atmosphere

MDPI

Article

Forecasting of Local Lightning Using Spatial–Channel-Enhanced Recurrent Convolutional Neural Network

Wei Zhou, Jinliang Li *, Hongjie Wang, Donglai Zhang and Xupeng Wang

School of Information and Control Engineering, Qingdao University of Technology, 777 Jialingjiang Road, Huangdao District, Qingdao 266520, China; zhouwei@qut.edu.cn (W.Z.)
* Correspondence: lijinliang@stu.qut.edu.cn; Tel.: +86-1766-038-0948

Abstract: Lightning is a hazardous weather phenomenon, characterized by sudden occurrences and complex local distributions. It poses significant challenges for accurate forecasting, which is crucial for public safety and economic stability. Deep learning methods are often better than traditional numerical weather prediction (NWP) models at capturing the spatiotemporal predictors of lightning events. However, these methods struggle to integrate predictors from diverse data sources, which leads to lower accuracy and interpretability. To address these challenges, the Multi-Scale Spatial–Channel-Enhanced Recurrent Convolutional Neural Network (SCE-RCNN) is proposed to improve forecasting accuracy and timeliness by utilizing multi-source data and enhanced attention mechanisms. The proposed model incorporates a multi-scale spatial–channel attention module and a cross-scale fusion module, which facilitates the integration of data from diverse sources. The multi-scale spatial–channel attention module utilizes a multi-scale convolutional network to extract spatial features at different spatial scales and employs a spatial–channel attention mechanism to focus on the most relevant regions for lightning prediction. Experimental results show that the SCE-RCNN model achieved a critical success index (CSI) of 0.83, a probability of detection (POD) of 0.991, and a false alarm rate (FAR) reduced to 0.351, outperforming conventional deep learning models across multiple prediction metrics. This research provides reliable lightning forecasts to support real-time decision-making, making significant contributions to aviation safety, outdoor event planning, and disaster risk management. The model's high accuracy and low false alarm rate highlight its value in both academic research and practical applications.

Keywords: lightning forecasting; recurrent convolutional neural network; multi-scale feature extraction; spatial–channel attention mechanism; hazardous weather events

Citation: Zhou, W.; Li, J.; Wang, H.; Zhang, D.; Wang, X. Forecasting of Local Lightning Using Spatial–Channel-Enhanced Recurrent Convolutional Neural Network. *Atmosphere* **2024**, *15*, 1478. https://doi.org/10.3390/atmos15121478

Academic Editor: Kreso Pandzic and Tanja Likso

Received: 11 November 2024
Revised: 9 December 2024
Accepted: 10 December 2024
Published: 11 December 2024

1. Introduction

Lightning is a critical atmospheric phenomenon with widespread impacts on public safety, transportation, agriculture, aviation, and energy infrastructure. Its sudden and intense nature necessitates accurate forecasting to mitigate risks since lightning strikes can lead to wildfires, infrastructure damage, service disruptions, and significant economic losses [1–4]. Precise and timely lightning forecasts enable emergency response teams, aviation operators, and utility companies to implement proactive measures that reduce potential damages, protect human life, and improve public safety [5–8].

Traditional lightning forecasting methods generally include NWP models and data-driven machine learning approaches [9–12]. NWP models rely on complex physical simulations to provide large-scale atmospheric predictions that capture general weather patterns reliably. However, these models face challenges in delivering high-resolution, short-term forecasts essential for capturing lightning's localized and transient nature, given their computational requirements [13,14]. Operational constraints further limit NWP models in real-time applications due to the substantial computational resources they demand.

In contrast, data-driven machine learning methods, especially deep learning methods, has shown promise in using large historical datasets to identify complex patterns

and facilitate near-term predictions. Architectures such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are effective in modeling the spatiotemporal dependencies inherent in weather systems, making them suitable for lightning nowcasting [15–19]. However, traditional deep learning models often emphasize temporal over spatial features, limiting their ability to predict localized phenomena such as lightning. Additionally, integrating features from heterogeneous data sources remains challenging, like radar, satellite, and environmental sensors, thus impacting both model performance and interpretability.

To overcome these limitations, the Multi-Scale Spatial–Channel-Enhanced Recurrent Convolutional Neural Network (SCE-RCNN) is introduced to improve both the accuracy and timeliness of lightning forecasting. The model utilizes spatial and channel attention mechanisms [20–23], which enhance its ability to represent spatial distributions and temporal evolution in regions prone to lightning. By effectively integrating radar data with NWP-derived predictions, the SCE-RCNN extracts multi-scale spatial features and employs a dynamic gating mechanism to fuse them, enabling precise localization of target areas and improving training efficiency. This model is effective at capturing subtle spatial variations and addressing imbalances in multi-source data through its unique multi-scale spatial–channel attention mechanism.

The SCE-RCNN provides several key contributions that address current challenges in lightning forecasting. First, the model provides localized prediction capability by using multi-scale convolutional kernels (e.g., 3×3, 5×5, 7×7), which are specifically designed to capture different spatial features in the data. Smaller kernels, such as 3×3, capture fine-grained local changes and high-frequency details, while larger kernels, like 7×7, focus on broader spatial dynamics and low-frequency global patterns. The intermediate size, 5×5, bridges the gap by capturing mid-range spatial dependencies, enabling the model to understand both local and global interactions. This multi-scale approach balances computational efficiency and spatial coverage, allowing the model to fully utilize complementary spatial information. This capability is essential for modeling the sudden and spatially confined nature of lightning. Additionally, the cross-scale cooperative fusion (CSCF) module integrates features across different spatial scales, facilitating complementary interactions among features at various resolutions. This integration enhances the model's ability to accurately capture complex spatial patterns, further improving its performance in local and large-scale forecasting. Second, it incorporates an enhanced feature fusion mechanism, where a spatial–channel attention mechanism dynamically prioritizes critical features from diverse data sources. This not only enhances the model's interpretability but also enables it to focus on the most relevant information, improving prediction quality. Finally, the model achieves improved predictive accuracy by combining spatial and channel attention mechanisms. This approach results in superior predictive performance, particularly under complex and rapidly evolving weather conditions, where the model is better able to focus on the most critical features for accurate forecasting.

In summary, our focus is on how to better address the challenge of weather and climate prediction, and we hope that our work can contribute to improving lightning prediction. By addressing limitations of existing forecasting approaches, this model provides an effective solution for improving public safety and mitigating the economic impacts of lightning-related events.

2. Related Work

Lightning forecasting has traditionally relied on NWP models, such as the European Centre for Medium-Range Weather Forecasts (ECMWF) and the Weather Research and Forecasting (WRF) model developed by the U.S. National Center for Atmospheric Research (NCAR) [24]. These models are effective at capturing large-scale atmospheric phenomena, such as frontal systems and tropical cyclones, providing essential meteorological context for lightning prediction [25,26]. However, their coarse spatial resolution, typically several kilometers, limits their ability to accurately represent the localized and transient characteristics

of lightning events. For example, Lynn et al. [27] enhanced the Lightning Potential Index (LPI) within the WRF model to predict hourly lightning flash density at a grid resolution of 1.33 km, demonstrating its applicability in high-resolution scenarios. Despite these efforts, simulated flash counts showed a weak correlation with observational data, highlighting challenges in operational real-time applications due to computational demands and data quality limitations. These constraints point to the need for alternative methods capable of improving both the precision and efficiency of lightning forecasting.

To address the limitations of NWP models, data-driven approaches, particularly deep learning, have gained attention for their ability to leverage extensive historical datasets to identify complex meteorological patterns and enable near-real-time predictions [28,29]. Deep learning models, such as CNNs and RNNs, are frequently used due to their complementary strengths in capturing spatial and temporal dependencies within meteorological data [30–32]. CNNs, for instance, excel at extracting spatial features from grid-based datasets, such as radar or satellite observations. For example, Sebastian Brodehl et al. [33] developed a CNN lightning prediction framework utilizing geostationary satellite images. Inspired by U-Net [34] and enhanced with ResNet-v2 [35] residual blocks, their model effectively captured spatial lightning patterns across three-dimensional datasets, including temporal sequences, leading to an improvement in prediction accuracy. However, CNNs primarily focus on spatial features and are less effective in modeling the temporal evolution of lightning events.

RNNs, particularly long short-term memory (LSTM) networks, complement CNNs by excelling at modeling temporal sequences [36–38]. For instance, Geng et al. (2019) [39] introduced a hybrid model, LightNet, combining CNNs and LSTM to predict lightning occurrences. In this framework, the CNNs extracted spatial features, while LSTM modeled temporal dependencies based on radar observations. This hybrid approach showed better performance than standalone CNNs or RNNs. However, traditional RNN architectures often encounter issues such as vanishing or exploding gradients when processing large-scale spatiotemporal datasets, which limits their scalability and robustness.

