

Article

AutoST-Net: A Spatiotemporal Feature-Driven Approach for Accurate Forest Fire Spread Prediction from Remote Sensing Data

Xuexue Chen ^{1,†}, Ye Tian ^{1,†}, Change Zheng ^{1,2,*}  and Xiaodong Liu ³

¹ School of Technology, Beijing Forestry University, Beijing 100083, China; xuexuechen@bjfu.edu.cn (X.C.); tytoemail@bjfu.edu.cn (Y.T.)

² State Key Laboratory of Efficient Production of Forest Resources, Beijing Forestry University, Beijing 100083, China

³ School of Ecology and Nature Conservation, Beijing Forestry University, Beijing 100083, China; xd_liu@bjfu.edu.cn

* Correspondence: zhengchange@bjfu.edu.cn

† These authors contributed equally to this work.

Abstract: Forest fires, as severe natural disasters, pose significant threats to ecosystems and human societies, and their spread is characterized by constant evolution over time and space. This complexity presents an immense challenge in predicting the course of forest fire spread. Traditional methods of forest fire spread prediction are constrained by their ability to process multidimensional fire-related data, particularly in the integration of spatiotemporal information. To address these limitations and enhance the accuracy of forest fire spread prediction, we proposed the AutoST-Net model. This innovative encoder–decoder architecture combines a three-dimensional Convolutional Neural Network (3DCNN) with a transformer to effectively capture the dynamic local and global spatiotemporal features of forest fire spread. The model also features a specially designed attention mechanism that works to increase predictive precision. Additionally, to effectively guide the firefighting work in the southwestern forest regions of China, we constructed a forest fire spread dataset, including forest fire status, weather conditions, terrain features, and vegetation status based on Google Earth Engine (GEE) and Himawari-8 satellite. On this dataset, compared to the CNN-LSTM combined model, AutoST-Net exhibits performance improvements of 5.06% in MIou and 6.29% in F1-score. These results demonstrate the superior performance of AutoST-Net in the task of forest fire spread prediction from remote sensing images.

Keywords: forest fire spread; prediction; deep learning; spatiotemporal features; attention mechanism; GEE; Himawari-8 satellite



Citation: Chen, X.; Tian, Y.; Zheng, C.; Liu, X. AutoST-Net: A Spatiotemporal Feature-Driven Approach for Accurate Forest Fire Spread Prediction from Remote Sensing Data. *Forests* **2024**, *15*, 705. <https://doi.org/10.3390/f15040705>

Academic Editor: Víctor Fernández-García

Received: 7 March 2024

Revised: 11 April 2024

Accepted: 12 April 2024

Published: 17 April 2024



Copyright: © 2024 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/>).

1. Introduction

Forests serve as vital natural resources, offering a wide range of ecosystem services such as providing habitats for diverse plant and animal species, mitigating climate change, and promoting biodiversity. However, not only do forest fires destroy ecosystems, but they also pose significant threats to property, buildings, and human lives. For instance, in June 2023, a fire in the Toronto area lasted for about two months. The smoke from the fire traveled a distance of 7000 km, reaching Spain. In August of the same year, a severe fire broke out on Ma'i Island, Hawaii, in the United States, resulting in the deaths of hundreds of people [1]. However, it is worth noting that although the severity of fires may be escalating in certain regions [2,3], recent studies reveal a global decrease in fire frequency as a result of human activity [4,5]. Proactive preventive measures, coupled with monitoring fire occurrences and predicting their spread, can significantly contribute to minimizing disaster losses [6,7]. Satellite remote sensing and computer vision provide appropriate methods for prevention.

The prediction of forest fire spread is a complex and challenging task. In addition to the fire itself, it is necessary to consider various factors such as weather conditions and terrain features [8]. However, poor quality and noisy data often result from incoherence in methods, data acquisition platforms, and data formats. Moreover, these datasets often lack comprehensive consideration of influencing factors, with most only involving fire data from specific regions. There is therefore a lack of globally inclusive forest fire spread datasets [9–12].

Google Earth Engine (GEE) integrates abundant remote sensing data, including meteorological, topographical, and satellite images, with powerful data processing, analysis, and visualization capabilities [13–15]. The Earth stationary meteorological satellite Himawari-8, located approximately 35,800 km above the Earth's equator, covers the research area (60 S–60 N, 80 E–160 W) at intervals of every 10 min, providing data on meteorological conditions, aerosol distribution, and fire detection [16–19], thus serving as a crucial tool for studying the spread of forest fire. Equipped with the Advanced Himawari Imager (AHI), Himawari-8 is capable of capturing a wide spectrum ranging from visible light to thermal infrared, enabling the monitoring of various environmental events. Additionally, Himawari-8 satellite and Google Earth Engine (GEE) data are accessible upon request at no cost, with downloaded data available in TIFF format, facilitating convenient utilization. Therefore, we used the Himawari-8 satellite and GEE [20] to collect fire and other data such as weather, wind, and terrain.

Researchers have explored various prediction methods of forest fire spread, categorized into three groups. The first focuses on physical models, such as cellular automata (CA) models [21–24], requiring large amounts of ground-based measurements and incorporating fire spread theory. Zhang et al. improved upon the Wang Zhengfei model of forest fire spread by incorporating the impact of fuel moisture on the speed of fire spread and refining the calculation of the initial fire spread rate [25]. Meng et al. proposed a theoretical model of forest fire spread based on an improved principle of Huygens that also factors in weather elements such as wind speed, wind direction, and precipitation [26]. Zhang et al. simplified the Rothermel forest fire spread rate formula and, combined with cellular automata, established a multi-dimensional cellular automata (MD-CA) model for forest fire spread containing different burning characteristics within each cell [27]. Li et al. integrated multifactor analysis methods with the FARSITE model to develop a 3D forest fire spread simulation system, FFSimulator, which visualizes the impact of multiple factors on forest fire spread [28]. While these models help us to understand the theoretical mechanisms of forest fire spread, they require extensive field data for calibration and validation, and their high complexity can affect prediction results in practical applications.

The second type employs machine learning and data mining methods, such as Random Forest and Logistic Regression, to construct predictive models by analyzing historical fire data and environmental factors [29–32]. Zheng employed Extreme Learning Machines to compute the probability of fire spots igniting, which were then used by cellular automata to define transition rules, simulating the forest fire spread process [33]. Xu combined Least Squares Support Vector Machines with a 3D forest fire cellular automata framework for forest fire spread predictions [34]. Rubi used climate characteristic observations from monitoring stations and two decades worth of satellite data on forest fires to predict and estimate the spread behavior of forest fires in Brazil's Federal District using Support Vector Machines, Random Forest, and AdaBoost, among others [35]. Janiec utilized Maxent models and Random Forest to study forest fire spread using satellite images, their various spatial and spectral resolution products, vector data, and bioclimatic variables, finding that the Random Forest approach showed better results on a macro scale, whereas the Maxent model was more effective on a micro scale [36]. Data-driven methods can avoid complex physical process simulations, but they are heavily reliant on the quality and completeness of datasets. A lack of high-quality training data could impact the accuracy of prediction outcomes. The quality and accuracy of the dataset are critical for the effectiveness of these models, but historical data frequently contain various noise sources [37]. The third is based

on deep learning and extensive fire remote sensing information gathered by monitoring and analyzing fire hotspots and relevant data, thereby predicting fire spread trends [38,39].

Convolutional neural networks (CNN) based on deep learning can extract features from data. Some scholars have attempted to use CNN for forest fire spread prediction [40,41]. Yang et al. used CNN to verify forest fire spread prediction, and the results showed that CNN can extract relevant data features effectively [42]. Zou et al. proposed an attention-based convolutional neural network that can learn the complex behaviors of wildfires across different fire-prone regions [43]. Ding et al. proposed a fully connected convolutional neural network for identifying the location and intensity of wildfires, which greatly exceeds that of other machine learning algorithms, such as support vector machine and k-means clustering [44]. Although these methods have effectively improved the prediction accuracy of forest fire spread, they mainly utilize 2DCNN to process the spatial features of forest fire spread, neglecting temporal dynamics. Hoai et al. proposed a fire detection method that includes a CNN to extract spatial features and a Long Short-Term Memory (LSTM) used to extract temporal features in videos [45]. Bhowmik proposed a U-Convolutional-LSTM neural network to extract key spatial and temporal features from environmental parameters inherent within contiguous weather data that are indicative of impending wildfires [46]. Nevertheless, these methods still have drawbacks, such as the separate processing of each frame image using 2D convolution and the requirement of processing each time step separately, resulting in prolonged prediction times. Considering these limitations, 3D convolution that accounts for time and spatial information simultaneously provides a better modeling ability for image-based spatiotemporal sequences.

While convolution-based deep learning architectures have demonstrated exceptional performance in image segmentation tasks, they are limited in capturing long-range spatial dependencies. Recently, self-attention mechanisms and transformers architectures have achieved substantial success in natural language processing (NLP) [47–49], addressing issues associated with long-range dependencies within sequences [50]. In 2021, Dosovitskiy et al. proposed the game-changing Vision Transformer (ViT) model, which outperformed traditional CNNs in image classification tasks, marking a substantial shift in the image processing domain [51]. This discovery has led to studies integrating the local feature extraction capabilities of CNNs with the global sequence processing strengths of transformers models. For instance, Zheng et al. have merged CNNs with transformers [52], not only achieving remarkable improvements in image segmentation performance but also demonstrating an enhanced understanding of images with complex spatial structures. Chen et al. put forward the TransUNet model, showcasing the powerful synergy obtained by fusing transformers with the widely recognized UNet architecture for medical image segmentation [53]. This hybrid methodology has demonstrated exceptional performance across various datasets, highlighting the potential of combined convolutional and transformer-based approaches in advancing image analysis.

