



Article

Contribution Disparity and Key Factor Screening of Lightning Identification and Nowcasting with Multi-Source Data

Xinjue Wang ^{1,2,3,†}, Zirui Xu ^{1,2,3,†}, Yujian Zhang ^{1,2}, Yuanpeng Han ^{1,2}, Lin Song ⁴ , Qilin Zhang ^{1,2} and Yi Liu ^{1,2,*}

¹ Collaborative Innovation Center on Atmospheric Environment and Equipment Technology, B-DAT, Nanjing University of Information Science and Technology, Nanjing 210044, China; hp803841@student.reading.ac.uk (X.W.); ny803789@student.reading.ac.uk (Z.X.); 202412492032@nuist.edu.cn (Y.Z.); 202512490858@nuist.edu.cn (Y.H.); qlzhang@nuist.edu.cn (Q.Z.)

² State Key Laboratory of Climate System Prediction and Risk Management, Key Laboratory of Meteorological Disaster, Ministry of Education, Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science and Technology, Nanjing 210044, China

³ Department of Computer Science, University of Reading, Whiteknights, Reading RG6 6DH, UK

⁴ Qingdao Meteorological Bureau, Qingdao 266003, China; qdsonglinyy@163.com

* Correspondence: 003766@nuist.edu.cn

† These authors contributed equally to this work.

Highlights

What are the main findings?

- A radar–satellite–lightning multi-source spatiotemporally aligned dataset is constructed for lightning identification and 0–1 h nowcasting by integrating radar composite reflectivity, Himawari-8/9 AHI observations, derived satellite features, and VLF-LLN lightning labels.
- An explainability-guided and collinearity-aware channel-selection workflow is developed by combining PFI/SHAP evidence, channel clustering, validation-constrained wrapper search, and backward redundancy elimination.
- Task-specific channel dependencies are identified: lightning identification is mainly associated with current convective-core and cloud-top-state information, whereas 0–1 h nowcasting requires additional features describing cloud-top structure, inter-channel differences, and short-term evolution.

What are the implications of the main findings?

- The unified multi-source data organization provides consistent inputs, sequence samples, and pixel-level labels for deep-learning-based lightning identification and nowcasting.
- The proposed selection strategy reduces redundant information in high-dimensional, collinear remote-sensing inputs without relying on a single importance ranking as the final criterion.
- The compact task-specific channel subsets provide physically interpretable guidance for predictor selection at different warning lead times and support lightweight, operationally interpretable lightning nowcasting models.



Academic Editor: Hao Chen

Received: 11 May 2026

Revised: 8 July 2026

Accepted: 9 July 2026

Published: 17 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

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

Abstract

Radar and geostationary satellite observations provide essential information for lightning identification and nowcasting. However, multi-source channels commonly exhibit correlation and redundancy, and directly using all channels increases computational cost and may degrade model performance. This paper focuses on lightning identification and 0–1 h nowcasting, integrating radar composite reflectivity, Himawari-8/9 AHI multi-channel observations, and VLF-LLN lightning location data to construct a multi-source spatiotemporally aligned dataset. Using baseline deep-learning-based identification and nowcasting

models, this paper proposes a workflow for analyzing multi-source channel contribution differences and selecting key factors by integrating permutation feature importance (PFI), SHAP attribution, channel collinearity grouping, and validation-set wrapper search. The results show that radar composite reflectivity has a stable dominant contribution in both tasks. The final selected channels are radar composite reflectivity, Δ Band 13, and Band 14 + Band 09 for the identification task, while they are radar composite reflectivity, Band 14, (Band 11 – Band 13) – (Band 13 – Band 15), and Δ (Band 08 – Band 13) for the nowcasting task. Independent test results show that under fixed thresholds, the three-channel identification model achieves a CSI of 0.433, outperforming the full 28-channel model with a CSI of 0.415, while the four-channel nowcasting model achieves a CSI of 0.284, outperforming the full-channel model with a CSI of 0.274. These results indicate that the proposed method can reduce input dimensionality while maintaining or improving model performance, and it can support multi-source predictor selection, lightweight deployment, and operational interpretation in lightning nowcasting.

Keywords: lightning nowcasting; lightning identification; Himawari-8/9; radar; explainable artificial intelligence; channel selection; PFI; SHAP

1. Introduction

Lightning is one of the most typical high-impact weather phenomena associated with deep convective cloud systems. Strong lightning activity can cause casualties, trigger forest fires, damage buildings and power facilities, and threaten aviation operations, communication systems, wind power generation, and transportation safety [1,2]. In mainland China, lightning disasters have long been closely related to severe convective weather and can cause substantial casualties and economic losses [3]. Lightning occurrence is jointly affected by charge separation, ice-phase microphysical processes, updraft intensity, precipitation particle distribution, and convective organization, and it is therefore highly localized, abrupt, and spatially sparse. Because lightning often occurs together with heavy rainfall, hail, thunderstorm gusts, and other hazardous weather, accurate minute-scale identification and nowcasting of lightning occurrence areas are important for improving severe-convective-weather warning capability and disaster prevention.

Existing lightning monitoring and nowcasting mainly rely on lightning location networks, radar, geostationary satellite observations, and numerical weather prediction data. Lightning location data directly record lightning events, radar reflectivity characterizes precipitation particle distribution and convective structures, and geostationary satellites provide continuous observations of cloud-top and water vapor features. However, each data source has inherent limitations, such as discrete event representation, radar coverage and scanning constraints, satellite indirectness, and numerical-model uncertainty. Therefore, a single data source is often insufficient to fully characterize the dynamical, microphysical, and cloud-precipitation conditions required for lightning occurrence, and multi-source remote-sensing fusion has become an important pathway for improving pixel-level lightning identification and nowcasting.

In recent years, machine learning and deep learning have provided new technical routes for lightning and severe-convection nowcasting. Mostajabi et al. [1] constructed a machine-learning model using conventional meteorological variables for short-term lightning occurrence warning. LightningCast, proposed by Cintineo et al. [4], uses GOES-16 ABI multi-channel satellite imagery and a semantic-segmentation convolutional neural network to predict lightning activity within the next 0–1 h. Leinonen et al. [5] systematically

compared the contributions of radar, satellite, lightning observations, numerical forecasts, and topographic data to thunderstorm-hazard nowcasting, while Leinonen et al. [2] further developed a multi-hazard data-fusion model based on deep learning. These studies demonstrate the value of multi-source observations and deep learning for nonlinear modeling and high-resolution risk mapping.

For East Asia, the Advanced Himawari Imager (AHI) onboard Himawari-8/9 provides 16 spectral channels, a 10 min full-disk scanning capability, and a spatial resolution of approximately 0.5–2 km [6]. Its visible, near-infrared, water vapor, and infrared-window channels can characterize cloud-top temperature, water vapor transport, cloud phase, cloud-top height, and convective development. Previous studies have shown that infrared brightness temperature, brightness-temperature differences, multi-channel brightness-temperature-derived variables, and cloud-top cooling features are useful for identifying cloud-top physical changes related to convective initiation [7,8]. However, Himawari-8/9 multi-channel data contain clear physical relationships and statistical collinearity. Directly feeding many raw and derived channels into deep learning models may increase training and inference costs and reduce the physical interpretability of model results.

Although previous studies have shown that multi-source remote-sensing data and deep learning can improve lightning and thunderstorm nowcasting, two issues still require further investigation. First, input variables are often selected according to experience, data availability, or existing operational products, and the relative contributions of different radar and satellite channels remain insufficiently quantified. Second, current lightning identification and future lightning nowcasting are physically related but not identical tasks. The former depends more on mature cloud-precipitation and electrical-activity features, whereas the latter requires an inference of future convective-development trends from current observations. Thus, task-specific channel contribution analysis and redundancy control are necessary for identifying stable, independent, and physically meaningful predictors. To address these issues, this paper develops a data-processing, deep-learning modeling, and explainable channel-selection framework for lightning identification and nowcasting using radar, Himawari-8/9 multi-channel satellite data, and Very Low Frequency Lightning Location Network (VLF-LLN) lightning location data. The focus is to obtain compact, stable, and physically meaningful key channel subsets under channel collinearity and information redundancy, thereby providing references for predictor selection, input-variable simplification, and severe-convection monitoring interpretation in meteorological operations.

The main contributions of this paper are threefold. First, a radar-satellite-lightning multi-source spatiotemporally aligned dataset is constructed for lightning identification and nowcasting. Radar composite reflectivity, Himawari-8/9 AHI multi-channel observations and derived features, and VLF-LLN lightning location data are unified to a 10 min temporal interval, 2 km spatial resolution, and 256×256 grid size. Discrete lightning events are converted into spatially continuous pixel-level supervision labels. This data organization provides unified multi-source inputs, sequence samples, and labels for deep-learning-based lightning nowcasting, and it also supports further analysis of how different radar and satellite predictors affect lightning occurrence and evolution. Second, an explainable channel-selection workflow is proposed by integrating permutation feature importance (PFI)/SHapley Additive exPlanations (SHAP) evidence, channel-collinearity constraints, and validation-constrained wrapper search. The workflow does not treat a single importance ranking as the final channel selection. Instead, PFI and SHAP are first used to form candidate evidence; channel-correlation grouping is then used to control redundancy; forward selection verifies stable gains from candidate channels; same-cluster coexistence tests examine whether highly correlated reserve channels still provide additional gains; and back-

ward elimination evaluates the independent necessity of selected channels. This provides an interpretable technical route for redundancy reduction, key-predictor identification, and operational input-variable simplification in high-dimensional, collinear meteorological remote-sensing features. Third, for current lightning identification and future 0–1 h lightning nowcasting, this paper compares the contribution differences of radar composite reflectivity, satellite channels, and derived features. The identification task focuses on the correspondence between current observations and lightning occurrence areas, whereas the nowcasting task uses continuous historical observations to infer future lightning evolution. Their differences in channel dependence provide physical guidance for lightning monitoring, warning-factor selection, and operational multi-source data applications.

2. Data Sources and Preprocessing

This paper uses three types of observations—namely, radar, geostationary satellite, and VLF-LLN ground-based lightning location data—to construct a multi-source dataset for lightning identification and nowcasting. The experimental dataset covers June–September 2023. The study domain is centered at 113.5°E and 24.0°N, and a 256×256 grid with 2 km spacing is constructed outward from this center as the spatial input domain. The overall workflow and the task-specific input and output formulations are summarized in Figure 1. Different data sources are first temporally aligned, spatially resampled, and gridded into samples with a 10 min temporal interval, 2 km spatial resolution, and 256×256 spatial size. Lightning labels and sequence samples are then constructed for current-time identification and future-period nowcasting tasks, respectively.

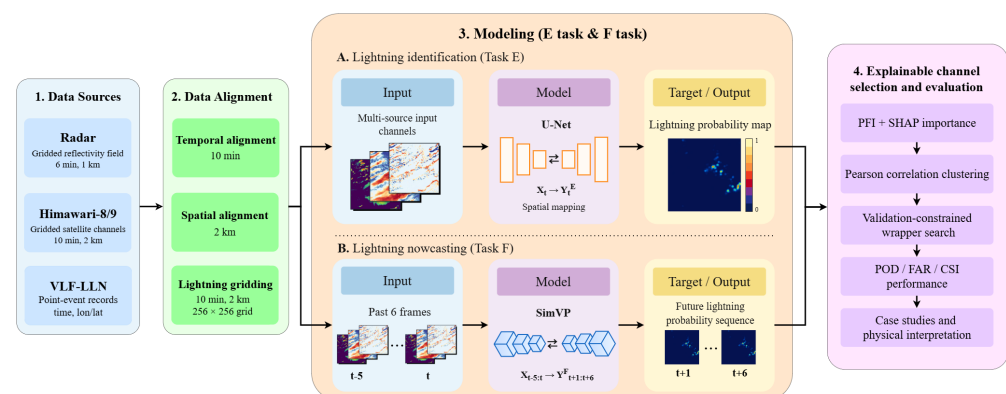


Figure 1. Schematic overview of the proposed multi-source lightning identification and nowcasting framework. The figure summarizes the spatiotemporal harmonization, Task E and Task F input and output formulations, task-specific modeling, and explainability-guided channel selection and evaluation procedures described in the corresponding Sections 2 and 3. Colors distinguish the major workflow modules and task components, while arrows indicate the direction of data processing and information flow. The ellipses in the Task F input and output sequences denote the intervening temporal frames between the displayed endpoints rather than incomplete content.

More specifically, the multi-source observations and detection hardware comprise the following: (1) radar observations from the China New Generation Doppler Radar (S-band) network, which provide radar composite reflectivity products; (2) satellite observations from the Advanced Himawari Imager (AHI) onboard the Himawari-8/9 geostationary meteorological satellites; and (3) lightning observations from the ground-based Very Low Frequency Lightning Location Network (VLF-LLN), which uses orthogonal magnetic-loop antennas to capture lightning stroke signals.

2.1. Radar Data

Radar composite reflectivity products from the national radar dataset of the China Meteorological Administration are used as radar inputs. The China New Generation Radar network provides high-spatiotemporal-resolution radar observations for severe-convection monitoring [9]. Composite reflectivity represents the strongest radar echo in the vertical direction and reflects precipitation particle distribution, strong echo cores, and convective-system intensity. For lightning occurrence, strong echo regions usually correspond to stronger updrafts, the growth of ice-phase particles, and mixed-phase particle collisions, and therefore they have clear physical links to in-cloud charge separation and lightning activity [10,11].

The original radar data have a spatial resolution of 1 km, a temporal resolution of 6 min, and an image size of 512×512 . To unify them with satellite and lightning data, radar images are downsampled to 256×256 and temporally interpolated from the 6 min radar sequence to the 10 min temporal framework. The interpolation method is based on optical-flow fields and spatiotemporal weighting constraints, and it is used to reduce inconsistencies among radar observation times, satellite scanning cycles, and lightning aggregation windows, ensuring temporal comparability among multi-source inputs. A similar preprocessing strategy has also been adopted in pixel-level lightning-identification research, where China Meteorological Administration (CMA) radar data with 1 km spatial resolution and 6 min temporal resolution were downsampled and temporally reconstructed using an optical-flow-based interpolation method to reduce radar-satellite temporal mismatch [12].

2.2. Satellite Data

Satellite data are obtained from the Advanced Himawari Imager (AHI) onboard Himawari-8/9. AHI provides full-disk observations in 16 spectral channels covering visible, near-infrared, water vapor, and infrared bands. Its full-disk scanning interval is 10 min, and the spatial resolution at the sub-satellite point is approximately 0.5–2.0 km [6]. Because this paper requires continuous day-night monitoring and focuses on deep-convective cloud tops and water vapor structures, infrared and water vapor-related channels are mainly used. Visible and near-infrared single channels (Bands 1–7) are not used because their reflectance-based signals show strong day–night differences and become largely unavailable or physically invalid at night; excluding them ensures consistent all-day inputs from a physical perspective.