To overcome these challenges, recent studies have explored advanced hybrid models that integrate multi-scale feature extraction and attention mechanisms [40,41]. For example, U-Net architectures embedded within recurrent convolutional neural networks (RCNNs) mitigate gradient issues and enhance temporal modeling capabilities. However, these RCNNs often rely on ResNet structures for spatial feature extraction, which lack effective mechanisms for multi-scale feature interaction and fail to comprehensively capture the complex spatiotemporal patterns required for lightning forecasting [10].

Building on these advancements, we propose a novel model, the SCE-RCNN, to address these limitations. At its core, SCE-RCNN introduces a multi-scale spatial–channel attention mechanism, which is designed to enhance the model's ability to capture complex spatiotemporal patterns in lightning forecasting. This mechanism comprises two key components: the Intra-Scale Joint Attention (ISJA) module and the CSCF module. The ISJA module operates within individual scales by integrating spatial and channel attention to refine feature representations. It dynamically identifies critical spatial regions and assigns adaptive weights to feature channels, allowing the model to prioritize relevant aspects of meteorological data efficiently. On the other hand, the CSCF module facilitates interactions across multiple scales, enabling the integration of fine-grained local features with broader global context. By using a query-key-value attention framework, the CSCF module balances contributions from different scales, capturing both localized lightning activity and large-scale atmospheric patterns. Together, these modules form the unified multi-scale spatial–channel attention mechanism, which equips SCE-RCNN with the capacity to model hierarchical meteorological phenomena more effectively, offering incremental improvements in the accuracy of lightning forecasting.

3. Lightning Forecasting System

This section provides a comprehensive overview of the lightning forecasting system, designed to harness multi-source data through a structured workflow, from data acquisition to forecast generation. The system utilizes diverse datasets to capture various atmospheric factors contributing to lightning formation, enhancing the depth and precision of predictions. As shown in Figure 1, the system operates in four main stages, creating a pipeline for real-time lightning forecasting: data preprocessing, model training, feature extraction, and prediction generation. The forecasting system relies on a range of data inputs that each capture essential aspects of atmospheric behavior. The Digital Elevation Model (DEM) provides critical topographical information, helping the model to understand how landscape variations influence storm behavior. Complex terrains, such as mountains or coastal areas, can affect weather patterns significantly, impacting both storm intensity and movement. These topographical data are complemented by radar reflectivity measurements, which capture precipitation patterns and intensity, providing insight into storm cell structures. Radar data are crucial for identifying convective activity, a key factor in lightning prediction, as increased precipitation reflectivity often correlates with thunderstorm formation.

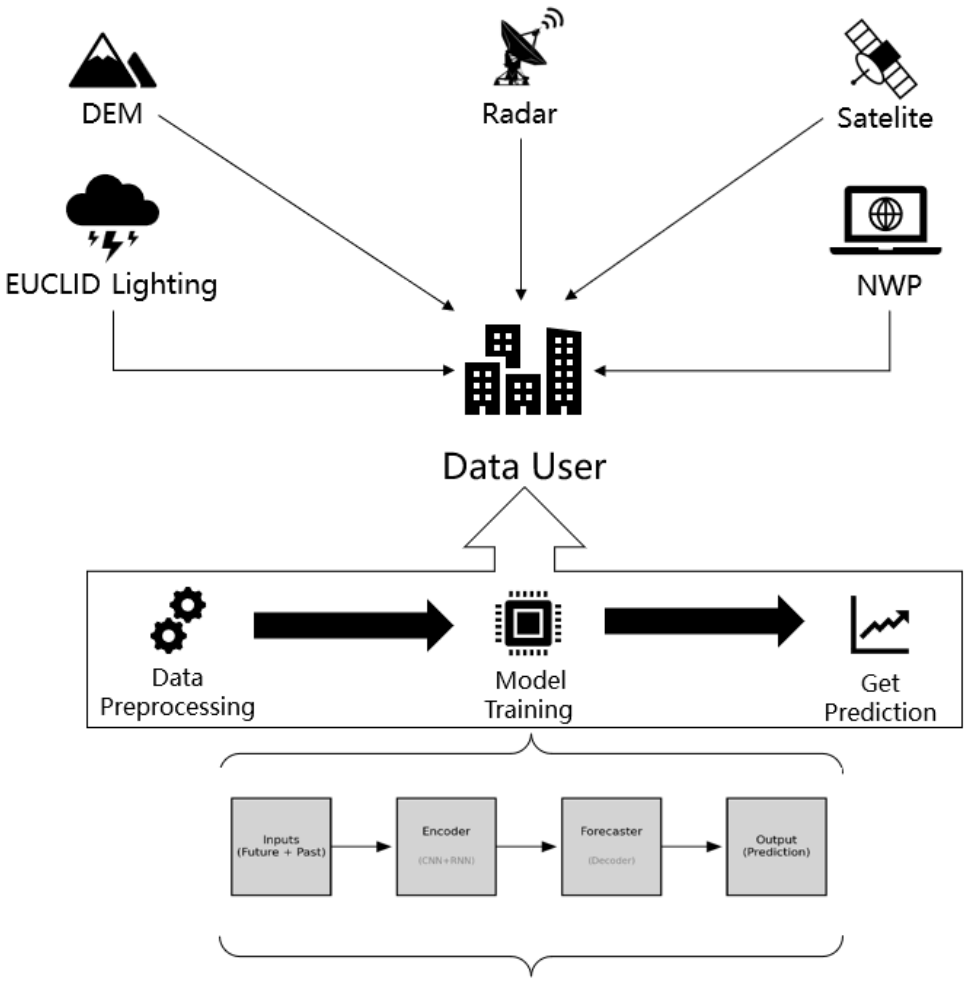

Figure 1. Model workflow diagram.

Satellite observations further enhance the system by supplying high-resolution data on cloud properties. Using instruments like Spinning Enhanced Visible and Infrared Imager (SEVIRI), which provide visible and infrared data, the model receives details on cloud-top height and thickness, both of which are strong indicators of the vigor of convective processes within the atmosphere. Taller, denser clouds generally signal stronger convective activity, suggesting a higher probability of lightning. Additionally, NWP data from models

such as Consortium for Small-scale Modeling (COSMO) add predictive power by supplying variables related to atmospheric instability, such as Convective Available Potential Energy (CAPE) and Convective Inhibition (CIN). CAPE quantifies the energy available for storm growth, while CIN represents the resistance to convection, helping the model to gauge the likelihood of thunderstorm development.

Ground-truth data from the European Cooperation for Lightning Detection (EUCLID) network, which records the timing and precise location of lightning events, plays a critical role in supervising the model during training. This dataset serves as a reliable benchmark, ensuring that the model aligns with observed lightning occurrences and thus improves the accuracy of future predictions. Utilizing such a diverse set of data sources poses challenges in terms of spatial and temporal consistency. Therefore, to ensure uniformity, all data are resampled to a common grid resolution of 1 km and aligned to a standardized temporal scale. This preprocessing step is essential for allowing the model to process the data coherently, enhancing feature compatibility across datasets.

Feature extraction is a pivotal part of the system, transforming raw data into meaningful inputs that capture the complexities of atmospheric conditions associated with lightning. From the DEM, the system derives topographic indicators to account for terrain impacts on storm formation and movement. Radar data contribute dynamic features such as reflectivity gradients and storm growth rates, capturing the evolving nature of storms linked to lightning events. Satellite data provide cloud characteristics, such as changes in cloud-top height and optical thickness, which are critical for assessing storm intensity and identifying the likelihood of convective lightning activity. Additionally, NWP data offer indices of atmospheric instability, helping the model to better understand environmental factors that may trigger lightning.

Following feature extraction, the system enters the model training phase, where a deep learning framework, such as an SCE-RCNN, is applied. During training, the model learns patterns within the extracted features, identifying correlations that are predictive of lightning events. The training process includes hyperparameter tuning, where parameters like learning rate, batch size, and network depth are optimized for maximum predictive accuracy and model stability. Attention mechanisms within the model dynamically focus on essential features, such as localized convective patterns, ensuring that critical aspects of each dataset are emphasized in predictions.

To capture complex spatial and temporal relationships, the SCE-RCNN employs multi-scale convolutional kernels that allow it to detect both local, fine-grained features and broader atmospheric trends. This cross-scale capability is crucial for recognizing the multi-dimensional characteristics of lightning, enabling the model to make more accurate predictions under varying weather conditions. By utilizing a cross-scale cooperative fusion module, the model facilitates feature interactions across scales, allowing information at each level to complement and reinforce predictions.

The final stage in the workflow involves generating real-time predictions based on the trained model. Outputs include the probability of lightning within specific time frames and geographic locations, providing actionable insights for stakeholders like emergency response teams, aviation operators, and utility companies. These forecasts empower decision-makers to implement timely precautions, minimizing the risks and damages associated with lightning.