In summary, the purposes of the paper are as follows:

1. Introduce the novel AutoST-Net model, which is based on the dynamics of fire behavior and employs a 3D convolutional neural network to capture the spatiotemporal features of forest fire spread. The model incorporates a transformer to extract global features and includes an attention mechanism to improve performance and accuracy.
2. By creating a forest fire spread dataset based on the GEE and Himawari-8 satellites, evaluate and compare the performance of the AutoST-Net model with other models such as Zhengfei Wang-CA, Random Forest, and a CNN-LSTM combination.

2. Data

2.1. Study Regions

The study regions chosen for this research are Sichuan and Yunnan Province in China. Due to the unique geographic location and climatic conditions, the study area is characterized by rich vegetation, complex topography, and humid climate, resulting in frequent forest fires in spring and winter. Fires typically spread rapidly along steep slopes, deep

valleys, and dense forests, posing a challenge to local fire suppression efforts. Since at least three fatal forest fires occur annually in the above areas, they provide suitable research subjects for forest fire spread studies. This study focuses on the southwestern part of Sichuan Province and the northwestern and southwestern parts of Yunnan Province, shown in Figure 1, as these areas experience the highest frequency of forest fires.



Figure 1. China, Sichuan, and Yunnan Province.

2.2. Datasets

The spread of forest fire is driven by three main factors: weather conditions, terrain features, and vegetation status. In this study, we integrated wildfire, weather, topography, and vegetation data to predict the spread of forest fire. We used records from the China Statistical Yearbook between 2016 and 2021 to investigate forest fire events. This registry meticulously catalogs the chronological and geographical dimensions of fires throughout China [54], providing a solid base for assessing the spatiotemporal distribution of forest fires. Given the spatial and temporal resolution capabilities of current remote sensing satellite technology, as well as the convenience and cost-effectiveness of data acquisition, we focus on analyzing major forest fires (with affected areas ranging from 100 to 1000 hectares) and exceptional forest fires (with affected areas exceeding 1000 hectares).

To enhance the efficiency and accuracy of forest fire monitoring and early warning, we utilize the Himawari-8 geostationary meteorological satellite operated by the Japan Meteorological Agency. Launched on 7 October 2014, and operational since 7 July 2015, the satellite orbits approximately 35,800 km above the Earth's equator. As a geostationary satellite, Himawari-8 maintains a fixed position relative to the Earth's surface, hovering over the equator at approximately 140.7 degrees east longitude. It continuously monitors its designated coverage area, providing uninterrupted observations of the Asia–Pacific region. Himawari-8 provides data in two formats: NetCDF4 and Himawari Standard Data (HSD), with observation modes including Full Disk, Japan Northeast, Japan Southwest, Target Area, and Landmark Area. Due to the straightforward preprocessing of NetCDF4 data, easily accessible through the NetCDF4 library in Python, we focus solely on processing Himawari-8 data in NetCDF4 format.

Furthermore, the advanced Advanced Himawari Imager (AHI) carried by Himawari-8 covers a broad spectral range from visible light to thermal infrared, enabling the monitoring of various environmental events. AHI exhibits significant improvements in temporal and spatial resolution, achieving a temporal resolution of every 10 min and a spatial resolution of 2 km in Full Disk observation mode, supporting true-color imaging. The AHI instrument is equipped with 16 spectral channels, covering a range from approximately 0.47 μm (blue light) to 13.3 μm (longwave infrared), where channels 1–3 correspond to visible light, 4–6 to near-infrared, and 7–16 to thermal infrared.

In predicting the spread of forest fires, specific infrared channels (7 and 14) on AHI can sensitively capture hot areas (hotspots) on the ground, accurately identifying the location of fires. Thanks to its rapid image update frequency (every 10 min), the Himawari-8 satellite can provide timely and dynamic fire information for predicting the spread of forest fires. Despite its spatial resolution not being as fine as some low Earth orbit satellites, Himawari-8, with a 2 km spatial resolution in the infrared channel, can capture large-scale fire dynamics and monitor changes in fire behavior in real time. It continues to provide valuable data support for predicting the spread of forest fires.

In addition to Himawari-8 data, we utilize a suite of environmental variables that influence the spread of forest fires. These variables include a digital elevation model (DEM) for precise terrain details, which are crucial for analyzing fire behavior in areas with complex topography. The drought index is employed to quantify the level of dryness of the soil surface, as forest fires are more likely to occur in arid areas. KBDI computation relies on four data points: the latitude of the weather station, annual average precipitation, maximum dry bulb temperature, and the last 24 h of rainfall. It is a closed system ranging from 0 to 800 units and represents a moisture regime from 0 to 8 inches of water through the soil layer. At 8 inches of water, the KBDI assumes saturation. Zero is the point of no moisture deficiency and 800 is the maximum drought that is possible. Weather conditions such as soil humidity, temperature, and humidity, as well as wind direction data, are incorporated to estimate the intensity and likelihood of wildfires and predict their spread. Precipitation data provide insight into past weather conditions, where increased precipitation may indicate a reduction in potential fuel sources. The normalized difference vegetation index (NDVI) is an indicator of vegetation cover and is used to evaluate the rate and extent of fire spread after a fire has occurred. Additionally, geopotential height is utilized to gauge changes in topography, where complex terrain may induce localized airflow anomalies, such as valley winds, which could exacerbate or mitigate the spread of fire.

All the aforementioned data were obtained from the robust Google Earth Engine (GEE) platform. All data are available from the cited data sources to ensure the authenticity of the experiments, as detailed in Table 1.

Table 1. The influential factors.

Data	Source	Temporal Resolution	Spatial Resolution
DEM	Copernicus DEM GLO-30 dataset	-	30 m
Drought	Keetch–Byram Drought Index (KBDI) dataset [55]	1 day	4 km
Geopotential Height	NCEP Climate Forecast System dataset [56]	6 h	22 km
Humidity, Soil Humidity	GLDAS-2.1 dataset [57]	3 h	27 km
Temperature, vwind, uwind	ERA5-Land dataset [58]	1 h	11 km
Precipitation	GSMaP dataset [59]	1 h	11 km
NDVI	MODIS Terra Daily NDVI dataset	1 day	0.4 km
Fire	Himawari-8 NetCDF data [60]	10 min	2 km

In our study, we meticulously constructed a dataset from seven distinct forest fire events, resulting in 2000 samples. Each forest fire image from our dataset has a temporal resolution of 1 h and a spatial resolution of 2 km per pixel, with an image size of 32×32 pixels, ensuring detailed representation of fire extent and shape. For our training dataset, we included 1500 samples across five forest fire events, primarily focusing on the diverse ter-

rains of Yunnan and Sichuan provinces to improve the model's ability to generalize across the different forest conditions found in these areas. For the testing dataset, we specifically selected 500 samples from two complete and distinct forest fire events to robustly assess the model's predictive performance in real-world spread scenarios for forest fires.

2.3. Data Processing

We labeled each fire image with a corresponding fire mask. The resolution of each fire image is 32×32 pixels. A threshold [61–63], shown in Equations (1)–(4), was set to identify possible active fires and eliminate false alarms, such as water and clouds. Pixels that match the Equations (5) and (6) were categorized as active fire points and marked as 1. We extracted the fire mask at t (called “pre-fire mask”) and the fire mask at $t + 1$ (called “fire mask”), both acting as the fire spread markers.

$$(abs(A_{0.64}) < 0.01) \cap (abs(A_{0.86}) < 0.01) : \text{night} - \text{time} \quad (1)$$

$$\left(\frac{T_{3.9} - \text{mean}(T_{3.9})}{\text{std}(T_{3.9})} > 0.8 \right) \cap \left(\frac{T_{3.9} - \text{mean}(T_{3.9})}{\text{std}(T_{3.9})} - \frac{T_{11.2} - \text{mean}(T_{11.2})}{\text{std}(T_{11.2})} < 1.5 \right) : \text{possible} - \text{AF} \quad (2)$$

$$A_{2.3} > 0.05 : \text{not} - \text{water} \quad (3)$$

$$(A_{0.64} + A_{0.86} < 1.2) \cap (T_{12.4} < 265K) \cap ((A_{0.64} + A_{0.86} < 0.7) \cup (T_{12.4} > 285K)) : \text{not} - \text{cloud} \quad (4)$$

$$\text{possible} - \text{AF} \cap \text{not} - \text{water} \cap \text{not} - \text{cloud} : \text{AF} \quad (5)$$

$$\text{possible} - \text{AF} \cap \text{night} - \text{time} : \text{AF} \quad (6)$$

To ensure consistent spatial resolution of the fire masks and capture the details of forest fire spread, we aligned all data sources to a unified standard where each pixel represents a ground surface area of 2 km. This resolution was chosen to account for the rapid spread of fires as well as the typical sizes of forest fire impacts in the study area, which often range from over 100 hectares (1 square kilometer) to no more than 5000 hectares (50 square kilometers). At a 2 km spatial resolution, each pixel's area is sufficient to cover these fire events, ensuring that our analysis captures the spatial variability and extent of fires without missing key details. To achieve this resolution in our prediction model, we utilized the Google Earth Engine (GEE) platform. By using the `img.getDownloadUrl` function provided by the Google Earth Engine (GEE) platform and setting the scale parameter to 2226, we obtained GEE image data with a pixel resolution of 2 km, consistent with the resolution of Himawari-8 satellite data. Specific details of the dataset can be seen in Figure 2.