In addition to basic brightness–temperature channels, this paper further constructs brightness–temperature differences, multi-channel brightness–temperature-derived variables, short-term temporal differences, and ratio variables to enhance the representation of cloud-top height, cloud phase, cloud-top cooling, vertical water vapor structure, and ice-phase microphysical processes. Here, TBB denotes the single-channel brightness temperature, BTD denotes a two-channel brightness–temperature difference, and TTD denotes a tri-spectral brightness–temperature difference constructed from three infrared channels. The final input variables consist of one radar variable and 27 satellite variables, including radar composite reflectivity, Himawari-8/9 water vapor and infrared brightness–temperature channels, and derived brightness–temperature difference, multi-channel calculation, temporal difference, and ratio features. The input variables and their physical meanings are listed in Table 1.

All satellite data are cropped to the study area and resampled to the same 2 km, 256×256 grid as the radar data and lightning labels, ensuring pixel-level spatial correspondence among different data sources.

Table 1. Input Variables and Physical Meanings.

Index	Type	Input Variable	Band or Formula	Physical Meaning
C4	Radar	Composite reflectivity	Radar composite reflectivity	Represents precipitation particle distribution, strong echo cores, and convective intensity
C5	Satellite	TBB08	Band 8, 6.2 μm	Upper- to mid-tropospheric water vapor information near deep-convective cloud tops
C6	Satellite	TBB09	Band 9, 6.9 μm	Mid-level water vapor, convective cloud extraction, and radar-reflectivity-retrieval-related feature
C7	Satellite	TBB10	Band 10, 7.3 μm	Lower- to mid-level water vapor related to moisture conditions during convective development
C8	Satellite	TBB11	Band 11, 8.6 μm	Cloud phase and thin cirrus identification; sensitive to ice-cloud and cloud-top microphysical features
C9	Satellite	TBB13	Band 13, 10.4 μm	Cloud-top brightness temperature and cloud-top height; lower brightness temperature usually indicates higher cloud top
C10	Satellite	TBB14	Band 14, 11.2 μm	Infrared-window brightness temperature for clear-sky moisture monitoring, convective initiation, and overshooting-top detection
C11	Satellite	TBB15	Band 15, 12.4 μm	Split-window channel for clear-sky moisture retrieval, convective-cloud extraction, and cloud microphysics
C12	Satellite	TBB16	Band 16, 13.3 μm	CO ₂ absorption channel for cloud-top height retrieval and severe-convection extraction
C13	Satellite	TTD11-13-15	$(TBB11 - TBB13) - (TBB13 - TBB15)$	Multi-channel variable from Bands 11, 13, and 15, representing cloud-top glaciation and mixed-phase microphysical structure
C14	Satellite	TTD11-14-15	$(TBB11 - TBB14) - (TBB14 - TBB15)$	Multi-channel variable for convective initiation and enhanced cloud-top phase differences
C15	Satellite	Delta(B08-B13)	$(TBB08 - TBB13)_t - (TBB08 - TBB13)_{t-1}$	Short-term change in cloud-top height relative to the tropopause; a proxy for convective vertical-development tendency
C16	Satellite	Delta(B16-B13)	$(TBB16 - TBB13)_t - (TBB16 - TBB13)_{t-1}$	Short-term trend of cloud-top height changes, describing convective vertical growth
C17	Satellite	Delta TBB13	$TBB13_t - TBB13_{t-1}$	Cloud-top cooling rate, indicating rapid cloud-top uplift and convective intensification
C18	Satellite	BTD08-10	$TBB08 - TBB10$	Related to precipitation and total column water vapor retrieval; represents vertical water vapor structure
C19	Satellite	BTD08-13	$TBB08 - TBB13$	Related to cloud-top height relative to the tropopause and precipitation retrieval
C20	Satellite	BTD09-10	$TBB09 - TBB10$	Mid-level water vapor difference representing mid-tropospheric humidity structure
C21	Satellite	BTD09-14	$TBB09 - TBB14$	Difference between water vapor and infrared-window channels for identifying overshooting tops and severe convection
C22	Satellite	Ratio10-14	$TBB10/TBB14$	Ratio between water vapor and infrared-window brightness temperature, assisting high-level ice-cloud and deep-convection identification
C23	Satellite	BTD10-13	$TBB10 - TBB13$	Difference between lower- to mid-level water vapor and cloud-top brightness temperature, reflecting water vapor distribution and cloud-top structure
C24	Satellite	Ratio11-14	$TBB11/TBB14$	Ratio between cloud-phase-related and infrared-window channels, describing cloud-emissivity differences
C25	Satellite	BTD11-14	$TBB11 - TBB14$	8.6 μm minus 11.2 μm brightness–temperature difference for distinguishing ice cloud, water cloud, and clear sky
C26	Satellite	BTD13-15	$TBB13 - TBB15$	Split-window brightness–temperature difference representing cloud microphysical properties and particle size
C27	Satellite	Sum14-09	$TBB14 + TBB09$	Combined infrared-window and water vapor feature representing high clouds and water vapor environment
C28	Satellite	BTD14-15	$TBB14 - TBB15$	Split-window brightness–temperature difference for distinguishing water cloud, ice cloud, and particle size
C29	Satellite	Ratio15-14	$TBB15/TBB14$	Split-window ratio enhancing ice-cloud recognition and cloud optical-thickness characterization
C30	Satellite	BTD16-13	$TBB16 - TBB13$	Difference between CO ₂ absorption and infrared-window channels for identifying deep/shallow convection and cloud-top height
C31	Satellite	BTD16-14	$TBB16 - TBB14$	Difference between CO ₂ absorption and infrared-window channels, reflecting deep-convective intensity and cloud-top vertical structure

For reproducibility and convenient reporting of channel-selection experiments, unified experimental indices are assigned to variables in the data files. VLF-LLN lightning-label-related intermediate variables are denoted as C0–C3; these variables are used only for label construction and data organization and are not included as model input channels. Radar composite reflectivity is denoted as C4, and Himawari-8/9 AHI direct brightness–temperature channels and their derived variables are denoted sequentially as C5–C31. The Cxx notation is only an experimental index for data organization and result tracking, and it does not represent a new physical channel name. The real observational channels, calculation forms, and physical meanings of model input variables C4–C31 are given in Table 1.

2.3. Lightning Data and Label Construction

The lightning location data used as supervision labels are derived from the Very Low Frequency Lightning Location Network (VLF-LLN) established by Nanjing University of Information Science and Technology. Since 2021, this network has deployed an observation array across China with 24 stations extending from west to east and covering plateaus, plains, basins, and major urban agglomerations [13]. Each station uses two orthogonal magnetic-loop antennas to receive VLF magnetic-field signals in the east–west and north–south directions, and real-time continuous acquisition is performed using a pre-triggering strategy in which a lightning event is recorded once the signal intensity exceeds a preset threshold [14]. Observational results indicate that VLF-LLN can achieve a detection radius exceeding 3000 km. Its relative detection efficiency is approximately 53%, and after propagation correction and peak-time-delay compensation, the median positioning accuracy reaches approximately 1.81 km [15,16]. The spatial distribution of the 24 VLF-LLN stations is shown in Figure 2.

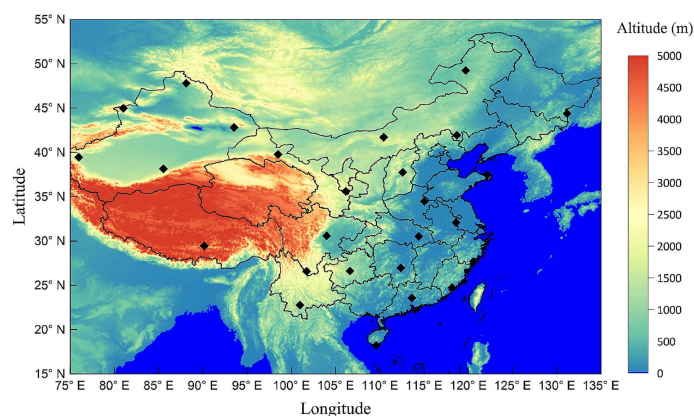


Figure 2. Spatial distribution of VLF-LLN stations, adapted from Liu et al. [13]. Black diamonds denote the VLF-LLN station locations, and the background shading represents terrain elevation in meters.

The raw lightning data record the occurrence time and longitude–latitude position of discrete lightning events. To match the 10 min scanning cycle of Himawari-8/9, lightning events are first aggregated within 10 min windows and projected onto the unified 2 km grid. For the 10 min window corresponding to time t , the lightning frequency at grid point (i, j) is defined as

$$L_t(i, j) = \sum_{n=1}^{N_t} \mathbb{I}[(x_n, y_n) \in G_{i,j}], \quad (1)$$

where N_t is the total number of lightning events within the time window, (x_n, y_n) is the projected grid coordinate of the n -th lightning event, $G_{i,j}$ is the (i, j) -th grid cell, and

$\mathbb{I}(\cdot)$ is the indicator function. This process converts discrete lightning points into a two-dimensional lightning-frequency field on a regular grid.

Because lightning activity is spatially sparse and locally clustered, directly using point-like labels can make the supervision signal too fragmented and aggravate positive–negative sample imbalance. Previous lightning-nowcasting research expanded discrete lightning locations into continuous high-density lightning areas using a Gaussian distribution, thereby representing the spatial influence range and internal frequency information of lightning activity [13]. Therefore, this paper applies a circularly constrained two-dimensional Gaussian kernel to spatially smooth the lightning-frequency field:

$$S_t = \text{clip}(L_t * K_\sigma, 0, M), \quad (2)$$

where $*$ denotes two-dimensional convolution, K_σ is a Gaussian kernel with standard deviation σ , $\text{clip}(\cdot)$ denotes numerical truncation, and M is the upper truncation limit. The Gaussian kernel is constrained by a circular mask and normalized:

$$K_\sigma(u, v) = \frac{1}{Z} \exp\left(-\frac{u^2 + v^2}{2\sigma^2}\right) \cdot \mathbb{I}(u^2 + v^2 \leq r^2), \quad (3)$$

where r is the circular-mask radius and Z is the normalization factor. This operation creates a continuous local influence area around lightning events and facilitates convolutional neural networks in learning the spatial correspondence between radar-satellite observations and lightning activity.

To mitigate the potential label noise caused by the uneven detection efficiency of the VLF-LLN network, three measures are adopted. First, the study domain is restricted to the high-efficiency core coverage zone. Second, spatial Gaussian smoothing is applied to buffer grid-scale positioning fluctuations; this density-blending strategy has been shown by Liu et al. (2025) [13] to compensate for systematic errors in lightning location data. Third, a spatially weighted loss function is used to focus model optimization on convective cores, thereby reducing sensitivity to individual missing lightning strokes.

For the current-time lightning identification task, the single-time lightning-frequency field is smoothed and converted into a binary lightning-occurrence label. Specifically, an 11×11 circular Gaussian kernel with a standard deviation of 3 grid points is used, and the smoothed result is truncated to $[0, 15]$. A binary label is then constructed according to whether the smoothed field is greater than zero:

$$Y_t^E(i, j) = \begin{cases} 1, & S_t(i, j) > 0, \\ 0, & S_t(i, j) = 0. \end{cases} \quad (4)$$

For the nowcasting task, 12 consecutive time steps are organized as one sequence sample. The first six frames of multi-source observations are used as input, and the following six frames of lightning activity fields are used as prediction targets. The interval between adjacent frames is 10 min; therefore, the model uses the past 60 min of observations to predict lightning activity in the next 60 min. The same spatial Gaussian smoothing, clipping, and normalization procedure is applied to each of the six target frames, and no additional lead-time-dependent temporal weighting is imposed on the nowcasting labels. These label-generation parameters are fixed before channel selection and kept identical for all channel subsets so that performance differences mainly reflect changes in input channels rather than changes in target construction. During training, processed lightning-activity fields are normalized to $[0, 1]$. During evaluation, target-event labels are fixed before any prediction-threshold scan. For Task E, the target-event label is the binary lightning-occurrence label defined in Equation (4). For Task F, the normalized

observed lightning-activity field is converted into binary target-event labels using the fixed target-occurrence threshold used for fixed-threshold evaluation. Threshold scanning is then applied only to the model-predicted fields.

This unweighted six-frame target design is consistent with the aggregate evaluation protocol used in precipitation-nowcasting studies in which all predicted frames within an output window are treated equally when computing skill scores [17]. In this paper, this design keeps the channel-selection objective aligned with the full 0–1 h nowcasting window rather than favoring a specific lead time during feature screening.

The preprocessing costs of the three observation sources also differ when they are converted into unified samples with a 10 min temporal interval, 2 km spatial resolution, and 256×256 grid size. For radar data, preprocessing includes radar composite reflectivity product decoding, study-area cropping, spatial-grid unification, optical-flow-based interpolation from the 6 min radar sequence to the 10 min temporal grid, and writing to .mat files; the representative single-frame processing time is approximately 1.472 s. For Himawari-8/9 AHI satellite data, preprocessing includes NC-file reading, study-area cropping, 16-channel array caching, C5–C31 direct and derived channel calculation, and .mat file writing, with a single-frame processing time of 1.302 s. For VLF-LLN lightning location data, preprocessing includes 10 min window aggregation, longitude-latitude-to-grid mapping, Gaussian-kernel smoothing, numerical clipping, binary-label generation, and .mat file writing with a single-frame processing time of 0.023 s. The representative single-frame preprocessing times for the three data sources are summarized in Table 2.

Table 2. Single-frame preprocessing time for the three data sources.

Data Source	Single-Frame Processing Time
Radar data	1.472 s frame ^{−1}
Himawari-8/9 AHI satellite data	1.302 s frame ^{−1}
VLF-LLN lightning location data	0.023 s frame ^{−1}

These timing results correspond to the preprocessing steps before sample construction and do not include data downloading, model training, or result visualization. Overall, the single-frame preprocessing costs of all three data sources are on the order of seconds or less. Large-area radar and satellite products are decoded, cropped, and cached before sample assembly, and the cached products are reused in subsequent sample organization, thereby reducing repeated reading and repeated computation over the original large-area products.

Figure 3 shows an example of multi-source data aligned at the same time, including radar composite reflectivity, satellite cloud-top brightness temperature, a cloud-top-cooling proxy, and the lightning occurrence label.

