

## Deep learning model for heavy rainfall nowcasting in South Korea

Seok-Geun Oh<sup>a</sup>, Seok-Woo Son<sup>a,b,\*</sup>, Young-Ha Kim<sup>a</sup>, Chanil Park<sup>a,c</sup>, Jihoon Ko<sup>d</sup>, Kijung Shin<sup>d</sup>, Ji-Hoon Ha<sup>e</sup>, Hyesook Lee<sup>e</sup>

<sup>a</sup> School of Earth and Environmental Sciences, Seoul National University, Seoul, Republic of Korea

<sup>b</sup> Interdisciplinary Program in Artificial Intelligence, Seoul National University, Seoul, Republic of Korea

<sup>c</sup> Department of Earth and Environmental Sciences, Boston College, Chestnut Hill, MA, United States

<sup>d</sup> Kim Jaechul Graduate School of AI, KAIST, Seoul, Republic of Korea

<sup>e</sup> National Institute of Meteorological Sciences, Jeju, Republic of Korea

### ARTICLE INFO

#### Keywords:

Deep learning nowcasting

Heavy rainfall events

Numerical weather prediction

### ABSTRACT

Accurate nowcasting is critical for preemptive action in response to heavy rainfall events (HREs). However, operational numerical weather prediction models have difficulty predicting HREs in the short term, especially for rapidly and sporadically developing cases. Here, we present multi-year evaluation statistics showing that deep-learning-based HRE nowcasting, trained with radar images and ground measurements, outperforms short-term numerical weather prediction at lead times of up to 6 h. The deep learning nowcasting shows an improved accuracy of 162%–31% over numerical prediction, at the 1-h to 6-h lead times, for predicting HREs in South Korea during the Asian summer monsoon. The spatial distribution and diurnal cycle of HREs are also well predicted. Isolated HRE predictions in the late afternoon to early evening which mostly result from convective processes associated with surface heating are particularly useful. This result suggests that the deep learning algorithm may be available for HRE nowcasting, potentially serving as an alternative to the operational numerical weather prediction model.

### 1. Introduction

Global damages from natural disasters have dramatically increased in recent decades (Kharin et al., 2013; Zhang et al., 2018; WMO et al., 2021). Among others, weather and climate hazards account for 50% of all disasters, 45% of all reported deaths, and 74% of all reported economic losses (WMO et al., 2021). Of the top 10 weather and climate hazards, rainfall-related disasters, such as storms and floods, occupy the first and second positions (WMO et al., 2021). Their damage cost is expected to even rise in the future, as the frequency and intensity of rainfall extremes are anticipated to increase under global warming (Zhang et al., 2018; Fowler et al., 2021). To mitigate such risks, it is imperative to have precise rainfall prediction across the lead times. Among others, the rainfall nowcasting, which provides detailed information about the locations and timings of heavy rainfall events (HREs) with lead times of up to several hours, is particularly crucial for taking proactive steps to minimize damages within an appropriate time frame.

Several approaches to rainfall nowcasting have been proposed. Traditional and common approaches include statistical predictions

based on rain gauge data (Johnson and Bras, 1980; Browning and Collier, 1989) and extrapolation of weather radar and satellite measurements (Turner et al., 2004; Germann et al., 2006; Kim and Cho, 2012). With increasing computing resources, numerical weather prediction (NWP) models, which assimilate all available observations and solve fundamental physical-law equations, have become increasingly important (Shahrban et al., 2016; Pu and Kalnay, 2018). However, NWP models have substantial uncertainties in predicting rapidly developing local convective systems, necessitating fine-scale observations and advanced techniques for assimilating them into initial fields, which may involve managing thermodynamically unstable flow states (Ducrocq et al., 2000; Klemp, 2006). Accurate parameterizations of subgrid-scale processes (e.g., turbulence, microphysics) are also essential. As a result, there have been challenges for NWP models to be successful in the prediction of localized precipitation systems (e.g., Ducrocq et al., 2000; Lin et al., 2005; Kim and Cho, 2012; Brotzge et al., 2023).

Recently, deep learning algorithms have been applied to rainfall nowcasting. Several studies, conducted for North America (Chen and Wang, 2022; Espeholt et al., 2022), Europe (Ayzel et al., 2020; Ravuri

\* Corresponding author. School of Earth and Environmental Sciences, Seoul National University, Republic of Korea.

E-mail address: [seokwooson@snu.ac.kr](mailto:seokwooson@snu.ac.kr) (S.-W. Son).

et al., 2021), and Asia (Yen et al., 2019; Zhang et al., 2021; Ko et al., 2022; Oh et al., 2022), have demonstrated their capability and strength in predicting weak and moderate rainfall events at lead times of 1–2 h. Despite such successes, challenges still persist when it comes to the prediction of HREs (Ravuri et al., 2021; Chen and Wang, 2022; Oh et al., 2022). Only a few studies have reported promising results in the nowcasting of HREs (Espeholt et al., 2022; Zhang et al., 2023), and thus the usefulness of deep learning models remains not generalized.

Here, we evaluate the performance of deep-learning-based HRE nowcasting for the summer season of 2020–2022 in South Korea, compared to the 6-h numerical weather predictions. Various types of HREs develop during the Asian summer monsoon. They include HREs associated with monsoon rain bands, mesoscale disturbances, localized thermodynamic instability, and typhoons (Park et al., 2021a; Oh et al., 2022). Their properties vary immensely depending on the region and timing of occurrences (Jo et al., 2020; Park et al., 2021a). As such, our evaluation also centers on different HRE types. Specifically, this study compares the DEEP-learning-based RAin Nowcasting and Estimation (DEEPRANE, Ko et al., 2022; Oh et al., 2022) to the Korea Local Analysis and Prediction System (KLAPS, Shin et al., 2022; Song and Roh, 2023) of the Korea Meteorological Administration (KMA). Only the summer season (June–August), which accounts for ~66% of the annual precipitation with frequent HREs (Park et al., 2021a; Oh et al., 2022), is considered. As shown below, the DEEPRANE outperforms the KLAPS for HRE nowcasting for all lead times from 1 to 6 h, regardless of HRE types. A greater improvement is particularly reported in predicting spatially localized HREs which are typically not well predicted by NWP models.

