Near-Real-Time Integration of Multisource Precipitation Products Using a Multiscale Convolutional Neural Network

DUONG THUY BUI,[a] TRUONG XUAN NGO,[a] THANH THI NHAT NGUYEN,[a] ANH DUC HOANG GIA,[a] BUI THI KHANH HOA,[b] HOANG ANH NGUYEN-THI,[b] TAN VAN PHAN,[c] FEDERICO PORCU,[d] AND AN HUNG NGUYEN[e]

[a] *University of Engineering and Technology, Vietnam National University, Hanoi, Vietnam*
[b] *National Centre for Hydro-Meteorological Network (NCN), Viet Nam Meteorological and Hydrological Administration, Hanoi, Vietnam*
[c] *University of Science, Vietnam National University, Hanoi, Vietnam*
[d] *Department of Physics and Astronomy "Augusto Righi", University of Bologna, Bologna, Italy*
[e] *Le Quy Don Technical University, Hanoi, Vietnam*

(Manuscript received 13 September 2024, in final form 29 April 2025, accepted 14 May 2025)

ABSTRACT: Merging multisource precipitation data based on deep learning models to create an accurate rainfall dataset has received significant interest in recent years. This article proposes a deep learning model to produce a high-accuracy, near-real-time precipitation product for the north-central region of Vietnam during the period 2019–23, with a spatial resolution of 0.04° and a temporal resolution of 1 h. The input multisource data including near-real-time satellite-derived precipitation products [PERSIANN-CCS, Global Satellite Mapping of Precipitation Near–Real Time (GSMaP-NRT), and IMERG-Early Run], radar precipitation, and gauge observations and spatial features neighbor estimation (NE) and precipitation occurrence probability (POP) are merged by a multiscale convolutional neural network (CNN)–based model with focal loss function and mean-square-error loss function for classification and regression tasks, respectively. Extensive experiments demonstrate that the proposed precipitation product outperforms all the input precipitation products and the post-real-time global precipitation products including GSMaP-moving vector with Kalman filter (MVK)-Gauge and IMERG-Final Run. It achieves classification metrics with a critical success index (CSI) of 0.65 and a BIAS of 1.03, with improvements from 31.58% to 54.8% in CSI and from 17.47% to 105.82% in BIAS, compared to radar, GSMaP-MVK-Gauge, and IMERG-Final Run products. For regression metrics, it achieves an RMSE of 3.34 mm h^{-1}, and a modified Kling–Gupta efficiency (mKGE) of 0.70, with improvements from 10.18% to 100% in RMSE and from 15.71% to 71.43% in mKGE over the same reference products. These results indicate that the merged product has a greater capability to detect rainfall events and significantly better overall performance, with lower systematic and random errors compared to the same reference products. Moreover, the proposed method outperforms the other methods, including random forest, long short-term memory, and the original multiscale CNN.

SIGNIFICANCE STATEMENT: We propose a high-accuracy hourly precipitation estimation method by merging multisource near-real-time precipitation products and testing on a dataset from 2019 to 2023 in the north-central region of Vietnam. The merged product has demonstrated its superiority over input products, the post-real-time global products, and other baseline models' results. This study paves the way for merging precipitation products, especially through the use of a multiscale convolutional neural network (CNN) model enhanced with focal loss and spatial input features from rain gauge observation. Moreover, our work is necessary since the study area is often heavily affected by floods under increasingly complicated conditions.

KEYWORDS: Algorithms; Data processing/distribution; Databases; Remote sensing; Neural networks; Regression analysis

1. Introduction

Precipitation is of utmost importance in different aspects of life as it plays a crucial role in the global water cycle as well as human life (Kidd and Huffman 2011). Accurate precipitation estimation is crucial for monitoring and effectively using this resource. To address this challenge, merging various precipitation data sources has gained significant attention. These sources typically include rain gauges, weather radars, and satellites. By merging them, the accuracy of ground-based measurements is combined with the broad coverage of radars and satellites, resulting in a higher-resolution precipitation product with improved accuracy compared to the individual sources. In addition to the three main sources, auxiliary data such as meteorological characteristics, topography, and reanalysis atmospheric data are also utilized to enhance the accuracy of the final merged precipitation product (Assiri and Qureshi 2022; Baez-Villanueva et al. 2020; Zhang et al. 2011).

Various methods have been proposed to address this merging issue. These methods can be categorized into two main groups: traditional approaches involving interpolation and statistical techniques and modern ones applying machine learning techniques. The former usually establishes a linear relationship between the merging inputs and the integrated precipitation product (Mastrantonas et al. 2019; Shen et al. 2014), which may

Corresponding author: An Hung Nguyen, hungan@lqdtu.edu.vn

DOI: 10.1175/JHM-D-24-0120.1

limit the ability to identify and estimate rainfall of the merged output accurately (Baez-Villanueva et al. 2020). Besides linear averaging, interpolation techniques often assume dependence of the error variance on precipitation intensity, a nonlinear correlation between the dependent variable, normal (Gaussian) distribution of precipitation, etc., which may not hold true in reality (Zhang et al. 2021). The latter using machine learning approach is becoming increasingly popular in this issue (Fang et al. 2023; Lei et al. 2022), because of its capability to process various types of data, especially multidimensional data, and capture complex nonlinear relationships from input data. Regarding the traditional machine learning techniques, Assiri and Qureshi (2022) used random forest (RF) to merge satellite retrievals, rain gauge data, and surface parameters in Saudi Arabia. Wehbe et al. (2020) combined the IMERG V06B precipitation product, radar precipitation data, and rain gauge observations using artificial neural networks (ANNs) in a hyperarid region of the United Arab Emirates. Zhang et al. (2021) proposed a double machine learning (DML) approach for integrating multisource precipitation products, combining RF for classification with various machine learning algorithms—including RF, ANN, support vector machine, and extreme learning machine—for regression.

Besides, deep learning methods have been widely used due to their more powerful ability to automatically capture spatial and temporal patterns hidden in the data (Tang et al. 2021). Shen et al. (2022) utilized long short-term memory (LSTM) to merge four satellite precipitation products with ground-based measurements over the Han River basin of China from 2007 to 2018. Lyu and Yong (2024) merged multiple precipitation products over the Tibetan Plateau using two models: extreme gradient boosting (XGBoost) for the classification task and LSTM for the regression task. Nan et al. (2023) used a convolutional neural network (CNN) model to merge multisource precipitation (i.e., gauged, gridded, satellite-derived, and dynamic downscaling products) over the Tibetan Plateau. The CNN achieved better performance than the ANN and the statistical extended triple collocation (ETC) method. To leverage the advantages of CNN with LSTM, Fang et al. (2023) proposed a convolutional LSTM (ConvLSTM) model with an attention mechanism to merge various precipitation sources, including ground observations, IMERG, Climate Hazards Infrared Precipitation with Stations (CHIRPS), ERA5 data, normalized difference vegetation index (NDVI), DEM, and streamflow series of five hydrological stations from 2011 to 2020 over the Yalong River of the Tibetan Plateau. Compared to the original precipitation products, the ConvLSTM model achieved higher accuracy [0.853 in correlation coefficient (CC) and 3.53 mm day^{-1} in root-mean-square error (RMSE)]. Although ConvLSTM utilizes both spatial and temporal factors, its ability to exploit spatial information is not as effective as that of CNN-based architecture (Sadeghi et al. 2020).

In the process of merging precipitation products, spatial factors are often considered more important than temporal ones due to their critical role in capturing the distribution and spatial variation of precipitation. Moraux et al. (2021) proposed a multiscale CNN architecture to merge rain gauge observations, precipitation radar, and thermal satellite infrared radiometer observations over Belgium, the Netherlands, and Germany for the period 2015–19. This model achieved good performance with a critical success index (CSI) of 0.629, a CC of 0.688, and an RMSE of 1.488 mm h^{-1}. Given the high variability of terrain and climate in the north-central region of Vietnam, the multiscale CNN approach could be a viable solution for merging rainfall products in this area.

In recent years, Vietnam has witnessed several extreme weather events, such as storms, floods, and droughts, which heavily affect agricultural production efficiency (Do et al. 2021). Therefore, accurate weather forecasting, particularly in estimating and predicting precipitation, is undoubtedly necessary for natural disaster precautions and management. Studies on the precipitation for the Vietnam region have so far mainly concentrated on evaluating and correcting precipitation products (Ngo-Duc et al. 2013; Thanh 2019; Trinh-Tuan et al. 2019; Kimpara et al. 2023; Tran et al. 2023; Roversi et al. 2024). Despite certain contributions, these studies have not yet exploited the advantages of merging multisource precipitation products.

In this research, we proposed a multiscale CNN-based model enhanced with additional spatial features to merge hourly multisource precipitation data, including gauge observations, radar precipitation, and three widely used near-real-time satellite precipitation products [Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks–Cloud Classification System (PERSIANN-CCS), Global Satellite Mapping of Precipitation Near–Real Time (GSMaP-NRT), and IMERG-Early Run] over the north-central region of Vietnam in the period of 2019–23. The final purpose is to create a near-real-time precipitation product with high accuracy, comparable to post-real-time precipitation products. The idea behind this proposal is to employ a multiscale CNN (M-CNN) architecture suggested by Moraux et al. (2021) due to its strong spatial feature extraction capabilities. However, this algorithm might be ineffective for point data such as rain gauge observations and imbalanced datasets. Therefore, to address this limitation, we exploit the focal loss function and enhance the spatial characteristics of rain gauge data by leveraging information from the surrounding areas of rain gauge stations rather than solely relying on point data at the gauge stations. Hence, the proposed model is a deployment of the modified M-CNN model with the additional precipitation spatial features from rain gauge data (hereafter referred to as PSM-CNN). Experimental results demonstrate that PSM-CNN can significantly improve accuracy in comparison with satellite and radar precipitation products and outperform other techniques.

The next sections of this paper are organized as follows. Section 2 introduces the study area and the data used. The methodology is presented in section 3, while the experimental results and further analysis are shown and discussed in sections 4 and 5, respectively. Finally, section 6 draws significant conclusions and future work.

2. Study area and data

a. Study area

The north-central region, including three provinces: Thanh Hóa, Nghe An, Hà Tĩnh, and its neighboring provinces in

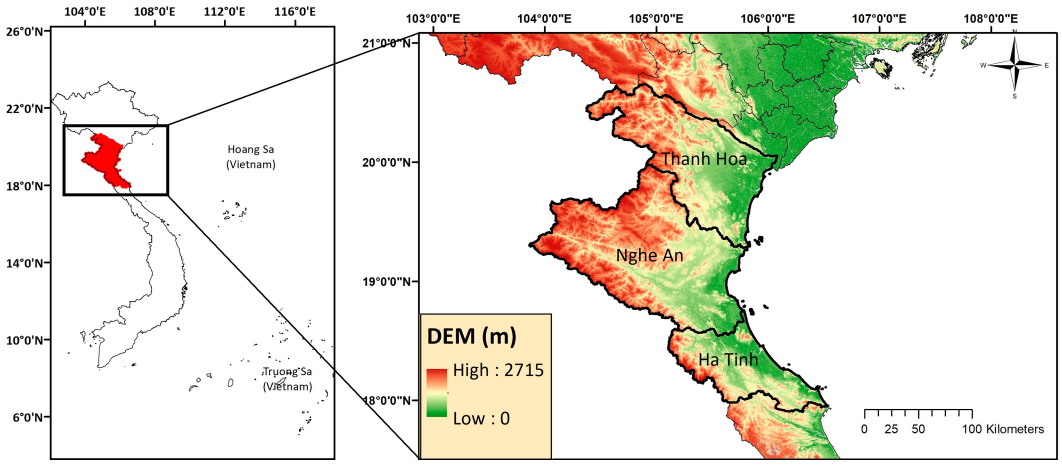

FIG. 1. Topography over the north-central region of Vietnam.

Vietnam, was selected as the study region. As illustrated in Fig. 1, the topography of the north-central region is fairly complex, with several mountains and rivers overlooking the sea. Due to the diverse topography with steep mountains and fast-flowing rivers, natural disasters, especially floods, are likely to have a severe impact on production activities and community life in the north-central region (Vu and Ranzi 2017). This area has a tropical-monsoon climate, with two distinct seasons, dry season and rainy season. The rainy season, accounting for 68%–75% of annual rainfall, often causes severe flooding and significant damage to production, property, people, and the environment (Casse et al. 2015). As a result, the process of gathering and measuring rainfall in this region is impeded.

b. Data

1) GAUGE OBSERVATIONS

Gauge precipitation data used in this study are collected from 726 automatic weather stations in the north-central region for the period 2019–23. The rain gauges measure precipitation accumulation every 10 min, from which we aggregated hourly precipitation observations. The data are provided by the National Centre for Hydro-Meteorological Network (NCN) under the Vietnam Meteorological and Hydrological Administration (VNMHA), suitable for various fields of scientific research and in practical life.