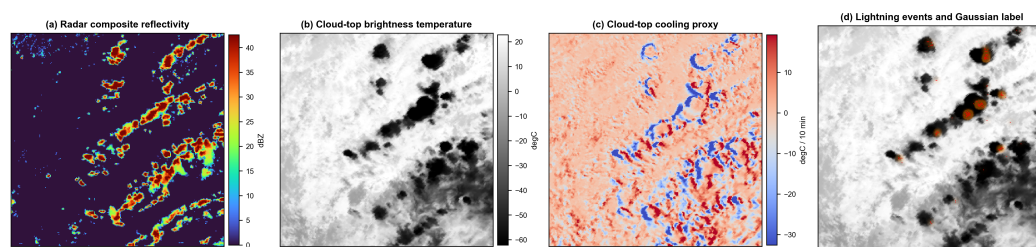


Figure 3. Example of spatially aligned radar, satellite, and lightning-label fields. In panel (d), the red dots denote the original discrete lightning-event locations overlaid on the Gaussian-smoothed lightning label.

3. Experiments and Methods

3.1. Task Definition for Lightning Identification and Nowcasting

Lightning occurrence modeling is divided into two related but physically different tasks: current-time lightning identification and future-period lightning nowcasting, as schematically summarized in Figure 1. The former determines whether lightning activity already exists at pixels corresponding to current radar and satellite observations, while the latter predicts the future spatial distribution of lightning activity based on a recent sequence of multi-source observations. This task division separately characterizes the current diagnostic features and short-term evolution features of deep convective systems.

For brevity, the current-time lightning identification task is denoted as Task E, and the future 0–1 h lightning nowcasting task is denoted as Task F throughout this paper.

In the model implementation, U-Net [18] is used as the baseline segmentation model for current lightning identification, and SimVP [19] is used as the baseline spatiotemporal prediction model for future 0–1 h lightning nowcasting. Because model-architecture improvement is not the main contribution of this paper, the two models are used as fixed baseline frameworks to compare the relative contributions of different radar and satellite channel combinations under identical training conditions.

The two task branches summarized in Figure 1 are formalized below using the same 10 min temporal interval, 2 km spatial resolution, and 256×256 grid introduced in the preprocessing workflow.

For the lightning identification task, the current multi-source input is denoted as

$$\mathbf{X}_t \in \mathbb{R}^{C \times H \times W}, \quad (5)$$

where C is the number of input channels, and H and W are the image height and width, respectively. All samples are unified to $H = W = 256$. The identification model learns the mapping from current multi-source observations $\mathbf{X}_t$ to the current lightning-occurrence label $\mathbf{Y}_t^E$:

$$\hat{\mathbf{Y}}_t^E = f_E(\mathbf{X}_t), \quad (6)$$

where $\hat{\mathbf{Y}}_t^E \in [0, 1]^{1 \times H \times W}$ is the pixel-level lightning occurrence probability map. Binary lightning occurrence areas are obtained after thresholding.

For the lightning nowcasting task, the model input is a sequence of multi-source observations from the previous six time steps with a 10 min interval between adjacent time steps:

$$\mathbf{X}_{t-5:t} \in \mathbb{R}^{T_{in} \times C \times H \times W}, \quad (7)$$

where $T_{in} = 6$. The nowcasting model outputs lightning occurrence probability fields for the next six consecutive time steps:

$$\hat{\mathbf{Y}}_{t+1:t+6}^F = f_F(\mathbf{X}_{t-5:t}), \quad (8)$$

where $\hat{\mathbf{Y}}_{t+1:t+6}^F \in [0, 1]^{T_{out} \times 1 \times H \times W}$ and $T_{out} = 6$. Thus, the nowcasting task corresponds to multi-step pixel-level lightning risk prediction within the 0–1 h lead-time window.

It should be emphasized that this paper does not focus on direct lightning-frequency regression. Instead, it focuses on identifying and nowcasting lightning occurrence areas. The lightning-frequency field is mainly used to construct spatially continuous training labels and is converted into lightning occurrence labels suitable for model training and evaluation through smoothing, truncation, and thresholding.

3.2. Dataset Organization, Experimental Settings, and Evaluation Metrics

3.2.1. Dataset Organization

After preprocessing, three data organizations are formed. Basic single-time samples are saved as multi-channel .mat files on the unified grid; identification-task data add current-time lightning occurrence labels to the basic samples; and nowcasting-task data organize 12 consecutive frames into one sequence sample. Only samples with lightning occurrence and complete radar–satellite–lightning fields during June–September 2023 are used for model training and evaluation. The eligible samples are randomly divided into training, validation, and test subsets. For the identification task, each timestamp is assigned to only one subset. For the nowcasting task, each sample consists of 12 consecutive 10 min frames, and samples assigned to different subsets are separated by at least 60 min (six frames) to avoid duplicated adjacent-frame windows across different subsets. The current dataset size is shown in Table 3.

Table 3. Dataset organization and sample numbers after lightning-occurrence filtering.

Dataset	Train	Validation	Test	Organization	Task Usage
Single-frame identification dataset	5321	1141	1140	one .mat per timestamp	current lightning identification
12-frame nowcasting dataset	958	205	206	one folder per 12-frame sequence	future lightning occurrence nowcasting

For the identification task, each sample contains multi-source inputs and a current-time binary lightning occurrence label. The model input consists of radar composite reflectivity and satellite-related variables, and the output is a pixel-level lightning occurrence probability map. For the nowcasting task, each sample contains 12 consecutive time steps; the first six frames are used as the input sequence, and the last six frames are used as supervision targets for predicting future 0–1 h lightning activity.

Channel selection requires many repeated training runs. To reduce computational cost and ensure comparability among candidate combinations, the selection stage mainly uses the training set for model-parameter learning and a fixed validation set for threshold scanning and channel-combination comparison. Validation metrics in this stage are used to compare the relative performance of different channel combinations and are not directly equivalent to final generalization performance on the independent test set.

3.2.2. Model Training and Hyperparameter Settings

Experiments are implemented using PyTorch 2.5.0 with CUDA 11.8 and performed in parallel on two NVIDIA RTX 3090 GPUs. When multiple CUDA devices are detected, DataParallel is enabled for data-parallel training. To remove the influence of different units and numerical ranges on model training and feature-importance analysis, all radar and satellite input channels are standardized channel by channel. For the c -th input channel, its mean μ_c and standard deviation σ_c are first computed on the training set, and then the same channel in the training, validation, and test sets is standardized using

$$X'_c = \frac{X_c - \mu_c}{\sigma_c + \epsilon}, \quad (9)$$

where ϵ is a small constant to avoid division by zero. Validation and test statistics are not recomputed; training-set statistics are used to avoid data leakage and ensure consistent processing across data splits.

For the identification task, U-Net is used as the baseline model, the initial channel width is set to 16, and the number of input channels changes with the channel combination. Training uses the Adam optimizer, an initial learning rate of 0.001, a batch size of 10, a maximum of 100 epochs, and a random seed of 42. The loss is a combination of weighted

binary cross-entropy and Dice loss, which is commonly used to alleviate class imbalance in sparse segmentation targets [20,21]:

$$\mathcal{L}_E = 0.7\mathcal{L}_{WBCE} + 0.3\mathcal{L}_{Dice}. \quad (10)$$

The additional weights for false-alarm pixels and missed-detection pixels in the weighted binary cross-entropy are set to 2.0 and 1.5, respectively. The identification task uses ReduceLROnPlateau based on validation loss, with a decay factor of 0.5, scheduler patience of 3, and minimum learning rate of 10^{-6} . Early stopping is applied when validation loss does not improve for 10 consecutive epochs, and the weights with the lowest validation loss are retained.

For the nowcasting task, SimVP is used as the baseline model. The input shape is $[6, C, 256, 256]$, where C is the number of input channels in the current combination, and the spatial hidden dimension is 64. Training uses the Adam optimizer, an initial learning rate of 0.002, a batch size of 3, a maximum of 50 epochs, and a random seed of 42. A weighted mean squared error is used:

$$\mathcal{L}_F = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2 (32y_i + 1), \quad (11)$$

where $\hat{y}_i$ and y_i are the prediction and truth for the i -th pixel. This loss assigns larger weights to regions with larger label values, making the model focus more on lightning-related areas. The nowcasting task also saves the best weights based on validation loss and uses early stopping with patience of 10.

3.2.3. Evaluation Metrics and Thresholds

Because lightning occurrence areas are much fewer than background areas, accuracy alone cannot reflect a model's ability to identify lightning targets. This paper uses POD, FAR, and CSI, which are commonly used in lightning warning and severe-convection nowcasting [22,23]:

$$POD = \frac{TP}{TP + FN}, \quad (12)$$

$$FAR = \frac{FP}{TP + FP}, \quad (13)$$

$$CSI = \frac{TP}{TP + FP + FN}, \quad (14)$$

where TP , FP , and FN denote hit, false-alarm, and missed-detection pixels, respectively. POD represents the ability to capture real lightning areas; FAR represents the proportion of false alarms among predicted lightning areas; and CSI jointly considers hits, false alarms, and missed detections, providing a more comprehensive metric for sparse-event identification.

During channel selection, the maximum CSI on the validation set is used as the primary comparison metric. For both Task E and Task F, target-event labels are fixed before threshold scanning. Candidate prediction thresholds are scanned only on the model-predicted fields, and the prediction threshold τ_{pred}^* that maximizes validation CSI and its corresponding metrics is recorded as CSI^* . In this paper, candidate prediction thresholds are scanned from 0.01 to 0.99 with a step of 0.01. Metrics under fixed prediction thresholds are also retained as diagnostic results and operational threshold references. The fixed prediction thresholds used for independent test evaluation are determined from validation-set diagnostics and then kept unchanged when applied to the test set. In final evaluation, the full 28-channel

baseline models and selected compact-channel models are applied to the independent test set, and both fixed-prediction-threshold and diagnostic test-set prediction-threshold-scan results are reported. Fixed-threshold results reflect direct test performance under predefined prediction thresholds, whereas diagnostic test-set prediction-threshold scanning is used only as a reference for threshold sensitivity and potential upper-bound performance, and it is not used for channel selection, model selection, or main workflow optimization. Therefore, the forward-selection gains and backward-elimination losses described below are computed using validation-set threshold-optimized CSI* rather than fixed-threshold test CSI.

3.3. Explainable Channel Selection Workflow

3.3.1. Explainability Evidence and Channel Collinearity Analysis

Himawari-8/9 multi-channel observations and their derived brightness–temperature-difference variables describe the cloud-top temperature, water vapor structure, cloud phase, and convective development from different perspectives, but they also contain clear redundancy and collinearity. Directly using all channels increases training and inference costs and reduces interpretability. Therefore, this paper designs a channel-selection workflow composed of explainability analysis, collinearity constraints, and validation-set wrapper search.

First, identification and nowcasting models are trained under the full 28-channel setting, and PFI [24] and SHAP [25] attribution results are computed for each channel. PFI measures the contribution of a channel to the overall prediction performance by perturbing the channel and observing the resulting performance decrease [24]. If perturbing a channel causes a clear CSI drop, the channel contains useful information that is difficult to fully replace using other channels. SHAP attributes model outputs at the sample level and measures local dependence on each channel [25]. PFI is closer to a global performance contribution, whereas SHAP is closer to local attribution strength; the two provide complementary evidence.

To ensure statistical robustness, PFI calculations are evaluated on the entire validation dataset (1141 samples for Task E and 205 sequence samples for Task F). For SHAP, 500 and 205 validation instances are randomly sampled for Task E and Task F, respectively, using 100 training samples as the background set. To preserve spatial–temporal autocorrelation within the gridded fields (256×256 pixels), PFI uses sample-level permutation, in which the entire 2D field (or 6-frame sequence) of a target channel is swapped between samples. For SHAP, raw pixel-level attribution maps (256×256 or $6 \times 256 \times 256$) are averaged across spatial pixels, time frames, and evaluated samples to obtain a single importance score for each channel. PFI evaluations are repeated five times under different seeds, and the resulting standard deviations are negligible (<0.003), confirming the stability of the channel rankings.

Second, input variables are grouped according to Pearson correlation coefficients. Because Task E and Task F have different data organizations, channel correlation matrices are computed separately on the training sets of the two tasks. For each task, 300 training samples are randomly selected, and full-resolution 256×256 pixels are used for statistics. For the nowcasting task, one frame is randomly selected from the six input time steps of each sample for correlation estimation. Let $\mathbf{x}_{c_i}$ and $\mathbf{x}_{c_j}$ be the flattened pixel vectors of channels c_i and c_j . Their Pearson correlation coefficient is

$$r_{ij} = \frac{\sum_k (x_{c_i,k} - \bar{x}_{c_i})(x_{c_j,k} - \bar{x}_{c_j})}{\sqrt{\sum_k (x_{c_i,k} - \bar{x}_{c_i})^2} \sqrt{\sum_k (x_{c_j,k} - \bar{x}_{c_j})^2}}. \quad (15)$$

After obtaining the correlation matrix, the channel distance is defined as $d_{ij} = 1 - |r_{ij}|$, and average-linkage hierarchical clustering is used for grouping. The distance threshold is set to 0.15, approximately corresponding to a strong correlation of $|r| \geq 0.85$, to identify highly collinear channel clusters. Figures 4 and 5 show the hierarchical clustering trees for Task E and Task F, respectively, and the dashed line indicates the default cutoff. Low-distance branches formed by the brightness–temperature, brightness–temperature–difference, and multi-channel calculation variables are evident in both tasks, indicating obvious collinearity among some satellite-derived variables. Figure 6 further shows the clustering results in correlation-matrix form, where colors indicate signed Pearson correlations and black boxes denote channel clusters obtained at $d = 0.15$. This threshold is selected by analyzing the structure of the hierarchical clustering tree (Figures 4 and 5). Setting the threshold too low (e.g., $d < 0.15$, corresponding to $|r| > 0.85$) moves the cutoff line to the left, resulting in many fragmented and highly collinear clusters. This would fail to remove redundancy effectively and would leave an excessively large candidate space for the wrapper search. Conversely, setting the threshold too high (e.g., $d > 0.15$) moves the cutoff line to the right and merges channels with lower correlations into the same cluster. This may group physically distinct properties, such as separate water vapor and cloud-phase signatures, causing the selection algorithm to overlook independently valuable features during representative screening. Thus, $d = 0.15$ represents a practical trade-off between redundancy removal and feature diversity preservation. Overall, several brightness–temperature and brightness–temperature–difference variables form stable high-correlation clusters in both tasks, while radar composite reflectivity is relatively independent. This provides a basis for subsequent channel redundancy reduction. The Cxx experimental indices are retained in channel-selection result tables for reproducibility, and their real bands and formulas are given in Table 1.