System performance is evaluated through key metrics, including POD, FAR, and CSI. These metrics are used to assess the accuracy and reliability of the system, ensuring that it provides meaningful and actionable predictions. Extensive validation against historical lightning events enables a robust evaluation, allowing for ongoing refinements that further enhance the system's predictive capabilities and reliability.

In summary, this lightning forecasting system represents an advancement in predictive meteorology. By integrating multi-source data and using advanced deep learning techniques, it delivers high-resolution and timely forecasts. This system offers substantial improvements in forecast accuracy and interpretability, especially in diverse terrains

and complex atmospheric conditions, making it a valuable tool for proactive lightning risk management.

4. Multi-Scale Spatial–Channel Attention Mechanism

The recurrent convolutional model extracts spatial features via convolutional layers while capturing the temporal dynamics of lightning activity with RNN layers. Inspired by the U-Net architecture, it incorporates shortcut connections that enable efficient integration of spatial and temporal information. This design allows the model to process multi-source data, including radar, satellite remote sensing, ground-based lightning detection, and NWP data, facilitating accurate lightning forecasting. By combining local and global information across multiple resolutions, this architecture proves highly effective for multi-scale feature extraction in lightning prediction tasks.

To address the complexity of lightning activity and fully leverage multi-scale feature extraction, a multi-scale spatial–channel attention mechanism was developed and integrated into the base model. This module enhances the model's ability to capture intricate spatiotemporal features, thereby improving both interpretability and processing efficiency. The workflow of this module is illustrated in Figure 2.

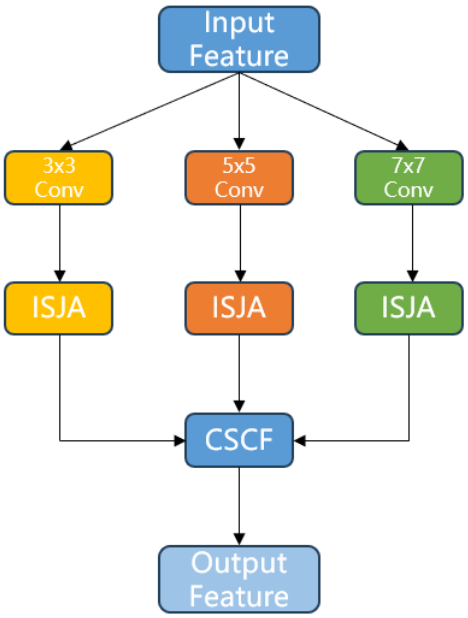

Figure 2. The workflow of the multi-scale spatial–channel attention mechanism.

This mechanism applies spatial and channel attention to convolutional features with kernel sizes of 3×3, 5×5, 7×7, enhancing the model's sensitivity to spatial and channel information across different scales. By allowing flexible weighting of multi-dimensional meteorological data, the model can prioritize key features more effectively. Recognizing the inherent relationships between spatial and channel dimensions, we designed an ISJA module to fuse spatial and channel features at the same scale, producing an initial fused feature representation. This approach promotes a deeper interaction of channel information within each scale, further refining feature accuracy.

To strengthen multi-scale fusion, we introduced a CSCF module that consolidates joint attention outputs from various scales. This preserves critical features at each level, reduces information loss, and supports a balanced fusion of global and local information. The progressive fusion approach enables the model to better adapt to complex terrains and diverse weather conditions, improving the precision and robustness of lightning forecasting. This multi-scale spatial–channel attention mechanism is a key innovation in our approach, highlighting its strong potential for practical lightning prediction applications.

4.1. Intra-Scale Joint Attention

The ISJA module integrates spatial and channel information within a single scale to enhance lightning prediction accuracy. The channel attention component dynamically assigns weights to each channel, identifying the most relevant meteorological data channels associated with lightning events. Recognizing that each type of meteorological datum represents distinct physical characteristics, this module automatically prioritizes channels that offer the most valuable insights for lightning prediction.

The spatial attention component, in contrast, emphasizes spatial regions linked to lightning activity by analyzing the spatial distribution within the input feature map. Given that lightning often follows specific regional patterns, this module enables the model to focus selectively on these critical areas while filtering out irrelevant information. The workflow of this module is shown in Figure 3.

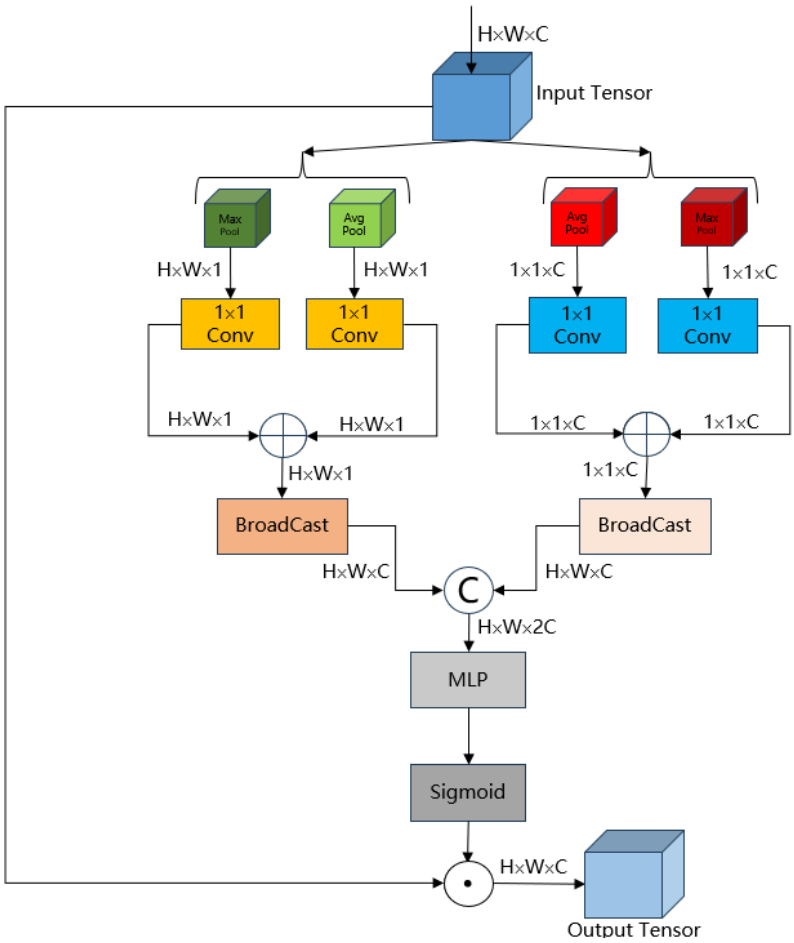

Figure 3. Intra-scale joint attention module.

4.1.1. Compressed Features

In the intra-scale joint attention module, the primary function of the compression operation is to perform global pooling on the input feature map, capturing global information in the form of channel and spatial descriptors. By compressing along different dimensions, the model extracts essential features related to both channels and spatial structure, providing effective global context information for subsequent attention mechanisms.

To generate global spatial features, we apply Global Average Pooling (GAP) and Global Max Pooling (GMP) along the spatial dimension of the input feature map. Let the input feature map be $X \in \mathbb{R}^{H \times W \times C}$, where H and W represent the height and width of the feature map, respectively, and C represents the number of channels. The calculation formula of average value is listed as follows:

$$X_{\text{avg}}^{c} = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_{i,j,c} \tag{1}$$

where, $X_{\text{avg}}^{c} \in \mathbb{R}^{1\times1\times C}$ represents the average value of the c-th channel, reflecting the mean activation strength of that channel across the entire spatial extent. The calculation formula of maximum value is listed as follows:

$$X_{\text{max}}^{c} = \max_{i=1,\ldots,H;j=1,\ldots,W} X_{i,j,c} \tag{2}$$

where, $X_{\text{max}}^{c} \in \mathbb{R}^{1\times1\times C}$ represents the maximum value of the c-th channel, capturing the most significant activation strength within that channel. In the multi-scale spatial–channel attention mechanism, to meet the real-time requirements and computational constraints of lightning nowcasting, we use a 1×1 convolution in place of the traditional multilayer perceptron (MLP) for processing global spatial features X_{avg} and X_{max}. The 1×1 convolution effectively generates channel weights while reducing computational complexity, making it more suitable for real-time applications. The global spatial features X_{avg} and X_{max} are each passed through 1×1 convolutions to obtain the corresponding weight matrices W_{avg} and W_{max}:

$$W_{\text{avg}} = \text{Conv}1 \times 1(X_{\text{avg}}) \tag{3}$$

$$W_{\text{max}} = \text{Conv}1 \times 1(X_{\text{max}}) \tag{4}$$

The weight matrices W_{avg} and W_{max} are combined through element-wise addition to obtain the final channel attention weight matrix X_{channel}:

$$X_{\text{channel}} = W_{\text{avg}} + W_{\text{max}} \tag{5}$$

In the intra-scale joint attention module, global channel features are generated by applying global pooling operations along the channel dimension of the input feature map. This process extracts global feature information for each spatial location across all channels, which is then used to generate attention weights for the subsequent spatial attention step. Specifically, global average pooling and global max pooling are applied to compress the input feature map, producing the global channel features. Let the input feature map be $X \in \mathbb{R}^{H\times W\times C}$, where H and W represent the height and width of the feature map, respectively, and C denotes the number of channels. The global channel features are calculated as follows:

$$X_{\text{avg}}^{(i,j)} = \frac{1}{C} \sum_{c=1}^{C} X_{i,j,c} \tag{6}$$

where $X_{\text{avg}}^{(i,j)}$ represents the average activation value of all channels at each spatial location (i, j). This average value represents the overall channel response at that spatial position.

$$X_{\text{max}}^{(i,j)} = \max_{c=1,\ldots,C} X_{i,j,c} \tag{7}$$

where $X_{\text{max}}^{(i,j)}$ represents the maximum activation value across all channels at each spatial location. This value highlights the strongest activation signal for each spatial position, thereby capturing significant spatial regions.