2) WEATHER RADAR PRECIPITATION

The radar precipitation product used in this study has a temporal resolution of 1 h and a spatial resolution of 0.01°. It was developed and provided by NCN, VNMHA. The input data of the radar precipitation calculation method applied by NCN are the 2-km pseudoconstant altitude plan position indicator (PCAPPI) data extracted from quality-controlled reflectivity intensity of 10 radars in Vietnam every 10 min. Depending on the scanning strategy, the 2-km PCAPPI data are generated by combining reflectivity data of the 3–5 lowest elevation angles of each radar in general. The 2-km PCAPPI data are then converted into rain rate through the Marshall–Palmer relationship. Moreover, radar precipitation was calibrated using rain gauge data (Kimpara et al. 2020; Makihara 1996).

3) SATELLITE-DERIVED PRECIPITATION PRODUCTS

PERSIANN-CCS was developed by the Center for Hydrometeorology and Remote Sensing (CHRS) at the University of California, Irvine (UCI). Its algorithm exploits the relationship between geostationary infrared brightness temperature images and rainfall rates (Hong et al. 2004; Hsu et al. 2013). The product is provided with a spatial resolution of 0.04°, a temporal resolution of 1 h, and a data latency of 1 h.

GSMaP is a high-resolution global precipitation product developed by the Japan Science and Technology Agency (JST) and the Japan Aerospace Exploration Agency (JAXA). The product uses precipitation retrieval algorithms based on physical models, combining passive microwave data from low-Earth-orbiting (LEO) satellites and infrared data from geostationary satellites (Okamoto et al. 2008). GSMaP has several products to suit different user requirements and research purposes. In this study, the product GSMaP-NRT version 8 with a spatial resolution of 0.1°, a temporal resolution of 1 h, and a data latency of 4 h is selected for integration, while the product GSMaP-moving vector with Kalman filter (MVK)-Gauge version 7, with a data latency of 3 days is selected to use for performance comparison with the merged product.

Integrated Multi-satellitE Retrievals for Global Precipitation Measurement (GPM) (IMERG) is a level-3 precipitation product of GPM, operated based on the collaboration between the National Aeronautics and Space Administration (NASA) and JAXA. IMERG uses various algorithms to estimate precipitation from GPM microwave sensors, geostationary satellite infrared sensors, and rain gauge observations (Huffman et al. 2023, 2020). The IMERG-Early Run (IMERG-E) version 6 product, with a spatial resolution of 0.1°, a temporal resolution of 30 min, and a data latency of 4 h, was used in this study for integration, while the IMERG-Final Run (IMERG-F) version 7 product, calibrated with gauge data, has a 0.1° spatial resolution, 30-min temporal resolution, and 3.5-month data latency and was used for performance comparison with the merged product.

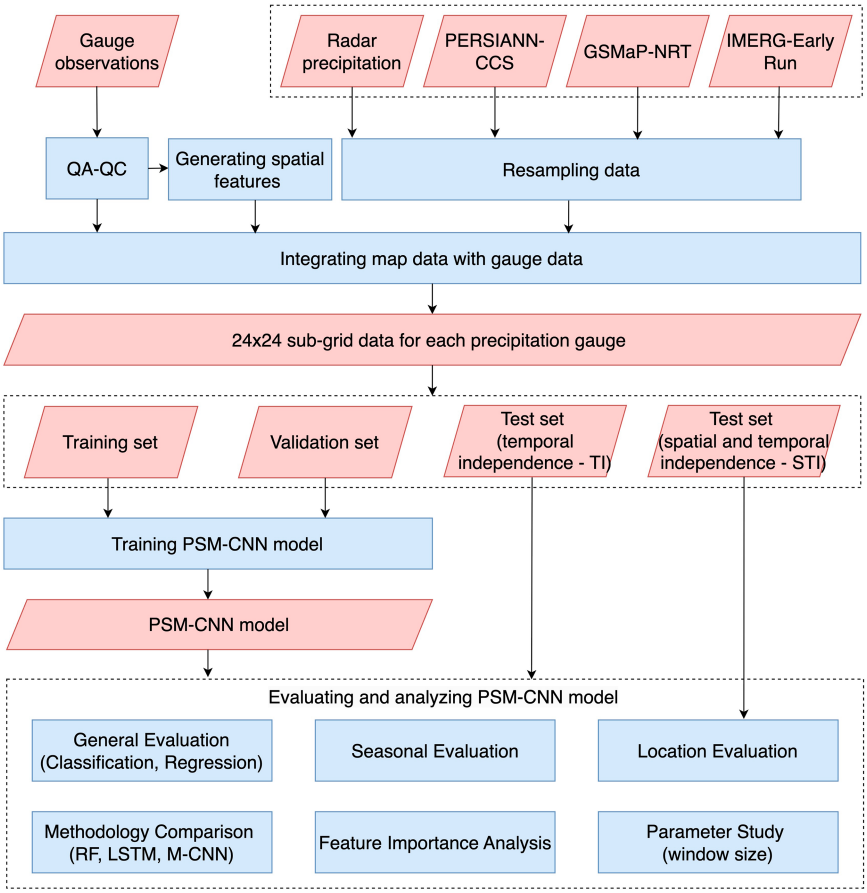

FIG. 2. Workflow of the study.

3. Methodology

The workflow of the study is illustrated in Fig. 2. First, preprocessing is applied on all datasets. After that, they are merged into samples and then divided into four distinct sets for model training and evaluation. Along with general, seasonal, and location-based evaluations of the proposed PSM-CNN, it is also compared with other algorithms such as RF, LSTM, and the original M-CNN. Besides investigating the impacts of the additional input spatial features and the loss function, the feature importance analysis and window parameter selection are conducted to provide insights into the PSM-CNN.

a. Data preprocessing

The rain gauge precipitation data are qualified by applying gross check, temporal check, and spatial check (Shafer et al. 2000). The gross check ensures that each rainfall observation data are not less than 0 and not greater than a maximum acceptable value determined by Vietnam's natural conditions standards used in construction for average maximum rainfall in a 60-min period by province. The temporal check compares the variation between consecutive observations to detect outliers. Meanwhile, the spatial test compares gauge observations with the estimated values from neighboring gauges to find outliers.

In addition, inspired by Zhang et al. (2021), the study also defines two complementary spatial features generated from gauge data to show the correlation between the data point and the surrounding stations. The two characteristics are calculated as follows:

$$\mathrm{NE} = \frac{\sum_{i=1}^{N} r_i P_i}{\sum_{i=1}^{N} r_i}, \quad r_i = \frac{1}{d_i^2}, \tag{1}$$

$$\mathrm{POP} = \frac{\sum_{i=1}^{N} r_i \mathrm{PO}_i}{\sum_{i=1}^{N} r_i}, \quad r_i = \frac{1}{d_i^2}, \quad \mathrm{PO}_i = \begin{cases} 0 & P_i = 0 \\ 1 & P_i > 0 \end{cases}, \tag{2}$$

where neighbor estimation (NE) is calculated based on the rainfall measured from neighboring gauges and precipitation occurrence probability (POP) represents the probability of rain occurrence based on neighboring gauges (Thornton et al. 1997); d_i is the distance to the ith neighboring gauge; and P_i is the rainfall recorded at the ith neighboring measuring station, while

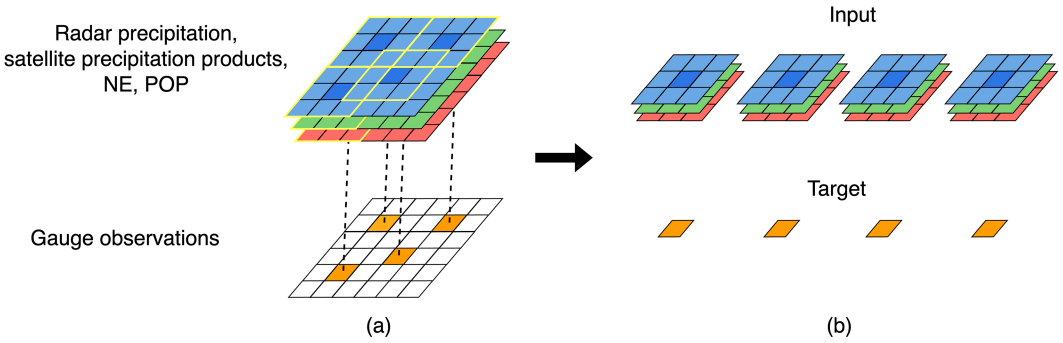

FIG. 3. An example of input cropping process. (a) Initial data; (b) data after cropping images around the station.

PO_i is a binary variable representing rain occurrence at the ith neighboring station.

Radar and satellite precipitation products need to be resampled to the same resolution with a 0.04° spatial resolution and an hourly time scale. In this study, the nearest neighbor method was used for resampling satellite data, and the average method was used for radar data. Since IMERG data have a 30-min temporal resolution, hourly precipitation data were obtained by accumulation.

b. *Data merging*

Model input is constructed as image patches of a 24 × 24 window, corresponding to a range of about 96 km × 96 km around the target station of the sample. Figure 3 illustrates an example of the process of clipping input image datasets into 3 × 3 patches. In addition, with respect to model training and evaluation, the measured rainfall from the target gauge of a patch will not be used to calculate NE and POP on this patch. The datasets are combined as a six-channel image input, including radar precipitation, satellite precipitation products (PERSIANN-CCS, GSMaP-NRT, and IMERG-E), and gauge-related features (NE and POP). For each sample, a gauge record is selected as the target.

c. *Convolutional neural network*

1) MODEL STRUCTURE

The model used in this study, simply referred to as PSM-CNN, can be divided into three main parts: the encoder, the decoder, and two subnetworks that perform two separate tasks, calculating precipitation probability and estimating rain-rate value.

The overall architecture of the model is illustrated in Fig. 4. First, the encoder gradually reduces the resolution of the feature map, allowing the filters to cover a larger spatial extent. Specifically, the input goes through two convolutional filters, each comprised of a 3 × 3 kernel filter, a batch normalization, and a rectified linear unit (ReLU) activation function. Then, the maximum pooling layer is applied to halve the size of the feature map. After two downscaling times, the typical 3 × 3 convolution layer is replaced by the inception module (Szegedy et al. 2015) to boost the ability of multiscale feature extraction. In this study, the inception layer architecture is optimized according to Szegedy et al. (2016) to reduce the computational cost of the model. After the third maximum pooling layer, the decoder progressively restores the resolution of the feature map following the same process as the encoder with the maximum pooling layer replaced by the bilinear upsampling. Simultaneously,

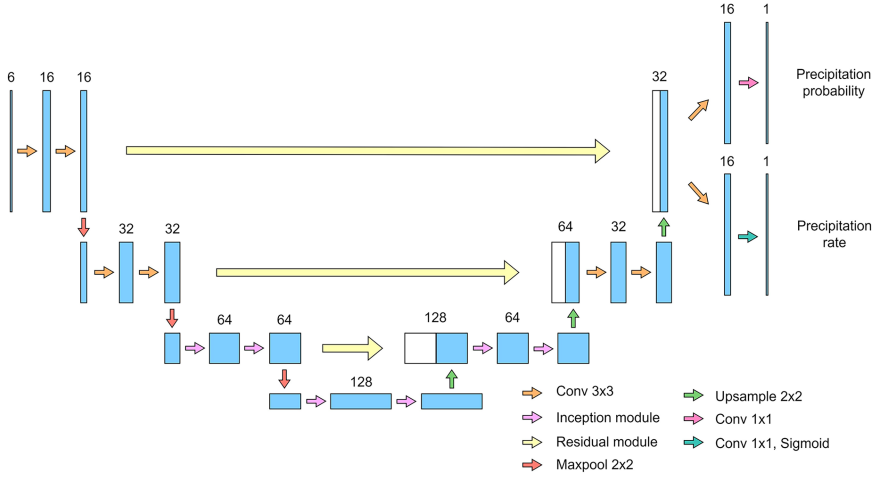

FIG. 4. Model structure.

the corresponding output from the encoder, passed through the residual layer (He et al. 2016), is combined with the feature map at the decoder. After undergoing three upsampling times, the feature map is restored to its original resolution.

Finally, at the end of the decoder, the feature map is passed through two different subnetworks, each consisting of a 3×3 convolutional layer, followed by a 1×1 convolutional layer. The sigmoid activation function is applied to the rain probability subnetwork. Instead of training two different models for each task, the study trains a multitask model to generate classification and regression results simultaneously. Besides saving training time, individual task performance can be improved compared to using single-task models (Kendall et al. 2018).

2) Loss function

In the study, mean-square-error loss is used to optimize the precipitation rate estimation branch, and focal loss (Lin et al. 2017) is applied to optimize the binary classification branch. Focal loss is an extension of cross-entropy loss, specialized for imbalanced data. Therefore, this loss function is selected for the precipitation classification task with norain data being the majority. Focal loss is defined as follows:

$$L_f(t) = -\alpha_t(1 - p_t)^{\gamma}\log(p_t), \tag{3}$$

where

$$p_t = \begin{cases} p, & y = 1 \\ 1 - p, & \text{otherwise} \end{cases},$$

$$\alpha_t = \begin{cases} \alpha & y = 1 \\ 1 - \alpha & \text{otherwise} \end{cases},$$

$y \in \{0, 1\}$ specifies the data label; $p \in [0, 1]$ is the model's estimated rain probability ($y = 1$); α is the weighting factor; and γ is the focusing parameter.