For both Task E and Task F, the PFI and SHAP evidence is normalized using the same robust procedure before aggregation. Negative PFI values are retained in the diagnostic table but are clipped to zero for positive-contribution scoring; i.e., $\max(\text{PFI}, 0)$ is transformed using a median absolute deviation (MAD) robust z-score. Because raw SHAP magnitudes can have a long-tailed distribution, SHAP is first compressed as $\log(1 + \text{SHAP})$ and then transformed using the same MAD robust z-score. If MAD is zero, the interquartile range (IQR)/1.349 or the standard deviation is used as a fallback. The combined evidence score is defined as $0.5 \times \text{PFI}_{\text{robust}} + 0.5 \times \text{SHAP}_{\text{robust}}$.

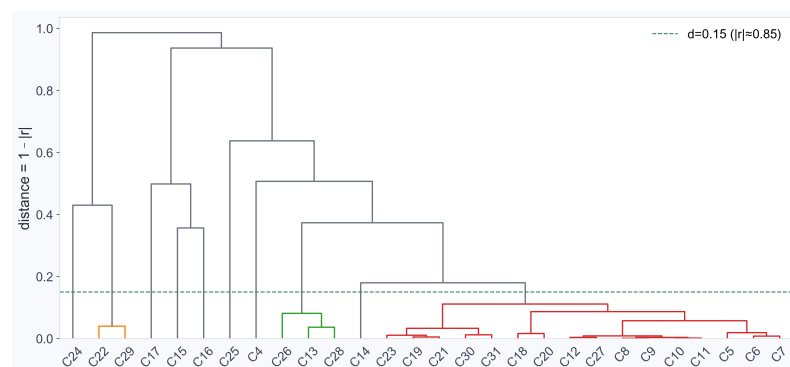


Figure 4. Average-linkage hierarchical clustering tree of the 28 input channels for Task E. The clustering distance is defined as $d = 1 - |r|$, and the dashed line indicates the default cutoff $d = 0.15$. Different colors distinguish separate channel clusters formed below the cutoff, whereas gray branches indicate higher-level linkages above the cutoff; the colors do not encode an additional quantitative variable.

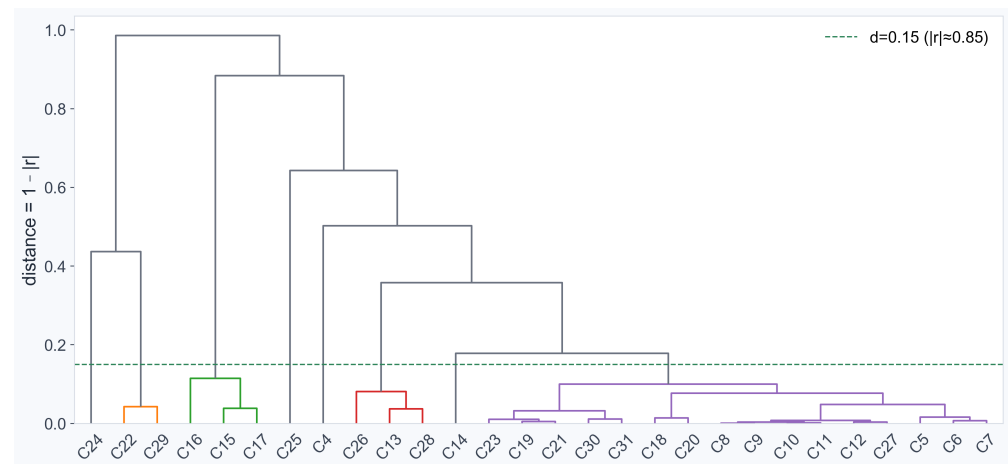


Figure 5. Average-linkage hierarchical clustering tree of the 28 input channels for Task F. The dashed line indicates the default cutoff $d = 0.15$, approximately corresponding to $|r| \geq 0.85$. Different colors distinguish separate channel clusters formed below the cutoff, whereas gray branches indicate higher-level linkages above the cutoff; the colors do not encode an additional quantitative variable.

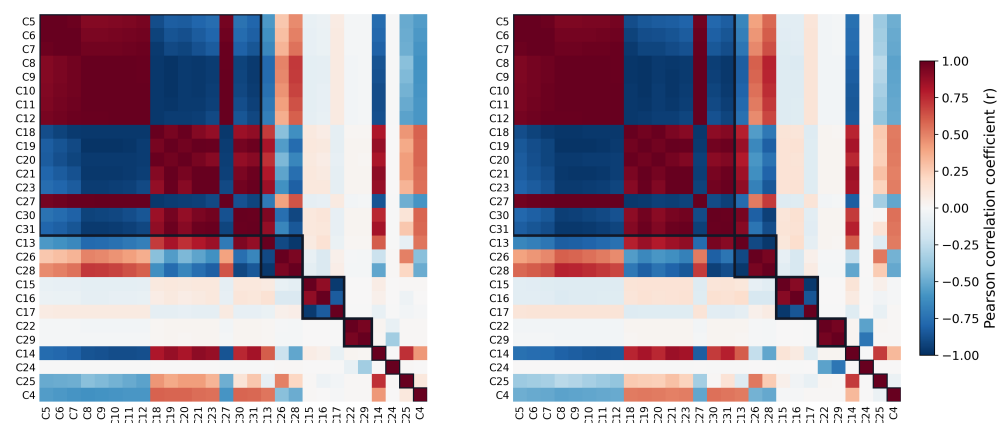


Figure 6. Pearson-correlation clustering matrix of the 28 input channels for Task E and Task F. Channels are reordered by average-linkage clustering using $d = 1 - |r|$, and black rectangles indicate clusters obtained with the default cutoff $d = 0.15$.

After collinearity clustering, PFI and SHAP results are projected onto the same channel-cluster framework to jointly examine the independent performance contribution, local attribution strength, and same-cluster redundancy. Figures 7 and 8 show the cluster-grouped PFI/SHAP evidence for Task E and Task F, respectively. The upper panel is the combined signed score after the robust standardization of PFI and SHAP, and the lower panel gives the $0.5 \times$ PFI robust and $0.5 \times$ SHAP robust components. Vertical dashed lines indicate cluster boundaries, while black dots and red stars indicate main channels and anchor channels entering the candidate process. The purpose of these figures is not to directly provide the final channel combination but to support the classification of core channels, diversity candidates, same-cluster reserve channels, and low-priority channels.

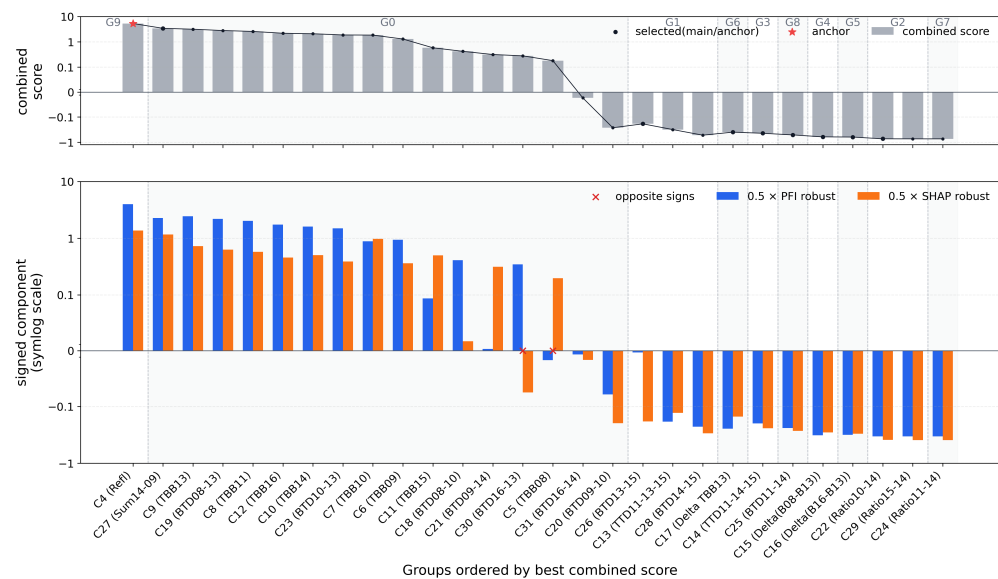


Figure 7. Cluster-grouped signed PFI/SHAP components for Task E. Channels are grouped by Pearson-correlation clusters and ordered by the best combined evidence within each cluster.

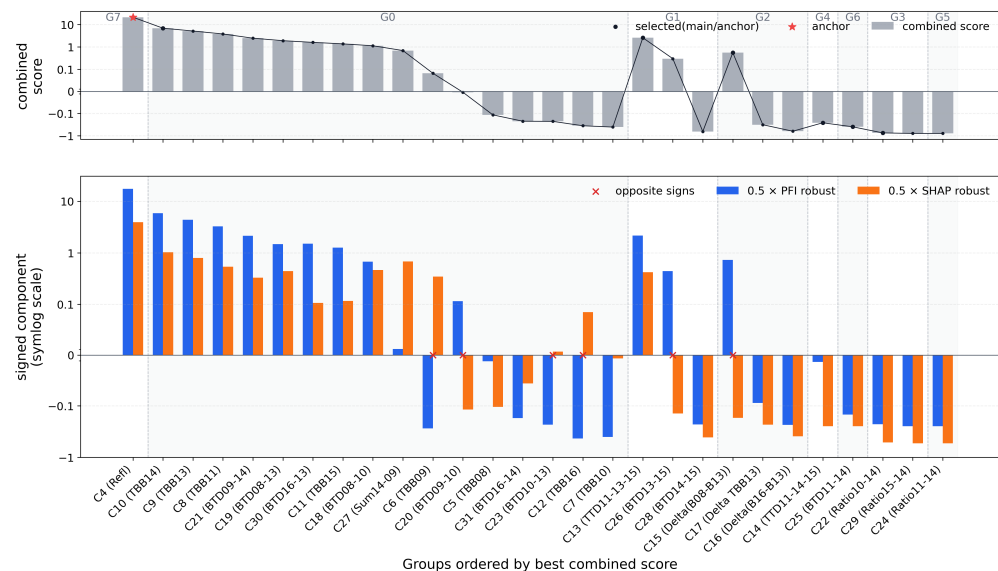


Figure 8. Cluster-grouped signed PFI/SHAP components for Task F. The figure highlights task-dependent differences in channel contribution and redundancy under the same collinearity-aware grouping strategy.

3.3.2. Candidate Pool Construction

For highly correlated channels, if multiple variables carry similar information, they should not be retained simply according to a single importance ranking. This paper combines PFI, SHAP, and correlation grouping to classify channels into core channels, diversity candidate channels, same-cluster reserve channels, and negative-contribution or low-priority channels. Core channels are variables with stable explainability contributions and clear physical meaning. Diversity candidates may provide information complementary to the core channels. Same-cluster reserve channels are highly correlated with selected channels but may still provide supplementary information under local conditions. Negative-contribution or low-priority channels are generally excluded from routine expansion.

After constructing the candidate pool, validation-constrained wrapper forward selection and backward elimination are used to confirm the actual performance of channel combinations. The overall workflow is shown in Figure 9. This process does not add

channels one by one according to PFI or SHAP ranking; instead, it retrains models to test whether adding candidate channels can stably improve validation CSI and then tests in the backward stage whether removing selected channels causes no effective performance loss.

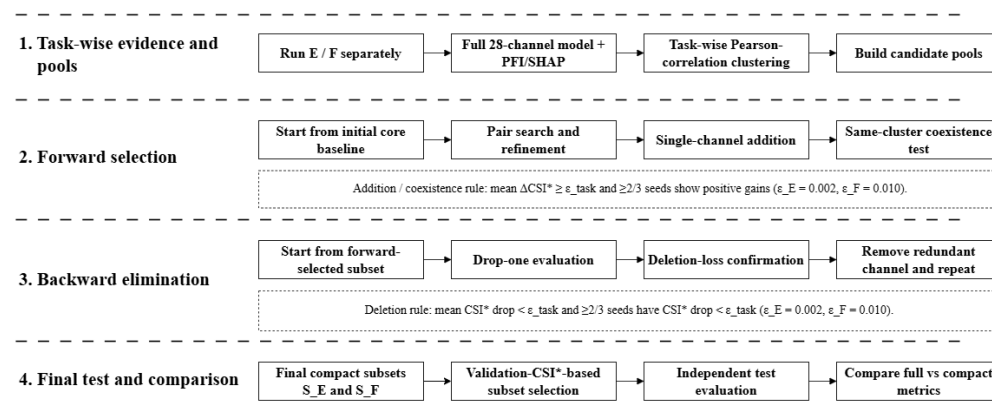


Figure 9. Schematic explainability-guided channel selection workflow. Task E and Task F are processed separately using full 28-channel PFI/SHAP evidence, Pearson-correlation clustering, validation-constrained forward wrapper search, backward redundancy elimination, and independent test comparison. The task-specific empirical thresholds are $\epsilon_E = 0.002$ for Task E and $\epsilon_F = 0.010$ for Task F.

3.3.3. Validation-Constrained Forward Wrapper Search

To reduce the influence of training randomness, a multi-random-seed confirmation mechanism is used. Specifically, the multi-random-seed confirmation uses three random seeds for repeated training, which are denoted as S1, S2, and S3 in the tables; accordingly, the stability criterion requires at least two out of three random seeds to satisfy the gain or deletion-loss condition. The minimum effective-change threshold is treated as a task-specific empirical screening threshold with $\epsilon_E = 0.002$ for Task E and $\epsilon_F = 0.010$ for Task F. These thresholds are not statistical-significance thresholds; instead, they are applied together with the multi-seed consistency rule to reduce the risk of mistaking seed-level fluctuations for stable channel contributions. CSI differences on the order of 10^{-3} have been reported and compared in precipitation-nowcasting studies [26,27], providing a practical scale reference for interpreting small CSI changes in the wrapper search. During the forward wrapper search, for each candidate A , the augmented set S_{cand} is first constructed as

$$S_{cand} = S_{current} \cup A. \quad (16)$$

The corresponding validation gain ΔCSI^* is then evaluated as

$$\Delta CSI^*(A | S_{current}) = CSI^*(S_{cand}) - CSI^*(S_{current}). \quad (17)$$

Promising candidates from rough screening are confirmed under multiple random seeds. If the average ΔCSI^* is $\geq \epsilon_{task}$ and at least two thirds of the seeds show positive gains, the current set is updated:

$$S_{current} \leftarrow S_{current} \cup A. \quad (18)$$

If candidate A does not meet these criteria, $S_{current}$ remains unchanged and other candidates are tested. Through this set-update procedure, the currently used channels and the candidate channels to be verified are explicitly distinguished, avoiding the direct equivalence of PFI/SHAP ranking and final channel selection.