## 2. Data and methods

### 2.1. Deep learning model

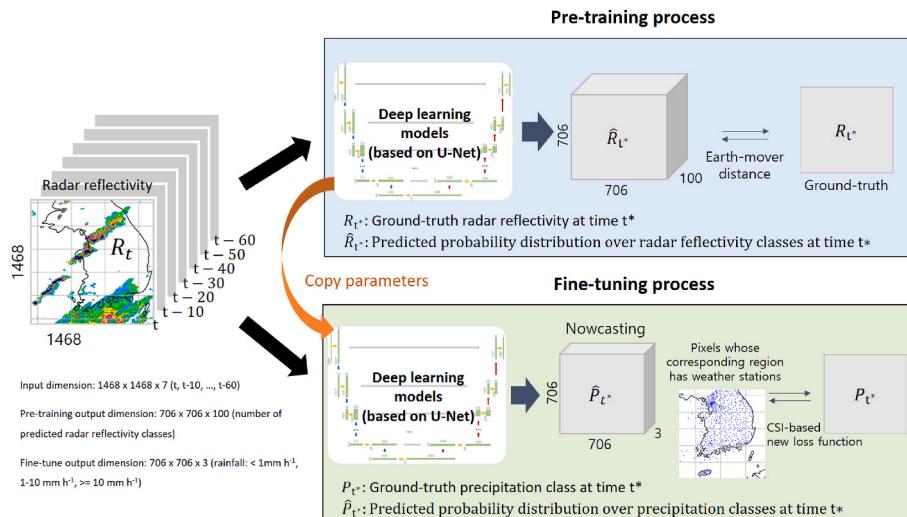
This study uses the deep learning model, DEEPRANE (Ko et al., 2022; Oh et al., 2022). The basic structure of this model follows that of U-Net (Ronneberger et al., 2015) with a contracting path to reduce input image resolution and an expanding path to re-expand contracted images (Fig. 1). In the contracting path, i.e., downsampling steps, the resolution of the input data is halved by merging each four incident pixels, while the number of latent features per pixel, which means the dimension of the vector representing each pixel, is doubled. In the expanding path, i.e., upsampling steps, the resolution of the input data is doubled, while

the number of latent features per pixel is halved. In addition, skip connections are used between the contracting path and expanding path. These connections match the downsampling steps and upsampling steps so that the intermediate output of each downsampling step, which indicates the vector representation of each pixel, is additionally considered to determine the output of the corresponding upsampling step.

Compared to the conventional U-Net, this model has two improved components (Ko et al., 2022). The first one is the pre-training process that optimizes the parameters for predicting radar reflectivity used in rainfall nowcasting. This process initializes the model with pre-trained parameters, which are then fine-tuned in the subsequent training process for rainfall nowcasting. The second one is a novel loss function based on critical success index (CSI) scores, which effectively addresses the issue of class imbalance in rainfall data. The loss function is constructed by true positives (Hit, see Method), false positives (False alarm), and false negatives (Miss), which are used to calculate CSI, based on the output probability distribution. This new loss function allows the deep learning model to achieve better predictive performance at relatively long lead times, 3–6 h (Ko et al., 2022), than the widely-used cross-entropy loss (Cox, 1958) and focal loss (Lin et al., 2017). See more details for the DEEPRANE in Ko et al. (2022).

An input process of this study consists of seven channels of radar reflectivity images, while six additional channels encode the target time. The seven input channels are the seven most recent radar reflectivity images taken at 10-min intervals, i.e., timestamps  $t-60$ ,  $t-50$ , ..., and  $t$ . The six output channels represent target times at 60-min intervals, which are given by  $t+60$ ,  $t+120$ , ..., and  $t+360$ . The input and output have dimensions of  $1468 \times 1468$  and  $706 \times 706$ , respectively, both with a horizontal resolution of 1 km. The final model output is a probability distribution for rainfall events at each pixel. It signifies the probability of rainfall events falling into the three categories, i.e.,  $< 1 \text{ mm h}^{-1}$ ,  $1-10 \text{ mm h}^{-1}$ , and  $> 10 \text{ mm h}^{-1}$ , for lead times ranging from 1 h to 6 h.

In this study, the radar reflectivity images taken at 10-min intervals are collected for eight years from 2014 to 2022 (Fig. S1a). For the same period, rainfall observations from 714 AWSs are also gathered (Fig. S1b). The pre-training is performed on all radar images available from 2014 to 2018 (5 years). During the fine-tuning process, radar and AWS datasets are only utilized for the months of June through September within the same timeframe (Fig. 1). This optimization aims to enhance model parameters specifically for predicting summer rainfall.



**Fig. 1.** Overview of deep-learning model. Deep-learning-based Rain Nowcasting and Estimation (DEEPRANE) is based on U-Net but is equipped with a new pre-training scheme and new loss function, both tailored specifically for rainfall nowcasting. In the pre-training phase, DEEPRANE is trained to predict radar reflectivity in the near future. In the fine-tuning phase, DEEPRANE is initialized with the pre-trained parameters and then fine-tuned for rainfall nowcasting. A new loss function based on critical success index (CSI) is used to mitigate the class imbalance problem. See the section on deep-learning model and method for the details.

The year 2019 is used for verification. Previous studies (Ko et al., 2022; Oh et al., 2022) have demonstrated that this approach effectively predicts summer rainfall events in South Korea. Using this model, rainfall nowcasting is conducted for three years from 2020 to 2022 and its performance is compared with the KLAPS.

## 2.2. KLAPS

This study uses the KMA-operational short-term numerical weather prediction model, KLAPS (Fig. S2, Shin et al., 2022). This model is a KMA modified version of Local Analysis and Prediction System (LAPS) developed by the US National Oceanic and Atmospheric Administration/Forecast System Laboratory (NOAA/FSL; McGinley et al., 1991; Albers, 1995). Recently, the dynamics and physics processes of the KLAPS were updated based on Weather Research and Forecasting (WRF) model (Shin et al., 2022; Song and Roh, 2023). The physics suites other than radiative transfer parameterization were the WRF double moment 7-class microphysics (Bae et al., 2019), unified Noah land surface model (Skamarock et al., 2019), simplified Arakawa-Schubert cumulus modified by the Korea Institute of Atmospheric Prediction Systems (Kwon and Hong, 2017), Shin and Hong boundary layer (Shin and Hong, 2015), and revised MM5 Monin-Obukhov surface layer (Jiménez et al., 2012).

