and differences in historical and recent preprocessing can affect apparent event frequency. Land cover is fixed at its 2018 state during projection, although vegetation, urbanisation, and land management can change fuel and ignition conditions. Historical climate relationships may also change, especially during extreme dry or wet years. A single 2018 test year and a single December 2018 forecast origin do not establish robust performance across years and lead times. Earlier-cutoff and rolling-origin validation, matched baselines, and uncertainty estimates are needed. The current pixelwise model also omits explicit interactions and fire spread between neighbouring cells.

## 6. Conclusions

This study develops a two-stage framework for decadal wildfire risk projection in Guangxi, China. FactorLSTM recursively estimates nine dynamic factors, and an LSTM–Transformer combines these estimates with four projection-static factors to generate monthly wildfire probabilities from January 2019 to December 2028. Seasonal presentation uses January–April, May–September, and October–December [48]. The framework separates environmental extrapolation from wildfire probability estimation, making it possible to evaluate the complete predicted-driver workflow in addition to the 2018 observed-driver test [49]. The additional 2019–2025 retrospective experiment provides empirical assessment over part of the projection interval, but limitations in calibration, rare-event detection, and label coverage remain. Consistent superiority over architectural baselines and reliability over the full ten-year horizon require further temporal validation and uncertainty assessment [49]. The maps are therefore interpreted as conditional projections, whose numerical probabilities and management applications require evaluation beyond visual pattern persistence.

We developed a two-stage workflow that combines recursive environmental prediction with monthly wildfire probability estimation for Guangxi. On the 2018 observed-driver test set, the hybrid classifier achieved an AUC of 0.967, close to the Transformer-only baseline. Ablation experiments showed the largest performance reductions when temporal convolution and the Skip-LSTM branch were removed; attention did not improve AUC in the reported experiment.

The separate 2019–2025 predicted-driver evaluation achieved a pooled AUC of 0.7835, but low average precision and probability overestimation limited its practical usefulness. The 2019–2028 maps are therefore conditional projections rather than validated monthly forecasts for the full decade. Further work should evaluate the workflow from multiple

forecast origins, compare it with simple reference forecasts, and assess uncertainty under predicted environmental inputs.