Forward selection consists of three outer loops. The first outer loop establishes the current baseline set. Based on PFI/SHAP evidence and physical meaning, an initial candidate set is constructed from the core-channel pool and used as $S_{current}$ for training and validation. This step does not exhaustively search for the optimal starting point among all channels; instead, it confirms that the core observational information can form stable initial prediction capability and provides a unified baseline for subsequent gain tests.

The second outer loop tests the incremental contribution of cross-cluster diversity candidates and is further divided into pair-combination expansion and single-channel addition. Given $S_{current}$ from the first loop, all two-channel combinations $A = \{d_i, d_j\}$ are enumerated from the diversity candidate pool D , where $d_i, d_j \in D$ and $i \neq j$, and $S_{cand} = S_{current} \cup A$ is constructed. All pair candidates are roughly screened under the same training settings, and promising pairs enter multi-random-seed confirmation. If a pair satisfies the task-specific average ΔCSI^* threshold ϵ_{task} and at least two thirds positive gains, it is considered to provide stable cross-cluster complementary information. If multiple pairs meet the criteria, the pair with the highest average CSI^* is added to $S_{current}$.

After pair expansion, the second loop continues to test the remaining diversity candidate channels one by one. For each remaining candidate d , $A = \{d\}$ and $S_{cand} = S_{current} \cup \{d\}$ are constructed, and the same rough-screening and multi-random-seed confirmation rules are applied. If a single-channel candidate satisfies the addition criteria, it is merged into $S_{current}$. If none of the remaining candidates meets the gain and stability requirements, the second loop stops. This design first tests possible combined complementary effects among diversity channels and then determines whether an individual candidate can still provide independent gains on the updated set, avoiding both the missed selection and redundant retention caused by single-channel ranking.

The third outer loop tests coexistence gains from same-cluster reserve channels. For the high-correlation clusters containing selected channels in $S_{current}$, this paper does not directly retain all channels in the cluster. Instead, reserve channels r are taken one at a time from the same-cluster reserve pool, $S_{cand} = S_{current} \cup \{r\}$ is constructed, and $\Delta CSI^*(\{r\} | S_{current})$ is calculated. This step tests whether a reserve channel can still provide additional gains when jointly input with the current set rather than replacing the selected cluster representative. Only reserve channels satisfying the multi-random-seed addition rules are included. This loop avoids the redundant retention caused by collinearity while allowing a small number of same-cluster channels with complementary local information to enter subsequent evaluation.

3.3.4. Backward Redundancy Elimination and Search Scope

In the backward-elimination stage, the deletion of each channel $c \in S_{current}$ is evaluated. The channel set after deletion is

$$S_{remove} = S_{current} \setminus \{c\}. \quad (19)$$

The resulting performance loss $drop^*$ is calculated as

$$drop^*(c | S_{current}) = CSI^*(S_{current}) - CSI^*(S_{remove}). \quad (20)$$

If the mean $drop^*$ across multiple random seeds is $< \epsilon_{task}$ and at least two thirds of the seeds show no effective performance degradation, the channel is removed as redundant. After each deletion, the next outer loop re-examines the remaining channels based on the updated $S_{current}$; otherwise, backward elimination stops and the current set is used as the final compact subset.

The purpose of this strategy is not to exhaustively search all possible channel combinations but to find a stable, compact, and physically meaningful subset at low computational cost based on explainability evidence and collinearity constraints. It should be noted

that this paper does not enumerate high-order combinations of multiple reserve channels within the same cluster. For example, in a high-correlation cluster, two non-representative channels might theoretically outperform a single cluster representative when used together. Such combinations are not included in the main workflow because same-cluster channels are already highly collinear, and the main experiment focuses on verifying cluster representatives and the effectiveness of a single same-cluster reserve channel coexisting with the current set. Exhaustively testing same-cluster pairs or higher-order combinations would rapidly expand the search space and weaken the purpose of clustering-based redundancy reduction. Therefore, the channel combinations obtained in this paper should be understood as empirically preferred subsets derived from PFI/SHAP evidence, collinearity constraints, and validation-set wrapper search rather than proof of the global optimum among all possible combinations of the 28 channels.

4. Results

4.1. Contribution Differences Between Identification and Nowcasting Tasks (PFI & SHAP)

To investigate how models utilize multi-source features across different tasks, PFI [24] and SHAP [25] are applied to compare channel contributions in lightning identification (Task E) and 0–1 h nowcasting (Task F). PFI measures global performance degradation under channel perturbation, whereas SHAP quantifies local attribution at the sample level. Combining both reveals task-dependent feature-dependence patterns.

Radar composite reflectivity consistently dominates in both tasks. This is physically reasonable because radar echoes directly represent convective cores and ice-phase hydrometeors essential for charging and lightning [10]. This confirms that the baseline models capture physically meaningful meteorological relations rather than statistical artifacts. Some figures and tables retain the Cxx indices defined in Table 1 for tracking consistency.

For Task E, the model relies on channels describing the instantaneous absolute state of convective systems. In addition to radar composite reflectivity, both PFI and SHAP assign high importance to Band 13, Band 14 + Band 09, Band 08 – Band 13, and Band 16. These variables characterize cloud-top radiative intensity, height, and ice cloud proxies, indicating that current-time identification is mainly controlled by the current absolute convective state. The corresponding PFI and SHAP results are shown in Figures 10 and 11, respectively.

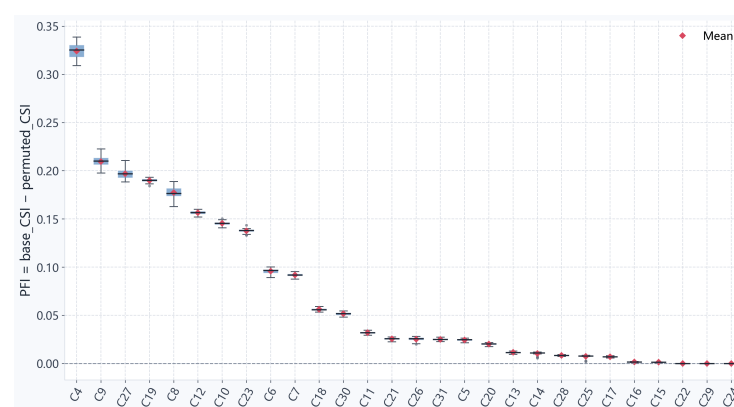


Figure 10. PFI results for Task E under the full 28-channel setting.

that its local attribution is high but its global contribution is largely redundant and compensated by related channels. This shows that relying on a single explainability method can overestimate the independent contribution of correlated features [28,29]. Therefore, robust channel selection should jointly evaluate SHAP and PFI, motivating our multi-step wrapper selection workflow.

4.2. Channel Selection Results for the Identification Task

4.2.1. Candidate Channel Set Obtained by Forward Selection for Task E

For Task E, validation-constrained forward wrapper search is applied using PFI/SHAP candidates and collinearity grouping under multiple random seeds to ensure stability. The baseline core set $\{C4, C27\}$ (radar composite reflectivity and Band 14 + Band 09) achieves a mean validation CSI^* of 0.458110 ± 0.002479 across random seeds (Table 4). The diversity candidate pair expansion results are shown in Table 4.

Table 4. Task E diversity pair expansion results in forward selection.

Stage	$S_{current}$	Candidate A	S_{cand}	S1 CSI^*	S2 CSI^*	S3 CSI^*	Mean ΔCSI^*	Pass Ratio	Decision
Core baseline	-	-	$\{C4, C27\}$	0.460033	0.458984	0.455313	-	-	Baseline for loop 2
Pair expansion	$\{C4, C27\}$	$\{C17, C15\}$	$\{C4, C15, C17, C27\}$	0.464545	0.456579	0.464199	0.003664	2/3	Passed; highest mean CSI^* ; added
Pair expansion	$\{C4, C27\}$	$\{C14, C22\}$	$\{C4, C14, C22, C27\}$	0.462615	0.459722	0.458211	0.002073	3/3	Passed but lower than best pair
Pair expansion	$\{C4, C27\}$	$\{C25, C15\}$	$\{C4, C15, C25, C27\}$	0.462448	0.456425	0.462410	0.002318	2/3	Passed but lower than best pair
Pair expansion	$\{C4, C27\}$	$\{C26, C15\}$	$\{C4, C15, C26, C27\}$	0.462681	0.455558	0.455182	-0.000303	1/3	Not passed

The combination of Δ Band 13 (C17) and Δ (Band 08 – Band 13) (C15) yields the highest average gain and is added, updating the set to $\{C4, C15, C17, C27\}$. Subsequent single-channel expansion (Table 5) shows negative average gains; hence, no additional diversity channels are selected.

Table 5. Task E single-channel diversity addition after pair expansion.

Stage	$S_{current}$	Candidate A	S_{cand}	S1 CSI^*	S2 CSI^*	S3 CSI^*	Mean ΔCSI^*	Pass Ratio	Decision
Single addition	$\{C4, C15, C17, C27\}$	$\{C14\}$	$\{C4, C14, C15, C17, C27\}$	0.464576	0.440376	0.458350	-0.007340	1/3	Not passed
Single addition	$\{C4, C15, C17, C27\}$	$\{C26\}$	$\{C4, C15, C17, C26, C27\}$	0.464556	0.447204	0.436947	-0.012205	1/3	Not passed

Same-cluster coexistence tests (Table 6) show that no reserve channel achieves a stable gain above $\epsilon_E = 0.002$.

Table 6. Task E same-cluster coexistence tests in forward selection.

Stage	$S_{current}$	Candidate A	S_{cand}	S1 CSI^*	S2 CSI^*	S3 CSI^*	Mean ΔCSI^*	Pass Ratio	Decision
Coexistence	$\{C4, C15, C17, C27\}$	$\{C23\}$	$\{C4, C15, C17, C23, C27\}$	0.468401	0.454699	0.464504	0.000760	2/3	Average gain insufficient
Coexistence	$\{C4, C15, C17, C27\}$	$\{C5\}$	$\{C4, C5, C15, C17, C27\}$	0.464420	0.467413	0.456263	0.000924	1/3	Average gain insufficient
Coexistence	$\{C4, C15, C17, C27\}$	$\{C8\}$	$\{C4, C8, C15, C17, C27\}$	0.468821	0.445566	0.467692	-0.001081	2/3	Not passed
Coexistence	$\{C4, C15, C17, C27\}$	$\{C6\}$	$\{C4, C6, C15, C17, C27\}$	0.465290	0.465626	0.449919	-0.001496	2/3	Not passed
Coexistence	$\{C4, C15, C17, C27\}$	$\{C11\}$	$\{C4, C11, C15, C17, C27\}$	0.465233	0.446516	0.464750	-0.002942	2/3	Not passed
Coexistence	$\{C4, C15, C17, C27\}$	$\{C9\}$	$\{C4, C9, C15, C17, C27\}$	0.463993	0.445242	0.461963	-0.004709	0/3	Not passed

Consequently, the forward-selected candidate set for Task E is $\{C4, C15, C17, C27\}$. The subsequent backward-elimination process for Task E is summarized in Figure 14.

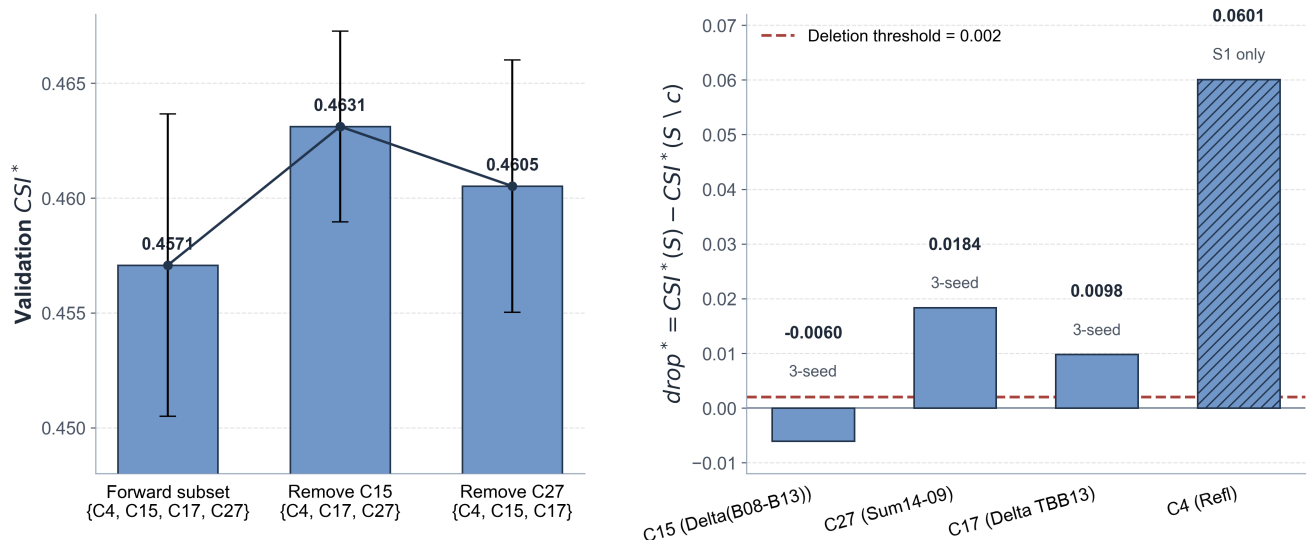


Figure 14. Backward elimination process for Task E. The (left) panel compares the forward-selected subset with deletion candidates confirmed under multiple random seeds, and the (right) panel shows deletion losses relative to the threshold. The Cxx indices are defined in Table 1.

4.2.2. Backward Elimination for Task E

Backward elimination starting from $\{C4, C15, C17, C27\}$ evaluates channel deletion losses (Table 7). Deleting $\Delta(\text{Band } 08 - \text{Band } 13)$ (C15) causes no effective performance loss (mean $\text{drop}^* < \epsilon_E = 0.002$) and is removed first.

Table 7. Task E first backward-elimination loop.