For the temporal resolution, we set a baseline refresh rate of 1 h. This approach not only enabled us to promptly access wildfire event data and precisely track the spread speed, direction, and coverage range of fires but also provided hourly updates of meteorological data, thereby safeguarding the timeliness and accuracy of the data. Considering the variations in spatial and temporal resolutions among different data sources, for slowly changing variables we selected the closest available data to represent the value at a specific time, regardless of whether the data were updated every 3 h, 6 h, or once a day. For instance, when collecting fire point data at 10 a.m. on 1 April 2020, for variables like drought index, we gathered data from 1 April 2020, to observe daily trends; for elevation, we collected data from 6 a.m. on that day, as the next data point would be at 12 p.m.; for humidity and soil moisture, we collected data from 9 a.m. on that day. This approach effectively reduces data noise and redundancy caused by inconsistencies across sources, enhancing the performance and stability of the model in monitoring fire propagation at a global scale.

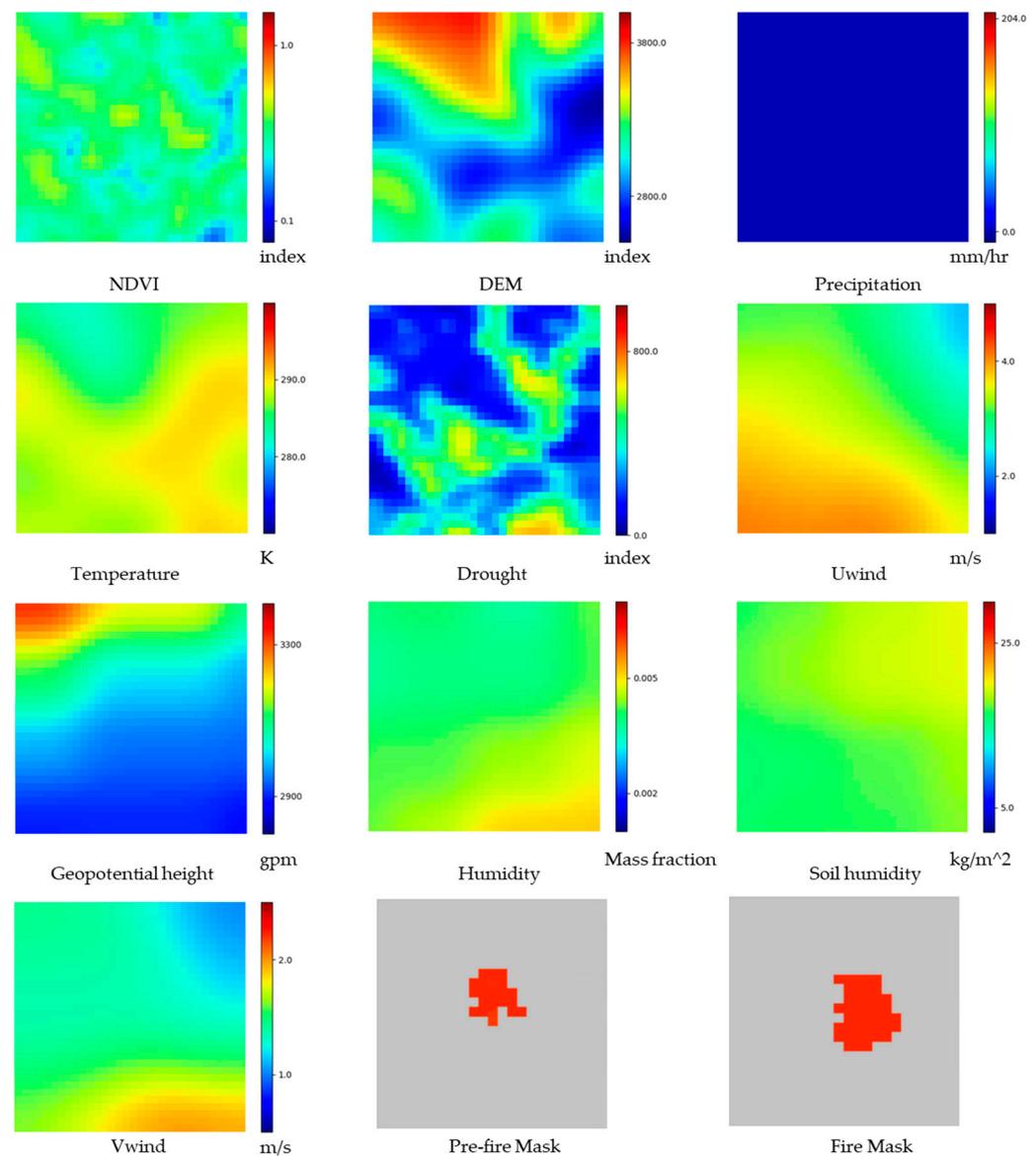


Figure 2. Forest fire spread dataset. Each image is 32×32 in size, with a spatial resolution of 2 km and a temporal resolution of 1 h. Within this dataset, the NDVI, DEM, terrain height, precipitation, uwind, vwind, humidity, soil humidity, temperature, drought, and the pre-fire mask were all collected at time “t”, whereas the fire mask was extracted at time “t + 1”. In the fire mask, red indicates the presence of fire, while gray signifies the absence of fire.

3. Methodology

3.1. Problem Definition

The task of fire spread prediction is formally defined as follows: given a sequence $\{x_1, x_2, \dots, x_t\}$ representing the fire images from the first moment to the t th moment and a sequence $\{y_1, y_2, \dots, y_t\}$ representing the corresponding factors influencing the fire occurrence, our goal is to generate the next frame of fire spread image x_{t+1} using the prediction model Γ . Each frame x_t represents a fire image of size $n * n$, and y_t represents the sum of all factors influencing the fire in each frame. Individual factors for each fire occurrence are defined as f_1, f_2, \dots, f_m . The input fire sequence and the corresponding sequence of influencing factors are used to predict the next frame of the fire spread image, utilizing the prediction model Γ . We have now defined the problem of fire spread prediction exactly as follows:

$$x_{t+1} = \Gamma(\{x_1, x_2, \dots, x_t\} | \{y_1, y_2, \dots, y_t\}) \quad (7)$$

$$y_t = f_1 \cup f_2 \cup \dots \cup f_m \quad (8)$$

3.2. AutoST-Net

Inspired by the U-Net [64], AutoST-Net combines encoder–decoder [65] architecture with an attention mechanism. The detailed architecture is shown in Figure 3.