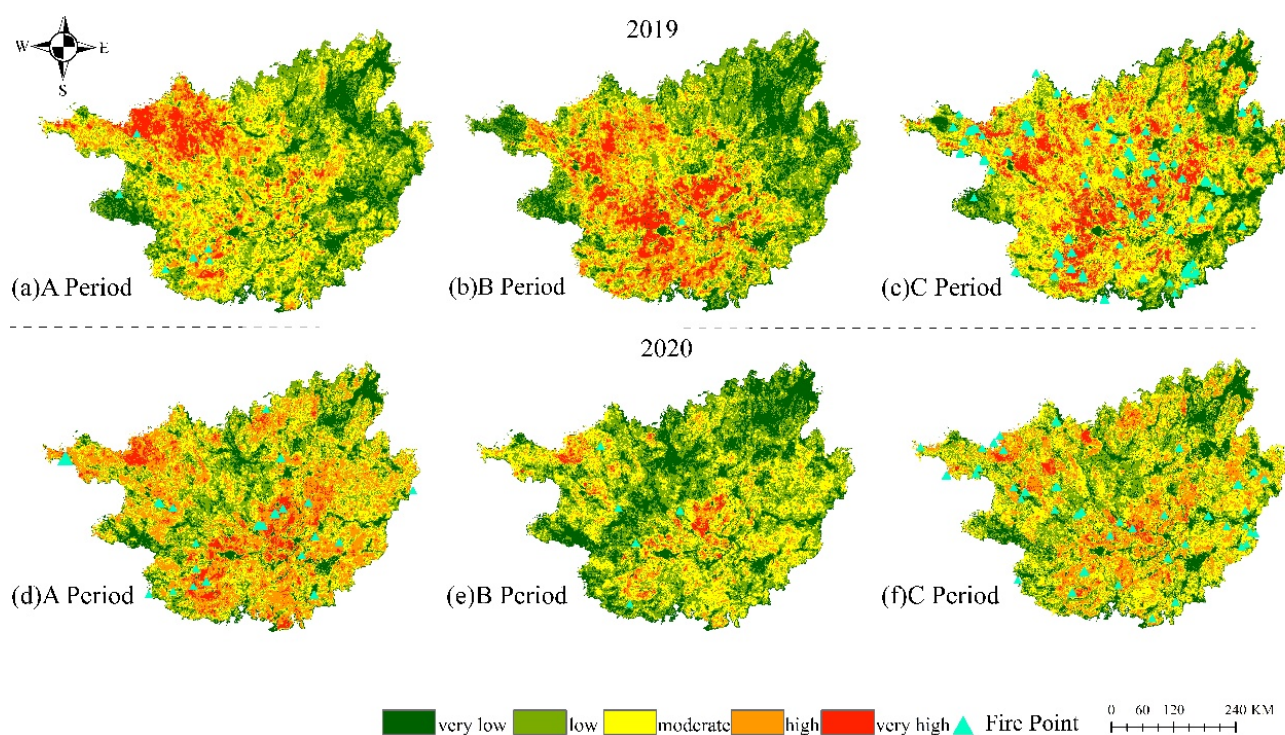
**Author Contributions:** Conceptualisation, Z.L., X.P. and K.T.; methodology, K.X. and K.T.; software, K.X. and A.A.; validation, K.X. and Z.Y.; formal analysis, K.X. and Z.Y.; investigation, R.G.; resources, X.P. and Y.Y.; data curation, K.X. and R.G.; writing—original draft preparation, K.X. and Z.L.; writing—review and editing, X.P., K.T. and A.A.; visualisation, K.X. and Y.Y.; supervision, Z.L.; project administration, Z.L.; funding acquisition, X.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Key Research and Development Programme of China [2025YFF1702301], Royal Society IEC\NSFC\223292—International Exchanges 2022 Cost Share (NSFC) grant of the UK, Special Funding Project for Independent Research