The optimization is performed on the central pixel of the patch, corresponding to the target gauge location. In addition, for the regression branch, the loss is calculated on positive samples with precipitation events recorded. The study applies the method proposed by Kendall et al. (2018) that considers task-specific uncertainty to combine different loss functions:

$$L = \frac{1}{\sigma_f^2}L_f + \frac{1}{2\sigma_{\text{mse}}^2}L_{\text{mse}} + \log\sigma_f + \log\sigma_{\text{mse}}, \tag{4}$$

The parameters σ_f and σ_{mse} are trainable and automatically optimized during model training by backpropagation.

In comparison with the original M-CNN proposed by Moraux et al. (2021), PSM-CNN has the same model structure but uses a different loss function. The focal loss is exploited instead of the binary cross-entropy loss for the classification task.

d. Experimental design

In this study, numerous experiments are conducted to validate the performance of the PSM-CNN model. The train dataset was collected from 518 gauges, while the validation dataset was collected from 70 gauges during the period from 2019 to 2022. The test dataset was collected in 2023 and separated into two datasets: the temporal independent (TI) test and the spatial and TI (STI) test. The former collected from all 588 rain gauges, representative of locations at trained and validated stations, is used only for comparing the methods in section 4e. Meanwhile, the latter, collected from 138 rain gauges (19% of the total number of stations) and representative of locations without stations, was used as reference data in sections 4b–4e to comprehensively evaluate the performance of the proposed method, other methods, and the investigated rainfall products. It should be noted that these 138 gauges are not part of the 588 gauges used for training and validation and were not used for computing the features NE and POP.

The general evaluation focuses on the performance of PSM-CNN for classification and regression tasks based on evaluation metrics, scatterplots, map visualization, and analysis of precipitation categories, including light rain ($[0.2, 1)$ mm h^{-1}), moderate rain ($[1, 5)$ mm h^{-1}), heavy rain ($[5, 30)$ mm h^{-1}), and very heavy rain ($[30, +\infty)$ mm h^{-1}). Moreover, the PSM-CNN model is also evaluated separately in two distinct seasons in 2023, particularly the dry season (January, February, March, April, November, and December) and the rainy season (May, June, July, August, September, and October). In addition, the study investigates the effectiveness of PSM-CNN in the method aspect by comparing it with RF, LSTM, and M-CNN based on both STI and TI test sets. For further discussion, the experiment on input selection, especially two additional gauge-based features (NE and POP), is conducted. We also investigate different window sizes $w \in \{8, 12, 16, 24, 28, 32\}$ to find the optimal value for calculating NE and POP features.

e. Evaluation metrics

In this study, four metrics are used to quantify the ability of the model to detect precipitation events (Table 1), namely, probability of detection (POD), false alarm rate (FAR), BIAS, and CSI. Rain/norain classification is performed by applying a threshold (0.2 mm h^{-1}) to the measured and estimated rainfall data. Classification metrics are presented on a single diagram for visual illustration purposes (Roebber 2009). Optimal performance is achieved when POD = 1 − FAR = BIAS = CSI = 1.

Also, four regression metrics are used to evaluate the model precipitation estimation capability (Table 1). Mean absolute error (MAE) and RMSE are employed to measure errors between predicted model values and actual observations. Pearson's CC represents the linear correlation between estimated precipitation and recorded gauge rainfall. Besides, modified Kling–Gupta efficiency (mKGE) (Kling et al. 2012) is composed of multiple indices: CC, bias ratio β, and variability ratio γ.

4. Experimental results

a. Input data statistics

In this study, we analyze five precipitation datasets in the period of 2019–23 to gain insight into the correlation between input data and gauge observations. Figure 5 presents the hourly seasonal precipitation of all datasets over the north-central region. It is notable that while IMERG-E shows little difference between the two seasons, the other datasets follow

TABLE 1. List of classification and regression metrics used in the evaluation. True positive (TP) is the number of precipitation events the model predicts correctly; false negative (FN) is the number of precipitation events the model misses; false positive (FP) is the number of events the model predicts as rainy when in fact there is no rain; $\beta = \mu_{\hat{y}}/\mu_y$, $\gamma = (\sigma_{\hat{y}}/\mu_{\hat{y}})/(\sigma_y/\mu_y)$; y is the gauge observation; $\hat{y}$ is the model's estimated value; μ is the mean value; and σ is the standard deviation.

Metrics		Equation	Optimal value
Classification	POD	$\text{POD} = \frac{\text{TP}}{\text{TP} + \text{FN}}$	1
	FAR	$\text{FAR} = \frac{\text{FP}}{\text{TP} + \text{FP}}$	0
	BIAS	$\text{BIAS} = \frac{\text{TP} + \text{FP}}{\text{TP} + \text{FN}}$	1
	CSI	$\text{CSI} = \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}}$	1
Regression	MAE	$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}\lvert\hat{y}_i - y_i\rvert$	1
	RMSE	$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}$	0
	CC	$\text{CC} = \frac{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - \mu_{\hat{y}})(y_i - \mu_y)}{\sigma_{\hat{y}}}\sigma_y$	1
	mKGE	$\text{mKGE} = 1 - \sqrt{(1 - \text{CC})^2 + (1 - \beta)^2 + (1 - \gamma)^2}$	1

a similar trend: The rainfall rate in the rainy season is greater than that in the dry one, with median values increasing from 1.71 to 2.50 times. Among satellite-based products, PERSIANN-CCS is the least similar to ground-based measurements with overestimated precipitation values. Figure 5 also suggests that GSMaP-NRT and radar precipitation align more closely with gauge observations than satellite retrievals in both seasons. Moreover, the radar dataset records more heavy precipitation events (>100 mm h^{-1}) than other datasets.

b. *General evaluation*

Figure 6 presents the classification results of radar precipitation and three near-real-time satellite products—PERSIANN-CCS, GSMaP-NRT, and IMERG-E—used as input data, with PSM-CNN as the output through the performance diagram. Additionally, the output is compared with post-real-time global satellite precipitation products, namely, IMERG-F and GSMaP-MVK-Gauge, through this diagram. The classification performance evaluation of all these rainfall products is based on the STI dataset.

It is evident that the proposed PSM-CNN model outperforms both the input rainfall products and the post-real-time rainfall products across all metrics, with a POD of 0.73, a FAR of 0.29, a BIAS of 1.03, and a CSI of 0.57. Specifically, the POD, FAR, BIAS, and CSI indices of the proposed model show improvements of 28.76%, 34.48%, 17.47%, and 31.58% for radar; 10.96%, 137.93%, 105.82%, and 54.38% for GSMaP-MVK-Gauge; and 28.76%, 124.13%, 43.68%, and 52.63% for IMERG-F, respectively.

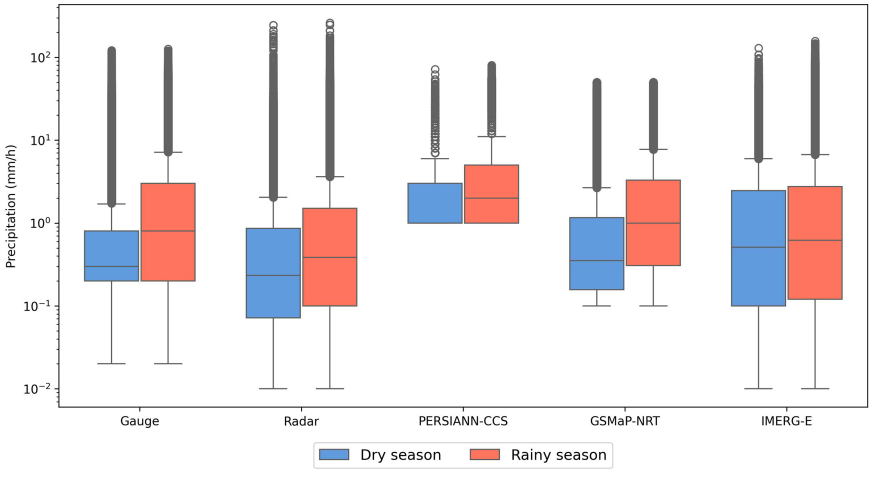

FIG. 5. Boxplot of hourly precipitation for five datasets in dry (red color) and rainy (blue color) seasons.

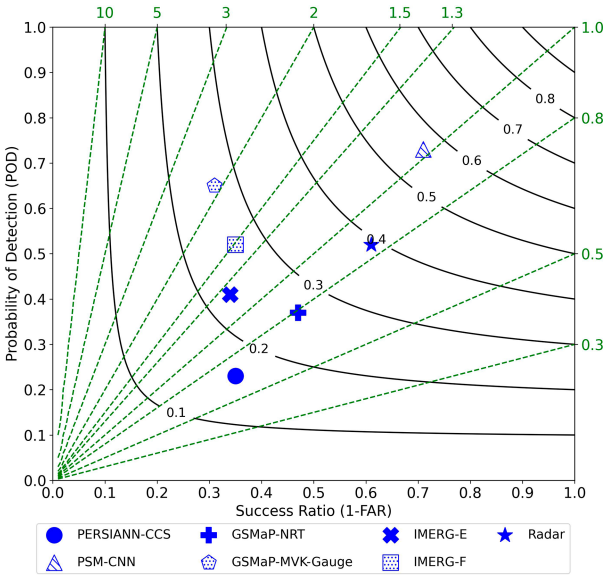

FIG. 6. Performance diagram for the merged precipitation product (PSM-CNN), the input NRT precipitation products (PERSIANN-CCS, GSMaP-NRT, IMERG-E, and radar), and post-real-time precipitation products (GSMaP-MVK-Gauge and IMERG-F). The green dashed line represents BIAS, while the black solid line represents CSI.

The addition of extra input features like NE and POP, along with the use of a focal loss function, likely contributes to PSM-CNN's enhanced performance in the precipitation identification task.

Besides the rain detection task, PSM-CNN is also validated in terms of the rain estimation task with statistics computed only on precipitation events. Figure 7 shows comparison results of precipitation obtained from radar, the three input satellite products, the two post-real-time global precipitation products, and PSM-CNN with rain gauge measurements. In terms of error metrics (Figs. 7a,b), PSM-CNN demonstrates the lowest error values among the six precipitation products investigated, with an MAE of 1.44 mm h^{-1} and an RMSE of 3.34 mm h^{-1}. When compared to the radar product, which has the second-highest accuracy (with an MAE of 1.54 mm h^{-1} and an RMSE of 3.68 mm h^{-1}), PSM-CNN shows an improvement of 6.94% in MAE and 10.18% in RMSE. Additionally, when compared to the post-real-time rainfall product GSMaP-MVK-Gauge, which ranks third in accuracy, PSM-CNN outperforms it with improvements of 49.31% in MAE and 48.80% in RMSE. Both IMERG-E and IMERG-F show lower performance compared to the other five products, with IMERG-E having an MAE of 3.37 mm h^{-1} and an RMSE of 7.14 mm, while IMERG-F has an MAE of 3.41 mm h^{-1} and an RMSE of 6.68 mm.

Similarly, in the two remaining metrics (Figs. 7c,d), PSM-CNN demonstrates its superiority over other precipitation products. Even the radar precipitation product shows lower values in both CC and mKGE compared to PSM-CNN, with a CC of 0.73 versus 0.77 for PSM-CNN and an mKGE of 0.59 versus 0.70, respectively. Notably, PSM-CNN achieves significant improvements of 48.05% and 51.95% in CC and 71.43% and 62.86% in mKGE compared to the post-real-time precipitation products, including GSMaP-MVK-Gauge and IMERG-F, respectively. PERSIANN-CCS shows the lowest performance, with a CC of 0.26 and a negative mKGE score of −0.03, which suggests that it may not effectively capture the actual precipitation distribution, even though it performs competitively with GSMaP-NRT in other metrics (Figs. 7a,b).