The three-dimension output of the Unified Model (UM) with a horizontal resolution of  $1.875^\circ$ , a KMA-operational global weather prediction system, is used as the first guess input data for KLAPS (Fig. S2). The various meteorological observations, such as surface observation data obtained by buoys, Meteorological Aerodrome Reports (METARs), AWS, upper-air sounding (TEMP) data from the Global Telecommunication System (GTS), wind data obtained by the KMA wind profiler, Aircraft Meteorological Data Relay (AMDAR) data, GK-2A satellite data, Global Navigation Satellite System (GNSS) data, and radar reflectivity, are also used (Shin et al., 2022). This model is designed to predict all possible meteorological elements over the Korean Peninsula with a 5-km horizontal resolution. Forecasts are generated every hour with lead times of up to 12 h. More information regarding the KLAPS procedure can be found in Shin et al. (2022) and Song and Roh (2023).

## 2.3. Evaluation strategies

The performance of the nowcasting system is evaluated against hourly AWS rainfall data for June–August 2020–2022 (Fig. S1b). Currently, the KMA is implementing a real-time quality control system for meteorological observation data, which includes missing value test, physical limit test, step test, internal consistency test, persistence test, climate range test, and others (Zahumenský 2004). The AWS data are also produced using this quality control process. To facilitate the one-to-one comparison, the DEEPRANE and KLAPS grid data are interpolated to the nearest AWS.

We use the F1 score as the primary evaluation metric. This score, which takes into account both the precision and recall metrics, has been widely used in machine learning to evaluate the accuracy of the model (e.g., Umer et al., 2020; Hicks et al., 2022; Ko et al., 2022). The precision is defined as the number of true positives divided by the number of all samples predicted to be positive, including those not correctly identified, i.e.,  $H/(H + F)$ , while the recall is defined as the number of true positives divided by the number of all samples that should have been identified as positive, i.e.,  $H/(H + M)$ .

$$F1 = \frac{2}{precision^{-1} + recall^{-1}} = 2 \frac{precision \times recall}{precision + recall} = \frac{2H}{2H + F + M} \quad (1)$$

Here, H is the number of correctly predicted events that match the observed events, called Hits, F is the number of predicted events that do not match the observed events, called False alarms, and M is the number of predictions missing the observed events, called Misses. The highest possible value of F1 score is 1, indicating perfect precision and recall,

and the lowest possible value is 0, when either precision or recall is zero.

In addition to the F1 score, the CSI, probability of detection (POD), false alarm rate (FAR), and bias score (BS) are also used for evaluation.

$$CSI = \frac{H}{H + F + M} \quad (2)$$

$$POD = \frac{H}{H + M} \quad (3)$$

$$FAR = \frac{F}{H + F} \quad (4)$$

$$BS = \frac{H + F}{H + M} \quad (5)$$

The CSI score measures the fraction of correctly predicted rainfall events to the total rainfall events except correct negatives (i.e., no rainfall in both observation and prediction). It ranges from 0 to 1 with a perfect score of 1. This score is not affected by the number of no-rainfall event predictions and widely used in weather and climate predictions (Ravuri et al., 2021; Oh et al., 2022). The POD denotes the fraction of correctly predicted rainfall events to the observed rainfall events. Its range is 0–1, and a perfect score is 1. The FAR is the fraction of false alarms to all predicted rainfall events, ranging from 0 to 1, with a perfect score of 0. The BS is the ratio of the total number of predicted rainfall events to the observed rainfall events. Its range is 0 to  $\infty$ , and a perfect score is 1. To enable comprehensive evaluation using the four metrics being represented together, a performance diagram (Roeber, 2009) is utilized. The bootstrapping method, repeating 1 000 re-samplings and evaluations from the verification data, is applied to estimate the sampling uncertainty about the evaluated value in the performance diagram.

In order to investigate the relative improvement in accuracy of the DEEPRANE against the KLAPS in terms of F1 score, the improvement rate (IR, unit: %) is defined as follows:

$$IR (\%) = \left( \frac{F1_D - F1_K}{F1_K} \right) \times 100 \quad (6)$$

Subscripts D and K denote DEEPRANE and KLAPS. This study examines the degree of IR with varying lead times, rainfall intensities, and rainfall types.