of Jiangsu Hydraulic Research Institute: 2025z066 and the National Nature Science Foundation of China [42230112, 41701487].

**Data Availability Statement:** The data presented in this study are available on request from the corresponding authors. The MODIS burned area product (MCD64A1), MODIS land cover product (MCD12Q1), and other remote sensing datasets analysed in this study are publicly available from the respective data providers. The CLDAS data were obtained from the National Meteorological Information Centre of China, and the hydrometeorological observation and EC data were obtained from the National Tibetan Plateau Data Centre.

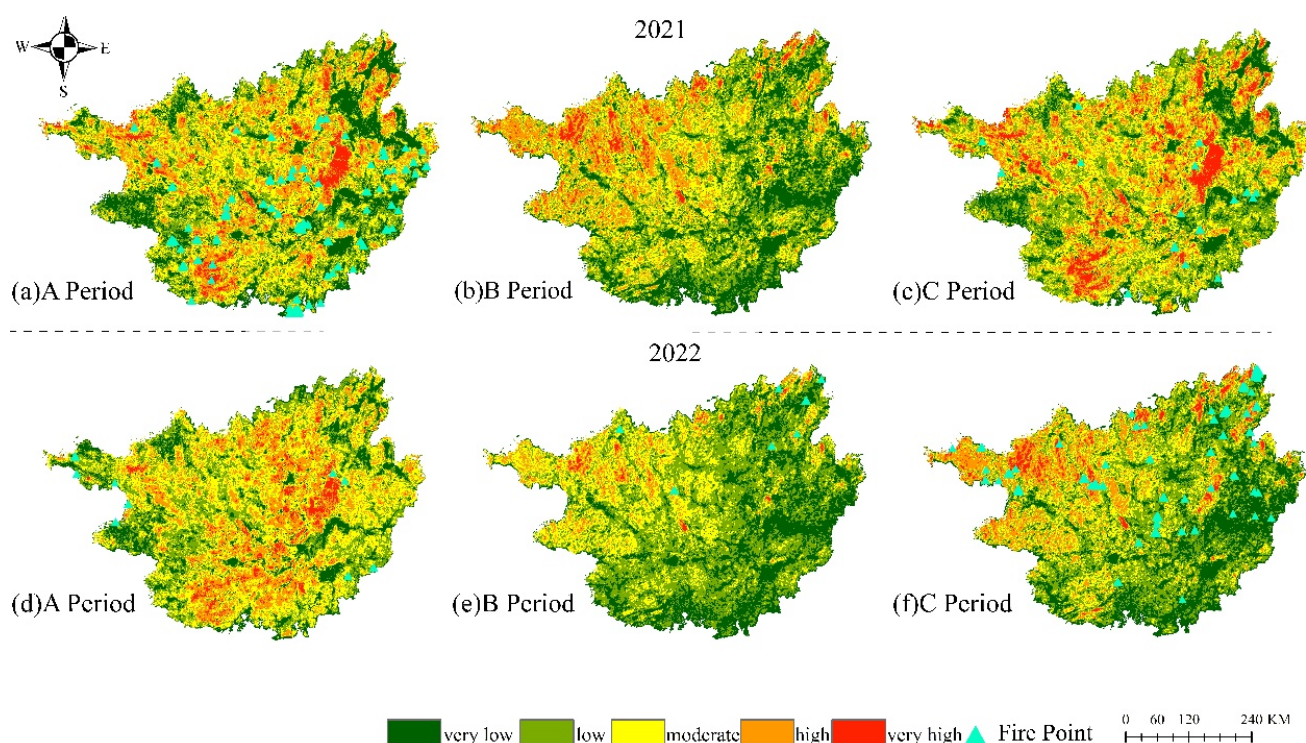
**Conflicts of Interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A