Overall, PSM-CNN has demonstrated superior accuracy and better alignment with ground-based measurements compared to

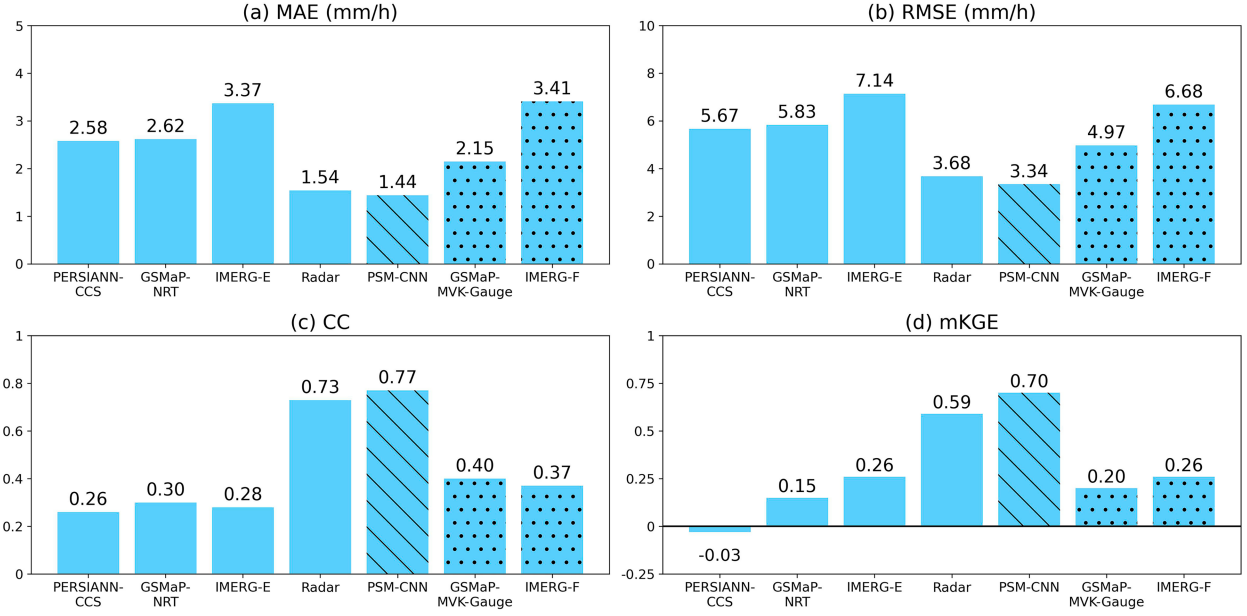

FIG. 7. Bar plots of regression metrics for the merged precipitation product (PSM-CNN), the input NRT precipitation products (PERSIANN-CCS, GSMaP-NRT, IMERG-E, and radar), and post-real-time precipitation products (GSMaP-MVK-Gauge and IMERG-F).

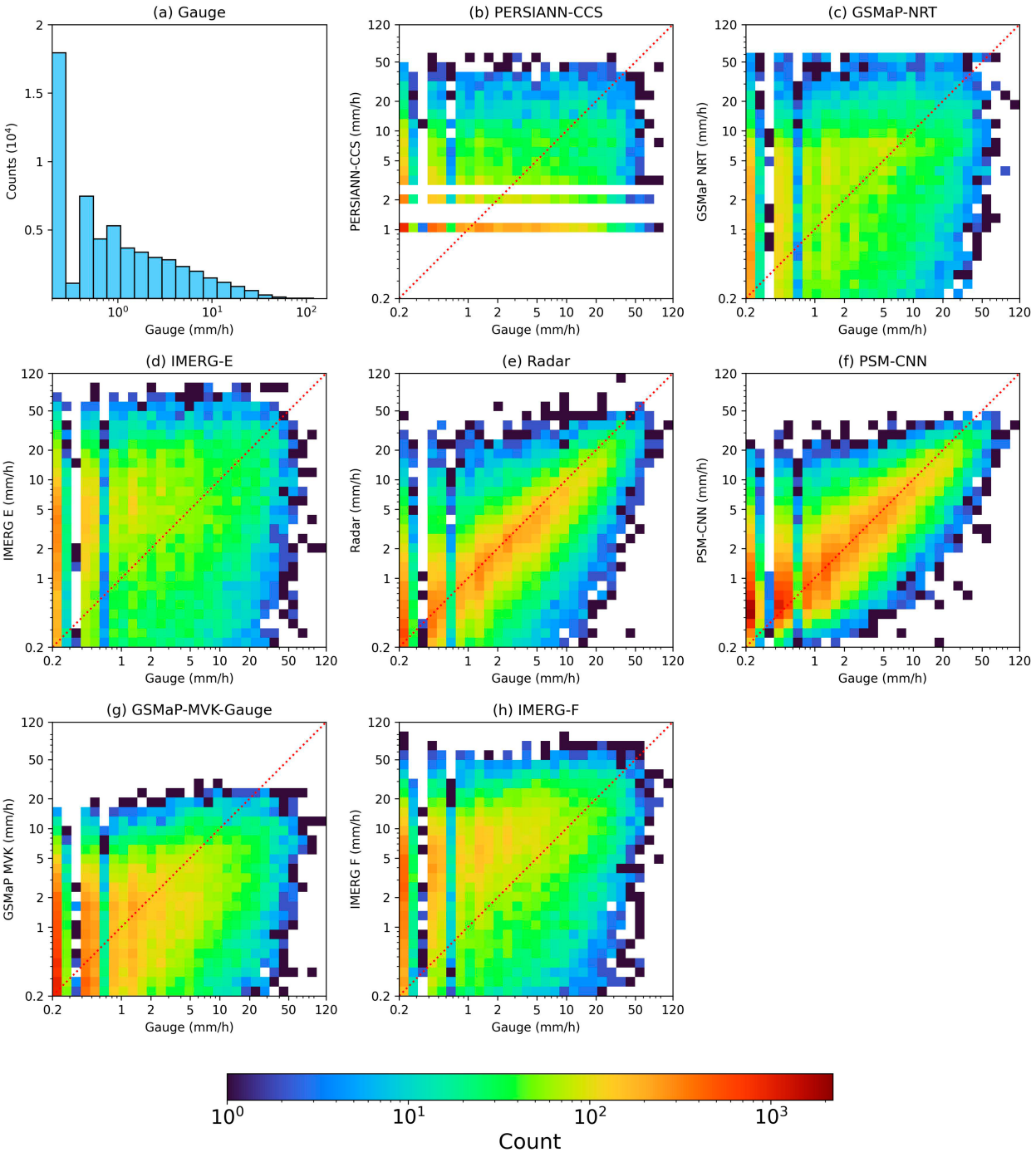

FIG. 8. (a) Histogram of gauge data and (b)–(h) scatterplots for the relationship between the precipitation products (input, merged, and post-real-time ones) and gauge observations, excluding nonprecipitating gauge samples.

the input precipitation products. It also shows significant improvements in both classification and regression performance when compared to the post-real-time precipitation products.

For further comparison, scatterplots in Fig. 8 visualize the correlation between gauge observations and seven precipitation products, including PSM-CNN, four input products, and two post-real-time products. For rain rate below 1 mm h^{-1}, the distribution of ground measurements is coarser than other products (Roversi et al. 2024), resulting in distinct peaks around 0.2 and 0.4 mm h^{-1} in Fig. 8a and higher vertical bands on the left in Figs. 8b–h. It is apparent that all seven products overestimate rainfall below 1 mm h^{-1} and underestimate rainfall above 20 mm h^{-1}.

Among the input precipitation products, PERSIANN-CCS displays data points scattered across the range of 1–60 mm h^{-1},

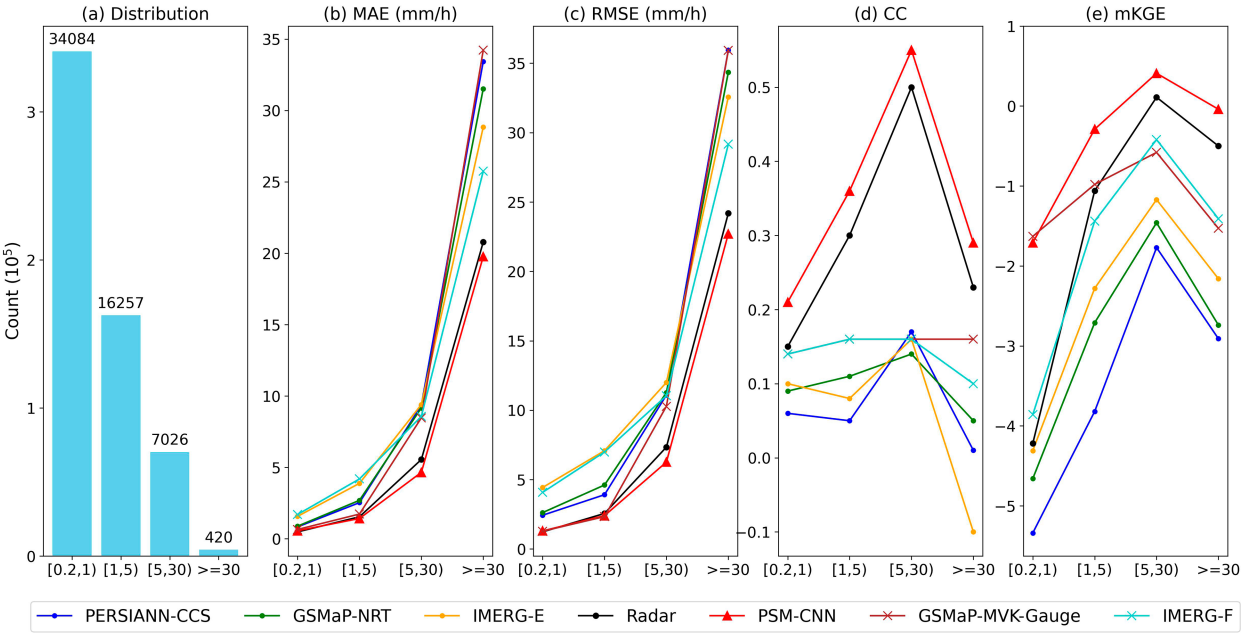

FIG. 9. (a) Data distribution per rain intensity class and (b)–(e) line plots of regression metrics for the merged precipitation product, PSM-CNN, and the other precipitation products at the four classes (mm h^{-1}).

with most values falling below 35 mm h^{-1}, showing minimal correlation with ground-based measurements. However, due to the skewed distribution of gauge observations, which are dominated by light and moderate rainfall and significantly lack heavy rainfall (Fig. 8a), PERSIANN-CCS, with its low-range estimates, exhibits error statistics comparable to GSMaP-NRT and even lower than IMERG-E (Figs. 7a,b). Meanwhile, GSMaP-NRT and IMERG-E data points are more tightly aligned along the diagonal compared to those of PERSIANN-CCS. However, both tend to overestimate values below 5 mm h^{-1}, with IMERG-E showing more pronounced overestimations, as its data points can exceed 90 mm h^{-1} within this range and are mostly concentrated above the diagonal. As a result, despite showing good agreement with gauge observations, IMERG-E exhibits the highest errors compared to the other input products (Fig. 7a). Radar data show the best linear correlation with gauge observations among the input precipitation products, with its data points densely clustered along the diagonal. However, some points are oriented toward the axes, indicating a bias in precipitation estimation.

The post-real-time precipitation products, including GSMaP-MVK-Gauge and IMERG-F, show a better level of concentration around the diagonal than the three input satellite precipitation products. However, GSMaP-MVK-Gauge tends to underpredict, with its predicted values not exceeding 30 mm h^{-1} and underestimating values above 2 mm h^{-1}. IMERG-F, on the other hand, tends to overestimate values below 10 mm h^{-1} and underestimate values above 20 mm h^{-1}. As shown in Figs. 7e–h, the degree of scatter around the main diagonal across the entire precipitation range for these two products is worse compared to radar and the proposed precipitation product, PSM-CNN.

The PSM-CNN significantly improves the correlation with ground-based measurements. It features a narrower and more diagonal-concentrated distribution compared with radar as well as minimizes the bias problem associated with radar. However, it still overestimates values below 0.5 mm h^{-1} and underestimates severe rain events exceeding 30 mm h^{-1}.

To assess the products' ability to estimate precipitation across different types, we conduct an experiment on samples from the four intensity ranges. The regression indices of seven products are shown in Fig. 9. As seen in Fig. 9a, most data fall into the light rain category ([0.2, 1) mm h^{-1}), while the very heavy rain category contains only 420 samples or 1.23%.

Regarding error metrics (MAE and RMSE), all the products exhibit a similar increasing trend across rising rainfall intensity ranges. PERSIANN-CCS has comparable results for light ([0.2, 1) mm h^{-1}), moderate ([1, 5) mm h^{-1}), and heavy ([5, 30) mm h^{-1}) rain events, gradually increasing from 0.83 to 2.53 to 9.35 in MAE and from 2.43 to 3.91 to 11.09 in RMSE. It then spikes to 33.43 in MAE and 35.94 in RMSE for very heavy rainfall (≥30 mm h^{-1}), recording the highest errors among all products. Similarly, GSMaP-NRT errors range from 0.88 to 9.17 in MAE and from 2.62 to 11.27 in RMSE for intensity classes below 30 mm h^{-1} and suddenly increase to 31.52 and 34.24, respectively. In contrast, IMERG-E maintains the lowest ranking for the three lower ranges (1.57, 3.88, and 9.40 in MAE and 4.43, 7.08, and 12.00 in RMSE) and performs fairly well for very heavy rainfall (28.83 in MAE and 32.57 in RMSE). It is obvious that radar exhibits the best results for all intensity classes among the input products, ranging from 0.48 to 20.78 in MAE and from 1.22 to 24.19 in RMSE.