## 2.4. Definition of rainfall events

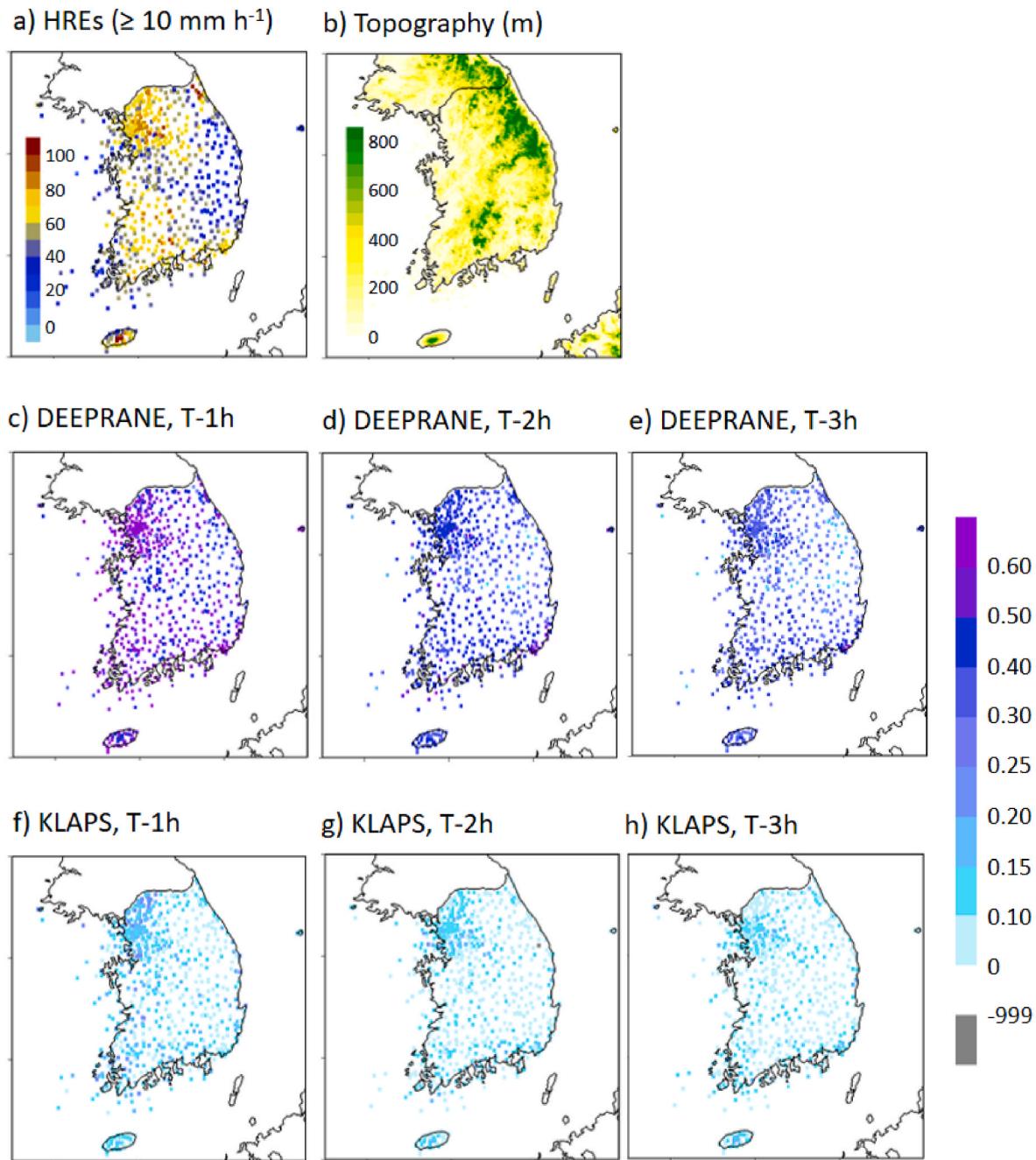
This study uses two rainfall classes: moderate rainfall events (MREs) with intensity greater than  $1 \text{ mm h}^{-1}$ , and heavy rainfall events (HREs) with intensity greater than  $10 \text{ mm h}^{-1}$ . Note that MREs include HREs to ensure enough sample size. To explore how the model performance is influenced by rainfall types, HREs are classified into three types by employing self-organizing map (SOM) algorithm (Kohonen, 1998, 2013). The SOM is an unsupervised neural network algorithm that compresses high-dimensional information into a low-dimensional array based on iterative training and Euclidean distance (Jo et al., 2020; Ko et al., 2022; Oh et al., 2022). The SOM algorithm is applied to 2 306 HREs that were identified during summer season of 2020–2022 (Fig. 5a–c). The parameters of the SOM algorithm used in this study are listed in Table S1. The KMA hourly precipitation reanalysis (KMAPR) data, derived from AWS precipitation and weather radar reflectivity (Jo et al., 2020; Oh et al., 2022), are used for clustering rainfall events. The KMAPR data provides a  $5 \text{ km} \times 5 \text{ km}$  horizontal resolution covering the Korean Peninsula (Fig. S1c).

### 3. Results

#### 3.1. Heavy rainfall nowcasting

The HREs, defined as rainfall events with 1-h accumulated rainfall amount greater than 10 mm, were prevalent in the central and southern parts of South Korea with a higher frequency in the western half of the country during the summer season of 2020–2022 (Fig. 2a). The westward concentration of HREs is related to the eastward-moving nature of rainstorms from the Yellow Sea to the Korean Peninsula because of the midlatitude westerly and their impinging on topography (Fig. 2b) (Park et al., 2021a, 2021b). In general, the DEEPRANE shows more accurate

rainfall nowcasting compared to the KLAPS across the country, as measured by the F1 score. This superiority is observed for all lead times from 1 h to 6 h. The DEEPRANE and KLAPS show the F1 scores of approximately 0.4–0.6 and 0.1–0.3, respectively, at the lead time of 1 h (Fig. 2c–f). Although the F1 scores of the DEEPRANE decrease with increasing lead time, they are still 0.3–0.5 at lead times of 2–3 h, double the KLAPS's scores (Fig. 2d,e,g,h). Even for longer lead times of 4–6 h, the DEEPRANE shows relatively higher F1 scores than the KLAPS (Fig. S3). Such superiority of the DEEPRANE over the KLAPS is also found in the CSI score (Figs. S4 and S5). Note that the two models predict similar occurrence frequency of HREs, as indicated by their comparable bias scores mostly falling between 0.8 and 2.0 (Figs. S6 and S7).



**Fig. 2.** Model performance for heavy rainfall event (HRE) nowcasting in South Korea for summer season of 2020–2022. Spatial distribution of (a) the number of HREs observed by automatic weather station (AWS) during June–August 2020–2022 in South Korea and (b) topography (m). (c–h) The F1 score of the DEEPRANE and Korea Local Analysis and Prediction System (KLAPS, see Method and Fig. S1) for HRE nowcasting at lead times of 1–3 h. F1 score of “-999” indicates no correctly predicted.

Therefore, the superior accuracy of the DEEPRANE against the KLAPS results from a more precise prediction of the HRE timing, rather than the HRE occurrence itself.

The regional-averaged accuracy of HRE nowcasting with respect to lead time is investigated in Fig. 3. The F1 scores of the DEEPRANE and KLAPS decrease from 0.55 to 0.17 and from 0.21 to 0.13, respectively, as the lead time increases from 1 h to 6 h. It highlights that the DEEPRANE outperforms the KLAPS for all lead times (compare red and black lines), with significant improvement rates against the KLAPS ranging from 31% to 162% (blue bars). The improvement rates are higher at shorter lead times. Similar results are also found for moderate rainfall events ( $\geq 1 \text{ mm h}^{-1}$ ) (Fig. S8). These results suggest that the DEEPRANE could offer improved rainfall nowcasting than the KLAPS, potentially enabling more effective preparation for hydrological disasters.

### 3.2. Diurnal variation of heavy rainfall events

The diurnal variation of summer monsoon rainfall in South Korea is characterized by a bi-modal distribution with one peak occurring in the early morning (approximately 4–8 local time) and the other in the late afternoon to early evening (approximately 15–20 local time) (Oh and Suh, 2018; Jo et al., 2020). These two peaks are attributed to different physical processes: the early morning peak is driven by nocturnal cooling of cloud tops, while the late afternoon-early evening peak is associated with surface heating by solar insolation during the daytime (Zhou et al., 2008; Jo et al., 2020). This bi-modal distribution is distinctly observed in HREs in the summer season of 2020–2022 (Fig. 4a).