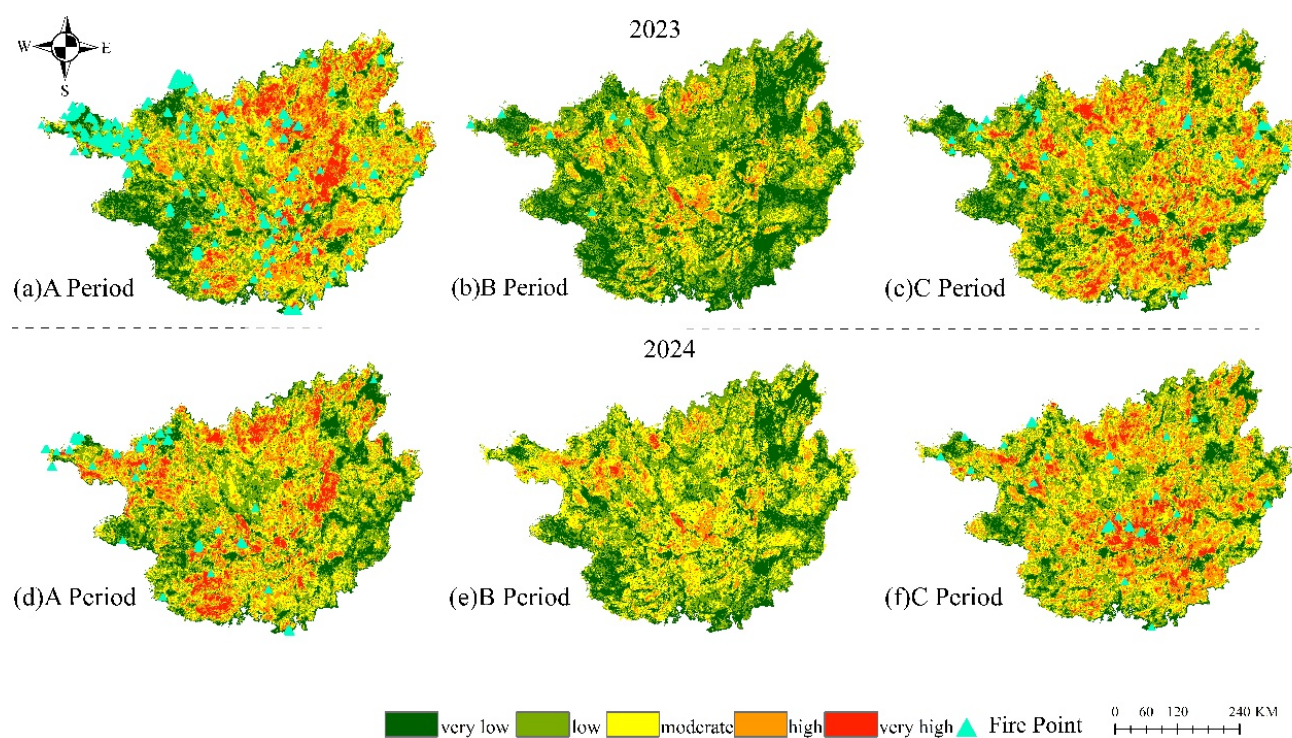


**Figure A1.** Seasonal summaries of monthly wildfire probability for 2019 and 2020: (a) 2019 Period A, (b) 2019 Period B, (c) 2019 Period C, (d) 2020 Period A, (e) 2020 Period B, and (f) 2020 Period C.

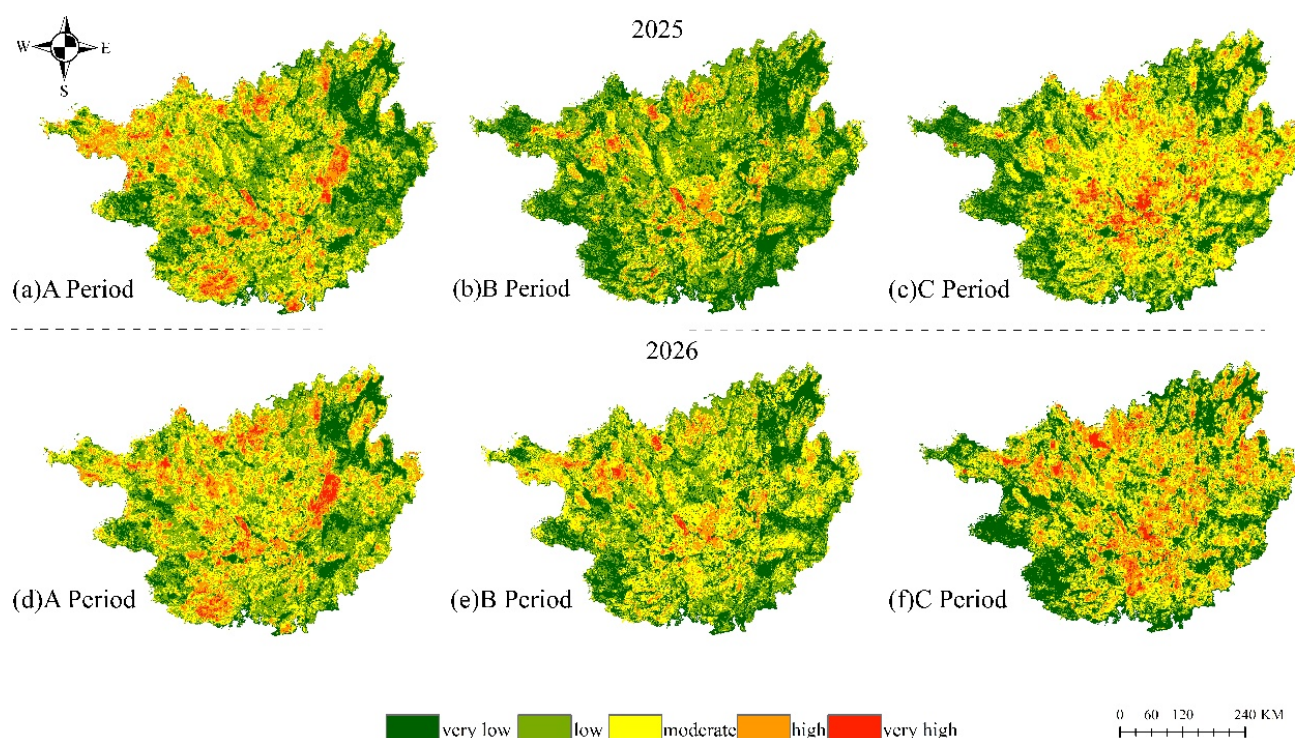
Period A denotes January–April (early dry season), Period B denotes May–September (rainy season), and Period C denotes October–December (late dry season). Cyan triangles indicate observed burned-pixel centres.