Loop	S_{current}	Removed Channel	S_{remove}	S1 CSI*	S2 CSI*	S3 CSI*	Mean drop^*	Ratio $\text{drop}^* < \epsilon_E$	Decision
1	$\{C4, C15, C17, C27\}$	{C15}	$\{C4, C17, C27\}$	0.467617	0.462274	0.459450	−0.006035	2/3	Remove
1	$\{C4, C15, C17, C27\}$	{C27}	$\{C4, C15, C17\}$	0.464578	0.462704	0.454269	−0.003439	2/3	Removable but lower priority than C15
1	$\{C4, C15, C17, C27\}$	{C17}	$\{C4, C15, C27\}$	0.458954	-	-	−0.003096 (S1)	-	Not confirmed
1	$\{C4, C15, C17, C27\}$	{C4}	$\{C15, C17, C27\}$	0.413286	-	-	0.042572 (S1)	-	Large S1 loss; retain

The second loop (Table 8) shows that deleting any remaining channel in $\{C4, C17, C27\}$ degrades performance. Thus, the backward search stops, yielding the final compact set $S_E = \{C4, C17, C27\}$ (radar composite reflectivity, $\Delta\text{Band } 13$, and Band 14 + Band 09), achieving a validation CSI^* of 0.463114 ± 0.004148 . This combination represents convective cores, cloud-top cooling, and water vapor structure, showing that C15 provides limited independent gain when combined with C17 and C27.

Table 8. Task E second backward-elimination loop.

Loop	S_{current}	Removed Channel	S_{remove}	S1 CSI*	S2 CSI*	S3 CSI*	Mean drop^*	Ratio $\text{drop}^* < \epsilon_E$	Decision
2	$\{C4, C17, C27\}$	{C17}	$\{C4, C27\}$	0.449210	0.458396	0.452281	0.009818	0/3	Retain C17
2	$\{C4, C17, C27\}$	{C27}	$\{C4, C17\}$	0.449286	0.445385	0.439561	0.018370	0/3	Retain C27
2	$\{C4, C17, C27\}$	{C4}	$\{C17, C27\}$	0.407520	-	-	0.060098 (S1)	-	Large S1 loss; retain C4

The physical interpretation of the final Task E channel subset is summarized in Figure 15.

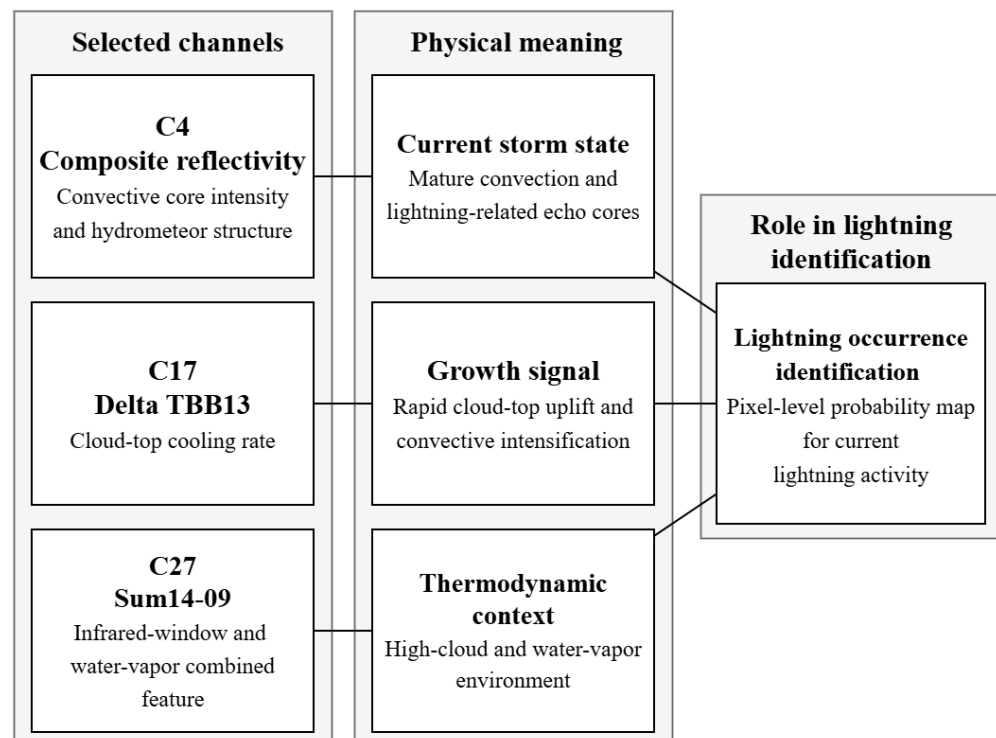


Figure 15. Physical interpretation of the final Task E channel subset derived from the backward-elimination results. The Cxx indices and their physical meanings are defined in Table 1.

4.3. Channel Selection Results for the Nowcasting Task

For Task F, the same selection logic is applied to historical sequence inputs. Predictors are expected to capture cloud-top state, cloud-phase/cloud-top structure, inter-channel differences, and short-term evolution to predict future 0–1 h lightning trends [7,8,30–32].

4.3.1. Candidate Channel Set Obtained by Forward Selection for Task F

The core baseline set {C4, C10, C13, C15} (radar composite reflectivity, Band 14, (Band 11 – Band 13) – (Band 13 – Band 15), and Δ (Band 08 – Band 13)) achieves a mean validation CSI^* of 0.350121 ± 0.014968 (Table 9). Compared with the Task E core baseline (0.458110 ± 0.002479), the Task F core baseline shows larger multi-seed variability, supporting the use of the more conservative threshold $\epsilon_F = 0.010$. Diversity pair expansions (Table 9) yield negative average gains, keeping the set unchanged.

Table 9. Task F diversity pair expansion results in forward selection.

Stage	$S_{current}$	Candidate A	S_{cand}	S1 CSI^*	S2 CSI^*	S3 CSI^*	Mean ΔCSI^*	Pass Ratio	Decision
Core baseline	-	-	{C4, C10, C13, C15}	0.341156	0.367400	0.341808	-	-	Baseline for loop 2
Pair expansion	{C4, C10, C13, C15}	{C14, C22}	{C4, C10, C13, C14, C15, C22}	0.339422	0.337979	0.318422	-0.018180	0/3	Not passed
Pair expansion	{C4, C10, C13, C15}	{C25, C22}	{C4, C10, C13, C15, C22, C25}	0.338820	0.285490	0.327441	-0.032871	0/3	Not passed
Pair expansion	{C4, C10, C13, C15}	{C14, C25}	{C4, C10, C13, C14, C15, C25}	0.317845	-	-	-	-	Not confirmed after rough screening

Subsequent single-channel additions (Table 10) also show negative gains, so no diversity channels are added.

Table 10. Task F single-channel diversity addition after pair expansion.

Stage	$S_{current}$	Candidate A	S_{cand}	S1 CSI*	S2 CSI*	S3 CSI*	Mean ΔCSI^*	Pass Ratio	Decision
Single addition	{C4, C10, C13, C15}	{C25}	{C4, C10, C13, C15, C25}	0.338395	0.338518	0.337183	−0.012089	0/3	Not passed
Single addition	{C4, C10, C13, C15}	{C14}	{C4, C10, C13, C14, C15}	0.338031	0.324592	0.332625	−0.018372	0/3	Not passed
Single addition	{C4, C10, C13, C15}	{C22}	{C4, C10, C13, C15, C22}	0.321172	-	-	-	-	Not confirmed after rough screening

Same-cluster coexistence tests (Table 11) show no stable gains for reserve channels.

Table 11. Task F same-cluster coexistence tests in forward selection.

Stage	$S_{current}$	Candidate A	S_{cand}	S1 CSI*	S2 CSI*	S3 CSI*	Mean ΔCSI^*	Pass Ratio	Decision
Coexistence	{C4, C10, C13, C15}	{C5}	{C4, C5, C10, C13, C15}	0.349808	0.316875	0.313719	−0.023320	1/3	Not passed
Coexistence	{C4, C10, C13, C15}	{C27}	{C4, C10, C13, C15, C27}	0.345485	0.344568	0.308148	−0.017388	1/3	Not passed
Coexistence	{C4, C10, C13, C15}	{C8}	{C4, C8, C10, C13, C15}	0.341721	0.335381	0.342033	−0.010410	2/3	Negative average gain; not added
Coexistence	{C4, C10, C13, C15}	{C19}	{C4, C10, C13, C15, C19}	0.335993	0.347263	0.301759	−0.021783	0/3	Not passed
Coexistence	{C4, C10, C13, C15}	{C31}	{C4, C10, C13, C15, C31}	0.335300	0.355353	0.299200	−0.020171	0/3	Not passed
Coexistence	{C4, C10, C13, C15}	{C21}	{C4, C10, C13, C15, C21}	0.335182	0.311735	0.308901	−0.031515	0/3	Not passed

Thus, the forward-selected candidate set is {C4, C10, C13, C15}, showing that additional channels provide no stable complementarity under wrapper constraints.

4.3.2. Backward Elimination for Task F

Backward elimination starting from {C4, C10, C13, C15} evaluates deletion losses under $\epsilon_F = 0.010$ (Table 12). Removing any channel leads to substantial average performance loss with radar reflectivity (C4) showing the largest drop.

Table 12. Task F backward-elimination loop.

Loop	$S_{current}$	Removed Channel	S_{remove}	S1 CSI*	S2 CSI*	S3 CSI*	Mean drop*	Ratio $drop^* < \epsilon_F$	Decision
1	{C4, C10, C13, C15}	{C13}	{C4, C10, C15}	0.321393	0.313482	0.320051	0.014285	1/3	Retain C13
1	{C4, C10, C13, C15}	{C15}	{C4, C10, C13}	0.298692	0.333941	0.318477	0.015557	1/3	Retain C15
1	{C4, C10, C13, C15}	{C10}	{C4, C13, C15}	0.296578	-	-	0.020359 (S1)	-	Large S1 loss; retain C10
1	{C4, C10, C13, C15}	{C4}	{C10, C13, C15}	0.224763	-	-	0.092174 (S1)	-	Large S1 loss; retain C4

The backward-elimination process for Task F is summarized in Figure 16.

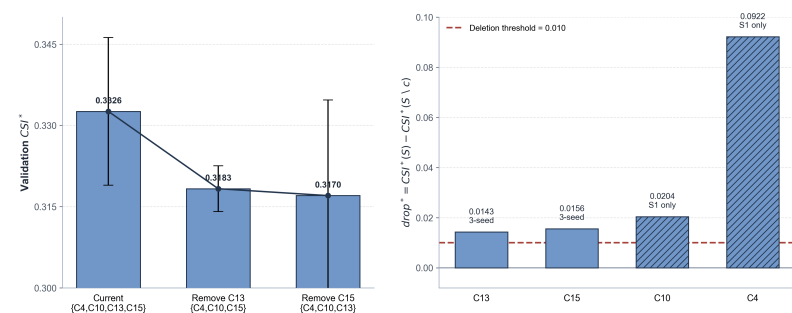


Figure 16. Backward elimination process for Task F. The (left) panel compares the current four-channel subset with the main deletion candidates confirmed under multiple random seeds, and the (right) panel shows deletion losses for all tested channels. The Cxx indices are defined in Table 1.

Consequently, backward elimination stops, and the final set is $S_F = \{C4, C10, C13, C15\}$. This suggests that nowcasting requires both convective-core (C4) and satellite features representing cloud-top state (C10), phase/structure (C13), and short-term change (C15) to predict lightning evolution. The physical interpretation of the final Task F channel subset is summarized in Figure 17.

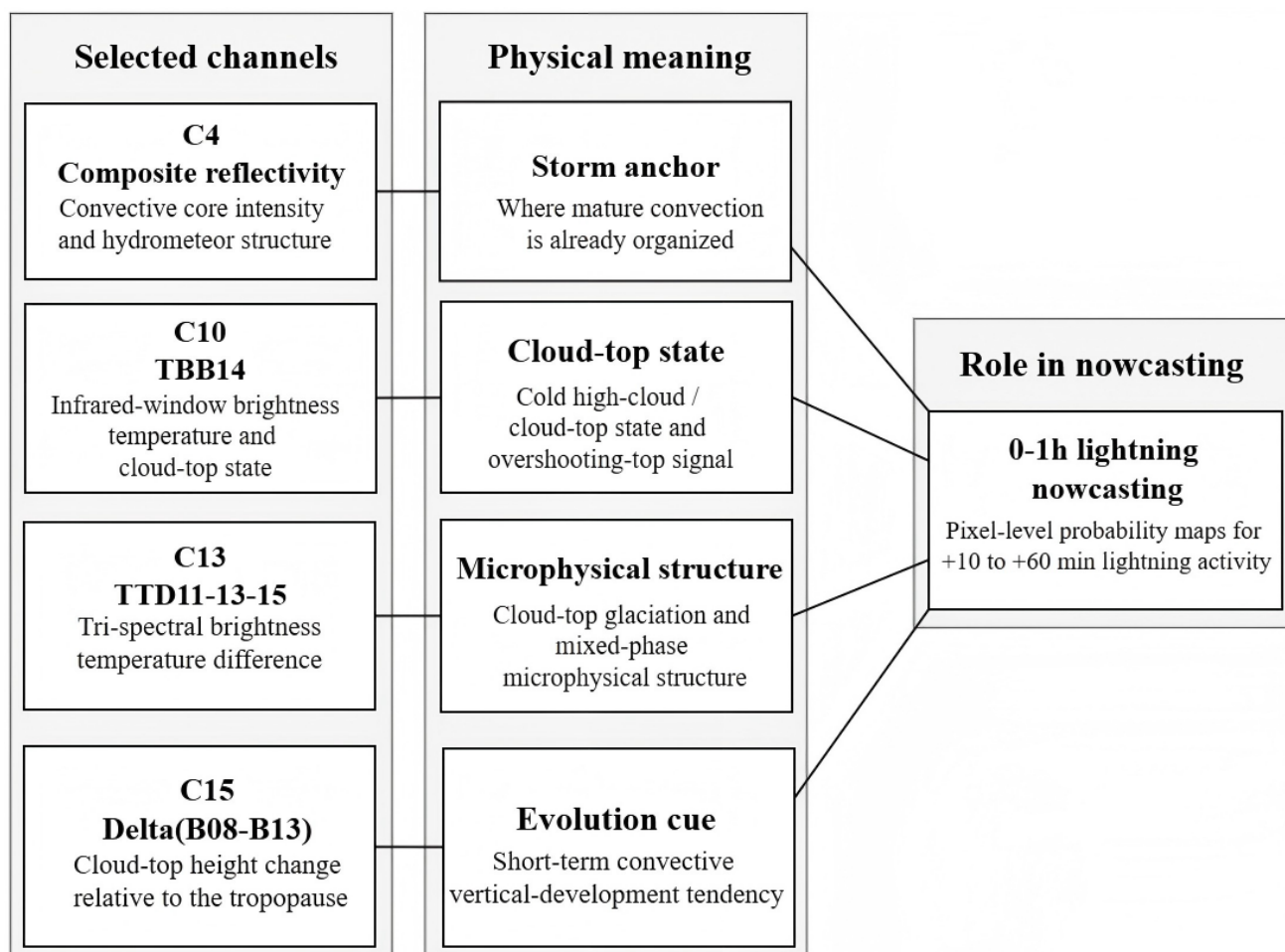


Figure 17. Physical interpretation of the final Task F channel subset derived from the backward-elimination results. The Cxx indices and their physical meanings are defined in Table 1.