Through global average pooling and max pooling, the resulting global channel features X_{avg} and X_{max} represent the average and maximum channel activations for each spatial location, respectively, providing global information for each position. This helps the attention mechanism focus on spatial regions with key roles.

Similar to the treatment of global spatial features, we use 1×1 convolution to process the features here as well. We input X_{avg} and $X_{\max}$ into 1×1 convolution to obtain the corresponding weight matrices $W_{\text{avg}}^{\text{spatial}}$ and $W_{\max}^{\text{spatial}}$:

$$W_{\text{avg}}^{\text{spatial}} = \text{Conv}1 \times 1(X_{\text{avg}}) \tag{8}$$

$$W_{\max}^{\text{spatial}} = \text{Conv}1 \times 1(X_{\max}) \tag{9}$$

In the intra-scale joint attention module, the weight matrices $W_{\text{avg}}^{\text{spatial}}$ and $W_{\max}^{\text{spatial}}$ are also element-wise added to generate the final global spatial feature:

$$X_{\text{spatial}} = W_{\text{avg}}^{\text{spatial}} + W_{\max}^{\text{spatial}} \tag{10}$$

This combined feature effectively captures the critical spatial characteristics derived from both the average and maximum activations, helping to guide the attention mechanism in highlighting the most relevant regions for the model's prediction.

4.1.2. Intra-Scale Fusion

In the intra-scale fusion module, we designed a gating mechanism to dynamically balance the contribution between the global spatial and global channel features. Since spatial and channel features often express different levels of information, simply concatenating them may amplify irrelevant features, introducing noise—especially in complex or unevenly distributed data. The gating mechanism allows us to balance these two types of information, ensuring the model simultaneously focuses on both without being overwhelmed by noise.

The gating mechanism also enhances model interpretability by dynamically adjusting the weights of channel and spatial features, making the decision-making process more transparent. This helps us understand if the model is focusing more on channel or spatial features in different scenarios. Given that the global spatial feature has dimensions $1 \times 1 \times C$, while the global channel feature has dimensions $H \times W \times 1$, they cannot be directly concatenated. To solve this issue, we broadcast both descriptors to the same shape of $H \times W \times C$:

$$X_{\text{channel\_expanded}} = X_{\text{channel}} \text{ (broadcasted to) } H \times W \times C \tag{11}$$

$$X_{\text{spatial\_expanded}} = X_{\text{spatial}} \text{ (broadcasted to) } H \times W \times C \tag{12}$$

To fuse the expanded global spatial and channel features, we designed a gating weight matrix G to control their contributions. Specifically, the expanded features are concatenated and passed through an MLP, followed by a Sigmoid activation function to generate the gating weights in the range [0, 1]:

$$G = \sigma(\text{MLP}(\text{Concat}(X_{\text{channel\_expanded}}, X_{\text{spatial\_expanded}}))) \tag{13}$$

where $G \in \mathbb{R}^{H \times W \times C}$ represents the contribution ratio for each spatial position and channel. The Sigmoid activation ensures that G values are between 0 and 1, allowing the model to automatically adjust the balance between channel and spatial features. We then generate the fused feature X_{fused}:

$$X_{\text{fused}} = G \cdot X_{\text{channel\_expanded}} + (1 - G) \cdot X_{\text{spatial\_expanded}} \tag{14}$$

Finally, the fused feature X_{fused} is applied as an attention weight to the original input feature map X, producing the final output X_{output}:

$$X_{\text{output}} = X \odot X_{\text{fused}} \tag{15}$$

This approach effectively fuses spatial and channel information, ensuring that the most relevant features are highlighted, leading to more accurate predictions.

4.2. Cross-Scale Cooperative Fusion

The complexity of lightning forecasting tasks stems primarily from the highly dynamic spatiotemporal characteristics of lightning, which cannot be effectively captured by a single scale capable of handling both local details and global features simultaneously. Existing multi-scale approaches, such as multi-branch inception networks and bilateral information exchange modules, have made progress in multi-scale feature extraction and interaction. However, these methods often either concatenate multi-scale features along the channel dimension or restrict feature exchange within a single scale, limiting their ability to fully exploit interactions between features across scales. To address these limitations, we propose the CSCF module, which introduces a novel mechanism for cooperative fusion between features at different scales. This module dynamically adjusts and fuses global and local information, thereby enhancing the model's capability to capture the complex spatiotemporal patterns of weather phenomena. The core concept of the CSCF module is to enable cooperative fusion across multiple scales, allowing features to mutually enhance each other and progressively fuse into a unified representation. The workflow of this module is illustrated in Figure 4.

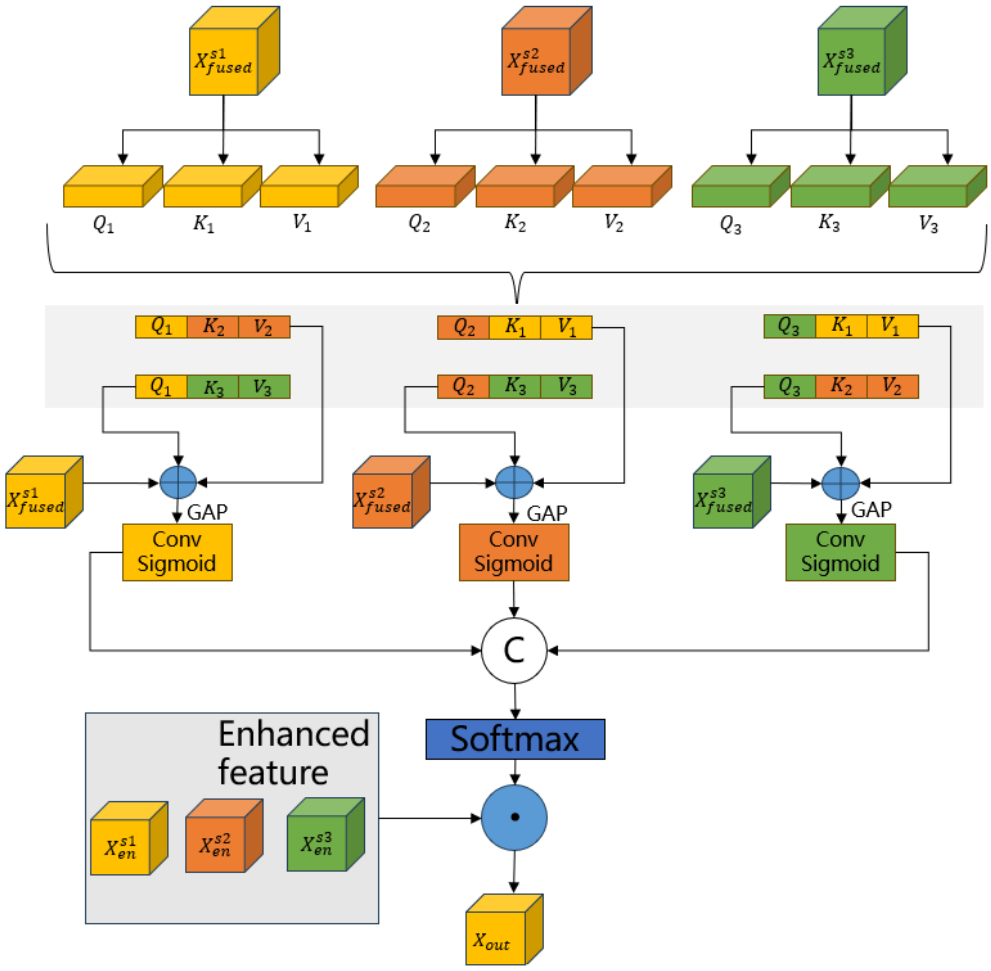

Figure 4. Cross-scale cooperative fusion module.