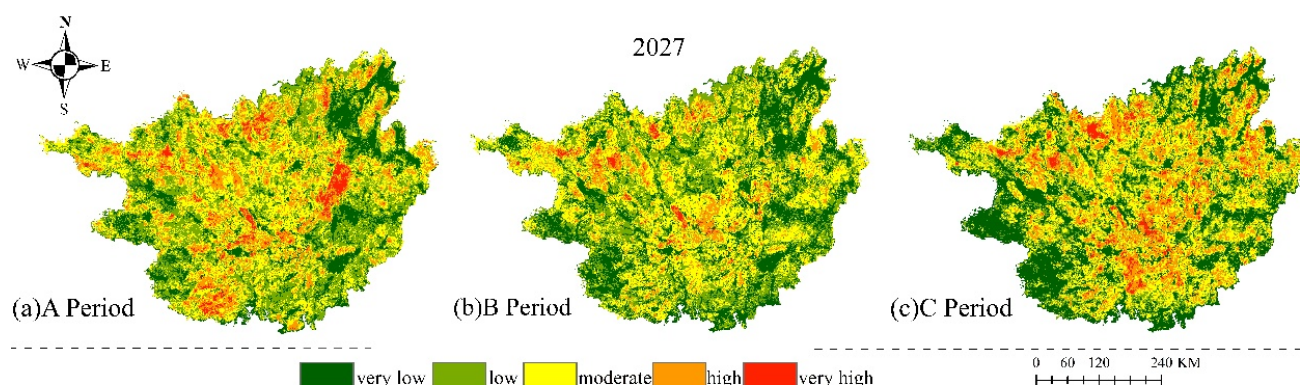
**Figure A2.** Seasonal summaries of monthly wildfire probability for 2021 and 2022: (a) 2021 Period A, (b) 2021 Period B, (c) 2021 Period C, (d) 2022 Period A, (e) 2022 Period B, and (f) 2022 Period C. Period A denotes January–April (early dry season), Period B denotes May–September (rainy season), and Period C denotes October–December (late dry season). Cyan triangles indicate observed burned-pixel centres.



**Figure A3.** Seasonal summaries of monthly wildfire probability for 2023 and 2024: (a) 2023 Period A, (b) 2023 Period B, (c) 2023 Period C, (d) 2024 Period A, (e) 2024 Period B, and (f) 2024 Period C. Period A denotes January–April (early dry season), Period B denotes May–September (rainy season), and Period C denotes October–December (late dry season). Cyan triangles indicate observed burned-pixel centres.



**Figure A4.** Seasonal summaries of monthly wildfire probability for 2025 and 2026: (a) 2025 Period A, (b) 2025 Period B, (c) 2025 Period C, (d) 2026 Period A, (e) 2026 Period B, and (f) 2026 Period C. Period A denotes January–April (early dry season), Period B denotes May–September (rainy season), and Period C denotes October–December (late dry season). Cyan triangles indicate observed burned-pixel centres.



**Figure A5.** Seasonal summaries of monthly wildfire probability for 2027: (a) 2027 Period A, (b) 2027 Period B, and (c) 2027 Period C. Period A denotes January–April (early dry season), Period B denotes May–September (rainy season), and Period C denotes October–December (late dry season). Cyan triangles indicate observed burned-pixel centres.