4.4. Independent Test-Set Performance Comparison

After channel selection, the full 28-channel models and final compact-channel models are compared on the independent test set. Because radar composite reflectivity is the only radar input in this paper and all other model input channels are Himawari-8/9 AHI satellite channels or satellite-derived variables, radar-only baselines using C4 are also included to quantify the contribution of the radar input alone and to clarify the incremental value of the selected radar–satellite channel subsets. Table 13 reports the probability of detection (POD), false alarm ratio (FAR), and CSI under fixed prediction thresholds and diagnostic test-set prediction-threshold scanning for the two tasks. Fixed-threshold results reflect direct test performance under predefined prediction thresholds, whereas diagnostic test-set prediction-threshold scanning with fixed target-event labels is used only as a diagnostic reference for threshold sensitivity and potential upper-bound performance, and it is not used for channel selection or model selection.

Table 13. Independent test performance of the full 28-channel models, selected compact channel subsets, and radar-only baselines.

Task/Model	Channels	Test Samples	Fixed τ_{pred}	POD	FAR	CSI	Test-Scan τ_{pred}^*	POD	FAR	CSI
Task E Radar-only (C4)	1	1140	0.50	0.561	0.373	0.421	0.35	0.594	0.398	0.427
Task E selected	3	1140	0.50	0.619	0.409	0.433	0.41	0.645	0.428	0.435
Task E full	28	1140	0.50	0.584	0.410	0.415	0.19	0.668	0.456	0.428
Task F Radar-only (C4)	1	206	0.23	0.628	0.666	0.279	0.29	0.847	0.663	0.318
Task F selected	4	206	0.23	0.666	0.669	0.284	0.31	0.805	0.613	0.354
Task F full	28	206	0.23	0.660	0.681	0.274	0.25	0.856	0.678	0.306

Figure 18 visualizes the selected-versus-full comparison using CSI as the core metric, and Table 13 additionally reports the radar-only baselines. For the identification task, the radar-only model obtains a fixed-threshold CSI of 0.421 and a diagnostic prediction-threshold-scan CSI of 0.427, indicating that radar composite reflectivity alone provides a strong baseline for current lightning identification. The selected three-channel model further increases the fixed-threshold CSI to 0.433 and the diagnostic prediction-threshold-scan CSI to 0.435, outperforming both the radar-only and full-channel models. For the identification task, the selected three-channel model achieves a fixed-threshold test CSI of 0.433, higher than the full-channel model (0.415), with an improvement of 0.018. Under diagnostic test-set prediction-threshold scanning, the selected model achieves a CSI of 0.435, which is also slightly higher than the full-channel model (0.428). For the nowcasting task, the radar-only model obtains a fixed-threshold CSI of 0.279 and a diagnostic prediction-threshold-scan CSI of 0.318, whereas the selected four-channel model improves these values to 0.284 and 0.354, respectively. The selected four-channel model also exceeds the full-channel model, which obtains a fixed-threshold CSI of 0.274 and a diagnostic prediction-threshold-scan CSI of 0.306. These results show that radar reflectivity is a strong single-source predictor, while selected satellite channels provide additional information for both current lightning identification and future lightning evolution. The diagnostic prediction-threshold-scan results provide supplementary information on threshold sensitivity, whereas the main model comparison is based on fixed-threshold performance. Overall, the selected compact channel subsets substantially reduce the number of input channels without degrading independent test performance and even achieve higher CSI values in both tasks.

For Task F, the fixed-threshold full-channel test results were further evaluated at each 10 min forecast lead time using the same fixed prediction threshold of 0.23 as in Table 13. From +10 to +60 min, the CSI decreases gradually from 0.3411 to 0.2220 with values of 0.3165, 0.2933, 0.2642, and 0.2461 at +20, +30, +40, and +50 min, respectively. Over the same lead-time range, the POD decreases from 0.7400 to 0.5966, whereas the FAR increases from 0.6124 to 0.7388. This temporal evolution indicates that forecast uncertainty accumulates with increasing lead time, while the model maintains measurable nowcasting skill throughout the 0–1 h window. Similar lead-time-dependent degradation has also been reported in precipitation-nowcasting studies, where the frame-wise CSI decreases or probabilistic error increases as lead time extends [27,33].

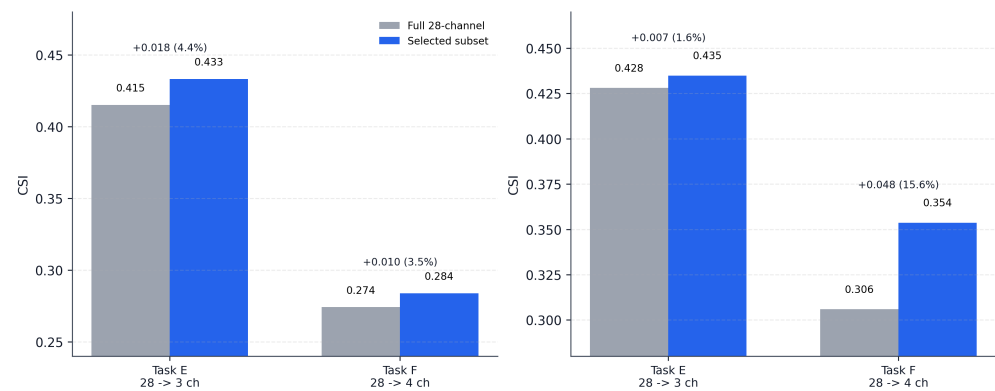


Figure 18. Independent test performance comparison of full and selected channel sets. The (left) panel reports fixed-prediction-threshold CSI, whereas the (right) panel reports diagnostic test-set prediction-threshold-scan CSI. In the diagnostic scan, the target-event labels are fixed and only the prediction threshold is scanned.

4.5. Qualitative Case Study for Nowcasting

To complement the aggregate test-set metrics, a representative Task F test sequence is visualized in Figure 19. The first row shows the four selected current-time input variables for the nowcasting task, including radar composite reflectivity, Band 14 brightness temperature, the tri-spectral brightness–temperature difference, and the short-term temporal-difference channel. The following rows compare the model-predicted lightning probability fields with the corresponding smoothed lightning labels from +10 min to +60 min. This case illustrates that the selected compact channel subset can preserve the main spatial organization and temporal evolution tendency of future lightning activity, supporting the quantitative finding that the four-channel Task F subset maintains useful nowcasting skill while reducing input dimensionality.

Comparing the predicted probability sequence with the true lightning distribution reveals that the selected four-channel model captures the eastward movement and spatial decay of the primary convective core. At early lead times (+10 to +30 min), the predicted high-probability regions exhibit strong spatial alignment with the observed lightning labels. As the lead time increases (+40 to +60 min), the predicted probability field becomes more diffuse and slightly overestimates the convective boundary, reflecting the accumulation of spatiotemporal prediction uncertainty inherent in convective development. Nonetheless, the core areas of active lightning remain well constrained, verifying the practical nowcasting capability of the selected channel subset.

To quantitatively and qualitatively analyze the forecasting errors, the predicted probability fields are compared against the smoothed lightning observation density. False alarms mainly occur near convective margins, where cold ice-phase cloud-top signals observed by Himawari AHI may extend outward, but strong radar composite reflectivity is absent or spatially displaced. Missed detections may occur in rapidly developing or localized convective cells, because lightning activity is sparse and abrupt and satellite infrared brightness–temperature features describe cloud-top evolution indirectly. Spatially smoothing the labels with a Gaussian kernel buffers these small spatial mismatches and helps maintain stable nowcasting skill.

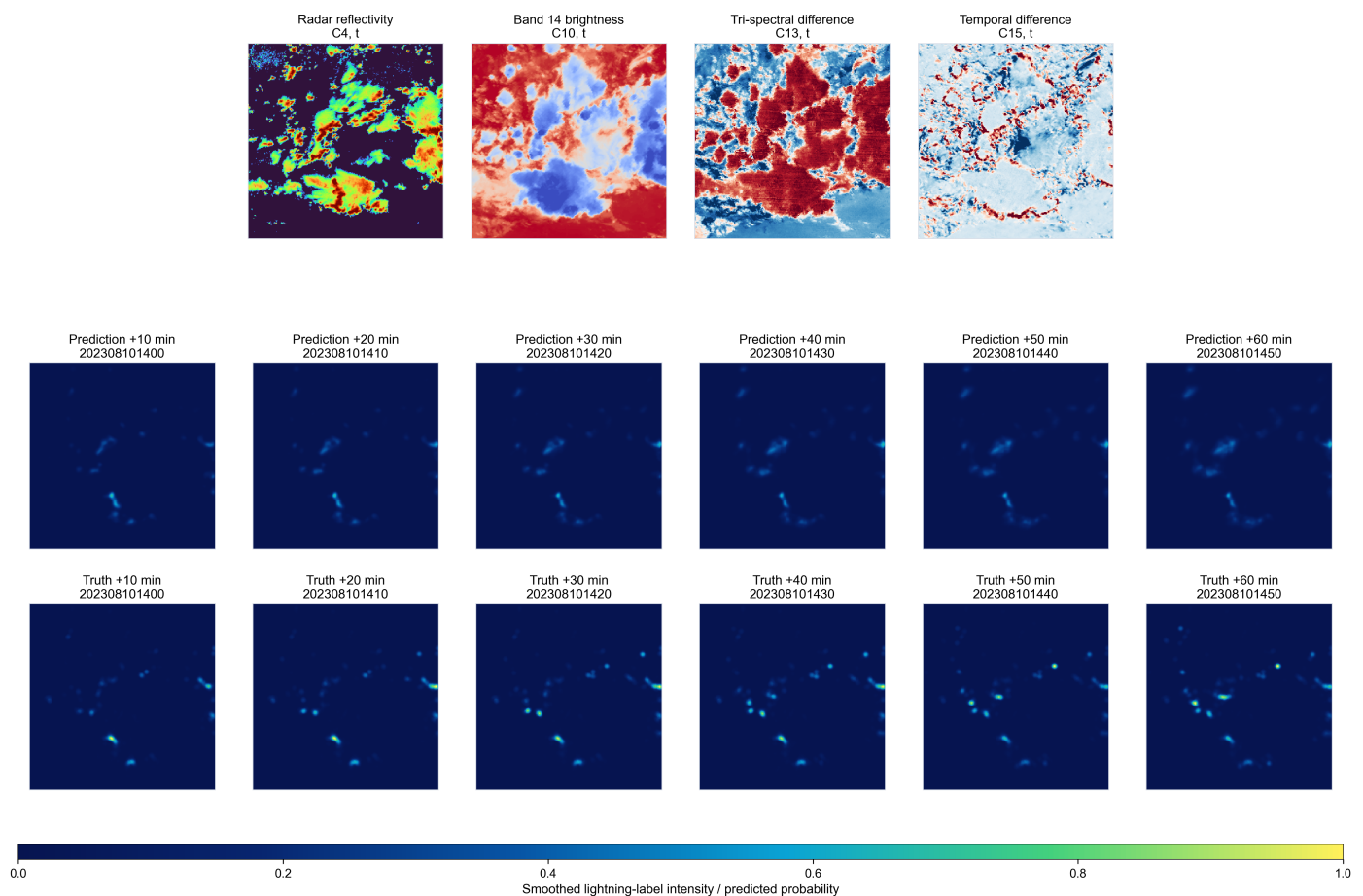


Figure 19. Qualitative case study of 0–1 h lightning nowcasting using the selected Task F channel subset. The upper row shows the selected current-time input channels, and the lower panels compare predicted lightning probability fields with the corresponding smoothed lightning-label fields at six 10 min lead times. The colors in the upper-row panels visualize the spatial patterns of input variables with different physical meanings and value ranges and are not directly comparable across channels. The prediction and truth panels share the common 0–1 color scale shown below.

5. Discussion

5.1. Influence of Feature Collinearity and Redundant Information on Interpretability in Remote-Sensing Deep Learning

Many deep-learning studies in remote sensing still treat models as black boxes or rely on a single explainability tool such as SHAP or attention weights to select variables. This paper explicitly highlights the strong physical collinearity and information redundancy among high-dimensional remote-sensing features such as Himawari-8/9 spectral bands and derived variables. By jointly using local attribution (SHAP) and global perturbation (PFI), this paper shows that a single explainability tool may overestimate the independent contribution of correlated features and lead to redundant channel retention. The case of Band 14 + Band 09 is representative. The proposed workflow therefore provides a robust, physically guided and data-driven white-box paradigm for channel-dimensionality reduction.

5.2. Physical Differences Between Lightning Identification and Nowcasting

The results reveal a clear physical difference between the two tasks. Task E finally retains three channels, radar composite reflectivity, Δ Band 13, and Band 14 + Band 09, indicating a stronger dependence on current convective-core structure, cloud-top evolution, and the mid- to upper-level water vapor environment. Task F finally retains four channels, radar composite reflectivity, Band 14, (Band 11 – Band 13) – (Band 13 – Band 15), and

Δ (Band 08 – Band 13), indicating that 0–1 h nowcasting requires both current convective-core information and satellite proxy features describing cloud-top state, cloud-phase/cloud-top structure, inter-channel differences, and short-term evolution. This suggests that the model captures statistical relationships consistent with convective development and lightning processes rather than only statistical noise. These results can guide input-feature design for different warning lead times.

5.3. Limitations and Sources of Uncertainty

Several uncertainties remain. First, multi-source spatiotemporal matching may be affected by satellite parallax and VLF-LLN positioning errors. However, the positioning deviations caused by these inherent observation errors are smaller than the spatial resolution of our network grid (2 km). Although spatial Gaussian smoothing (approximately 6 km) buffers these sub-grid scale physical variances, they may still introduce minor label noise. Second, the proposed selection strategy is a greedy search. Although clustering and explainability greatly reduce the search space, the method does not guarantee the absolute global optimum among all 2^{28} possible combinations. Third, this paper focuses on south China during the severe-convection season, and the results should be interpreted within this regional and seasonal context. When applied to other climate regions or convective weather types, the proposed workflow should be re-run on regional datasets to identify empirically suitable channel subsets. Additionally, to adapt to seasonal shifts in convective dynamics and enhance year-round nowcasting performance, it is possible to train unified multi-seasonal models or develop separate seasonal models for different periods of the year.