Through the intra-scale joint attention module, we obtain spatial–channel-fused features at three different scales, denoted as X_{fused}^{s1}, X_{fused}^{s2}, and X_{fused}^{s3}. We map the features of each scale into query, key, and value spaces using an attention mechanism to compute the

similarity and feature reconstruction between different scales. For scale $s1$, we define the following linear transformations to generate the query, key, and value:

$$Q_1 = W_Q X_{fused}^{s1}, \quad K_1 = W_K X_{fused}^{s1}, \quad V_1 = W_V X_{fused}^{s1} \tag{16}$$

Here, Q_{s1} is the query vector for scale $s1$, representing the information that small-scale features seek to obtain from other scales; K_{s1} is the key vector at the corresponding scale, representing the information it can provide; and V_{s1} is the value vector at this scale, representing the actual features provided. Similarly, we can perform the same operations for the remaining scales $s2$ and $s3$.

After obtaining the query, key, and value representations, we perform cross-scale dot-product attention to compute the similarity and exchange information between different scales. For interaction between scales si and sj, the attention calculation is given by

$$X_{\text{inter}}^{si,sj} = \text{Attention}(X_{fused}^{si}, X_{fused}^{sj}) = \text{Softmax}\left(\frac{Q_i K_j^T}{\sqrt{d_k}}\right) V_j \tag{17}$$

This fusion mechanism allows features from different scales to exchange information. Through this mechanism, each scale can acquire complementary information from others, such as small scales obtaining global information from large scales and vice versa. After completing cross-scale interaction, we need to combine the interaction results of each scale to generate enhanced features for each scale. For scale $s1$, the enhanced feature is expressed as

$$X_{\text{enhanced}}^{s1} = \alpha_{s1} X_{fused}^{s1} + \beta_{s1,s2} X_{\text{inter}}^{s1,s2} + \gamma_{s1,s3} X_{\text{inter}}^{s1,s3} \tag{18}$$

Here, α_{s1}, $\beta_{s1,s2}$, and $\gamma_{s1,s3}$ are learnable weight parameters used to control the contribution of each scale to the final feature.

Similarly, the enhanced representations for scales $s2$ and $s3$ are

$$X_{\text{enhanced}}^{s2} = \alpha_{s2} X_{fused}^{s2} + \beta_{s2,s1} X_{\text{inter}}^{s2,s1} + \gamma_{s2,s3} X_{\text{inter}}^{s2,s3} \tag{19}$$

$$X_{\text{enhanced}}^{s3} = \alpha_{s3} X_{fused}^{s3} + \beta_{s3,s1} X_{\text{inter}}^{s3,s1} + \gamma_{s3,s2} X_{\text{inter}}^{s3,s2} \tag{20}$$

Subsequently, we perform global average pooling on the enhanced features of the three scales to obtain global feature representations, then model the inter-channel correlations through a 1×1 convolution, and generate global spatial features through a Sigmoid activation function:

$$F_{si} = \text{Sigmoid}(\text{Conv}(\text{GAP}(X_{\text{enhanced}}^{si}))) \tag{21}$$

After obtaining the three global spatial features, we concatenate them along the second dimension to obtain F_{concat}, apply the Softmax function along the second dimension to obtain their respective weight representations, and finally perform element-wise multiplication with the corresponding inputs and sum them to obtain the final X_{output}:

$$F_{concat} = \text{Concat}(F_{s1}, F_{s2}, F_{s3}) \tag{22}$$

$$W = \text{Softmax}(F_{concat}) \tag{23}$$

$$X_{output} = W[:0:] \odot X_{\text{enhanced}}^{s1} + W[:1:] \odot X_{\text{enhanced}}^{s2} + W[:2:] \odot X_{\text{enhanced}}^{s3} \tag{24}$$

5. Experiment

5.1. Experimental Setup

Following the setup of the baseline model, we configure model training to predict the next $N_f = 12$ time steps using past $N_p = 6$ time steps and NWP data. Specifically, our training scheme is designed to forecast lightning occurrences over the next $N_f = 12$ time steps based on $N_p = 6$ historical time steps and NWP data. With a time resolution of 5 min, this setup uses data from the past 30 min to predict lightning activity over the next 60 min.

During each training epoch, the data generator randomly traverses the starting time points of all training sequences. For each starting point, only one training sample is generated per epoch, selected at random from all possible options. This approach minimizes overlap between training samples, reducing the risk of model overfitting to specific large-scale convective patterns.

The model training was conducted on a server equipped with NVIDIA Tesla V100 GPUs, meeting our computational requirements. Under this hardware environment, each training epoch takes approximately 60 min. To manage the training duration, we employed an early stopping strategy: if validation loss shows no significant improvement over three consecutive epochs, the learning rate is reduced to one-fifth of its current value; if no improvement occurs over six consecutive epochs, training is stopped, and the model parameters with the lowest validation loss are saved. Although each training epoch is lengthy, loading the model on the same device for testing and generating results takes only about 20 s, excluding dataset loading time.

5.2. Datasets and Models

The dataset used in this study was sourced from the Swiss Meteorological Office (MeteoSwiss) and the EUCLID. It includes various data types such as lightning observations, radar reflectivity, satellite remote sensing imagery, NWP data, and topographic data. The dataset covers Central and Western Europe, with a particular focus on Switzerland and the surrounding regions, spanning approximately 710 km east–west and 640 km north–south. Data collection occurred between April and September of 2020, during which time the data were selectively downsampled to prioritize areas and periods with active convective weather, ensuring the dataset's relevance and effectiveness for studying dynamic weather patterns. The selection of this region for the study was based on the high frequency of thunderstorms, particularly in the Alpine and Central European regions, where lightning occurrence rates are among the highest in Europe. The most intense thunderstorm activity typically occurs between April and August each year [42]. Additionally, the peak thunderstorm season in northern, eastern, and central Europe occurs in July and August, while the peak in western and southeastern Europe is observed in May and June. These characteristics make this region particularly suitable for studying thunderstorms and related convective weather processes.

The lightning observation data consist of ground-based measurements from the EUCLID network, and a high-precision detection system with a temporal resolution of 5 min and a spatial resolution of 1 km, capturing lightning density and intensity in the target area. The radar data, sourced from C-band dual-polarization Doppler radar, provide reflectivity information that details precipitation intensity and spatial distribution within the region. These radar data have been standardized and horizontally projected to align with the other data sources. Satellite remote sensing data include meteorological observations over Central and Western Europe, primarily from the SEVIRI instrument on the Meteosat Second Generation-3 (MSG-3) satellite, covering multiple visible and infrared bands. The NWP data come from the COSMO model, offering meteorological predictors such as CAPE and CIN. Additionally, topographic data, in the form of high-resolution DEM data, are used to account for the impact of terrain on convective processes. All data were resampled to a 1-km grid and normalized for consistency, with the detailed data selection and processing methods described in [10].

For model validation, we compared the proposed method with baseline models selected from [10]. The Eulerian Persistence Model serves as the simplest baseline, assuming that the spatial distribution and location of lightning activity remain constant over time steps. In other words, it assumes that lightning positions and intensities observed in one time step will persist into the next, without accounting for the movement or evolution of the lightning systems. The Lagrangian Persistence Model, an extension of the Eulerian model, updates the lightning field's position by extrapolating its motion. This is achieved through the Lucas–Kanade method using RZC radar data, which estimates the movement

of the lightning field and applies it to the previous time step's lightning distribution to predict activity in the subsequent time step.

Additionally, to demonstrate the superiority of our model, we included a state-of-the-art model for lightning forecasting as a comparison. This model combines CNNs and RNNs to capture the spatiotemporal characteristics of lightning by processing multi-source data and incorporates shortcut connections, similar to a U-Net structure, between the encoder and decoder.

These models were used as baselines in our experiments to evaluate the improvements offered by the proposed model in lightning forecasting tasks.

5.3. Experiment Results

Figure 5 presents the short-term forecast results of lightning activity within the next hour, generated using the SCE-RCNN model.