## References

1. Oliveira, S.; Gonçalves, A.; Zêzere, J.L. Reassessing Wildfire Susceptibility and Hazard for Mainland Portugal. *Sci. Total Environ.* **2021**, *762*, 143121. <https://doi.org/10.1016/j.scitotenv.2020.143121>.
2. Pechony, O.; Shindell, D.T. Driving Forces of Global Wildfires over the Past Millennium and the Forthcoming Century. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 19167–19170. <https://doi.org/10.1073/pnas.1003669107>.
3. Artés, T.; Oom, D.; de Rigo, D.; Durrant, T.H.; Maianti, P.; Libertà, G.; San-Miguel-Ayanz, J. A Global Wildfire Dataset for the Analysis of Fire Regimes and Fire Behaviour. *Sci. Data* **2019**, *6*, 296. <https://doi.org/10.1038/s41597-019-0312-2>.
4. Mbow, H.-O.P.; Reisinger, A.; Canadell, J.; O'Brien, P. *Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems (SR2)*; IPCC: Geneva, Switzerland, 2017; Volume 650.
5. Lizundia-Loiola, J.; Otón, G.; Ramo, R.; Chuvieco, E. A Spatio-Temporal Active-Fire Clustering Approach for Global Burned Area Mapping at 250 m from MODIS Data. *Remote Sens. Environ.* **2020**, *236*, 111493. <https://doi.org/10.1016/j.rse.2019.111493>.
6. Van Wagner, C. *Development and Structure of the Canadian Forest Fire Weather Index System*; Canadian Forestry Service: Ottawa, ON, Canada, 1987.
7. Keetch, J.J.; Byram, G.M. *A Drought Index for Forest Fire Control*; U.S. Department of Agriculture, Forest Service, Southeastern Forest Experiment Station: Asheville, NC, USA, 1968.
8. Jain, P.; Coogan, S.C.P.; Subramanian, S.G.; Crowley, M.; Taylor, S.; Flannigan, M.D. A Review of Machine Learning Applications in Wildfire Science and Management. *Environ. Rev.* **2020**, *28*, 478–505. <https://doi.org/10.1139/er-2020-0019>.
9. Jaafari, A.; Mafi-Gholami, D.; Thai Pham, B.; Tien Bui, D. Wildfire Probability Mapping: Bivariate vs. Multivariate Statistics. *Remote Sens.* **2019**, *11*, 618. <https://doi.org/10.3390/rs11060618>.
10. Sun, Y.; Zhang, F.; Lin, H.; Xu, S. A Forest Fire Susceptibility Modeling Approach Based on Light Gradient Boosting Machine Algorithm. *Remote Sens.* **2022**, *14*, 4362. <https://doi.org/10.3390/rs14174362>.
11. Zhang, G.; Wang, M.; Liu, K. Forest Fire Susceptibility Modeling Using a Convolutional Neural Network for Yunnan Province of China. *Int. J. Disaster Risk Sci.* **2019**, *10*, 386–403. <https://doi.org/10.1007/s13753-019-00233-1>.
12. Mambile, C.; Kaijage, S.; Leo, J. Application of Deep Learning in Forest Fire Prediction: A Systematic Review. *IEEE Access* **2024**, *12*, 190554–190581.
13. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat, F. Deep Learning and Process Understanding for Data-Driven Earth System Science. *Nature* **2019**, *566*, 195–204.
14. Graves, A. Long Short-Term Memory. In *Supervised Sequence Labelling with Recurrent Neural Networks*; Studies in Computational Intelligence; Springer: Berlin/Heidelberg, Germany, 2012; Volume 385, pp. 37–45, ISBN 978-3-642-24796-5.
15. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. In *Proceedings of the Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30.
16. Lai, G.; Chang, W.-C.; Yang, Y.; Liu, H. Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks. In *Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*; Association for Computing Machinery: New York, NY, USA, 27 June 2018; pp. 95–104.
17. Huot, F.; Hu, R.L.; Ihme, M.; Wang, Q.; Burge, J.; Lu, T.; Hickey, J.; Chen, Y.-F.; Anderson, J. Deep Learning Models for Predicting Wildfires from Historical Remote-Sensing Data. *arXiv* **2021**, arXiv:2010.07445.
18. Kondylatos, S.; Prapas, I.; Ronco, M.; Papoutsis, I.; Camps-Valls, G.; Piles, M.; Fernández-Torres, M.; Carvalhais, N. Wildfire Danger Prediction and Understanding with Deep Learning. *Geophys. Res. Lett.* **2022**, *49*, e2022GL099368. <https://doi.org/10.1029/2022GL099368>.
19. Xu, K.; Pan, X.; Tansey, K.; Abdulla, A.; Guluzade, R.; Yang, Z.; He, M.; Zhu, C.; Yang, S.; Yang, Y. Wildfire Probability Assessment and Analysis Based on Multi-Source Data in Guangxi Province, China. *Ecol. Indic.* **2025**, *179*, 114148. <https://doi.org/10.1016/j.ecolind.2025.114148>.
20. Giglio, L.; Boschetti, L.; Roy, D.P.; Humber, M.L.; Justice, C.O. The Collection 6 MODIS Burned Area Mapping Algorithm and Product. *Remote Sens. Environ.* **2018**, *217*, 72–85.
21. Schroeder, W.; Oliva, P.; Giglio, L.; Csiszar, I.A. The New VIIRS 375m Active Fire Detection Data Product: Algorithm Description and Initial Assessment. *Remote Sens. Environ.* **2014**, *143*, 85–96. <https://doi.org/10.1016/j.rse.2013.12.008>.