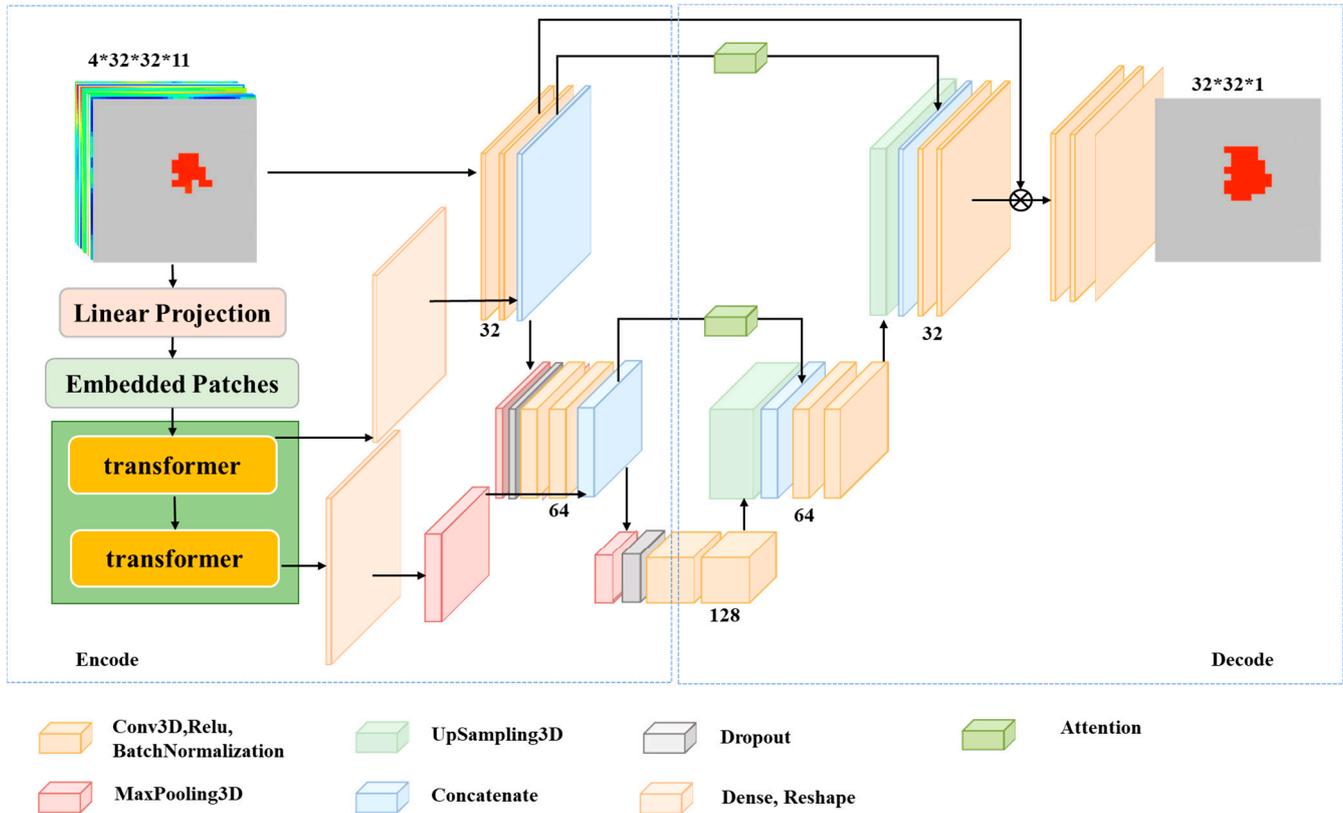


Figure 3. AutoST-Net model. The input data dimension is $32 \times 32 \times 4 \times 11$, indicating that the data are presented in the form of 4 frames of images, with each frame being 32×32 pixels in size and each pixel containing 11 different types of data. The output data dimension is $32 \times 32 \times 1$, which corresponds to a single 32×32 forest fire spread prediction image. The encoder of the model integrates both the transformer and 3DCNN to capture the spatiotemporal characteristics of forest fire spread. The decoder, on the other hand, employs upsampling, skip connections, and attention mechanisms to restore the feature maps to a size of 32×32 pixels, thereby generating the final prediction result.

A. Encoder

The AutoST-Net encoder can capture intricate dependencies in the input data, which is crucial for predicting the spread of forest fire. The encoder backbone network comprises a series of downsampling, 3D convolutional neural networks (3DCNN) [66]. The 3DCNN module extracted the temporal and spatial features of forest fire spread from consecutive fire sequences and relevant influential factors, while downsampling progressively decreased the spatial dimensions of the feature volume. Each encoder layer of AutoST-Net consists of two 3D convolutions (Conv3D) and one MaxPooling3D operation to capture features of the input data. In addition to the first layer, two dropout layers are added to the model to account for randomly dropping neurons during training, reducing the model's dependence on specific neurons and improving generalization.

An auxiliary branch parallel to the backbone network integrates a transformer [67] module to capture long-range dependencies in the sequence. Utilizing the multi-head self-attention mechanism of the transformer, the model can effectively identify direct

correlations between different points in temporal and spatial locations within the entire input data sequence. This mechanism allows the model to surpass the limitations of traditional convolution operations and analyze the global dynamics of forest fire spread, capturing inter-regional interaction features which are crucial for prediction tasks.

By fusing the two complementary features through a concatenation strategy, AutoST-Net generates a representation that includes both local detailed features and global features of forest fire spread. This fusion significantly enhances the model's understanding of local and global information and its ability to capture fire dynamics across various temporal scales, enabling accurate predictions of forest fire spread trends under diverse environmental conditions.

B. Decoder

The decoder in the AutoST-Net serves as the counterpart to the encoder. Its purpose is to restore the resolution of the low-resolution feature maps obtained from the encoder to match the resolution of the original input data. The decoder is achieved by a sequence of upsampling operations and corresponding skip connections to the encoder. Upsampling progressively increases the spatial size of the feature maps, while the skip connections merge the feature maps from the encoder and decoder.

In addition to upsampling and skip connections, a fusion strategy is used before the output layer, where the initial convolutional feature map is multiplied by the final upsampled feature map that combines the detailed spatial structure of the earlier layers with the abstract features of the deeper layers, enhancing the final feature representation and improving the accuracy of pixel classification. The segmentation output is produced using a Conv operation with a 1×1 kernel, reducing the channel dimensions without changing the spatial resolution to match the corresponding numbers of categories. This approach allows segmentation decisions to be made at the voxel level while preserving the resolution of the original input.

C. The attention mechanism

To accurately identify and locate areas of forest fire spread while differentiating them from unaffected regions, it is crucial for models to capture important local feature information. Traditional U-Net architectures address this by applying skip connections that directly concatenate feature maps from the encoder to the decoder, bridging the semantic gap between the two. However, this approach can introduce redundant low-level feature information. In this study, we propose a new attention model within the skip connections to enhance feature delivery to the decoder and improve the model's performance in forecasting forest fire spread. Drawing inspiration from channel-wise and spatial cross-attention mechanisms, we introduce the Channel Attention Module (CAM) [68] and the Spatial Attention Model (SAM) [69]. These independent attention modules selectively extract features from both the channel and spatial dimensions.

Considering the dynamic spatiotemporal and contributing factors in wildfire spreading, we combine the CAM and SAM through concatenation to obtain fine-grained feature representations to enhance the model performance. The fusion of these attention modules is depicted in Figure 4.

$$\text{Attention}(x) = C(\text{CAM}(x), \text{SAM}(x)) \quad (9)$$

In Formula (9), x represents the input feature, CAM and SAM represent the channel attention module and spatial attention model, and C represents the concatenation operation.

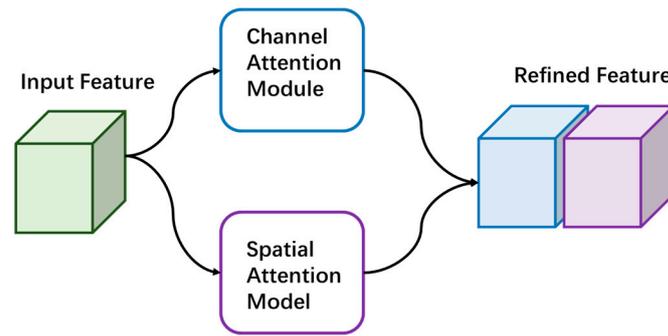


Figure 4. The attention mechanism.

Channel Attention Module

The Channel Attention Module (CAM) is designed to analyze the relationship between the forest fire and its associated factors, as illustrated in Figure 5. To this end, we implement two global pooling strategies—Average Pooling and Max Pooling—to extract comprehensive contextual information from different perspectives. Subsequently, the pooled features are fed into a fully connected (dense) layer with shared weights to further fine-tune the feature representation, thus enhancing the inter-channel correlation. After feature transformation, a sigmoid activation function is employed to produce channel attention maps, indicating the significance of each channel. This approach effectively streamlines the model’s parameters and fortifies the model’s sensitivity to crucial feature channels pertinent to the task.

$$CAM(x) = x \otimes \sigma (\text{MaxPool} (x) + \text{AvgPool}(x)) \tag{10}$$

In Formula (10), x represents the input feature, MaxPool and AvgPool represent Max Pooling and Average Pooling, and σ represents the sigmoid activation function.

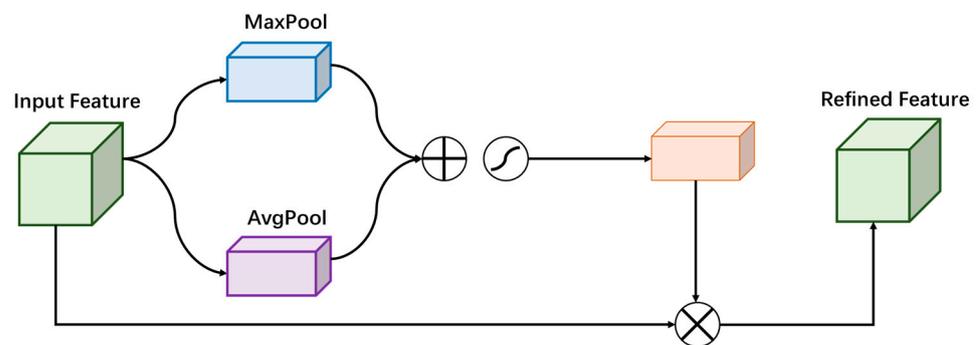


Figure 5. Channel attention module.

Spatial Attention Model

Given the inherent limitations of convolutional kernels in capturing only local features, we have introduced a transformer architecture during the encoder design phase to grasp global spatiotemporal dependencies. Consequently, the spatial attention model in this section focuses on extracting salient spatial features by executing Max Pooling operations. This is followed by utilizing a 3DCNN with an extensive receptive field to amplify the model’s ability to capture precise spatial location details. Finally, we apply a sigmoid activation function to dynamically modulate the responsiveness of each position in the original feature map, guiding the model to focus on areas more likely to represent fire zones, as shown in Figure 6.

$$SAM(x) = x \otimes \sigma (\text{Conv3D} (\text{MaxPool} (x))) \tag{11}$$

In Formula (11), x represents the input feature, MaxPool represents Max Pooling, Conv3D represents 3DCNN, and σ represents the sigmoid activation function.