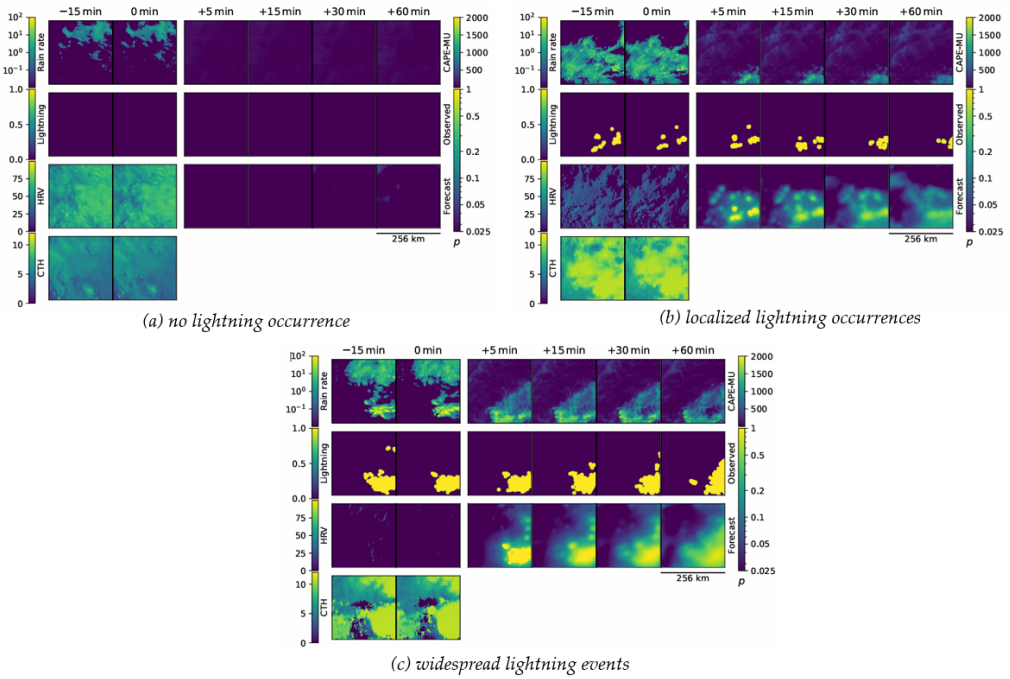

Figure 5. Predicted lightning for the next hour across regions and intensities.

Figure 5 comprises three sub-figures illustrating distinct spatial scenarios: (a) no lightning occurrence, (b) localized lightning occurrences, and (c) widespread lightning events. Each sub-figure demonstrates changes in rain rate, lightning, high-resolution visible (HRV) imagery, and cloud-top height (CTH) across multiple time points (from −15 min to +60 min). By jointly predicting these variables, the model effectively simulates the development and progression of thunderstorms and related meteorological phenomena under varying conditions.

Sub-figure (a) showcases the model's ability to predict the absence of lightning activity accurately, confirming its reliability in stable weather conditions. Sub-figure (b) highlights the model's capability to capture the spatial distribution and intensity of localized lightning occurrences, aligning well with observed data. This includes identifying changes such as gradual increases in rain rate and shifts in cloud structures, as depicted in the HRV imagery. Sub-figure (c) demonstrates the model's performance under a more complex scenario involving widespread lightning activity. The forecast maps effectively reflect both the spatial extent and intensity of lightning, emphasizing the model's robustness in dynamic and challenging weather situations.

By incorporating multi-scale convolutional kernels and spatial–channel attention mechanisms, the model captures key features at different spatial scales, effectively pre-

dicting the probability of lightning occurrence. Additionally, the cross-scale cooperative fusion module facilitates information exchange across different scales, ensuring high prediction accuracy and reliability under complex weather conditions. In the forecast map, the probability of lightning occurrence is indicated by color intensity, with brighter areas corresponding to higher probabilities. The model dynamically adjusts these probability outputs to achieve more accurate short-term lightning forecasts.

However, the model's predictive performance declines noticeably as forecast time increases. We use the CSI as the evaluation metric, as it effectively reflects the model's ability to accurately predict lightning events. Detailed CSI data are shown in Figure 6.

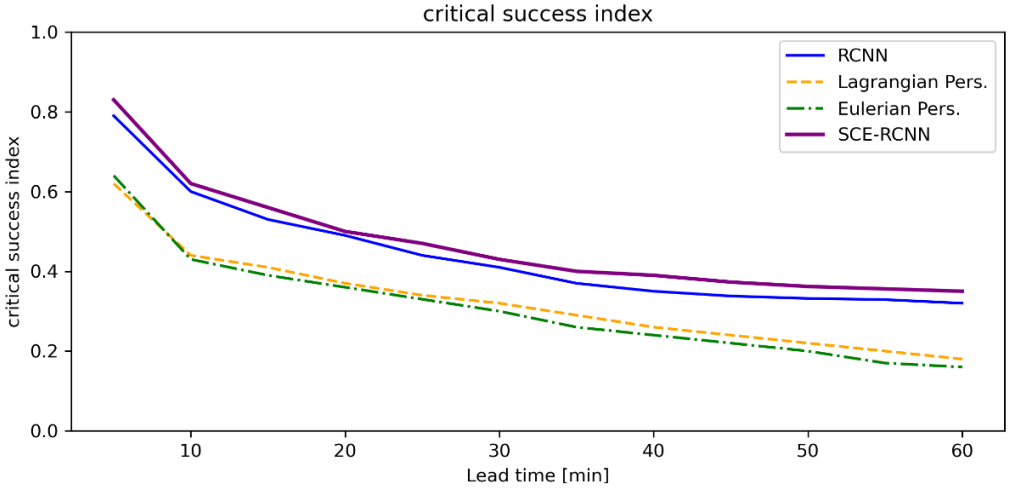

Figure 6. Critical success index (CSI) over time for different models.

This phenomenon indicates that although the SCE-RCNN model performs well in the short term (e.g., +5 min and +15 min), its accuracy faces challenges over longer time frames (e.g., +30 min and +60 min). However, compared to other models, the rate of decline is slightly mitigated, primarily due to the highly random and complex nature of convective weather phenomena like lightning. As time progresses, the uncertainty of weather systems increases, leading to a decline in predictive performance. Nonetheless, due to more comprehensive feature extraction and an emphasis on potential correlations across various data sources, our model experiences a less pronounced decline in accuracy over extended forecasts. Maintaining prediction accuracy at longer time scales remains a challenge that requires further research.

The performance of the SCE-RNN model was compared with the baseline models RCNN, Lagrangian, and Eulerian on the same test dataset to evaluate the effectiveness of the proposed model in lightning event prediction. The specific results are shown in Table 1, listing various performance metrics, including POD, FAR, CSI, Equitable Threat Score (ETS), Heidke Skill Score (HSS), Pierce Skill Score (PSS), Receiver Operating Characteristic–Area Under the Curve (ROC AUC), and Precision–Recall–Area Under the Curve (PR AUC).

Table 1. Performance metrics comparison of different models.

Model	T	POD	FAR	CSI	ETS	HSS	PSS	ROC AUC	PR AUC
RCNN	0.426	0.61	0.362	0.453	0.449	0.62	0.607	0.989	0.688
Lagrangian	nan	0.473	0.509	0.317	0.313	0.476	0.468	nan	nan
Eulerian	nan	0.439	0.581	0.273	0.268	0.422	0.432	nan	nan
SCE-RCNN	0.426	0.629	0.351	0.467	0.472	0.64	0.618	0.991	0.697

Using a consistent event occurrence threshold (T = 0.426), all models were trained to achieve their best possible performance. The SCE-RCNN model performed better than the baseline models across multiple evaluation metrics, including POD, FAR, CSI, and Area

Under the Curve (AUC). For example, compared to the RCNN model, the SCE-RCNN model improved the POD from 0.610 to 0.629, i.e., an increase of 3.1%, while the FAR decreased from 0.362 to 0.351, i.e., a reduction of 3.0%. These results show that the model is better at detecting lightning events and reducing false alarms, making it more reliable and stable in practical applications.

The improved performance of the SCE-RCNN model is mainly due to its updated architecture. The use of multi-scale convolutional kernels allows the model to extract features at different spatial levels. This helps it to identify both small local details and larger atmospheric patterns, leading to better detection accuracy. Additionally, the spatial–channel attention mechanism helps the model focus on key features for lightning prediction by dynamically assigning importance to different data channels and regions, which reduces the impact of noise and lowers the FAR.

Further improvements are reflected in metrics such as the ETS and CSI, which increased from 0.449 and 0.453 to 0.472 and 0.467, respectively. These scores indicate that the model has better overall predictive accuracy and reliability. The CSCF module plays an important role here, helping the model combine information from both global and local scales, improving its ability to understand complex weather patterns.

The robustness of the SCE-RCNN model is also evident in its higher HSS and PSS, which increased from 0.620 and 0.607 to 0.640 and 0.618, respectively. These metrics highlight the model's stronger classification accuracy under challenging weather conditions. Additionally, the ROC AUC and PR AUC metrics saw slight improvements, increasing from 0.989 to 0.991 and from 0.688 to 0.697, respectively. These gains show the model's improved ability to differentiate between lightning and non-lightning events across various thresholds.

Although the SCE-RCNN model has 15% more parameters than the RCNN model, it is more efficient during training. The SCE-RCNN model reached its best performance by epoch 21, while the RCNN model required 28 epochs. This efficiency is due to the SCE-RCNN's advanced feature extraction and integration strategies, which make the learning process faster and more effective.