22. Chuvieco, E.; Mouillot, F.; van der Werf, G.R.; San Miguel, J.; Tanase, M.; Koutsias, N.; García, M.; Yebra, M.; Padilla, M.; Gitas, I.; et al. Historical Background and Current Developments for Mapping Burned Area from Satellite Earth Observation. *Remote Sens. Environ.* **2019**, *225*, 45–64. <https://doi.org/10.1016/j.rse.2019.02.013>.
23. Zhang, Y.; Joiner, J.; Alemohammad, S.H.; Zhou, S.; Gentine, P. A Global Spatially Contiguous Solar-Induced Fluorescence (CSIF) Dataset Using Neural Networks. *Biogeosciences* **2018**, *15*, 5779–5800.
24. Pan, X.; Ding, X.; Tansey, K.; Guluzade, R.; Hu, P.; Yang, Z.; Liu, S.; Yuan, J.; Wang, Z.; Lv, Z.; et al. Synthesis of a Global Daily Average Gapless Land Surface Temperature Dataset from 2003 to 2018. *Sci. Data* **2025**, *12*, 1177. <https://doi.org/10.1038/s41597-025-05498-4>.
25. Quan, X.; Yebra, M.; Riaño, D.; He, B.; Lai, G.; Liu, X. Global Fuel Moisture Content Mapping from MODIS. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *101*, 102354.
26. Jing, W.; Yang, Y.; Yue, X.; Zhao, X. A Spatial Downscaling Algorithm for Satellite-Based Precipitation over the Tibetan Plateau Based on NDVI, DEM, and Land Surface Temperature. *Remote Sens.* **2016**, *8*, 655.
27. Sun, W.; Shao, Q.; Liu, J.; Zhai, J. Assessing the Effects of Land Use and Topography on Soil Erosion on the Loess Plateau in China. *Catena* **2014**, *121*, 151–163.
28. Mu, H.; Li, X.; Wen, Y.; Huang, J.; Du, P.; Su, W.; Miao, S.; Geng, M. A Global Record of Annual Terrestrial Human Footprint Dataset from 2000 to 2018. *Sci. Data* **2022**, *9*, 176.
29. Athmania, D.; Achour, H. External Validation of the ASTER GDEM2, GMTED2010 and CGIAR-CSI-SRTM v4. 1 Free Access Digital Elevation Models (DEMs) in Tunisia and Algeria. *Remote Sens.* **2014**, *6*, 4600–4620.
30. Farr, T.G.; Rosen, P.A.; Caro, E.; Crippen, R.; Duren, R.; Hensley, S.; Kobrick, M.; Paller, M.; Rodriguez, E.; Roth, L.; et al. The Shuttle Radar Topography Mission. *Rev. Geophys.* **2007**, *45*, 2005RG000183. <https://doi.org/10.1029/2005RG000183>.
31. Friedl, M.A.; Sulla-Menashe, D.; Tan, B.; Schneider, A.; Ramankutty, N.; Sibley, A.; Huang, X. MODIS Collection 5 Global Land Cover: Algorithm Refinements and Characterization of New Datasets. *Remote Sens. Environ.* **2010**, *114*, 168–182. <https://doi.org/10.1016/j.rse.2009.08.016>.
32. Guo, C.; Pleiss, G.; Sun, Y.; Weinberger, K.Q. On Calibration of Modern Neural Networks. In *Proceedings of the 34th International Conference on Machine Learning*; PMLR: Sydney, Australia, 17 July 2017; pp. 1321–1330.
33. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In *Proceedings of the 32nd International Conference on Machine Learning*; PMLR: Lille, France, 1 June 2015; pp. 448–456.
34. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