Among the two post-real-time precipitation products, GSMaP-MVK-Gauge outperforms IMERG-F in accuracy across

the three lower precipitation levels, with MAE errors ranging from 0.65 to 8.47 and RMSE errors from 1.28 to 10.28, and is only less accurate than the radar product and PSM-CNN. For precipitation values greater than 30 mm h^{-1}, the errors of GSMaP-MVK-Gauge are comparable to those of PERSIANN-CCS, with MAE of 34.24 and RMSE of 35.92. Although IMERG-F has lower accuracy in the three lower rainfall ranges compared to GSMaP-MVK-Gauge, it achieves accuracy (with MAE of 25.74 and RMSE of 29.16) higher than that of GSMaP-MVK-Gauge and the three input satellite rainfall products in the very heavy rainfall range and is only less accurate than the radar product and PSM-CNN.

Although PSM-CNN shows little noticeable difference in the error statistics with radar for light and moderate rain types, it achieves remarkable improvement from this best component product at heavy and very heavy rainfall, increasing by 19.35% (MAE) and 17.25% (RMSE) for the former and 5.16% (MAE) and 6.52% (RMSE) for the latter.

Besides, for CC and mKGE results, all the evaluated products gradually rise to a peak at heavy rainfall and then mostly decline at very heavy rainfall, except for the CC statistic of GSMaP-MVK-Gauge, which levels off. It has shown the increase in CC when the precipitation amount increases and the low accuracy for very heavy rainfall estimation of all considered precipitation products. Generally, all the satellite rainfall products have relatively low correlation (CC) with the range from −0.1 to 0.17. Among them, PERSIANN-CCS shows lower performance than the other products for all ranges, reporting CC between 0.01 and 0.17 and mKGE between −5.34 and −1.77. Radar yields the best results among the input precipitation products.

For the post-real-time rainfall products, GSMaP-MVK-Gauge and IMERG-F show similar CC values (from 0.14 to 0.16) for the three lower rainfall ranges, while for the very heavy rainfall, the former has a better CC value than the latter (0.16 compared to 0.10). For the mKGE indices, GSMaP-MVK-Gauge performs better than IMERG-F for light and moderate rainfall ranges, while for heavy and very heavy rainfall ranges, the mKGE values of the two products are quite similar. These two post-real-time products have lower CC than radar. However, they have lower mKGE values in the heavy and very heavy rainfall ranges, while showing higher values in the light, when compared to radar. In the moderate rainfall category, the mKGE value of GSMaP-MVK-Gauge is approximately equal to that of radar, while the value of IMERG-F is lower. From Fig. 9, it is clearly shown that PSM-CNN achieves the best CC and mKGE values among all the products compared.

In Fig. 10, hourly precipitation maps of precipitation products in a particular event of the rainy season are shown. The event began at 2000 UTC 29 October 2023. For gauge observations, it is apparent that rainfall is heavier and more abundant in the southern region than in the northern one, with several stations recording rainfall amount exceeding 60 mm.

Most input component precipitation products, especially PERSIANN-CCS, fail to capture rainfall events in the northern part. Meanwhile, GSMaP-NRT detects several precipitation events in this region, albeit overestimating precipitation occurrence. IMERG-E and radar maps share similar patterns with the latter having finer resolution as well as closer estimations for heavy rainfall than the former. Figures 10g and 10h show that GSMaP-MVK-Gauge underestimates rainfall events and fails to detect very heavy rainfall events, while IMERG-F tends to overestimate rainfall events, with an expanded very heavy rainfall area (indicated by red color).

It is evident that PSM-CNN performs fairly well in both rain classification and regression tasks on this map, identifying both the northern and southern rainy areas as well as localizing the heavy rainfall in the south. Moreover, PSM-CNN could provide estimations closely related to gauge observations for high-intensity values, showing potential to resolve common underestimation problems.

c. *Seasonal evaluation*

The PSM-CNN is evaluated using data from two distinct seasons: dry and rainy. First, the classification results are visualized in Fig. 11. In general, it is apparent that, for the rain classification task, all precipitation products perform less effectively in the dry season (symbol in blue) than in the rainy one (symbol in red), mainly due to a decline in detection capability.

Among the input rainfall products, radar shows the highest classification performance in both the rainy and dry seasons, with a POD of 0.22, CSI of 0.20, BIAS of 0.35, and FAR of 0.36 in the dry season and a POD of 0.70, CSI of 0.48, BIAS of 1.16, and FAR of 0.40 in the rainy season. GSMaP-NRT ranks second for all metrics except the rainy season POD, where it underperforms compared to IMERG-E (0.51 vs 0.59). PERSIANN-CCS shows the lowest performance in both seasons.

For the post-real-time rainfall products, both GSMaP-MVK-Gauge and IMERG-F exhibit relatively high BIAS and FAR values in both seasons, with BIAS ranging from 1.63 to 2.42 and FAR from 0.66 to 0.71 for the former and BIAS from 0.50 to 2.09 and FAR from 0.62 to 0.65 for the latter. There is a significant difference in POD between the rainy and dry seasons for the IMERG-F product (0.73 for the rainy season and 0.19 for the dry season), whereas this difference is smaller for the GSMaP-MVK-Gauge rainfall product (0.71 for the rainy season and 0.55 for the dry season).

It can be observed that PSM-CNN outperforms all other precipitation products, reaching a CSI of 0.61 in the rainy season and 0.50 in the dry season. This is contributed by a noticeable improvement in POD and FAR. Specifically, in the dry season, the POD shows significant improvements of 67.16%, 17.91%, and 71.64%, while the FAR improves by 9.09%, 100%, and 87.88% when compared to radar, GSMaP-MVK-Gauge, and IMERG-F, respectively. In contrast, during the rainy season, the POD improves by 9.09%, 7.79%, and 5.19%, and the FAR sees improvements of 53.85%, 173.08%, and 150%, relative to the same three products.

In addition, we also assess the ability of precipitation products to estimate rainfall in these two different seasons. The regression results are presented in Fig. 12. All products are reported to have higher errors in the rainy season than in the dry season (2.01–2.86 times in MAE and 1.73–1.95 times in RMSE) as rainfall in the rainy season tends to be heavier than in the dry one (Fig. 5). It can be seen that IMERG-E has

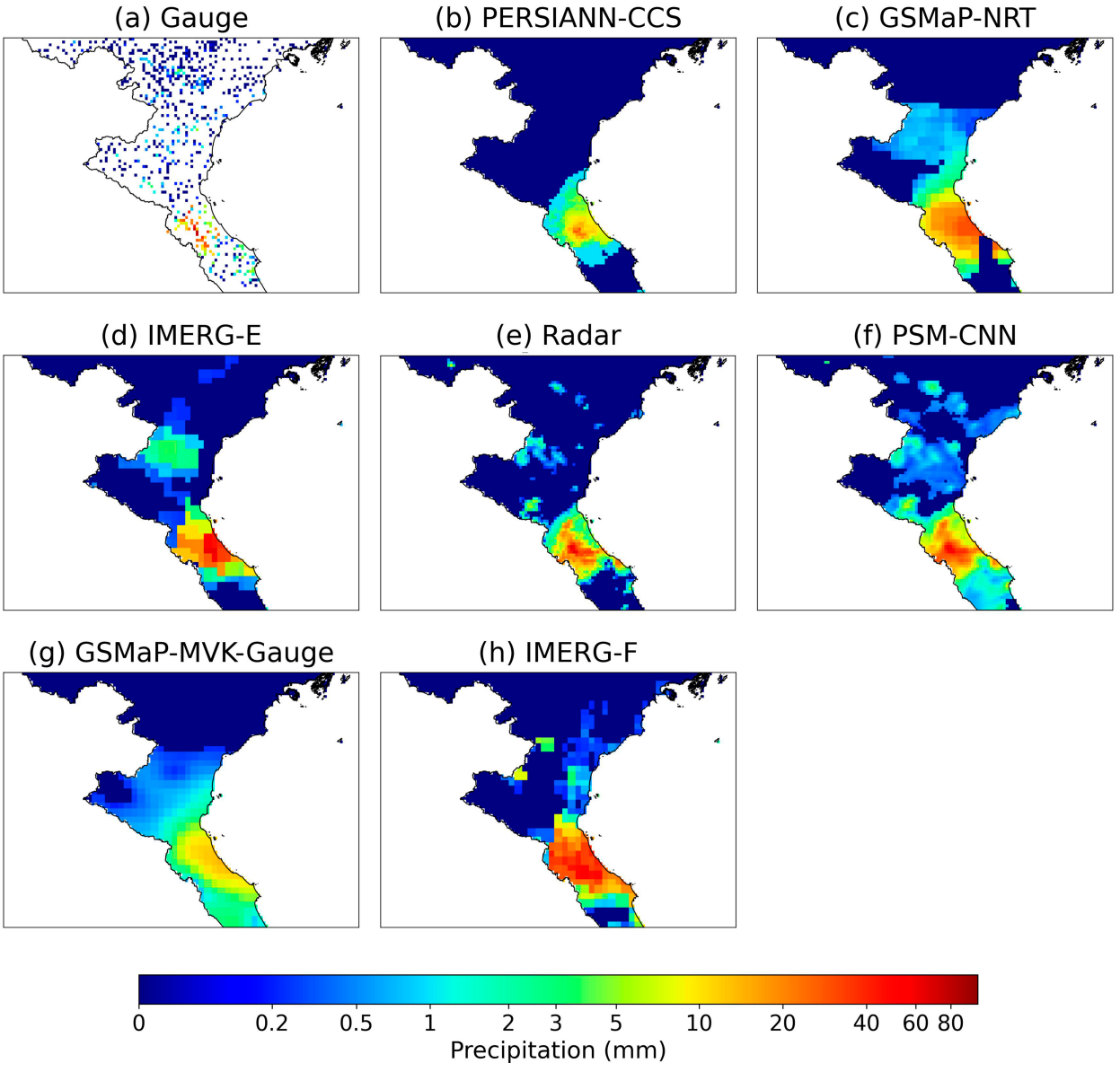

FIG. 10. Hourly precipitation maps of precipitation products in a particular case starting from 2000 UTC 29 Oct 2023.

the highest errors among the input products due to its severe overestimations, while radar yields the lowest error results (Figs. 12a,b). PERSIANN-CCS and GSMaP-NRT show relatively similar error statistics, but GSMaP-NRT exhibits higher CC and mKGE values than PERSIANN-CCS. Regarding these two metrics, radar also achieves the best agreement results among the input products. However, PSM-CNN still outperforms radar in all four of these metrics.

For the post-real-time rainfall products, IMERG-F has higher MAE and RMSE errors than GSMaP-MVK-Gauge in both seasons, and its errors are only lower than those of IMERG-E. In contrast, GSMaP-MVK-Gauge shows better accuracy than the three input satellite rainfall products but lower accuracy than radar. Additionally, GSMaP-MVK-Gauge achieves a slightly higher CC score than IMERG-F in both seasons, while the former obtains a larger mKGE score in the rainy season but a smaller mKGE value in the dry season compared to the latter. It is also clear that both post-real-time products have higher CC scores than the input satellite rainfall products in both seasons. This trend continues in the dry season for mKGE; however, their mKGE values in the rainy season are lower compared to those of GSMaP-NRT and IMERG-E. Similar to the results for error indices, both post-real-time rainfall products have lower CC and mKGE values compared to radar and PSM-CNN.