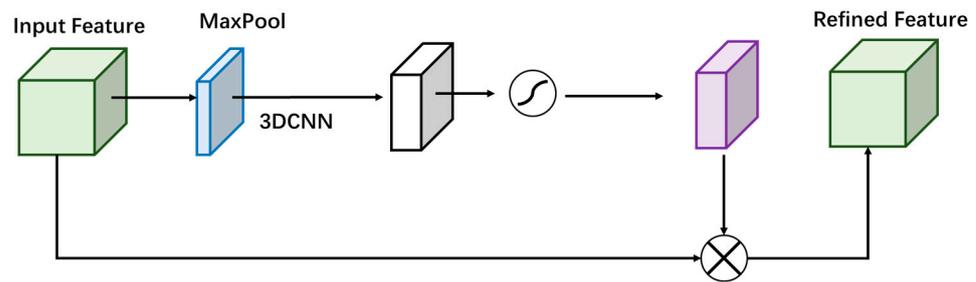


Figure 6. Spatial attention model.

After the completion of the overall design of the model, we employed the TensorFlow 2.0.0 deep learning framework and the Python 3.6.5 programming language, supplemented with imageio, matplotlib, and other image processing libraries, for practical operations. Our computational software is primarily automated, and once the parameters and training procedures of the AutoST-Net model are set, the software automatically performs steps such as data loading, preprocessing, model training, validation, and testing. Nevertheless, during the development and debugging stages of the model, we still need to manually adjust certain parameters and configurations to optimize the model's performance. Therefore, although most tasks are automated, manual intervention is still required for some critical steps. To visually demonstrate the model's performance, prediction results, and comparisons with other methods, we utilize various software tools for result visualization. Among them, Python libraries such as matplotlib and seaborn are our primary tools for creating various charts and images.

Hyperparameter tuning was performed sequentially, with the best result from one set of experiments used in subsequent experiments, in the following order: learning rate ($\{1 \times 10^{-4}, 1 \times 10^{-5}, 1 \times 10^{-6}\}$); image resizing ($\{100\%, 75\%, 50\%\}$ of 32×32); dropout ($\{0, 0.1, 0.2, 0.3, 0.4, 0.5\}$); frames of image ($\{2, 4, 8\}$). The final model was trained using an Adam optimizer with learning rate 1×10^{-4} , dropout 0.2, epoch 150, and batch size 2.

D. Loss function

Focal loss [70] is adopted as the loss function in this study. During the training of the AutoST-Net model, we also compared the results of focal loss with binary cross-entropy loss. The comparative experiment results are presented in Section 4.2.

$$p_t = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{otherwise} \end{cases} \quad (12)$$

$$\text{Focal Loss}(p_t) = -(1 - p_t)^\gamma \log(p_t) \quad (13)$$

$$\text{Binary - Cross Entropy loss} = -\frac{1}{N} \sum_{i=1}^N y_i * \log(p(y_i)) + (1 - y_i) * \log(1 - p(y_i)) \quad (14)$$

4. Experiments and Results

4.1. Evaluation Metrics

For each experiment, we report the following evaluation metrics typical for binary classification problems:

$$MIou = \left(\frac{TP}{TP + FP + FN} + \frac{TN}{TN + FP + FN} \right) * \frac{1}{2} \quad (15)$$

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

$$Precision = \frac{TP}{TP + FP} \quad (17)$$

$$F1 - score = \frac{2 * Recall * Precision}{Recall + Precision} \quad (18)$$

The above equations show an analysis of the model's mean intersection over union (MIOU) and *F1-score* metrics for the positive class (burned area). Specifically, MIOU measures the overlap between the predicted and ground truth regions, and *F1-score* is the harmonic mean of precision and recall. These metrics provide a comprehensive assessment of the model's performance to predict the spread of forest fires.

4.2. Comparative Experiments

The AutoST-Net model is ingeniously designed to operate in a near-real-time mode. It can accurately predict the trend of forest fire spread within the next hour based on the fire conditions that have already occurred, providing crucial support for a timely response. To comprehensively evaluate the superior performance of the AutoST-Net model in forest fire spread prediction, we conducted a series of experiments, comparing it with several other significant forest fire spread models. Firstly, the Wang Zhengfei-CA model [71], which is based on the principles of cellular automata, meticulously simulates the dynamics of fire spread across different cellular units. It integrates crucial factors such as wind speed, terrain, and vegetation type, providing a detailed prediction of fire spread. Secondly, the Random Forest model [35] leverages ensemble learning strategies, combining predictions from multiple decision trees to enhance stability and accuracy. This model excels in extracting underlying patterns of fire spread, especially when processing large amounts of training data. Furthermore, the CNN-LSTM combined model [46] integrates the strengths of image processing and sequence analysis. The CNN component focuses on extracting the spatial features of the fire, while the LSTM component captures the temporal dependencies of the fire spread. Together, they provide more precise prediction results. Additionally, we included the 3DUnet model and the 3DUnetTransformer model for performance comparison with the AutoST-Net model, with a focus on the influence of attention mechanisms and transformers.

In the testing phase of the AutoST-Net model, we adopted an approach of conducting a series of rigorous experiments using a comprehensive test dataset comprising data from two distinct forest fires. This approach aimed to enhance the model's robustness and generalizability in facing diverse testing conditions by amalgamating data from different fire scenarios. Such amalgamated testing methodology aids in mitigating incidental errors and provides performance evaluations closer to real-world scenarios. To ensure a fair comparison, we employed standardized techniques to fine-tune the hyperparameters of all models involved in the comparison. The experimental results were derived using the optimal parameters, and the number of epochs for model training was appropriately determined based on the performance of the validation set.

The results of the comparative experiments are consolidated in Table 2 and feature a performance juxtaposition between the focal loss and cross-entropy loss as implemented in the AutoST-Net model, which is thoroughly documented in Table 3.

Table 2. Comparative experimental results. Best results are bold.

Model	F1-Score	MIou	Execution Time (s)
Wang Zhengfei-CA	0.7041	0.7570	0.5
Random Forest	0.6975	0.7232	2
CNN-LSTM combined	0.7421	0.7792	13
3DUnet	0.7715	0.8090	12
3DUnetTransformer	0.7769	0.8114	16
AutoST-Net	0.8050	0.8298	24

Table 3. Comparison of focal loss and cross-entropy loss results. Best results are bold.

Model	Loss	F1-Score	MIou
AutoST-Net	Focal Loss	0.8050	0.8298
	Binary Cross-Entropy Loss	0.7243	0.8001

As shown in Table 2, in terms of performance, the AutoST-Net model outperformed the other models, achieving an MIou (mean intersection over union) of 82% and an F1-score of 80% on the test set. It exhibited significant superiority over the Random Forest and the CNN-LSTM combined models. Compared to the CNN-LSTM combined model, the AutoST-Net model showed improvements of 5.06% in MIou and 6.29% in F1-score, which indicates the powerful capability of the AutoST-Net model in processing complex spatiotemporal data and its precision in predicting forest fire spread.

The AutoST-Net model outperforms the CNN-LSTM combined model in capturing the complex features from forest fire spread datasets, owing to its innovative integration of 3DCNN for fine-grained feature extraction and transformers for extensive global feature extraction. This method enhances the understanding of the spatial and temporal dynamics of forest fire spread. Figure 7 vividly illustrates the high performance of the AutoST-Net model, exposing its acute sensitivity to local details during a forest fire spread and its impressive accuracy in global predictions. The comparative analysis of the CNN-LSTM combined and 3DUnet models indicates that 3DCNN is far superior in detecting subtle variations within spatiotemporal predictive data. Such sensitivity likely stems from the 3DCNN's capability to discern complex interrelations across spatial and temporal aspects that might be missed by the sequential learning mechanism of 2DCNN and LSTM.

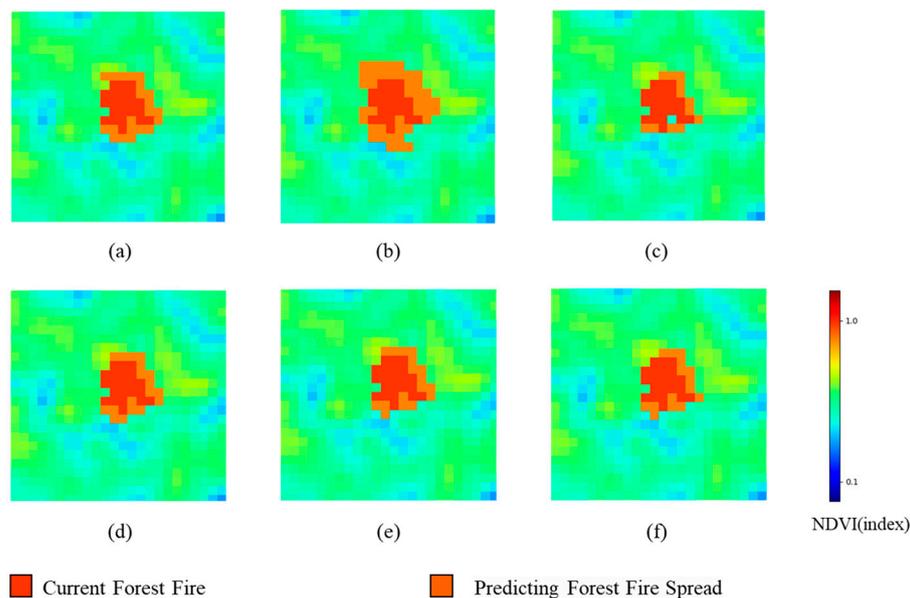


Figure 7. The forest fire spread prediction results of different models. (a) The actual spread of the forest fire that occurred in Muli County, Sichuan Province, in March 2020; (b) the prediction results of the Wang Zhengfei-CA model; (c) the prediction results of the Random Forest model; (d) the prediction results of the AutoST-Net model; (e) the prediction results of the 3DUnet model; (f) the prediction results of the CNN-LSTM combined model. All images adopt a consistent color scheme, with red representing the areas where forest fire had already occurred, orange indicating the predicted fire spread trends, and the background depicting vegetation coverage through NDVI.