The DEEPRANE consistently outperforms the KLAPS throughout the diurnal cycle. Its F1 score is higher than  $\sim 0.5$  at a lead time of 1 h for all local times, which is better than that of the KLAPS with  $\sim 0.2$  (Fig. 4b and c). Unlike the KLAPS that shows poor performance in the afternoon and evening nowcasting (Fig. 4c), the DEEPRANE performs well for both early morning and late afternoon-early evening HREs (Fig. 4b). As a result, the performance gap between the two models is relatively larger in late afternoon-early evening HRE nowcasting (compare Fig. 4b and c). A similar result is obtained for moderate rainfall events (Fig. S9) and for the CSI-based assessment (Fig. S10).

A relatively poor performance of the KLAPS at late afternoon-early evening hours results from significant over-prediction of HREs (Fig. 4e). The late afternoon-early evening HREs are mainly associated with deep convection caused by solar heating of the surface (Zhou et al., 2008; Jo et al., 2020). Since convection involves complex small-scale

processes, many of which occur at a spatial scale smaller than the numerical model's grid scale and their representation in the model depends on subgrid-scale process parameterization. Those parameterizations can lead to systematic errors in convective rainfall prediction (Lin et al., 2005; Oh and Suh, 2018; Brotzge et al., 2023). Given this difficulty of NWP models, recent studies have highlighted the potential usefulness of deep learning algorithms in convective rainfall prediction (Zhou et al., 2019; Su et al., 2020; Zhang et al., 2023). Our result provides a piece of additional evidence that the deep learning model could be a great alternative in convective rainfall nowcasting.

### 3.3. Heavy rainfall types

The HREs are often spatially and temporally heterogeneous, exhibiting localized characteristics varying with regions (Lee and Seo, 2008; Jo et al., 2020; Park et al., 2021b). Such natures should be taken into account when devising effective preparation and response strategies to HRE-related disasters. In this regard, we further evaluate the performance of the DEEPRANE by considering HRE types. Following Jo et al. (2020), HREs are classified into isolated, central, and southern rainfall types (Fig. 5a–c), using the SOM algorithm (Kohonen, 1998, 2013) as described in Section 2.4.

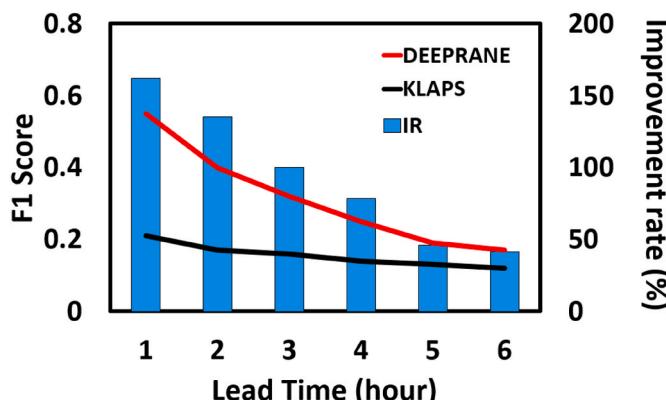
The central and southern rainfall types (Fig. 5b and c) are related to the synoptic-scale systems located over the central and southern parts of the country (Jo et al., 2020). In contrast, the isolated rainfall type (Fig. 5a) is caused by local processes which occur sporadically throughout the country (Jo et al., 2020; Oh et al., 2022). Although the isolated rainfall events account for the majority of HREs during the summer monsoon period ( $\sim 80\%$ , see Fig. 5a), NWP models struggle to accurately predict them due to their small spatial scale and short lifetime (Lin et al., 2005; Oh and Suh, 2018).

The performance of two nowcasting systems is summarized in Fig. 5d–f using a performance diagram which combines the POD (y-axis), success ratio (x-axis) defined by one minus FAR, BIAS (dashed lines), and CSI (solid contours). See Section 2.3 for the details. The result confirms that the DEEPRANE consistently outperforms the KLAPS for all HRE types, especially at shorter lead times (Fig. 5d and e). At the lead time of 1 h, the CSI scores of the DEEPRANE are greater than 0.35 across all HRE types, whereas those of the KLAPS are less than 0.2 (compare red marks). The bias scores are slightly smaller than one, but comparable between the two systems. The CSI scores of the DEEPRANE at lead times of 2–4 h are still more than double those of the KLAPS, although the DEEPRANE slightly over-predicts the central and southern HREs as indicated by the BIASs of  $\sim 2$ .