From the analysis of the results in Fig. 12, it can be concluded that PSM-CNN provides the highest accuracy according to the two criteria of MAE and RMSE compared to all the rainfall

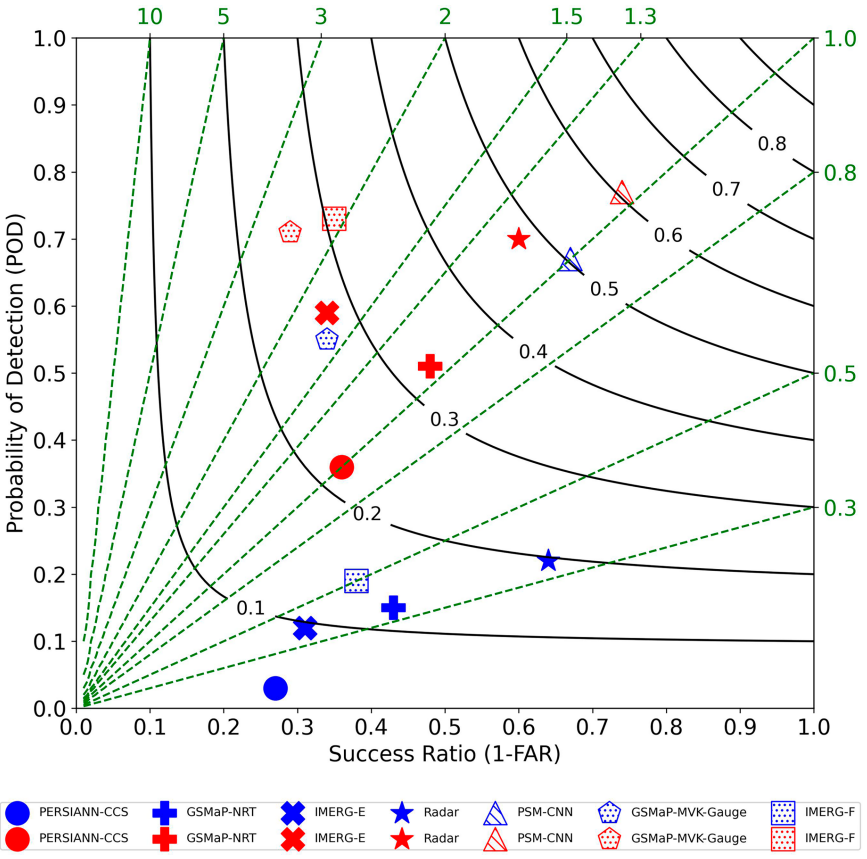

FIG. 11. Performance diagram for the merged precipitation product, PSM-CNN, and the original precipitation products in dry (blue color) and rainy (red color) seasons. The green dashed line represents BIAS, while the black solid line represents CSI.

products in both seasons. Also, it shows a noticeable performance gain in precipitation estimation, especially in CC and mKGE. In particular, the CC and mKGE values improved by 8% and 68.11% in the dry season and by 6.49% and 8.70% in the rainy season, respectively, compared to the second-best product, which is radar. For post-real-time products, the improvements are even more impressive, with CC and mKGE improving by 41.33% and 66.66% in the dry season and 53.25% and 75.36% in the rainy season, respectively, compared to GSMaP-MVK-Gauge, while for IMERG-F, the CC and mKGE scores improve by 44% and 56.52% in the dry season and 57.14% and 76.81% in the rainy season, respectively. Through seasonal validations, PSM-CNN has demonstrated its robustness in independent periods.

d. *Independent location evaluation*

Figures 13a and 13b present the spatial distribution of RMSE and CC for the merged product of PSM-CNN on the STI dataset described in section 3d. It can be seen from Fig. 13a that 91.18% of the STI datasets report RMSE values below 4.30. Meanwhile, 8.82% (12 points) have high RMSE values above 4.30, with only 2 points (1.47%) having RMSE values above 6.60. As shown in Fig. 13a, the distribution of RMSE errors is fairly uniform across the entire study area, with no region showing an excessive concentration of large errors. This demonstrates the stable accuracy of the proposed algorithm in the study area.

Similarly, 84.56% of the STI datasets report CC values above 0.63. Also, only 15.44% of the STI dataset (21 points) have CC values smaller than 0.63, with only 6.62% (9 points) having CC values smaller than 0.43. Thus, the points with high CC values dominate, while the number of points with low correlation is very small. It can be seen that the distribution of points with high or low CC values is evenly spread across the entire study area.

Overall, PSM-CNN predictions show a strong correlation with ground-based observations over the study area, exhibiting a close relationship between the precipitation amounts estimated by PSM-CNN and gauge observations.

e. *Methodology comparison*

To further investigate the capabilities of PSM-CNN, its results are compared with two other algorithms, RF and LSTM (without NE and POP), on the same experimental datasets (STI and TI described in section 3d). RF (Rigatti 2017) is an ensemble method combining multiple decision trees to improve performance and reduce overfitting. In this study, grid search is used to select the most suitable hyperparameter sets for RF. While RF primarily exploits the spatial aspect of data, LSTM (Hochreiter and Schmidhuber 1997) is designed to

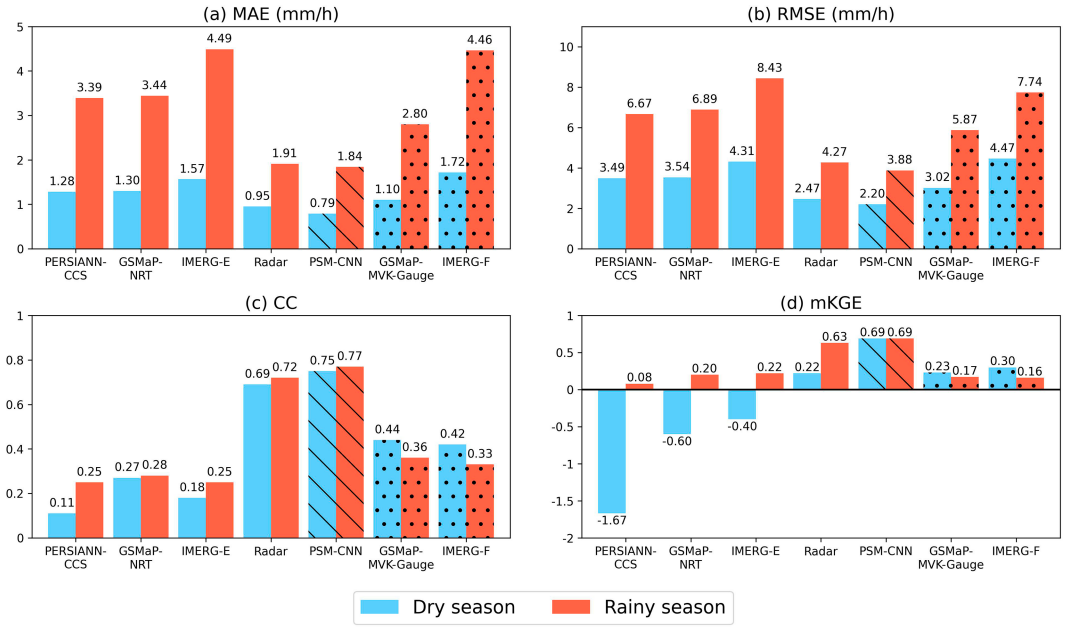

FIG. 12. Bar plots of regression metrics for the merged precipitation product, PSM-CNN, and the original precipitation products in the dry and rainy seasons.

process time-series data. It is capable of capturing long-term dependencies and mitigating the vanishing gradient problem of traditional recurrent neural network. LSTM is tested with different time steps, specifically ranging from one to seven historical hours in addition to the current hour. The model using data from the previous 3 h is selected to implement. RF and LSTM are selected to evaluate the effectiveness of PSM-CNN due to their efficiency in utilizing spatial and temporal aspects of data, respectively. Moreover, these two models use pixel-based input, while our method employs image patches. Therefore, comparing PSM-CNN with RF and LSTM could help assess the importance of incorporating extra spatial information into the input.

Moreover, to analyze the effectiveness of two additional features NE and POP, we also compare PSM-CNN with the original M-CNN which replaces NE and POP with two directly gauge-generated ones (Moraux et al. 2021). For the first feature, pixels with gauge measurements are assigned the recorded precipitation rate, while pixels without measurements are assigned the value of 0. To distinguish between pixels with a rain rate of 0 and empty pixels, the second feature represents gauge location information. Specifically, pixels with rain gauge measurements are marked as 1 and others are marked as 0. In addition, to evaluate the impact of the focal loss function and the rain gauge–based spatial features (NE and POP) on the proposed model, this model with and without using

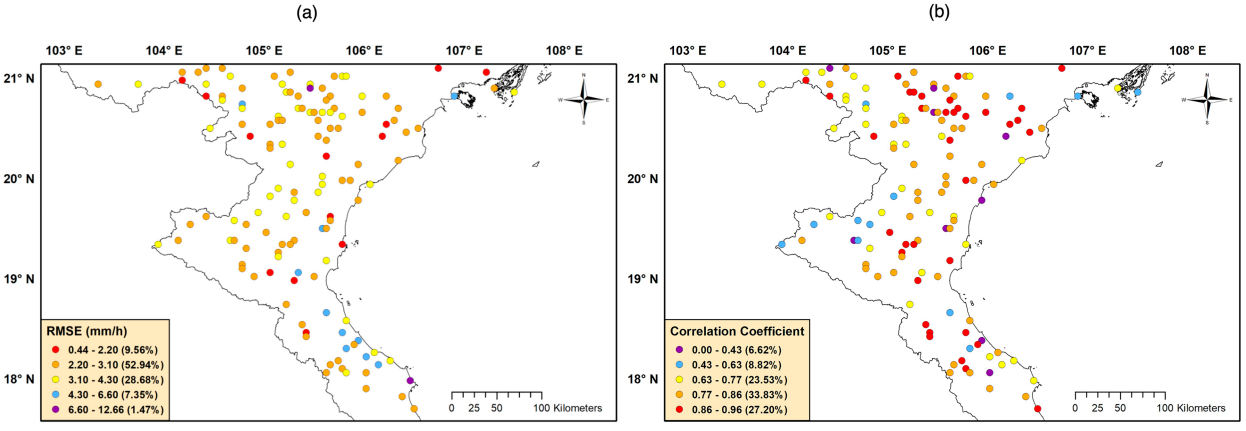

FIG. 13. Spatial distribution of (a) RMSE and (b) CC for the merged precipitation product, PSM-CNN, on the STI dataset. The legends present RMSE and CC ranges and the percentage of rain gauges at each range.

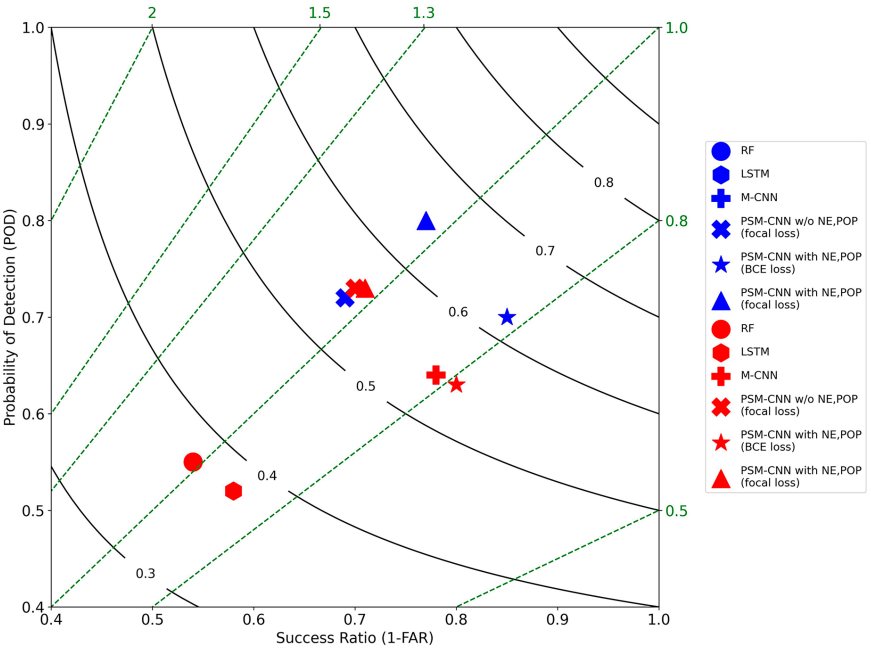

FIG. 14. Classification performance evaluation of different algorithms on the TI dataset (blue color) and STI dataset (red color).

them was deployed and compared with the M-CNN model and the other investigated algorithms.

The classification metric results of RF, LSTM, M-CNN, and PSM-CNN and different versions of PSM-CNN with and without using NE and POP features and focal loss function are presented in Fig. 14. As can be seen in this figure, classification results of RF, LSTM, and M-CNN are nearly unchanged on both TI and STI datasets. From the evaluation results on the TI dataset, it can be seen that PSM-CNN achieves the best classification quality among the products, with a POD of 0.79 and a CSI of 0.65, improving by 30.38%, 34.17%, and 18.98% in POD and 41.27%, 39.68%, and 14.28% in CSI compared to RF, LSTM, and M-CNN, respectively.

On the STI dataset, the classification performance decreases compared to the TI dataset, with POD (0.73) and CSI (0.57). However, it still outperforms all the compared products, with improvements of 24.65%, 28.76%, and 12.32% in POD and 35.08%, 33.33%, and 5.55% in CSI compared to RF, LSTM, and M-CNN, respectively. This lower result is completely reasonable, as the STI test dataset was randomly selected in both space and time and did not participate in the training process, while the TI dataset was only independent in terms of time and was involved in the training process.

As seen in Fig. 14, algorithms using the focal loss function, including PSM-CNN and PSM-CNN without using NE and POP features, tend to have BIAS values closer to the optimal line (BIAS = 1), whereas the algorithm PSM-CNN using only NE and POP and binary cross-entropy (BCE) loss function and M-CNN tend to have BIAS values further away from the BIAS optimal line in both STI and TI datasets. Moreover, there is a trade-off between POD and FAR values of the algorithms using the focal loss function and the BCE loss function. The higher values in POD are observed when using the focal loss, while the CSI values are quite similar for these algorithms. However, the two algorithms using the BCE loss function have a lower FAR compared to those using the focal loss function, with FAR values ranging from 0.16 to 0.22 for the former, and from 0.24 to 0.31 for the latter, on both TI and STI datasets. Therefore, for the classification task, it appears that the focal loss function has a more significant positive impact on the performance of the proposed rainfall product than the use of the input spatial features NE and POP.