In summary, the SCE-RCNN model shows clear advantages in lightning forecasting. It improves the accuracy and timeliness of predictions for localized and short-term weather events like lightning. By using multi-scale feature extraction, spatial–channel attention mechanism, and cross-scale fusion, the model achieves better detection rates and fewer false alarms, while maintaining reliability across different weather scenarios. These improvements demonstrate the model's potential for use in disaster prevention and weather forecasting, providing valuable support for risk management and emergency response.

5.4. Ablation Study

To evaluate the specific contributions of the proposed modules in enhancing the lightning forecasting model, we conducted detailed ablation experiments. The experimental design included three versions of the model: the baseline model (original RCNN), the baseline model augmented with the multi-scale spatial–channel attention mechanism (SC-RCNN), and the complete SCE-RCNN model. The results of the ablation experiments are summarized in Table 1, and the performance metrics of each model version were quantitatively analyzed using the POD, FAR, CSI, and ROC AUC. The specific results are shown in Figure 7.

To assess the specific contributions of the proposed modules in enhancing the performance of the lightning forecasting model, we performed a comprehensive ablation study. The experimental design included three model versions: the baseline model (original RCNN), the SC-RCNN model with the addition of the multi-scale spatial–channel attention mechanism to the baseline model, and the complete SCE-RCNN model. We used key metrics such as POD, FAR, CSI, and ROC AUC to quantitatively analyze the performance of each model version.

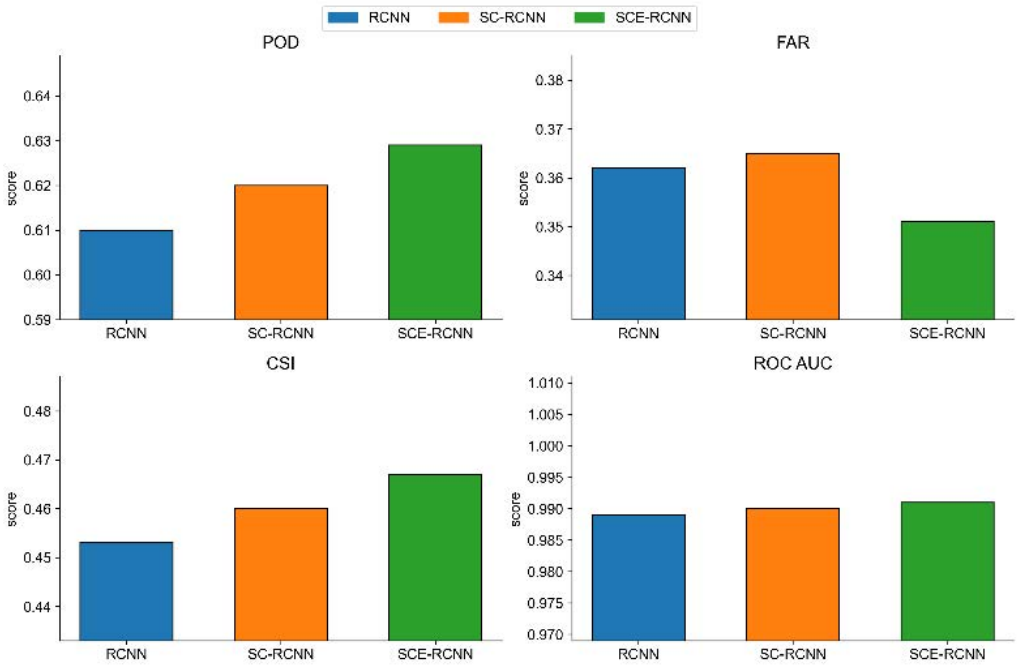

Figure 7. Performance metrics comparison among different model versions.

The experimental results show that the addition of the multi-scale spatial–channel attention mechanism improved the model's performance. Specifically, the POD of the SC-RCNN model increased from 0.610 in the baseline model to 0.620, indicating an enhanced capability of the model in accurately identifying lightning events. This improvement may be attributed to the attention module's ability to highlight key features and enhance the model's focus on important regions. However, the FAR slightly increased by 0.003 (from 0.362 to 0.365), possibly because the attention mechanism amplified some noise signals while emphasizing target features. The CSI improved from 0.453 to 0.460, indicating an overall increase in the model's predictive accuracy, and the ROC AUC increased from 0.989 to 0.990, suggesting an enhanced ability of the model to distinguish positive and negative samples.

Furthermore, the complete SCE-RCNN model achieved improvements across all metrics. The POD increased to 0.629, a 3.1% improvement over the baseline model, demonstrating the model's superior performance in capturing lightning events. This enhancement is attributed to the introduction of feature enhancement and multi-scale fusion strategies, which allow the model to extract and utilize multi-scale features more comprehensively. The FAR decreased to 0.351, indicating a noticeable reduction in false alarms and more reliable prediction results, likely due to the effective suppression of noise by the multi-scale fusion. The CSI improved to 0.467, reflecting an enhancement in the model's overall predictive capability. The ROC AUC reached 0.991, approaching an ideal state, further validating the model's discriminative power. These results demonstrate that the complete model, incorporating multi-scale attention, feature enhancement, and multi-scale fusion strategies, has an advantage in improving the accuracy and reliability of lightning forecasting, thus validating the practical value and superiority of the proposed method in lightning forecasting tasks.

6. Conclusions

The SCE-RCNN model improves the spatial and temporal resolution of lightning forecasts while reducing false alarm rates. By integrating multi-scale feature extraction, spatial–channel attention mechanism, and cross-scale fusion modules, the model captures spatiotemporal patterns and enhances detection performance. These improvements contribute to better stability and accuracy in lightning prediction, making the model applicable

to scenarios such as emergency management, aviation, and power systems. Experimental results indicate that the model provides measurable improvements over traditional deep learning models, demonstrating its potential for practical applications.

However, there is room for improvement. First, the multi-scale attention and fusion modules increase computational complexity, posing challenges for real-time deployment. Future work could focus on optimizing these modules to reduce redundant features and enhance efficiency. Second, the current model evaluation relies on specific meteorological datasets. Expanding validation to include globally diverse datasets across different climatic and regional conditions would help to improve the model's generalization and practical applicability.

Author Contributions: Conceptualization, J.L. under the guidance of W.Z.; methodology, J.L.; software, Windows; validation, J.L. and H.W.; formal analysis, J.L.; investigation, J.L.; resources, W.Z.; writing—original draft preparation, J.L.; writing—review and editing, X.W. and D.Z.; visualization, J.L.; supervision, W.Z. and X.W.; project administration, W.Z.; funding acquisition, not applicable. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the National Natural Science Foundation of China (61502262 and 42201506).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets, including the preprocessed training, validation, and testing datasets, along with the models and results, are based on the work of Leinonen et al. (2022a) [10] and are available for noncommercial use under the CC BY-NC-SA 4.0 license: https://doi.org/10.5281/zenodo.6802292, (accessed on 9 December 2024). The original data from the EUCLID lightning network are proprietary and cannot be shared in raw form. However, data from the Swiss radar network and the COSMO NWP model are available for research purposes upon request. Additional details can be found in the sources listed, including the EUMETSAT Data Store (https://data.eumetsat.int/), (accessed on 9 December 2024) and NWCSAF software (https://www.nwcsaf.org/), accessed on 9 December 2024.

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

References

1. Haidar, A.; Verma, B. Monthly Rainfall Forecasting Using One-Dimensional Deep Convolutional Neural Network. *IEEE Access* **2018**, *6*, 69053–69063. [CrossRef]
2. Holle, R.L. Some aspects of global lightning impacts. In Proceedings of the 2014 International Conference on Lightning Protection (ICLP), Shanghai, China, 1–18 October 2014; pp. 1390–1395. [CrossRef]
3. Holle, R.L. A Summary of Recent National-Scale Lightning Fatality Studies. *Weather Clim. Soc.* **2016**, *8*, 35–42. Available online: https://www.jstor.org/stable/24907479 (accessed on 8 October 2024). [CrossRef]
4. Battaglia, M.; Lee, C.; Thomason, W.; Fike, J.; Sadeghpour, A. Hail damage impacts on corn productivity: A review. *Crop Sci.* **2019**, *59*, 1–14. [CrossRef]
5. Maqsood, I.; Khan, M.R.; Abraham, A. An ensemble of neural networks for weather forecasting. *Neural Comput. Appl.* **2004**, *13*, 112–122. [CrossRef]
6. Xing, H.; Zhang, Q.; Ji, X. Study on Short-Term Forecasting of Thunderstorms Based on Ground Electric Field Data. *Acta Meteorol. Sin.* **2017**, *40*, 111–117.
7. Mo, C.; Wan, X.; Lei, X.; Chen, X.; Ma, R.; Huang, Y.; Sun, G. Hydrometeorological Insights into the Forecasting Performance of Multi-Source Weather over a Typical Hill-Karst Basin, Southwest China. *Atmosphere* **2024**, *15*, 236. [CrossRef]
8. Tan, X.; Ma, K.; Dou, F. A Convolutional Neural Network and Attention-Based Retrieval of Temperature Profile for a Satellite Hyperspectral Microwave Sensor. *Atmosphere* **2024**, *15*, 235. [CrossRef]
9. Sun, J.; Xue, M.; Wilson, J.W.; Zawadzki, I.; Ballard, S.P.; Onvlee-Hooimeyer, J.; Joe, P.; Barker, D.M.; Li, P.-W.; Golding, B. Use of NWP for nowcasting convective precipitation: Recent progress and challenges. *Bull. Am. Meteorol. Soc.* **2014**, *95*, 409–426. [CrossRef]
10. Leinonen, J.; Hamann, U.; Germann, U. Seamless lightning nowcasting with recurrent-convolutional deep learning. *Artif. Intell. Earth Syst.* **2022**, *1*, e220043. [CrossRef]
11. Lu, M.; Zhang, Y.; Ma, Z.; Yu, M.; Chen, M.; Zheng, J.; Wang, M. Lightning strike location identification based on 3D weather radar data. *Front. Environ. Sci.* **2021**, *9*, 714067. [CrossRef]