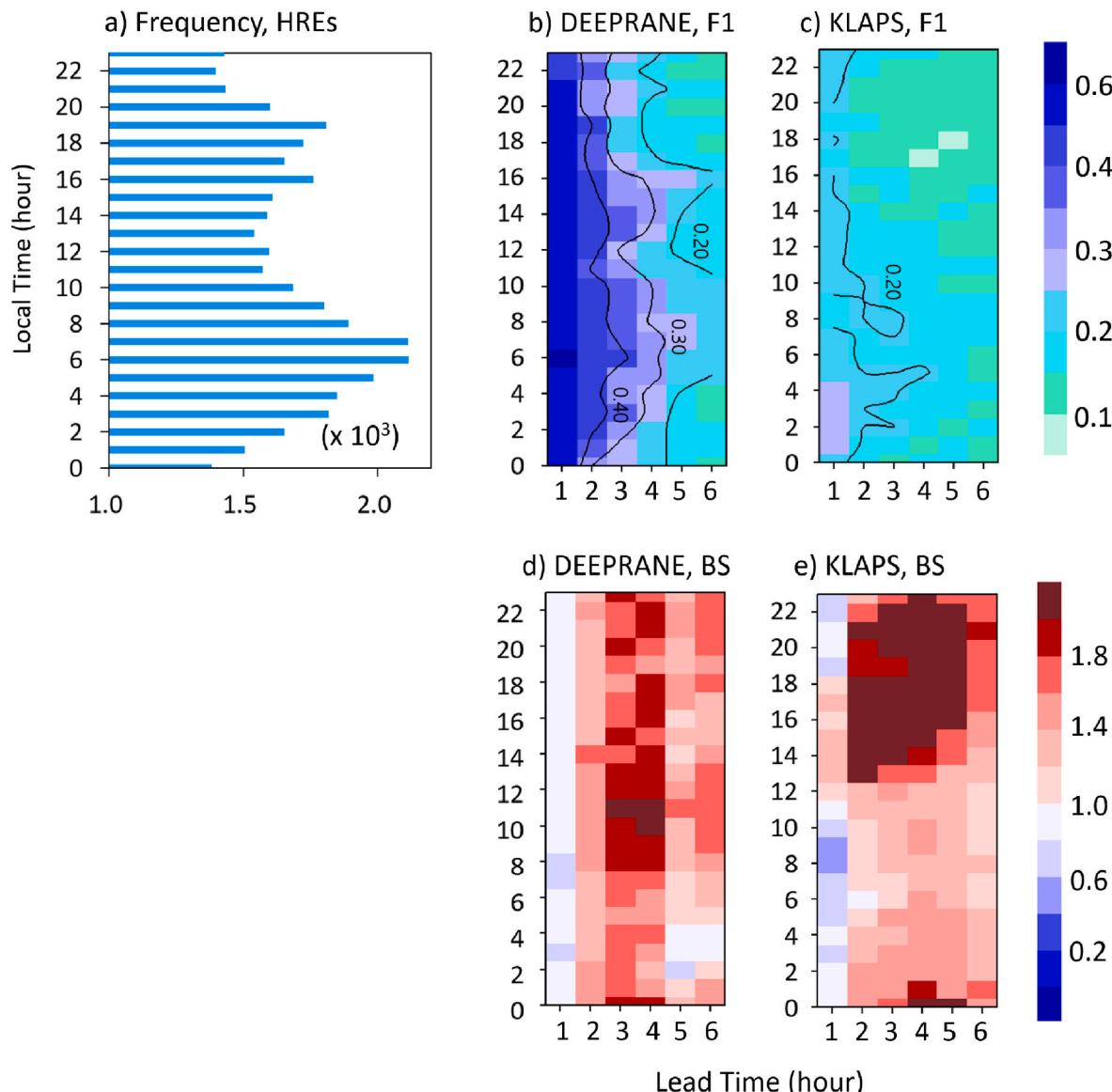
Upon examining the performance of the two systems for three HRE types, it is evident that they show relatively lower accuracy in predicting isolated HREs (represented by the circles in Fig. 5d and e; see also Fig. 5a), compared to the central and southern HREs (triangles and squares in Fig. 5d and e; see also Fig. 5b and c). However, it is worth noting that improvement of the DEEPRANE to the KLAPS is largest for predicting isolated HREs at all lead times (Fig. 5f, see also Fig. S11). This result can be linked to the better performance of the DEEPRANE than the KLAPS for the late afternoon-early evening HRE predictions as shown in Fig. 4. The isolated HREs are only weakly influenced by synoptic-scale environment. They are instead dominated by local convective processes and thus have a shorter lifespan than spatially-organized HREs (i.e., central and southern HREs), which pose challenges to NWP models in accurate nowcasting (Lin et al., 2005; Oh and Suh, 2018). This result suggests that the DEEPRANE can be an innovative alternative to enhance the nowcasting of isolated rainfall extremes with a short lifespan.

## 4. Summary and discussion

The present study evaluates the deep-learning-based rainfall nowcasting system, DEEPRANE, against the 6-h numerical weather



**Fig. 3.** Overall accuracy of HRE nowcasting with varying lead times. Red and black solid lines denote the F1 score of Deep-learning-based Rain Nowcasting and Estimation (DEEPRANE) and Korea Local Analysis and Prediction System (KLAPS), respectively, during summer season of 2020–2022 in South Korea. Blue bars represent the improvement rate (IR) of DEEPRANE over KLAPS (see Method). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)



**Fig. 4.** Diurnal variation of HREs and their predictions by deep-learning nowcasting and operational numerical weather prediction model (a) Observed diurnal variation of the number of HREs aggregated in South Korea during summer season of 2020–2022, and the performance of (b, d) Deep-learning-based Rain Nowcasting and Estimation (DEEPRANE) and (c, e) Korea Local Analysis and Prediction System (KLAPS) in terms of (b, c) the F1 score and (d, e) bias score (BS) with respect to the local time and lead time.

predictions of the KLAPS for the summer season of 2020–2022. We present that the DEEPRANE outperforms the KLAPS in heavy rainfall nowcasting for all lead times. Quantitatively, the accuracy is higher than the KLAPS by 16%–31% at the 1-h to 6-h lead times. The DEEPRANE is particularly useful for nowcasting of isolated HREs and HREs in the late afternoon to early evening which mostly result from convective processes associated with surface heating.

While previous studies have introduced several deep-learning-based rainfall nowcasting systems (e.g., [Ayzel et al., 2020](#); [Chen and Wang, 2022](#); [Espeholt et al., 2022](#)), there remains a scarcity of statistics regarding their performance over multiple years especially for heavy rainfall nowcasting. The present study quantitatively shows that the deep learning model has a clear advantage over the operational NWP model in moderate and heavy rainfall nowcasting. It supports the argument that the deep learning model can be a feasible alternative to the NWP model for rainfall nowcasting.

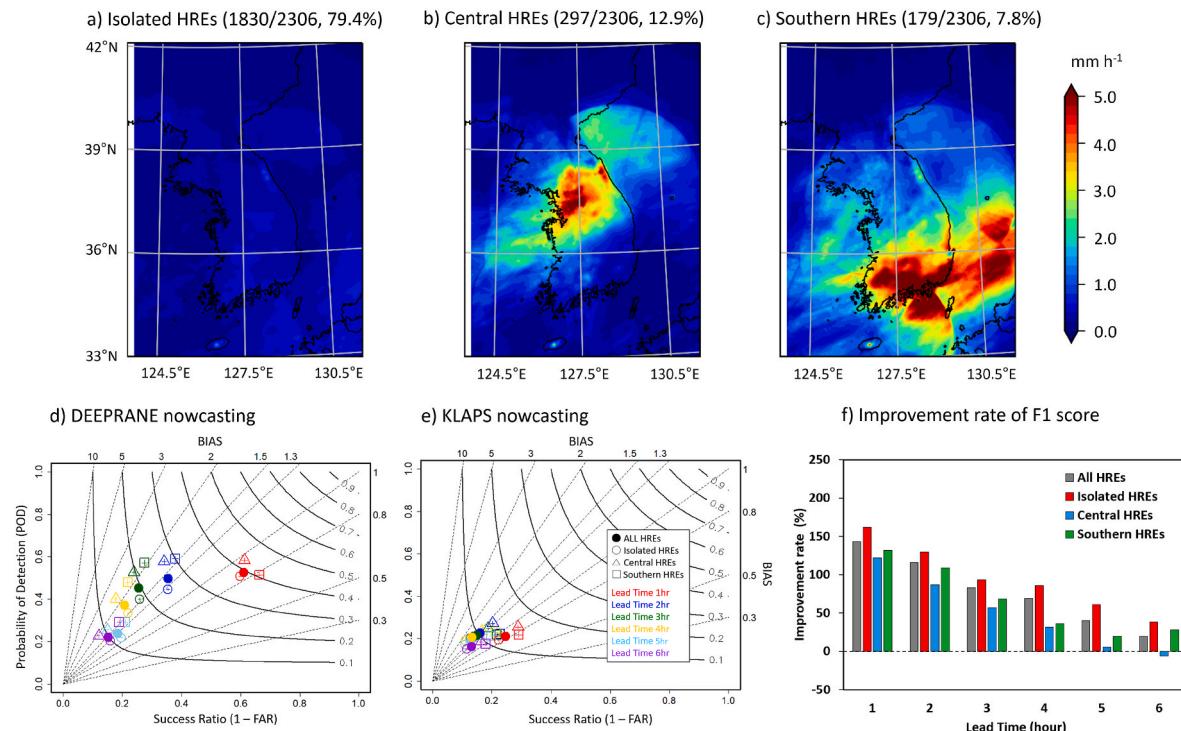
Several challenges remain to improve the deep learning model. The DEEPRANE is designed to predict three rainfall categories, i.e., rainfall