The regression metric results of the six above investigated algorithms are presented in Fig. 15 on both TI and STI datasets. From Figs. 15a and 15b, it can be seen that, for the regression error criteria including MAE and RMSE, the error on the STI dataset is slightly higher than on the TI dataset, with the largest differences being 0.13 mm h^{-1} in MAE and 0.24 mm h^{-1} in RMSE. Moreover, PSM-CNN achieves the smallest error values among the compared methods on both the TI and STI datasets, with 1.31 and 1.44 mm h^{-1} in MAE and 3.10 and 3.34 mm h^{-1} in RMSE, respectively. Compared to M-CNN and PSM-CNN without NE and POP but using the focal loss function, PSM-CNN with NE and POP and the BCE loss function exhibits smaller MAE and RMSE error values. This indicates that the impact of NE and POP on the accuracy of PSM-CNN is better than that of the focal loss function.

Similarly, the CC and mKGE values on the STI dataset tend to be slightly lower than those on the TI dataset across all methods, with the largest differences being 0.03 in CC and 0.02 in mKGE. PSM-CNN still achieves the highest CC and mKGE values among the compared methods, with a CC of 0.77 and an mKGE of 0.70 on the STI dataset and a CC of 0.80 and an mKGE of 0.72 on the TI dataset. Both RF and

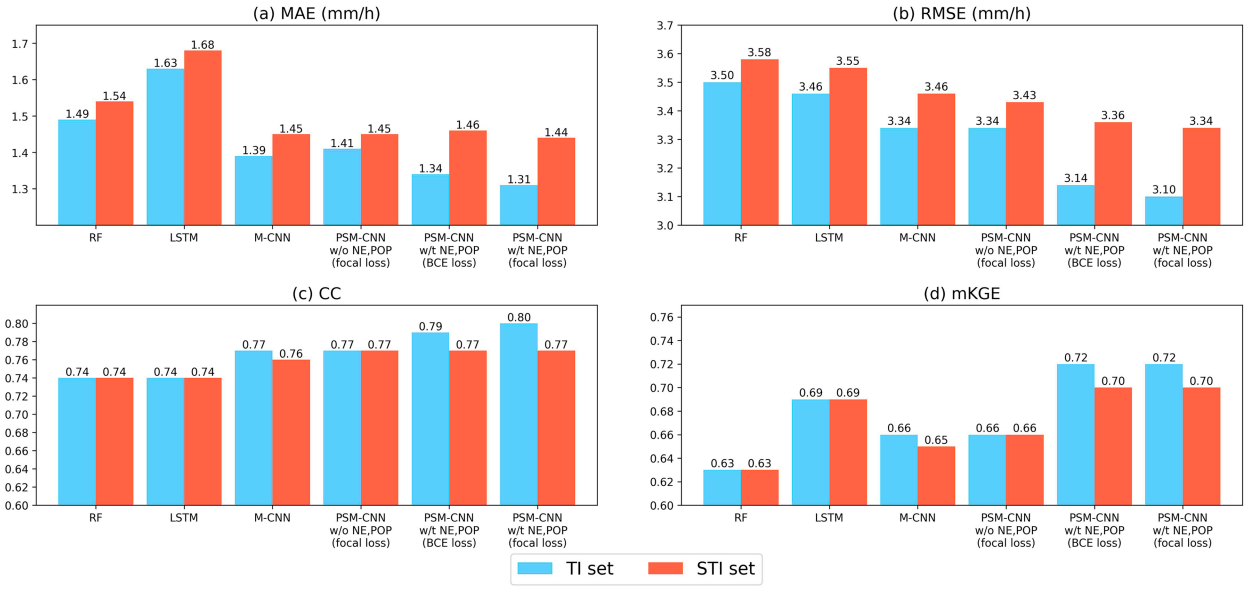

FIG. 15. Regression performance evaluation of different algorithms.

LSTM have a CC value of 0.74 on both the TI and STI datasets, while the former has an mKGE value of 0.63 and the latter has an mKGE value of 0.69 on both datasets. It can be observed that PSM-CNN using NE and POP along with the BCE loss function achieves slightly higher CC compared to M-CNN on both the TI and STI datasets and also outperforms PSM-CNN using the focal loss function without NE and POP on the TI dataset. However, for mKGE, PSM-CNN using NE and POP along with the BCE loss function shows a significantly better performance compared to both M-CNN and PSM-CNN using the focal loss function without NE and POP on both datasets. Therefore, for the regression task, it appears that NE and POP have a more positive impact on the performance of the proposed rainfall product compared to the focal loss function.

Overall, PSM-CNN has a better classification and regression performance than RF, LSTM, and M-CNN, with POD ranging from 0.73 to 0.79, CSI from 0.57 to 0.65, MAE from 1.31 to 1.44, RMSE from 3.10 to 3.34, CC from 0.77 to 0.80, and mKGE from 0.70 to 0.72. This enhanced performance is due to the contribution of the focal loss function for classification tasks and the additional input features, NE and POP, for regression tasks.

5. Further discussion

a. *Feature importance*

Analyzing feature importance could assist in better understanding the contribution of each feature to model performance. Permutation is the technique selected to estimate feature importance for the two distinct tasks of PSM-CNN, namely, classification and regression.

As shown in Fig. 16a, for the classification task, POP achieves the highest score of 0.65, followed by NE with 0.14 and radar with 0.10. POP is directly related to the precipitation identification task as it represents the precipitation occurrence probability of neighboring gauges and NE, calculated based on neighboring gauge measurements, could also provide useful

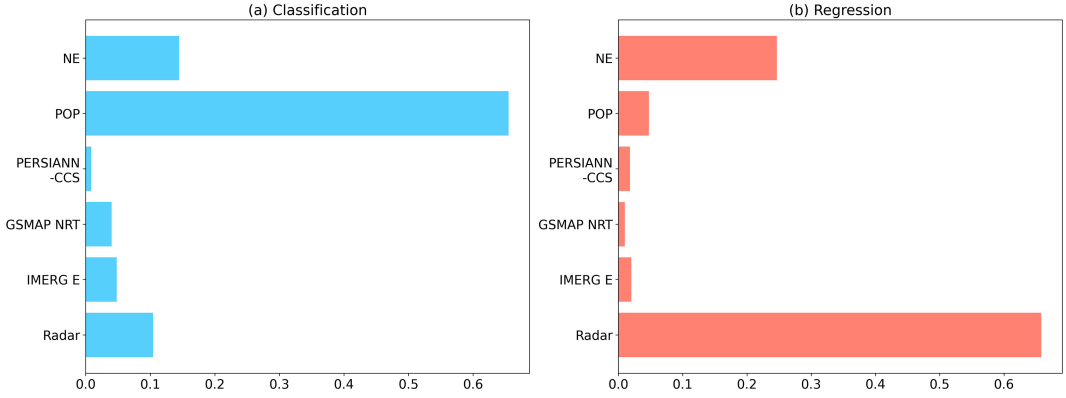

FIG. 16. Permutation feature importance of PSM-CNN for the (a) classification and (b) regression tasks.

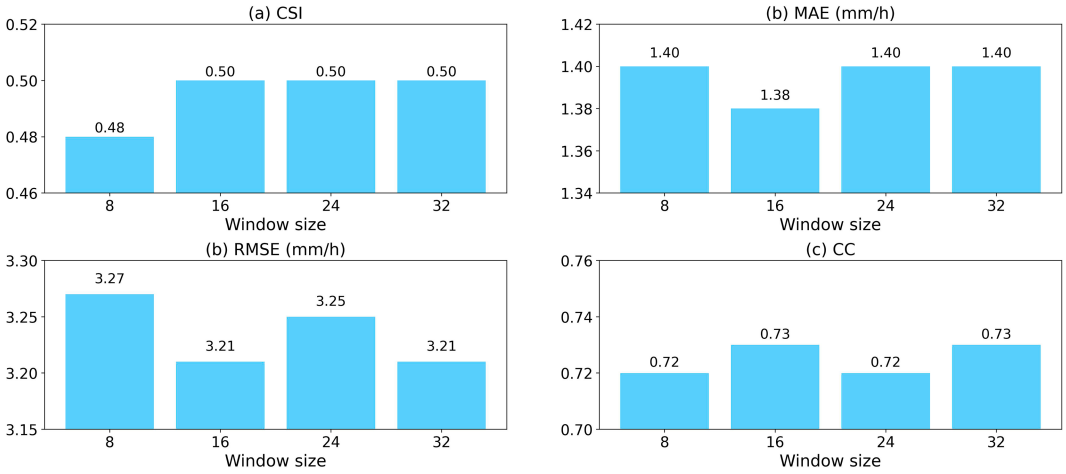

FIG. 17. Performance of PSM-CNN with different window sizes on the validation set.

information to the classification task. Meanwhile, PERSIANN-CCS ranks the lowest at 0.01 and the other products fall between 0.04 and 0.05. This is likely because the classification results of the input precipitation products are not sufficiently accurate as presented in sections 4b and 4c.

In addition, for the regression task (Fig. 16b), radar makes the largest contribution with a value of 0.66, suggesting a close relationship between radar precipitation and gauge observations. This is reasonable because radar performs relatively well for this task, superior to other products (sections 4b and 4c). Meanwhile, NE and POP rank second and third, with a relative importance score of 0.25 and 0.05, respectively, demonstrating the necessity of these two features in the precipitation estimation task. On the other hand, the other satellite rainfall products contribute minimally to the regression performance of the final proposed rainfall product, with GSMaP-NRT having the lowest importance score of 0.01, while IMERG-E and PERSIANN-CCS products have scores of 0.02. The low importance level of satellite rainfall products, compared to the very high importance level of radar in the regression task, may be attributed to the high accuracy of radar data, which covers the entire study area.

Also, the high importance scores of the NE and POP, and the radar, are also reflected in the highest cross-correlation values among all the cross-correlation values of the five input products. Specifically, the cross correlation between the radar and NE is the highest, with a value of 0.63, while the cross correlation between the radar and POP is the second highest, with a value of 0.58. Although the three input satellite rainfall products have low importance scores, their cross-correlation values with NE, POP, and radar are lowest at 0.19 (occurring between PERSIANN-CCS and POP), while the remaining values range from 0.22 to 0.46, which are still considered significant. As a result, these three satellite products are still viable and suitable choices for input.

b. Parameter study

The window size refers to the coverage range around a given point, used as the model input. In addition, it also determines the maximum distance from a neighboring gauge to the given point, applied for calculating NE and POP. The smaller window size might overlook valuable spatial details, while the larger one could introduce unnecessary information to the input. The experiment is conducted on the validation set. Figure 17 shows the result of testing this parameter across several metrics. It is evident that a window size of 16 yields the best performance in all metrics, with a CSI of 0.50, an MAE of 1.38, an RMSE of 3.21, and a CC of 0.73. Therefore, the window size of 16 is selected to use due to its optimal performance.

6. Conclusions

In this study, the multiscale CNN with additional spatial input features, named PSM-CNN, was proposed to merge hourly multisource precipitation products from gauge observations, radar precipitation, and three near-real-time satellite products (PERSIANN-CCS, GSMaP-NRT, and IMERG-Early Run) over the north-central region of Vietnam in the period from 2019 to 2023. The additional spatial input features, including NE and POP, were created from rain gauge observations to exploit the spatial correlation between the data point and the surrounding stations.

Experimental results on both classification and regression metrics have demonstrated the effectiveness of PSM-CNN compared to near-real-time component precipitation products, two post-real-time precipitation products (GSMaP-MVK-Gauge and IMERG-Final Run), and baseline models (RF, LSTM, and M-CNN). It achieves classification metric improvements of 10.96%–28.76% in POD, 34.48%–137.93% in FAR, 17.47%–105.82% in BIAS, and 31.58%–54.8% in CSI compared to the radar and the two post-real-time precipitation products. For regression metrics, it achieves corresponding improvements of 6.94%–136.80% in MAE, 10.18%–100% in RMSE, 5.19%–51.95% in CC, and 15.71%–71.43% in mKGE. In other words, PSM-CNN is capable of extracting useful information from input products to establish a high-accuracy product. Moreover, PSM-CNN, using NE and POP features,

and focal loss function, proves its efficiency compared to a spatial machine learning model, RF, a temporal deep learning model, LSTM, and the original multiscale CNN (M-CNN) without spatial feature enhancement. The study also builds a near-real-time merged precipitation map dataset with high accuracy for the north-central region in 2019 and 2023. These contributions are significant to the field of merging multisource precipitation products in Vietnam.

Finally, the limitation of this method is still the overestimation of rainfall values below 0.5 mm h^{-1} and the underestimation of very heavy rainfall events exceeding 30 mm h^{-1}, which is due to data imbalance. Despite that, this study serves as the foundation for further research on merging multisource precipitation products in Vietnam. With extensive experiments, PSM-CNN is thoroughly verified for wide applicability and has the potential for a broader-scale implementation. Furthermore, the approach presented in this paper opens future research directions for improving the accuracy of merging multisource precipitation data by enhancing deep learning architectures to deal with imbalanced datasets, improve spatial feature extraction, and increase input spatial features.