35. Ba, J.L.; Kiros, J.R.; Hinton, G.E. Layer Normalization. *arXiv* **2016**, arXiv:1607.06450.
36. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. *Focal Loss for Dense Object Detection*; ICCV: Venice, Italy, 2017; pp. 2980–2988.
37. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2017**, arXiv:1412.6980.
38. Loshchilov, I.; Hutter, F. SGDR: Stochastic Gradient Descent with Warm Restarts. *arXiv* **2017**, arXiv:1608.03983.
39. Tashman, L.J. Out-of-Sample Tests of Forecasting Accuracy: An Analysis and Review. *Int. J. Forecast.* **2000**, *16*, 437–450. [https://doi.org/10.1016/S0169-2070\(00\)00065-0](https://doi.org/10.1016/S0169-2070(00)00065-0).
40. He, H.; Garcia, E.A. Learning from Imbalanced Data. *IEEE Trans. Knowl. Data Eng.* **2009**, *21*, 1263–1284. <https://doi.org/10.1109/TKDE.2008.239>.
41. Xiong, R.; Yang, Y.; He, D.; Zheng, K.; Zheng, S.; Xing, C.; Zhang, H.; Lan, Y.; Wang, L.; Liu, T. On Layer Normalization in the Transformer Architecture. In *Proceedings of the International Conference on Machine Learning Virtual Event*; PMLR: 2020; pp. 10524–10533.
42. Bengio, Y.; Simard, P.; Frasconi, P. Learning Long-Term Dependencies with Gradient Descent Is Difficult. *IEEE Trans. Neural Netw.* **1994**, *5*, 157–166. <https://doi.org/10.1109/72.279181>.
43. Bai, S.; Kolter, J.Z.; Koltun, V. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. *arXiv* **2018**, arXiv:1803.01271.
44. Bowman, D.M.J.S.; Balch, J.K.; Artaxo, P.; Bond, W.J.; Carlson, J.M.; Cochrane, M.A.; D’Antonio, C.M.; DeFries, R.S.; Doyle, J.C.; Harrison, S.P.; et al. Fire in the Earth System. *Science* **2009**, *324*, 481–484. <https://doi.org/10.1126/science.1163886>.
45. Jolly, W.M.; Cochrane, M.A.; Freeborn, P.H.; Holden, Z.A.; Brown, T.J.; Williamson, G.J.; Bowman, D.M.J.S. Climate-Induced Variations in Global Wildfire Danger from 1979 to 2013. *Nat. Commun.* **2015**, *6*, 7537. <https://doi.org/10.1038/ncomms8537>.

46. Flannigan, M.D.; Krawchuk, M.A.; de Groot, W.J.; Wotton, B.M.; Gowman, L.M. Implications of Changing Climate for Global Wildland Fire. *Int. J. Wildland Fire* **2009**, *18*, 483–507. <https://doi.org/10.1071/WF08187>.
47. Moritz, M.A.; Parisien, M.-A.; Batllori, E.; Krawchuk, M.A.; Van Dorn, J.; Ganz, D.J.; Hayhoe, K. Climate Change and Disruptions to Global Fire Activity. *Ecosphere* **2012**, *3*, 1–22. <https://doi.org/10.1890/ES11-00345.1>.
48. Shi, X.; Chen, Z.; Wang, H.; Yeung, D.-Y.; Wong, W.; Woo, W. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. In *Proceedings of the Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2015; Volume 28.
49. Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; Yu, P.S. A Comprehensive Survey on Graph Neural Networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *32*, 4–24. <https://doi.org/10.1109/TNNLS.2020.2978386>.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.