The results show that training with focal loss yields a better average F1-score and MIou because focal loss is more effective in handling imbalanced data due to adjusting sample weights during training, which helps to avoid an over-emphasis on majority classes.

MIou metric can reflect the overlap between the predicted and ground truth masks, which can provide a more intuitive representation of the forest fire spread compared to the F1-score. The prediction results of all models are shown in Figure 7.

Figure 7 shows the predicting results of different models. In this illustration, the Ground Truth represents the actual fire spread. The red regions depict the burned areas, while the orange regions represent the predicted spread in the next time step. The blue pixels represent the areas with low NDVI values, indicating the regions without possibility of fire spread. Overall, the predictions from all models align closely with the actual fire spread, with variations primarily observed in the spread of the small-area fire.

Upon further observation of the aforementioned prediction results, it can be noted that the predictions from the AutoST-Net model closely resemble the Ground Truth, with only slight deviations observed in the bottom left spreading area. In contrast, the prediction from the CNN-LSTM combined model exhibits larger discrepancies. Furthermore, the Random Forest model displays noticeable disparities compared to the Ground Truth, with the issue of missing data in the bottom right area. Overall, our model performs the best and presents a relatively complete fire contour.

To verify the influence of different remote sensing bands on forest fire spread prediction, we conducted an experimental comparison among different remote sensing data. Table 4 illustrates the impact of each dataset on forest fire spread prediction. When the number of input channels is 11, the input data include fire, NDVI, DEM, terrain height, precipitation, uwind, vwind, humidity, soil humidity, temperature, and drought. When the number of input channels is 14, three visible light bands are added as the input data, including band 1, band 2, and band 3 of the Himawari-8 satellite. Additionally, when the number of input channels is 16, band 7 and band 14 of the Himawari-8 satellite are added.

Table 4. The impact of different remote sensing data. Best results are bold.

The Number of Input Channels	F1-Score	MIou
11 (Baseline)	0.7534	0.7976
14 (+Visible Light Bands)	0.8050	0.8298
16 (+Bands 7 and 14)	0.7681	0.7945

From Table 4, we can visually observe that visible light has a significant impact on predicting the spread of forest fires. The models that include visible light bands produce higher predictive performance compared to the models that exclude this type of feature. The visible light spectrum is useful for detecting the reflective characteristics of smoke, flames, and other features that can help to distinguish vegetation types, land use patterns, and the topography of the area. Therefore, incorporating visible light bands as a part of the feature selection is an effective strategy for improving the accuracy of the model's predictions for forest fire spread. Additionally, we observed that the inclusion of bands 7 and 14, which are related to the burning area, led to a decrease in the model's predictive performance. This could result from the introduction of new features that disrupted the balance of the model's underlying structure. Therefore, carefully selecting and evaluating feature sets to optimize model performance is crucial.

In further research, to verify the capability of our model, we obtained the fire data for Hanma Biosphere Reserve in Inner Mongolia on 1 June 2018, using the same data acquisition methods as for the original dataset. In introducing these data, we aimed to test the applicability and accuracy of our AutoST-Net model across different forest ecosystems.

The results of the tests from Table 5 were encouraging, demonstrating that our model exhibits good adaptability to new geographic locations. Despite significant differences in forest types, climatic conditions, and topographical features between Inner Mongolia and the previously studied regions of Yunnan and Sichuan provinces, our model maintained a high level of predictive accuracy. Specifically, when compared with Yunnan and Sichuan provinces, the accuracy of fire prediction in Inner Mongolia was controlled within an error

range of less than 10%. This indicates that our AutoST-Net model has a certain degree of regional adaptability and generalization capability.

Table 5. The impact of different regions. Best results are bold.

Region	F1-Score	MIou
Sichuan and Yunnan Province	0.8050	0.8298
Hanma Biosphere Reserve	0.7144	0.7362

However, we must acknowledge that this error margin might be in part attributable to geographical differences. The patterns of fire spread in the Greater Khingan Range area may be influenced by the local unique climate conditions, vegetation types, and topographical characteristics. Adapting the model to account for these regional discrepancies will be the focus of subsequent research. We plan to improve the model's predictive accuracy further through more refined feature engineering and incorporating environmental factors that are specific to the region.

5. Discussion

The AutoST-Net model has shown remarkable potential in predicting the spread of forest fires, skillfully leveraging deep learning technology to extract key features of fire spread directly from data. This approach moves away from the reliance on cumbersome ground measurement data required by traditional physical models [22,23] such as the CA model, thereby streamlining the data collection process and reducing costs. This transformation not only grants the model greater flexibility but also broadens its applicability across different geographical and climatic conditions, enabling outstanding performance.

Compared to traditional machine learning and data mining methods, AutoST-Net displays significant advantages. Traditional models typically predict based on simple statistical relationships between historical data and environmental factors [31,32]. In contrast, AutoST-Net utilizes its deep neural network structure to delve into more complex and subtle patterns of fire spread. Specifically, its integration of attention mechanisms allows the model to automatically focus on areas critical to predictive outcomes, like active fire zones and potential fire spread paths, thus significantly boosting prediction accuracy.

However, every model has its limitations, and AutoST-Net is no exception. Despite considerable efforts in model design and optimization, the inherent limitations of data characteristics and model structure currently prevent us from directly analyzing the importance of features. This is mainly due to our model structure and data characteristics, which do not readily provide insight into the contribution of each feature to the model's predictions. Although we recognize the importance of feature analysis, we are indeed unable to offer detailed interpretations within our current research framework.

To overcome this limitation, we plan to explore more diverse model architectures and data processing methods in future research to more accurately assess the impact of various features on model performance. We will also actively focus on and attempt to adopt new technological approaches, such as model interpretability methods [72] or feature selection techniques [73], to gain deeper insights into how the model works and the significance of its features.

Moreover, we acknowledge that the performance of the AutoST-Net model is still constrained by the quality and quantity of datasets. To further enhance the model's generalizability, we plan to introduce a more diverse range of datasets in future research and conduct comprehensive training and validation across different regions and seasons. This will help the model to better adapt to various environments and scenarios, thereby improving its robustness in practical applications.

We are also acutely aware of the challenges that the model faces in terms of computational cost. Despite many optimizations in model design, the demand for computational resources remains high when processing large-scale or high-resolution remote sensing data.

To effectively reduce computational costs, we will actively explore more efficient model architectures and algorithms, and fully utilize advanced technologies such as distributed computing to accelerate the model's training and inference processes.

6. Conclusions

In this study, we present the AutoST-Net model to predict the spread of forest fires. The model consists of an encoder–decoder structure, with the encoder incorporating a 3DCNN and a transformer module. These components capture local spatiotemporal features and global dependencies in the input data. A fusion stage combines these features to provide comprehensive information for the decoder to reconstruct forest fire spread imagery. To address the issue of feature redundancy in skip connections, we introduce a composite attention mechanism that combines channel and spatial attention. This mechanism selectively extracts key features from different perspectives, enhancing the model's ability to represent important information. The spatial attention focuses on prominent regions using Max Pooling, while the channel attention analyzes the impact of related factors.

To evaluate the performance of the AutoST-Net model, we establish a forest fire spread dataset based on data from the Himawari-8 satellite and GEE. This dataset includes the seven fires and ten influential factors. Through comparative experiments with advanced methods such as Wang Zhengfei-CA, Random Forest, and CNN-LSTM combined models, AutoST-Net demonstrated outstanding performance in terms of both MIou and F1 scores. The experimental results fully confirm the following conclusions:

1. The AutoST-Net model efficiently captures the complex spatiotemporal characteristics of forest fire spread through the skillful integration of 3DCNN and transformer.
2. The innovative attention mechanism significantly enhances the model's ability to precisely extract and utilize key features, thereby substantially improving prediction accuracy.
3. The high quality dataset constructed in this study lays a solid foundation for research on forest fire spread prediction.

Due to regional limitations in our study, the current model is more suitable for south-western China. However, since the dataset originates from the Himawari-8 satellite and GEE platform, there is a strong possibility of adapting the model to an entire state by collecting recent data from across the state and retraining the model. For forest service applications, only local meteorological factors and existing fire data need to be collected.

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

Funding: This work was supported in part by the National Key R&D Program of China under Grant 2023YFC3006805, and in part by the National Natural Science Foundation of China under Grant 31971668.