Acknowledgments. The authors thank the Vietnam-Italy bilateral project, titled "Research and Development of a Precipitation Estimation Methodology using the Integrated Rainfall Measurements Platform for Agriculture," named the INDRA project (code NDT/IT/22/07), and the National Research Project under the Ministry of Science and Technology of Vietnam, titled "Research on the Application of Optical Flow Technique to the Quantitative Estimation and Quantitative Forecast of Precipitation in Vietnam based on Himawari Satellite and Weather Radar Data" (Grant code: ĐTĐL.CN-58/21), for funding this research. The authors thank Prof. Marco Abbiati (Scientific Attaché of the Italian Embassy in Hanoi) for his continuous support and availability throughout the INDRA Project. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability statement. The integration data used in this study are available upon request. PERSIANN-CCS data can be downloaded freely from CHRS (https://chrsdata.eng.uci.edu/). GSMaP data could be accessed from the JAXA G-portal (https://gportal.jaxa.jp/gpr/). IMERG data can be retrieved from the Precipitation Data Directory of GPM (https://gpm.nasa.gov/data/directory). The authors do not have permission to share the gauge and radar data used in this study.

REFERENCES

Assiri, M. E., and S. Qureshi, 2022: A multi-source data fusion method to improve the accuracy of precipitation products: A machine learning algorithm. *Remote Sens.*, **14**, 6389, https://doi.org/10.3390/rs14246389.

Baez-Villanueva, O. M., and Coauthors, 2020: RF-MEP: A novel random forest method for merging gridded precipitation products and ground-based measurements. *Remote Sens. Environ.*, **239**, 111606, https://doi.org/10.1016/j.rse.2019.111606.

Casse, T., A. Milhøj, and T. Nguyen, 2015: Vulnerability in north-central Vietnam: Do natural hazards matter for everybody? *Nat. Hazards*, **79**, 2145–2162, https://doi.org/10.1007/s11069-015-1952-y.

Do, V. Q., M. L. Phung, D. T. Truong, T. T. T. Pham, V. T. Dang, and T. K. Nguyen, 2021: The impact of extreme events and climate change on agricultural and fishery enterprises in central Vietnam. *Sustainability*, **13**, 7121, https://doi.org/10.3390/su13137121.

Fang, W., H. Qin, G. Liu, X. Yang, Z. Xu, B. Jia, and Q. Zhang, 2023: A method for spatiotemporally merging multi-source precipitation based on deep learning. *Remote Sens.*, **15**, 4160, https://doi.org/10.3390/rs15174160.

He, K., X. Zhang, S. Ren, and J. Sun, 2016: Deep residual learning for image recognition. *2016 IEEE Conf. on Computer Vision and Pattern Recognition*, Las Vegas, NV, Institute of Electrical and Electronics Engineers, 770–778, https://doi.org/10.1109/CVPR.2016.90.

Hochreiter, S., and J. Schmidhuber, 1997: Long short-term memory. *Neural Comput.*, **9**, 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735.

Hong, Y., K.-L. Hsu, S. Sorooshian, and X. Gao, 2004: Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system. *J. Appl. Meteor.*, **43**, 1834–1853, https://doi.org/10.1175/JAM2173.1.

Hsu, K., S. Sellars, P. Nguyen, D. Braithwaite, and W. Chu, 2013: G-WADI PERSIANN-CCS GeoServer for extreme precipitation event monitoring. *Sci. Cold Arid Reg.*, **5**, 0006–0015.

Huffman, G. J., and Coauthors, 2020: Integrated multi-satellite retrievals for the Global Precipitation Measurement (GPM) Mission (IMERG). *Satellite Precipitation Measurement*, V. Levizzani, et al., Ed., Advances in Global Change Research, Vol. 67, Springer, 343–353, https://doi.org/10.1007/978-3-030-24568-9_19.

——, and Coauthors, 2023: NASA Global Precipitation Measurement (GPM) Integrated Multi-Satellite Retrievals for GPM (IMERG). Algorithm Theoretical Basis Doc. (ATBD), version 7, 52 pp., https://gpm.nasa.gov/resources/documents/imerg-v07-atbd.

Kendall, A., Y. Gal, and R. Cipolla, 2018: Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. *2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, UT, Institute of Electrical and Electronics Engineers, 7482–7491, https://doi.org/10.1109/CVPR.2018.00781.

Kidd, C., and G. Huffman, 2011: Global precipitation measurement. *Meteor. Appl.*, **18**, 334–353, https://doi.org/10.1002/met.284.

Kimpara, C., M. Tonouchi, B. T. K. Hoa, N. V. Hung, N. M. Cuong, and K. Akaeda, 2020: Quantitative precipitation estimation by combining rain gauge and meteorological radar network in Viet Nam. *Vietnam J. Hydrometeor.*, **5**, 36–50, https://doi.org/10.36335/VNJHM.2020(5).

——, ——, ——, ——, ——, and ——, 2023: Evaluation of the radar-based quantitative precipitation estimation composite in Viet Nam. *J. Hydrometeor.*, **15**, 28–39, https://doi.org/10.36335/VNJHM.2023(15).

Kling, H., M. Fuchs, and M. Paulin, 2012: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *J. Hydrol.*, **424–425**, 264–277, https://doi.org/10.1016/j.jhydrol.2012.01.011.

Lei, H., H. Zhao, and T. Ao, 2022: A two-step merging strategy for incorporating multi-source precipitation products and gauge observations using machine learning classification and regression over China. *Hydrol. Earth Syst. Sci.*, **26**, 2969–2995, https://doi.org/10.5194/hess-26-2969-2022.

Lin, T.-Y., P. Goyal, R. Girshick, K. He, and P. Dollár, 2017: Focal loss for dense object detection. *Proc. IEEE Int. Conf. on Computer Vision*, Institute of Electrical and Electronics Engineers, 2999–3007, https://doi.org/10.1109/ICCV.2017.324.

Lyu, Y., and B. Yong, 2024: A novel Double Machine Learning strategy for producing high-precision multi-source merging precipitation estimates over the Tibetan Plateau. *Water Resour. Res.*, **60**, e2023WR035643, https://doi.org/10.1029/2023WR035643.

Makihara, Y., 1996: A method for improving radar estimates of precipitation by comparing data from radars and raingauges. *J. Meteor. Soc. Japan*, **74**, 459–480, https://doi.org/10.2151/jmsj1965.74.4_459.

Mastrantonas, N., B. Bhattacharya, Y. Shibuo, M. Rasmy, G. Espinoza-Dávalos, and D. Solomatine, 2019: Evaluating the benefits of merging near-real-time satellite precipitation products: A case study in the Kinu basin region, Japan. *J. Hydrometeor.*, **20**, 1213–1233, https://doi.org/10.1175/JHM-D-18-0190.1.

Moraux, A., S. Dewitte, B. Cornelis, and A. Munteanu, 2021: A deep learning multimodal method for precipitation estimation. *Remote Sens.*, **13**, 3278, https://doi.org/10.3390/rs13163278.

Nan, T., J. Chen, Z. Ding, W. Li, and H. Chen, 2023: Deep learning-based multi-source precipitation merging for the Tibetan Plateau. *Sci. China Earth Sci.*, **66**, 852–870, https://doi.org/10.1007/s11430-022-1050-2.

Ngo-Duc, T., J. Matsumoto, H. Kamimera, and H.-H. Bui, 2013: Monthly adjustment of Global Satellite Mapping of Precipitation (GSMaP) data over the VuGia–ThuBon River Basin in Central Vietnam using an artificial neural network. *Hydrol. Res. Lett.*, **7**, 85–90, https://doi.org/10.3178/hrl.7.85.

Okamoto, K., N. Takahashi, K. Iwanami, S. Shige, and T. Kubota, 2008: High precision and high resolution global precipitation map from satellite data. *2008 Microwave Radiometry and Remote Sensing of the Environment*, Florence, Italy, Institute of Electrical and Electronics Engineers, 1–4, https://doi.org/10.1109/MICRAD.2008.4579485.

Rigatti, S. J., 2017: Random forest. *J. Insur. Med.*, **47**, 31–39, https://doi.org/10.17849/insm-47-01-31-39.1.

Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, **24**, 601–608, https://doi.org/10.1175/2008WAF2222159.1.

Roversi, G., M. Pancaldi, W. Cossich, D. Corradini, T. T. N. Nguyen, T. V. Nguyen, and F. Porcu', 2024: The extreme rainfall events of the 2020 typhoon season in Vietnam as seen by seven different precipitation products. *Remote Sens.*, **16**, 805, https://doi.org/10.3390/rs16050805.

Sadeghi, M., P. Nguyen, K. Hsu, and S. Sorooshian, 2020: Improving near real-time precipitation estimation using a U-Net convolutional neural network and geographical information. *Environ. Modell. Software*, **134**, 104856, https://doi.org/10.1016/j.envsoft.2020.104856.

Shafer, M. A., C. A. Fiebrich, D. S. Arndt, S. E. Fredrickson, and T. W. Hughes, 2000: Quality assurance procedures in the Oklahoma Mesonetwork. *J. Atmos. Oceanic Technol.*, **17**, 474–494, https://doi.org/10.1175/1520-0426(2000)017<0474:QAPITO>2.0.CO;2.

Shen, J., P. Liu, J. Xia, Y. Zhao, and Y. Dong, 2022: Merging multisatellite and gauge precipitation based on geographically weighted regression and long short-term memory network. *Remote Sens.*, **14**, 3939, https://doi.org/10.3390/rs14163939.

Shen, Y., A. Xiong, Y. Hong, J. Yu, Y. Pan, Z. Chen, and M. Saharia, 2014: Uncertainty analysis of five satellite-based precipitation products and evaluation of three optimally merged multi-algorithm products over the Tibetan Plateau. *Int. J. Remote Sens.*, **35**, 6843–6858, https://doi.org/10.1080/01431161.2014.960612.

Szegedy, C., and Coauthors, 2015: Going deeper with convolutions. *2015 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, Institute of Electrical and Electronics Engineers, 1–9, https://doi.org/10.1109/CVPR.2015.7298594.

——, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, 2016: Rethinking the inception architecture for computer vision. *2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, Institute of Electrical and Electronics Engineers, 2818–2826, https://doi.org/10.1109/CVPR.2016.308.

Tang, X., Z. Yin, G. Qin, L. Guo, and H. Li, 2021: Integration of satellite precipitation data and deep learning for improving flash flood simulation in a poor-gauged mountainous catchment. *Remote Sens.*, **13**, 5083, https://doi.org/10.3390/rs13245083.

Thanh, N. T., 2019: Evaluation of multi-precipitation products for multi-time scales and spatial distribution during 2007-2015. *Civ. Eng. J.*, **5**, 255–267, https://doi.org/10.28991/cej-2019-03091242.

Thornton, P. E., S. W. Running, and M. A. White, 1997: Generating surfaces of daily meteorological variables over large regions of complex terrain. *J. Hydrol.*, **190**, 214–251, https://doi.org/10.1016/S0022-1694(96)03128-9.

Tran, T.-N.-D., M.-H. Le, R. Zhang, B. Q. Nguyen, J. D. Bolten, and V. Lakshmi, 2023: Robustness of gridded precipitation products for Vietnam basins using the comprehensive assessment framework of rainfall. *Atmos. Res.*, **293**, 106923, https://doi.org/10.1016/j.atmosres.2023.106923.

Trinh-Tuan, L., J. Matsumoto, T. Ngo-Duc, M. I. Nodzu, and T. Inoue, 2019: Evaluation of satellite precipitation products over central Vietnam. *Prog. Earth Planet. Sci.*, **6**, 54, https://doi.org/10.1186/s40645-019-0297-7.

Vu, T. T., and R. Ranzi, 2017: Flood risk assessment and coping capacity of floods in central Vietnam. *J. Hydro-Environ. Res.*, **14**, 44–60, https://doi.org/10.1016/j.jher.2016.06.001.

Wehbe, Y., M. Temimi, and R. F. Adler, 2020: Enhancing precipitation estimates through the fusion of weather radar, satellite retrievals, and surface parameters. *Remote Sens.*, **12**, 1342, https://doi.org/10.3390/rs12081342.

Zhang, J., and Coauthors, 2011: National Mosaic and Multi-Sensor QPE (NMQ) system: Description, results, and future plans. *Bull. Amer. Meteor. Soc.*, **92**, 1321–1338, https://doi.org/10.1175/2011BAMS-D-11-00047.1.

Zhang, L., X. Li, D. Zheng, K. Zhang, Q. Ma, Y. Zhao, and Y. Ge, 2021: Merging multiple satellite-based precipitation products and gauge observations using a novel double machine learning approach. *J. Hydrol.*, **594**, 125969, https://doi.org/10.1016/j.jhydrol.2021.125969.