12. Song, G.; Li, S.; Xing, J. Lightning nowcasting with aerosol-informed machine learning and satellite-enriched dataset. *npj Clim. Atmos. Sci.* **2023**, *6*, 126. [CrossRef]
13. Price, C.; Rind, D. A simple lightning parameterization for calculating global lightning distributions. *J. Geophys. Res. Atmos.* **1992**, *97*, 9919–9933. [CrossRef]
14. Leinonen, J.; Hamann, U.; Sideris, I.V.; Germann, U. Thunderstorm Nowcasting With Deep Learning: A Multi-Hazard Data Fusion Model. *Geophys. Res. Lett.* **2023**, *50*, e2022GL101626. [CrossRef]
15. Pulkkinen, S.; Nerini, D.; Pérez Hortal, A.A.; Velasco-Forero, C.; Seed, A.; Germann, U.; Foresti, L. Pysteps: An open-source Python library for probabilistic precipitation nowcasting (v1. 0). *Geosci. Model Dev.* **2019**, *12*, 4185–4219. [CrossRef]
16. Rombeek, N.; Leinonen, J.; Hamann, U. Exploiting radar polarimetry for nowcasting thunderstorm hazards using deep learning. *Nat. Hazards Earth Syst. Sci.* **2024**, *24*, 133–144. [CrossRef]
17. Cuomo, J.; Chandrasekar, V. Developing deep learning models for storm nowcasting. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–13. [CrossRef]
18. Geng, Y.-A.; Li, Q.; Lin, T.; Yao, W.; Xu, L.; Zheng, D.; Zhou, X.; Zheng, L.; Lyu, W.; Zhang, Y. A deep learning framework for lightning forecasting with multi-source spatiotemporal data. *Q. J. R. Meteorol. Soc.* **2021**, *147*, 4048–4062. [CrossRef]
19. Lin, T.; Li, Q.; Geng, Y.-A.; Jiang, L.; Xu, L.; Zheng, D.; Yao, W.; Lyu, W.; Zhang, Y. Attention-based dual-source spatiotemporal neural network for lightning forecast. *IEEE Access* **2019**, *7*, 158296–158307. [CrossRef]
20. Ouyang, D.; He, S.; Zhang, G.; Luo, M.; Guo, H.; Zhan, J.; Huang, Z. Efficient multi-scale attention module with cross-spatial learning. In Proceedings of the ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing, Rhodes Island, Greece,4–10 June 2023; pp. 1–5.
21. Shao, H.; Zeng, Q.; Hou, Q.; Yang, J. MCANet: Medical Image Segmentation with Multi-Scale Cross-Axis Attention. *arXiv* **2023** arXiv:2312.08866.
22. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
23. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
24. Saleh, N.; Gharaylou, M.; Farahani, M.M.; Alizadeh, O. Performance of lightning potential index, lightning threat index, and the product of CAPE and precipitation in the WRF model. *Earth Space Sci.* **2023**, *10*, e2023EA003104. [CrossRef]
25. Skamarock, W.C.; Klemp, J.B.; Dudhia, J.; Gill, D.O.; Barker, D.M.; Duda, M.G.; Huang, X.-Y.; Wang, W.; Powers, J.G. A description of the advanced research WRF version 3. *NCAR Tech. Note* **2008**, *475*, 10–5065.
26. Al-Yahyai, S.; Charabi, Y.; Gastli, A. Review of the use of numerical weather prediction (NWP) models for wind energy assessment. *Renew. Sustain. Energy Rev.* **2010**, *14*, 3192–3198. [CrossRef]
27. Lynn, B.; Yair, Y. Prediction of lightning flash density with the WRF model. *Adv. Geosci.* **2010**, *23*, 11–16. [CrossRef]
28. Mostajabi, A.; Finney, D.L.; Rubinstein, M.; Rachidi, F. Nowcasting lightning occurrence from commonly available meteorological parameters using machine learning techniques. *npj Clim. Atmos. Sci.* **2019**, *2*, 41. [CrossRef]
29. Zeng, A.; Liu, S.; Yu, Y. Comparative study of data driven methods in building electricity use prediction. *Energy Build.* **2019**, *194*, 289–300. [CrossRef]
30. Cintineo, J.L.; Pavolonis, M.J.; Sieglaff, J.M. ProbSevere LightningCast: A deep-learning model for satellite-based lightning nowcasting. *Weather Forecast.* **2022**, *37*, 1239–1257. [CrossRef]
31. Bao, R.; Zhang, Y.; Ma, B.J.; Zhang, Z.; He, Z. An artificial neural network for lightning prediction based on atmospheric electric field observations. *Remote Sens.* **2022**, *14*, 4131. [CrossRef]
32. Speranza, D.V. Lightning Prediction Using Recurrent Neural Networks. Master's Thesis, Department of Operational Sciences, Air University, Wright-Patterson Air Force Base, OH, USA, 2019. Available online: https://scholar.afit.edu/etd/2317 (accessed on 9 December 2024).
33. Brodehl, S.; Müller, R.; Schömer, E.; Spichtinger, P.; Wand, M. End-to-End Prediction of Lightning Events from Geostationary Satellite Images. *Remote Sens.* **2022**, *14*, 3760. [CrossRef]
34. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015, 18th International Conference, Munich, Germany, 5–9 October 2015; Springer: Cham, Swizerland, 2015; pp. 234–241. [CrossRef]
35. He, K.; Zhang, X.; Ren, S.; Sun, J. Identity Mappings in Deep Residual Networks. In Proceedings of the Computer Vision–ECCV 2016, 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Cham, Swizerland, 2016; pp. 630–645. [CrossRef]
36. Leinonen, J. Spatiotemporal weather data predictions with shortcut recurrent-convolutional networks: A solution for the Weather4cast challenge. *arXiv* **2021** arXiv:2111.02121.
37. Zang, T.; Zou, J.; Li, Y.; Qiu, Z.; Wang, B.; Cui, C.; Li, Z.; Hu, T.; Guo, Y. Development and Evaluation of a Short-Term Ensemble Forecasting Model on Sea Surface Wind and Waves across the Bohai and Yellow Sea. *Atmosphere* **2024**, *15*, 197. [CrossRef]
38. Fukawa, M.; Deng, X.; Imai, S.; Horiguchi, T.; Ono, R.; Rachi, I.; Shinomura, K.; Niwa, S.; Kudo, T.; Ito, H.; et al. A Novel Method for Lightning Prediction by Direct Electric Field Measurements at the Ground Using Recurrent Neural Network. *IEICE Trans. Inf. Syst.* **2022**, *105*, 1624–1628. [CrossRef]

39. Geng, Y.; Li, Q.; Lin, T.; Jiang, L.; Xu, L.; Zheng, D.; Yao, W.; Lyu, W.; Zhang, Y. LightNet: A Dual Spatiotemporal Encoder Network Model for Lightning Prediction. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; ACM: New York, NY, USA, 2019; pp. 2439–2447. [CrossRef]
40. Hou, Q.; Zhou, D.; Feng, J. Coordinate attention for efficient mobile network design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13713–13722.
41. Fu, J.; Liu, J.; Tian, H.; Li, Y.; Bao, Y.; Fang, Z.; Lu, H. Dual attention network for scene segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3146–3154.
42. Taszarek, M.; Allen, J.; Púčik, T.; Groenemeijer, P.; Czernecki, B.; Kolendowicz, L.; Lagouvardos, K.; Kotroni, V.; Schulz, W. A Climatology of Thunderstorms across Europe from a Synthesis of Multiple Data Sources. *J. Clim.* **2019**, *32*, 1813–1837. [CrossRef]