Data Availability Statement: Data available on request due to privacy restrictions.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Marris, E. Hawaii wildfires: Did scientists expect Maui to burn? *Nature* **2023**, *620*, 708–709. [\[CrossRef\]](#)
2. John, T.A.; Williams, A.P. Impact of anthropogenic climate change on wildfire across western US forests. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 11770–11775.
3. Barbero, R.; Abatzoglou, J.T.; Pimont, F.; Ruffault, J.; Curt, T. Attributing Increases in Fire Weather to Anthropogenic Climate Change Over France. *Front. Earth Sci.* **2020**. [\[CrossRef\]](#)
4. Andela, N.; Morton, D.C.; Giglio, L.; Chen, Y.; van der Werf, G.R.; Kasibhatla, P.S.; DeFries, R.S.; Collatz, G.J.; Hantson, S.; Kloster, S.; et al. A human-driven decline in global burned area. *Science* **2017**, *356*, 1356–1362. [\[CrossRef\]](#)

5. Fernández-García, V.; Alonso-González, E. Global Patterns and Dynamics of Burned Area and Burn Severity. *Remote Sens.* **2023**, *15*, 3401. [[CrossRef](#)]
6. Jinkyu, R.; Dongkurl, K.; Seungmin, C. Position Estimation of Forest Fires Using an Infrared Camera Based on Pan Tilt Servo. *J. Korean Soc. Hazard Mitig.* **2022**, *22*, 97–103.
7. Gayathri, S.; Karthi, P.V.A.; Sunil, S. Prediction and Detection of Forest Fires Based on Deep Learning Approach. *J. Pharm. Negat. Results* **2022**, *13*, 429–433. [[CrossRef](#)]
8. Tian, Y.; Wu, Z.; Li, M.; Wang, B.; Zhang, Z. Forest Fire Spread Monitoring and Vegetation Dynamics Detection Based on Multi-Source Remote Sensing Images. *Remote Sens.* **2022**, *14*, 4431. [[CrossRef](#)]
9. Amatulli, G.; Camia, A.; San-Miguel-Ayanz, J. Estimating future burned areas under changing climate in the EU-Mediterranean countries. *Sci. Total Environ.* **2013**, *450*, 209–222. [[CrossRef](#)]
10. Verdú, F.; Salas, J.; Vega-García, C. A multivariate analysis of biophysical factors and forest fires in Spain 1991–2005. *Int. J. Wildland Fire* **2012**, *21*, 498–509. [[CrossRef](#)]
11. Vecín-Arias, D.; Castedo-Dorado, F.; Ordóñez, C.; Rodríguez-Pérez, J.R. Biophysical and lightning characteristics drive lightning-induced fire occurrence in the central plateau of the Iberian Peninsula. *Agric. For. Meteorol.* **2016**, *225*, 36–47. [[CrossRef](#)]
12. Bui, D.T.; Bui, Q.T.; Nguyen, Q.P.; Pradhan, B.; Nampak, H.; Trinh, P.T. A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area. *Agric. For. Meteorol.* **2017**, *233*, 32–44.
13. Zhao, Q.; Yu, L.; Li, X.; Peng, D.; Zhang, Y.; Gong, P. Progress and Trends in the Application of Google Earth and Google Earth Engine. *Remote Sens.* **2021**, *13*, 3778. [[CrossRef](#)]
14. Lasaponara, R.; Abate, N.; Fattore, C.; Aromando, A.; Cardettini, G.; Di Fonzo, M. On the Use of Sentinel-2 NDVI Time Series and Google Earth Engine to Detect Land-Use/Land-Cover Changes in Fire-Affected Areas. *Remote Sens.* **2022**, *14*, 4723. [[CrossRef](#)]
15. Yailymov, B.; Shelestov, A.; Yailymova, H.; Shumilo, L. Google Earth Engine Framework for Satellite Data-Driven Wildfire Monitoring in Ukraine. *Fire* **2023**, *6*, 411. [[CrossRef](#)]
16. Chen, J.; Lv, Q.C.; Wu, S.; Zeng, Y.L.; Li, M.C.; Chen, Z.Y.; Zhou, E.Z.; Zheng, W.; Liu, C.; Chen, X.; et al. An adapted hourly Himawari-8 fire product for China: Principle, methodology and verification. *Earth Syst. Sci. Data* **2023**, *15*, 1911–1931. [[CrossRef](#)]
17. Xu, H.; Zhang, G.; Zhou, Z.; Zhou, X.; Zhou, C. Forest Fire Monitoring and Positioning Improvement at Subpixel Level: Application to Himawari-8 Fire Products. *Remote Sens.* **2022**, *14*, 2460. [[CrossRef](#)]
18. Zhou, W.; Tang, B.H.; He, Z.W.; Huang, L.; Chen, J.Y. Identification of forest fire points under clear sky conditions with Himawari-8 satellite data. *Int. J. Remote Sens.* **2024**, *45*, 214–234. [[CrossRef](#)]
19. Zhang, D.; Huang, C.; Gu, J.; Hou, J.; Zhang, Y.; Han, W.; Dou, P.; Feng, Y. Real-Time Wildfire Detection Algorithm Based on VIIRS Fire Product and Himawari-8 Data. *Remote Sens.* **2023**, *15*, 1541. [[CrossRef](#)]
20. Gupta, S.K.; Kanga, S.; Meraj, G.; Kumar, P.; Singh, S.K. Uncovering the hydro-meteorological drivers responsible for forest fires utilizing geospatial techniques. *Theor. Appl. Climatol.* **2023**, *153*, 675–695. [[CrossRef](#)]
21. Mutthulakshmi, K.; Wee, M.R.E.; Wong, Y.C.K.; Lai, J.W.; Koh, J.M.; Acharya, U.R.; Cheong, K.H. Simulating Forest Fire Spread and Fire-fighting Using Cellular Automata. *Chin. J. Phys.* **2020**, *65*, 642–650. [[CrossRef](#)]
22. Freire, J.G.; DaCamara, C.C. Using cellular automata to simulate wildfire propagation and to assist in fire management. *Natural Hazards Earth Syst. Sci.* **2019**, *19*, 169–179. [[CrossRef](#)]
23. Sun, L.Y.; Xu, C.C.; He, Y.L.X.; Zhao, Y.J.; Xu, Y.; Rui, X.P.; Xu, H.W. Adaptive Forest Fire Spread Simulation Algorithm Based on Cellular Automata. *Forests* **2021**, *12*, 1431. [[CrossRef](#)]
24. Rui, X.P.; Hui, S.; Yu, X.T.; Zhang, G.Y.; Wu, B. Forest fire spread simulation algorithm based on cellular automata. *Nat. Hazards* **2018**, *91*, 309–319. [[CrossRef](#)]
25. Zhang, X.T.; Liu, P.S.; Wang, X.F. Research on the Improvement of Wang Zhengfei. *Shandong For. Sci. Technol.* **2020**, *50*, 1–6+40.
26. Meng, Q.K.; Huai, Y.J.; You, J.W.; Nie, X.Y. Visualization of 3D forest fire spread based on the coupling of multiple weather factors. *Comput. Graph.* **2023**, *110*, 58–68. [[CrossRef](#)]
27. Zhang, S.Y.; Liu, J.Q.; Gao, H.W.; Chen, X.D.; Li, X.D.; Hua, J. Study on Forest Fire spread Model of Multi-dimensional Cellular Automata based on Rothermel Speed Formula. *Cerne* **2021**, *27*, e-102932. [[CrossRef](#)]
28. Li, J.W.; Li, X.W.; Chen, C.C.; Zheng, H.R.; Liu, N.Y. Three-Dimensional Dynamic Simulation System for Forest Surface Fire Spreading Prediction. *Int. J. Pattern Recognit. Artif. Intell.* **2018**, *32*, 1850026. [[CrossRef](#)]
29. Khanmohammadi, S.; Arashpour, M.; Golafshani, E.M.; Cruz, M.G.; Rajabifard, A.; Bai, Y. Prediction of wildfire rate of spread in grasslands using machine learning methods. *Environ. Model. Softw.* **2022**, *156*, 105507. [[CrossRef](#)]
30. Bot, K.; Borges, J.G. A Systematic Review of Applications of Machine Learning Techniques for Wildfire Management Decision Support. *Inventions* **2022**, *7*, 15. [[CrossRef](#)]
31. Shmuel, A.; Heifetz, E. A Machine-Learning Approach to Predicting Daily Wildfire Expansion Rate. *Fire* **2023**, *6*, 319. [[CrossRef](#)]
32. Michael, Y.; Helman, D.; Glickman, O.; Gabay, D.; Brenner, S.; Lensky, I.M. Forecasting fire risk with machine learning and dynamic information derived from satellite vegetation index time-series. *Sci. Total Environ.* **2020**, *764*, 142844. [[CrossRef](#)]
33. Zheng, Z.; Huang, W.; Li, S.N.; Zeng, Y.N. Forest fire spread simulating model using cellular automaton with extreme learning machine. *Ecol. Model.* **2017**, *348*, 33–43. [[CrossRef](#)]
34. Xu, Y.Q.; Li, D.J.; Ma, H.; Lin, R.; Zhang, F.Q. Modeling Forest Fire Spread Using Machine Learning-Based Cellular Automata in a GIS Environment. *Forests* **2023**, *13*, 1974. [[CrossRef](#)]