events of  $< 1 \text{ mm h}^{-1}$ ,  $1\text{--}10 \text{ mm h}^{-1}$ , and  $> 10 \text{ mm h}^{-1}$ . However, the rainfall amount or intensity predictions are more important for effective pre-emptive action to minimize the socioeconomic damages caused by HREs. To address this issue, a more detailed subdivision of rainfall categories could be implemented in the next version. This update will enable the deep learning nowcasting to make more informed decisions with comprehensive information on rainfall amounts.

Another possible way to improve the model is to adopt the ensemble approach. The multiple deep learning models, with different algorithms, could be combined to construct the ensemble prediction system, similar to the multi-model ensemble method commonly utilized in climate predictions ([Xiao et al., 2018](#); [Mohammed and Kora, 2021](#)). Such approach could lead to the development of probabilistic rainfall nowcasting.

#### Data availability

The software of deep-learning model used in this study, DEEPRANE



**Fig. 5.** Three types of heavy rainfall events during summer season of 2020–2022 in South Korea and their nowcasting. Spatial distribution of rainfall ( $\text{mm h}^{-1}$ ) averaged for (a–c) each HRE cluster obtained from the self-organizing map (SOM) analysis. Performance diagram of (d) Deep-learning-based Rain Nowcasting and Estimation (DEEPRANE) and (e) Korea Local Analysis and Prediction System (KLAPS) nowcasting for all HREs and each HRE type, and (f) improvement rate (%) of DEEPRANE against KLAPS in terms of the F1 score. In performance diagram, dashed lines indicate bias scores (BIAS) and solid contours represent critical success index (CSI). Sampling uncertainty, estimated using a bootstrapping method, is denoted with crosshairs about the verification point. Note that mean rainfall amount of isolated HREs in (a) is small as isolated HREs have small spatial scale and spread across the country. The improvement rate in terms of CSI score can be also found in Fig. S11 of Supplementary Information.

version 1.0, is released under the GPL-3.0 license at <https://github.com/jihoonko/DeepRaNE>. To run the model, Python 3.6.9 or later, NumPy 1.19 or later, and Pytorch 1.6.0 or later are required to be installed in the system. The radar reflectivity and AWS data across South Korea used in this study are available on Korea Meteorological Administration data released website <https://data.kma.go.kr/cmmn/main.do>. The KLAPS prediction data can be also obtained from <https://www.data.go.kr/>.

#### CRediT authorship contribution statement

**Seok-Geun Oh:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Seok-Woo Son:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Funding acquisition, Conceptualization. **Young-Ha Kim:** Writing – review & editing, Writing – original draft, Validation, Methodology. **Chanil Park:** Writing – original draft, Visualization, Validation, Methodology. **Jihoon Ko:** Writing – review & editing, Software, Data curation. **Kijung Shin:** Writing – review & editing, Software, Data curation, Conceptualization. **Ji-Hoon Ha:** Writing – review & editing, Writing – original draft, Visualization. **Hyesook Lee:** Writing – review & editing, Supervision, Resources, Funding acquisition.

#### Declaration of competing interest

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

#### Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2023R1A2C3005607). This work was also funded by the Korea Meteorological Administration Research and Development Program “Advancing Severe Weather Analysis and Forecast Technology” under Grant (KMA2018-00121). This work was partly supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) [NO.2021-0-01343, Artificial Intelligence Graduate School Program (Seoul National University)]

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.wace.2024.100652>.

#### References

- Albers, S.C., 1995. The LAPS wind analysis. *Weather Forecast.* 10, 342–352.
- Ayzel, G., Scheffer, T., Heistermann, M., 2020. RainNet v1.0: a convolutional neural network for radar-based precipitation nowcasting. *Geosci. Model Dev. (GMD)* 13, 2631–2644.
- Bae, S.Y., Hong, S.Y., Tao, W.K., 2019. Development of single-moment cloud microphysics scheme with prognostic hail for the Weather Research and Forecasting (WRF) model. *Asia-Pac. J. Atmos. Sci.* 55, 233–246.
- Brotzge, J.A., Berchoff, D., Carlis, D.L., Carr, F.H., Carr, R.H., Gerth, J.J., Gross, B.D., Hamill, T.M., Haupt, S.E., Jacobs, N., McGovern, A., Stensrud, D.J., Szatkowski, G., Szunyogh, I., Wang, X., 2023. Challenges and opportunities in numerical weather prediction. *Bull. Am. Meteorol. Soc.* 104 (3) <https://doi.org/10.1175/BAMS-D-22-0172.1>.
- Browning, K.A., Collier, C.G., 1989. Nowcasting of precipitation systems. *Rev. Geophys.* 27, 345–370.