Furthermore, as Peng et al. (2020) [34] reported, different thunderstorm cells exhibit distinct VLF pulse characteristics, and the ratio of positive to negative flashes is highly indicative of the convective stage of a thunderstorm, such as developing or dissipating stages. The ground-based VLF-LLN primarily detects cloud-to-ground (CG) lightning flashes (ground flashes). Because CG lightning directly threatens human safety and infrastructure, focusing on CG-related lightning activity supports the practical relevance of the current warning-oriented framework. Nonetheless, the current framework does not distinguish polarities or CG versus intra-cloud (IC) strokes. Future studies should incorporate more diverse flash types and polarities as multi-task learning targets to further improve the model's ability to capture the dynamic development stage of severe thunderstorms.

5.4. Operational Prospects

The selected compact channel subsets may reduce the cost of satellite-data transmission, storage, and model inference, making real-time lightning nowcasting more feasible under computationally constrained operational conditions.

Persistence and extrapolation baselines remain important references for operational nowcasting; radar-echo extrapolation is also commonly used as a baseline or component in precipitation-nowcasting research [17]. Because the objective here is to isolate multi-source channel contribution under fixed deep-learning architectures, the radar-only and full-channel models are used as the main baselines in this manuscript, while direct persistence/extrapolation comparisons will be included in subsequent operational evaluation.

6. Conclusions

This paper constructs a multi-source spatiotemporally aligned dataset for lightning identification and nowcasting by integrating radar composite reflectivity, Himawari-8/9 AHI multi-channel satellite observations, and VLF-LLN lightning location data. Through temporal alignment, spatial resampling, lightning-label gridding, and smoothing, different data sources are unified into samples with a 10 min temporal interval, 2 km spatial resolu-

tion, and 256×256 grid size. This provides a unified data foundation for current lightning identification and future 0–1 h lightning nowcasting, and it also supports further analysis of how different radar and satellite predictors influence lightning occurrence and evolution.

Methodologically, this paper proposes an explainable channel-selection workflow for high-dimensional multi-source remote-sensing inputs by combining PFI, SHAP, channel-collinearity grouping, and validation-constrained wrapper search. The workflow avoids directly equating a single importance ranking with final channel selection. Instead, it uses explainability methods to form candidate evidence, reduces redundancy through correlation constraints, and validates stable gains through forward selection and backward elimination. The results show that this strategy can obtain validation-constrained, compact, and physically meaningful channel subsets under channel collinearity and information redundancy.

The channel-selection results show that radar composite reflectivity provides stable contributions in both identification and nowcasting tasks, indicating that current convective cores and precipitation structures remain important information sources for lightning occurrence discrimination and short-term prediction. For the identification task, the final selected channels are radar composite reflectivity, Δ Band 13, and Band 14 + Band 09, reflecting the joint roles of current convective structure, cloud-top evolution, and mid- to upper-level water vapor conditions. For the nowcasting task, the final selected channels are radar composite reflectivity, Band 14, (Band 11 – Band 13) – (Band 13 – Band 15), and Δ (Band 08 – Band 13), indicating that future lightning nowcasting requires not only current convective-core information but also satellite features representing cloud-top state, cloud-phase changes, inter-channel differences, and short-term evolution trends.

Independent test-set results further show that the compact channel subsets maintain or improve model performance while substantially reducing the number of input variables. In the identification task, the radar-only baseline achieves a fixed-threshold test CSI of 0.421, while the selected three-channel model achieves a CSI of 0.433, which is higher than both the radar-only baseline and the full 28-channel model (0.415). In the nowcasting task, the radar-only baseline achieves a fixed-threshold test CSI of 0.279, while the selected four-channel model achieves a CSI of 0.284, which is also higher than the radar-only and full-channel (0.274) models. These results indicate that not all available remote-sensing channels provide stable independent gains. Reasonable channel redundancy reduction helps improve model compactness, computational efficiency, and physical interpretability.

Overall, the results demonstrate that integrating explainability evidence, channel-collinearity constraints, and validation-set wrapper search can provide stable, compact, and meteorologically meaningful input channel combinations for lightning identification and nowcasting. The proposed method can provide a reference for multi-source predictor selection, lightweight model deployment, and operational interpretation in severe-convection nowcasting. Future studies should further evaluate the generalization ability of this channel-selection strategy across more regions, seasons, weather types, and lightning detection networks. In addition, the specific channel subsets identified in this paper are derived from summer convection cases in south China (Guangdong) and should be interpreted within this regional and seasonal setting. When applied to other regions or convective weather types, the proposed workflow can be used to screen and evaluate corresponding task-specific channel subsets. Furthermore, analyzing the temporal decay and lead-time-specific performance of selected channels under operational environments is an important direction for subsequent application research, because this paper primarily uses aggregated 0–1 h performance to avoid lead-time-specific overfitting during feature selection.

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

Funding: This research was funded by the Young Scientists Fund of the National Natural Science Foundation of China, grant number 42405148; the Startup Foundation for Introducing Talent of Nanjing University of Information Science and Technology, grant number 2023r124; the Open Research Project of the Key Laboratory of Lightning, China Meteorological Administration, grant number 2024KELL-B013; the Natural Science Foundation of Shandong Province, China, grant number ZR2023MD012; the Scientific and Technological Research Projects of Guangdong Meteorological Bureau, grant number GRMC2022Z06; and the Open Research Project of Key Laboratory of Intelligent Meteorological Observation Technology, China Meteorological Administration, grant number ZNGC2025MS09.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw datasets and processing code used in this paper are subject to confidentiality agreements and institutional restrictions, and therefore they are not publicly available. However, the channel-selection workflow proposed in this paper is designed as a generalized, dataset-independent pipeline that can be adapted to other deep-learning models using similar multi-source inputs. Researchers interested in replicating this screening pipeline or applying it to their own datasets can contact the corresponding author via email to request anonymized sample data and workflow code for academic research purposes.

Acknowledgments: The authors thank the data providers and collaborators for their support. Tools provided by OpenAI were used solely for language polishing, consistency checking, and reviewing wording and layout. They were not used to generate or alter scientific data, numerical results, model outputs, experimental results, research conclusions, or graphical elements in the final figures. All AI-assisted suggestions were reviewed and verified by the authors, who take full responsibility for the content of this publication.

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

Abbreviations

The following abbreviations are used in this manuscript:

AHI	Advanced Himawari Imager
BTD	Brightness Temperature Difference
CSI	Critical Success Index
FAR	False Alarm Ratio
PFI	Permutation Feature Importance
POD	Probability of Detection
SHAP	SHapley Additive exPlanations
TBB	Brightness Temperature
TTD	Tri-spectral Brightness–Temperature Difference
VLF-LLN	Very Low Frequency Lightning Location Network

References

- 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. [\[CrossRef\]](#)
- 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. [\[CrossRef\]](#)
- Zhang, D.; Holle, R.L. The epidemiology of lightning in mainland China: A review of two datasets from the 1950s to 2018. In *2022 36th International Conference on Lightning Protection (ICLP)*; IEEE: Piscataway, NJ, USA, 2022; pp. 597–602. [\[CrossRef\]](#)
- Cintineo, J.L.; Pavlonis, M.J.; Sieglaff, J.M. ProbSevere LightningCast: A deep-learning model for satellite-based lightning nowcasting. *Weather Forecast.* **2022**, *37*, 1131–1147. [\[CrossRef\]](#)
- Leinonen, J.; Hamann, U.; Germann, U.; Mecikalski, J.R. Nowcasting thunderstorm hazards using machine learning: The impact of data sources on performance. *Nat. Hazards Earth Syst. Sci.* **2022**, *22*, 577–597. [\[CrossRef\]](#)
- 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. Ser. II* **2016**, *94*, 151–183. [\[CrossRef\]](#)
- Mecikalski, J.R.; Bedka, K.M. Forecasting convective initiation by monitoring the evolution of moving cumulus in daytime GOES imagery. *Mon. Weather Rev.* **2006**, *134*, 49–78. [\[CrossRef\]](#)
- Mecikalski, J.R.; Bedka, K.M.; Paech, S.J.; Litten, L.A. A statistical evaluation of GOES cloud-top properties for nowcasting convective initiation. *Mon. Weather Rev.* **2008**, *136*, 4899–4914. [\[CrossRef\]](#)
- Min, C.; Chen, S.; Gourley, J.J.; Chen, H.; Zhang, A.; Huang, Y.; Huang, C. Coverage of China new generation radar network. *Adv. Meteorol.* **2019**, *2019*, 5789358. [\[CrossRef\]](#)
- MacGorman, D.R.; Rust, W.D. *The Electrical Nature of Storms*; Oxford University Press: Oxford, UK, 1998.
- Williams, E.R. The tripole structure of thunderstorms. *J. Geophys. Res. Atmos.* **1989**, *94*, 13151–13167. [\[CrossRef\]](#)
- Chen, Y.; Han, Y.; Zhang, Y.; Liu, Y.; Song, L.; Wang, J.; Wang, X.; Zhang, Q. MDE-UNet: A physically guided asymmetric fusion network for multi-source meteorological data lightning identification. *Remote Sens.* **2026**, *18*, 1027. [\[CrossRef\]](#)
- Liu, Y.; Wang, J.; Song, Y.; Liang, S.; Xia, M.; Zhang, Q. Lightning nowcasting based on high-density area and extrapolation utilizing long-range lightning location data. *Atmos. Res.* **2025**, *321*, 108070. [\[CrossRef\]](#)
- Zhang, J.; Zhou, J.; Li, J.; Gu, J.; Zhang, Q.; Dai, B.; Wang, Y.; Wang, J.; Zhong, Y.; Li, Q. Location accuracy improvement of long-range lightning detection network in China by compensating ground wave propagation delay. *Remote Sens.* **2022**, *14*, 3397. [\[CrossRef\]](#)
- Li, J.; Dai, B.; Zhou, J.; Zhang, J.; Zhang, Q.; Yang, J.; Wang, Y.; Gu, J.; Hou, W.; Zou, B.; et al. Preliminary application of long-range lightning location network with equivalent propagation velocity in China. *Remote Sens.* **2022**, *14*, 560. [\[CrossRef\]](#)
- Li, J.; Song, L.; Zhang, Q.; Wang, S.; Yang, J.; Ge, Q. Optimizing lightning location accuracy: A study of propagation velocity and time of arrival in long-range lightning location algorithms. *Measurement* **2024**, *234*, 114754. [\[CrossRef\]](#)
- Beutler, T.; Rudolph, A.; Goehring, D.; Vercauteren, N. Exploring the ability of regional extrapolation for precipitation nowcasting with deep learning. *Meteorol. Z.* **2024**, *33*, 305–321. [\[CrossRef\]](#)
- Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention, MICCAI 2015*; Springer: Cham, Switzerland, 2015; pp. 234–241. [\[CrossRef\]](#)
- Gao, Z.; Tan, C.; Wu, L.; Li, S.Z. SimVP: Simpler yet better video prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, 18–24 June 2022*; IEEE: Piscataway, NJ, USA, 2022; pp. 3170–3180. [\[CrossRef\]](#)
- Milletari, F.; Navab, N.; Ahmadi, S.A. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*; IEEE: Piscataway, NJ, USA, 2016; pp. 565–571. [\[CrossRef\]](#)
- Sudre, C.H.; Li, W.; Vercauteren, T.; Ourselin, S.; Cardoso, M.J. Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Springer: Cham, Switzerland, 2017; pp. 240–248. [\[CrossRef\]](#) [\[PubMed\]](#)
- Wilks, D.S. *Statistical Methods in the Atmospheric Sciences*, 3rd ed.; Academic Press: Cambridge, MA, USA, 2011.
- Gilleland, E.; Ahijevych, D.; Brown, B.G.; Casati, B.; Ebert, E.E. Intercomparison of spatial forecast verification methods. *Weather Forecast.* **2009**, *24*, 1416–1430. [\[CrossRef\]](#)
- Fisher, A.; Rudin, C.; Dominici, F. All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *J. Mach. Learn. Res.* **2019**, *20*, 1–81.
- Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30 (NIPS 2017)*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; pp. 4765–4774.
- Naz, F.; She, L.; Sinan, M.; Shao, J. Enhancing radar echo extrapolation by ConvLSTM2D for precipitation nowcasting. *Sensors* **2024**, *24*, 459. [\[CrossRef\]](#) [\[PubMed\]](#)
- She, L.; Zhang, C.; Man, X.; Shao, J. LLMDiff: Diffusion model using frozen LLM transformers for precipitation nowcasting. *Sensors* **2024**, *24*, 6049. [\[CrossRef\]](#) [\[PubMed\]](#)

28. Guyon, I.; Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182.
29. Molnar, C. *Interpretable Machine Learning*; Lulu.com: Morrisville, NC, USA, 2020. Available online: <https://books.google.com/books?id=jBm3DwAAQBAJ> (accessed on 10 April 2026).
30. Zhuge, X.; Zou, X. Summertime convective initiation nowcasting over southeastern China based on Advanced Himawari Imager observations. *J. Meteorol. Soc. Jpn. Ser. II* **2018**, *96*, 337–353. [[CrossRef](#)]
31. Iwabuchi, H.; Putri, N.S.; Saito, M.; Tokoro, Y.; Sekiguchi, M.; Yang, P.; Baum, B.A. Cloud property retrieval from multiband infrared measurements by Himawari-8. *J. Meteorol. Soc. Jpn. Ser. II* **2018**, *96B*, 27–42. [[CrossRef](#)]
32. Kim, M.; Im, J.; Park, H.; Park, S.; Lee, M.-I.; Ahn, M.-H. Detection of tropical overshooting cloud tops using Himawari-8 imagery. *Remote Sens.* **2017**, *9*, 685. [[CrossRef](#)]
33. Gong, J.; Bai, L.; Ye, P.; Xu, W.; Liu, N.; Dai, J.; Yang, X.; Ouyang, W. CasCast: Skillful high-resolution precipitation nowcasting via cascaded modelling. In *Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria, 21–27 July 2024*; PMLR: Cambridge, MA, USA, 2024; Volume 235, pp. 15809–15822. Available online: <https://proceedings.mlr.press/v235/gong24a.html> (accessed on 8 July 2026).
34. Peng, C.; Liu, F.; Zhu, B.; Ma, M.; Zhou, H.; Qin, Z.; Lu, G.; Wang, W.; Wang, Y. Observations of single-stroke flashes from five isolated small thunderstorms in East China. *J. Atmos. Sol.-Terr. Phys.* **2020**, *211*, 105441. [[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.