35. Rubi, J.N.S.; de Carvalho, P.H.P.; Gondim, P.R.L. Application of Machine Learning Models in the Behavioral Study of Forest Fires in the Brazilian Federal District region. *Eng. Appl. Artif. Intell.* **2023**, *118*, 105649. [[CrossRef](#)]
36. Janiec, P.; Gadal, S. A Comparison of Two Machine Learning Classification Methods for Remote Sensing Predictive Modeling of the Forest Fire in the North-Eastern Siberia. *Remote Sens.* **2021**, *12*, 4157. [[CrossRef](#)]
37. De Bem, P.P.; de Carvalho, O.A.; Matricardi, E.A.T.; Guimaraes, R.F.; Gomes, R.A.T. Predicting wildfire vulnerability using logistic regression and artificial neural networks: A case study in Brazil's Federal District. *J. Int. Assoc. Wildland Fire* **2019**, *28*, 35–45. [[CrossRef](#)]
38. Cardil, A.; Monedero, S.; Ramírez, J.; Silva, C.A. Assessing and reinitializing wildland fire simulations through satellite active fire data. *J. Environ. Manag.* **2019**, *231*, 996–1003. [[CrossRef](#)]
39. Luz, A.E.O.; Negri, R.G.; Massi, K.G.; Colnago, M.; Silva, E.A.; Casaca, W. Mapping Fire Susceptibility in the Brazilian Amazon Forests Using Multitemporal Remote Sensing and Time-Varying Unsupervised Anomaly Detection. *Remote Sens.* **2022**, *14*, 2429. [[CrossRef](#)]
40. Stankevich, T.S. The use of convolutional neural networks to forecast the dynamics of spreading forest fires in real time. *Bus. Inform.* **2018**, *46*, 17–27. [[CrossRef](#)]
41. Prapas, I.; Kondylatos, S.; Papoutsis, I.; Camps-Valls, G.; Ronco, M.; Fernandez-Torres, M.A.; Guillem, M.P.; Carvalhais, N. Deep Learning Methods for Daily Wildfire Danger Forecasting. *arXiv* **2021**, arXiv:2111.02736.
42. Yang, S.; Lupascu, M.; Meel, K.S. Predicting Forest Fire Using Remote Sensing Data and Machine Learning. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtually, 2–9 February 2021; Volume 35, pp. 14983–14990.
43. Zou, Y.; Sadeghi, M.; Liu, Y.; Puchko, A.; Le, S.; Chen, Y.; Andela, N.; Gentine, P. Attention-Based Wildland Fire Spread Modeling Using Fire-Tracking Satellite Observations. *Fire* **2023**, *6*, 289. [[CrossRef](#)]
44. Ding, C.; Zhang, X.; Chen, J.; Ma, S.; Lu, Y.; Han, W. Wildfire detection through deep learning based on Himawari-8 satellites platform. *Int. J. Remote Sens.* **2022**, *43*, 5040–5058. [[CrossRef](#)]
45. Hoai, N.V.; Anh, D.T.; Manh, D.N.; Bokgil, C.; Soonghwan, R. Investigation of Deep Learning Method for Fire Detection from Videos. In Proceedings of the 2021 International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, Republic of Korea, 20–22 October 2021; pp. 593–595.
46. Bhowmik, R.T.; Jung, Y.S.; Aguilera, J.A.; Prunicki, M.; Nadeau, K. A multi-modal wildfire prediction and early-warning system based on a novel machine learning framework. *J. Environ. Manag.* **2023**, *341*, 117908. [[CrossRef](#)]
47. Devlin, J.; Chang, M.W.; Lee, K.T.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2019**, arXiv:1810.04805.
48. Radford, A.; Narasimhan, K. Improving Language Understanding by Generative Pre-Training. 2018. Available online: <https://api.semanticscholar.org/CorpusID:49313245> (accessed on 1 October 2022).
49. Raffel, C.; Shazeer, N.M.; Roberts, A.; Lee, K.; Narang, S.; Matena, M.; Zhou, Y.; Li, W.; Liu, P.J. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *J. Mach. Learn. Res.* **2020**, *21*, 140:1–140:67.
50. Brown, T.B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language models are few-shot learners, Advances in Neural Information Processing Systems. *arXiv* **2020**, arXiv:2005.14165.
51. Alexey, D.; Lucas, B.; Alexander, K.; Dirk, W.; Zhai, X.; Thomas, U.; Mostafa, D.; Matthias, M.; Georg, H.; Sylvain, G.; et al. An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale. *arXiv* **2021**, arXiv:2010.11929.
52. Zheng, S.; Lu, J.; Zhao, H.; Zhu, X.; Luo, Z.; Wang, Y.; Fu, Y.; Feng, J.; Xiang, T.; Philip, H.S.T.; et al. Rethinking Semantic Segmentation from a Sequence-to-Sequence Perspective with Transformers. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Virtual, 19–25 June 2021; pp. 6877–6886.
53. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Ehsan, A.; Wang, Y.; Lu, L.; Alan, L.Y.; Zhou, Y. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. *arXiv* **2021**, arXiv:2102.04306.
54. National Bureau of Statistics of China. *China Statistical Yearbook*, 2023rd ed.; National Bureau of Statistics of China: Beijing, China, 2023.
55. Takeuchi, W.; Darmawan, S.; Shofiyati, R.; Khiem, M.V.; Oo, K.S.; Pimple, U.; Heng, S. Near-real time meteorological drought monitoring and early warning system for croplands in Asia. In Proceedings of the 36th Asian Conference on Remote Sensing (ACRS), Quezon City, Philippines, 24–28 October 2015.
56. Saha, S.; Moorthi, S.; Wu, X.; Wang, J.; Nadiga, S.; Tripp, P.; Behringer, D.; Hou, Y.; Chuang, H.; Iredell, M.; et al. *NCEP Climate Forecast System Version 2 (CFSv2) 6-Hourly Products*; Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory: Boulder, CO, USA, 2011. [[CrossRef](#)]
57. Rodell, M.; Houser, P.R.; Jambor, U.; Gottschalck, J.; Mitchell, K.; Meng, C.-J.; Arsenault, K.; Cosgrove, B.; Radakovich, J.; Bosilovich, M.; et al. The Global Land Data Assimilation System. *Bull. Am. Meteorol. Soc.* **2004**, *85*, 381–394. [[CrossRef](#)]
58. Muñoz Sabater, J. *ERA5-Land Monthly Averaged Data from 1981 to Present*; Copernicus Climate Change Service (C3S) Climate Data Store (CDS): Reading, UK, 2019.
59. Kubota, T.; Aonashi, K.; Ushio, T.; Shige, S.; Takayabu, Y.N.; Kachi, M.; Arai, Y.; Tashima, T.; Masaki, T.; Kawamoto, N.; et al. Global Satellite Mapping of Precipitation (GSMaP) Products in the GPM Era. *Adv. Glob. Chang. Res.* **2020**, *67*. [[CrossRef](#)]
60. Bessho, K.; Date, K.; Hayashi, M.; Ikeda, A.; Imai, T.; Inoue, H.; Kumagai, Y.; Miyakawa, T.; Murata, H.; Ohno, T.; et al. An introduction to Himawari-8/9—Japan's new-generation geostationary meteorological satellites. *J. Meteorol. Soc. Jpn.* **2016**, *94*, 151–183. [[CrossRef](#)]

61. Todd, J.H.; Melanie, K.V.; Gail, L.S.; Yen-Ju, B.; Joshua, J.P.; Joshua, D.T.; Jeff, T.F.; John, L.D. The Landsat Burned Area algorithm and products for the conterminous United States. *Remote Sens. Environ.* **2020**, *244*, 111801.
62. Xu, G.; Zhong, X. Real-time wildfire detection and tracking in Australia using geostationary satellite: Himawari-8. *Remote Sens. Lett.* **2017**, *8*, 1052–1061. [[CrossRef](#)]
63. Liu, X.; He, B.; Quan, X. Near Real-Time Extracting Wildfire Spread Rate from Himawari-8 Satellite Data. *Remote Sens.* **2018**, *10*, 1654. [[CrossRef](#)]
64. Olaf, R.; Philipp, F.; Homas, B. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Munich, Germany, 5–9 October 2015; pp. 234–241.
65. López-Sánchez, M.; Hernández-Ocaña, B.; Chávez-Bosquez, O.; Hernández-Torruco, J. Supervised Deep Learning Techniques for Image Description: A Systematic Review. *Entropy* **2023**, *25*, 553. [[CrossRef](#)]
66. Tran, D.; Bourdev, L.; Fergus, R.; Torresani, L.; Paluri, M. Learning Spatiotemporal Features with 3D Convolutional Networks. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 4489–4497.
67. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17), Long Beach, CA, USA, 4–9 December 2017; pp. 6000–6010.
68. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
69. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the Computer Vision—ECCV 2018, Munich, Germany, 8–14 September 2018.
70. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 318–327. [[CrossRef](#)]
71. Wu, Z.; Wang, B.; Li, M.; Tian, Y.; Quan, Y.; Liu, J. Simulation of forest fire spread based on artificial intelligence. *Ecol. Indic.* **2022**, *136*, 108653. [[CrossRef](#)]
72. Linardatos, P.; Papastefanopoulos, V.; Kotsiantis, S. Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy* **2021**, *23*, 18. [[CrossRef](#)]
73. Khaire, U.M.; Dhanalakshmi, R. Stability of feature selection algorithm: A review. *J. King Saud Univ.-Comput. Inf. Sci.* **2022**, *34*, 1060–1073. [[CrossRef](#)]

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.