- Chen, G., Wang, W.C., 2022. Short-term precipitation prediction for contiguous United States using deep learning. *Geophys. Res. Lett.* 49 <https://doi.org/10.1029/2022GL097904>.
- Cox, D.R., 1958. The regression analysis of binary sequences. *J. Roy. Stat. Soc. 20* (2), 215–242.
- Ducrocq, V., Laporte, J.P., Redelstperger, J.L., Orain, F., 2000. Initialization of a fine-scale model for convective-system prediction: a case study. *Q. J. R. Meteorol. Soc.* 126, 3041–3065. <https://doi.org/10.1002/qj.49712657004>.
- Espeholt, L., Agrawal, S., Sønderby, C., Kumar, M., Heek, J., Bromberg, C., Gazen, C., Carver, R., Andrychowicz, M., Hickey, J., Bell, A., Kalchbrenner, N., 2022. Deep learning for twelve hour precipitation forecasts. *Nat. Commun.* 13, 5145. <https://doi.org/10.1038/s41467-022-32483-x>.
- Fowler, H.J., Lenderink, G., Prein, A.F., Westra, S., Allan, R., Ban, N., Barbero, R., Berg, P., Blenkinsop, S., Do, H.X., Guerreiro, S., Haerter, J.O., Kendon, E.J., Lewis, E., Schaer, C., Sharma, A., Villarini, G., Wasko, C., Zhng, X., 2021. Anthropogenic intensification of short-duration rainfall extremes. *Nat. Rev. Earth Environ.* 2, 107–122.
- Germann, U., Zawadzki, I., Turner, B., 2006. Predictability of precipitation from continental radar images. Part IV: limits to prediction. *J. Atmos. Sci.* 63, 2092–2108.
- Hicks, S.A., Strümke, I., Thambarwita, V., Hammou, M., Riegler, M.A., Halvorsen, P., Parasa, S., 2022. On evaluation metrics for medical applications of artificial intelligence. *Sci. Rep.* 12, 5979.
- Jiménez, P.A., Dudhia, J., González-Rouco, J.F., Navarro, J., Montávez, J.P., García-Bustamante, E., 2012. A revised scheme for the WRF surface layer formulation. *Mon. Weather Rev.* 140, 898–918.
- Jo, E., Park, C., Son, S.W., Roh, J.W., Lee, G.W., Lee, Y.H., 2020. Classification of localized heavy rainfall events in South Korea. *Asia-Pac. J. Atmos. Sci.* 56, 77–88.
- Johnson, E.R., Bras, R.L., 1980. Multivariate short-term rainfall prediction. *Water Resour. Res.* 16, 173–185.
- Kharin, V.V., Zwiers, F.W., Zhang, X., Wehner, M., 2013. Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Clim. Change* 119, 345–357.
- Kim, G.S., Cho, S.H., 2012. Summer precipitation forecast using satellite data and numerical weather forecast model data. *J. Korean Water Resour. Assoc.* 45, 631–641.
- Klemp, J.B., 2006. Advances in the WRF model for convection-resolving forecasting. *Adv. Geosci.* 7, 25–29. <https://doi.org/10.5194/adgeo-7-25-2006>.
- Ko, J., Lee, K., Hwang, H., Oh, S.G., Son, S.W., Shin, K., 2022. Effective training strategies for deep-learning-based precipitation nowcasting and estimation. *Comput. Geosci.* <https://doi.org/10.1016/j.cageo.2022.105072>.
- Kohonen, T., 1998. The self-organizing map. *Neurocomputing* 21, 1–6.
- Kohonen, T., 2013. Essentials of the self-organizing map. *Neural Network*. 37, 52–65. <https://doi.org/10.1016/j.neunet.2012.09.018>.
- Kwon, Y.C., Hong, S.Y., 2017. A mass-flux cumulus parameterization scheme across gray-zone resolutions. *Mon. Weather Rev.* 145, 583–598.
- Lee, G.H., Seo, K.H., 2008. Analysis of diurnal and semidiurnal cycles of precipitation over South Korea. *Atmos* 18 (4), 475–483.
- Lin, C., Vasić, S., Kilambi, A., Turner, B., Zawadzki, I., 2005. Precipitation forecast skill of numerical weather prediction models and radar nowcasts. *Geophys. Res. Lett.* 32 <https://doi.org/10.1029/2005GL023451>.
- Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017. Focal loss for dense object detection. In: *IEEE International Conference on Computer Vision (ICCV)*. <https://arxiv.org/abs/1708.02002>.
- McGinley, J., Albers, S.C., Stamus, P.A., 1991. Validation of a composite convective index as defined by a real time local analysis system. *Weather Forecast.* 6, 337–356.
- Mohammed, A., Kora, R., 2021. An effective ensemble deep learning framework for text classification. *J. King Saud Univ. – Comput. Inf. Sci.* 10 <https://doi.org/10.1016/j.jksuci.2021.11.001>.
- Oh, S.G., Suh, M.S., 2018. Change in seasonal and diurnal precipitation types during summer over South Korea in the late twenty-first century (2081–2100) projected by the RegCM4.0 based on four RCP scenarios. *Clim. Dynam.* 51, 3041–3060.
- Oh, S.G., Park, C., Son, S.W., Ko, J., Shin, K., Kim, S., Park, J., 2022. Evaluation of deep-learning-based very short-term rainfall forecasts in South Korea. *Asia-Pac. J. Atmos. Sci.* published online. <https://doi.org/10.1007/s13143-022-00310-3>.
- Park, C., Son, S.W., Kim, J., Chang, E.C., Kim, J.H., Jo, E., Cha, D.H., Jeong, S., 2021a. Diverse synoptic weather patterns of warm-season heavy rainfall events in South Korea. *Mon. Weather Rev.* 149, 3875–3893.
- Park, C., Son, S.W., Kim, J.H., 2021b. Role of baroclinic trough in triggering vertical motion during summertime heavy rainfall events in Korea. *J. Atmos. Sci.* 78, 1687–1702.
- Pu, Z., Kalnay, E., 2018. Numerical weather prediction basics: models, numerical methods, and data assimilation. In: Duan, Q., Pappenberger, F., Thielen, J., Wood, A., Cloke, H., Schaake, J. (Eds.), *Handbook of Hydrometeorological Ensemble Forecasting*. Springer, Berlin, Heidelberg.
- Ravuri, S., Lenc, K., Willson, M., Kangin, D., Lam, R., Mirowski, P., Fitzsimons, M., Athanassiadou, M., Kashem, S., Madge, S., Prudden, R., Mandhane, A., Clark, A., Brock, A., Simonyan, K., Hadsell, R., Robinson, N., Clancy, E., Arribas, A., Mohamed, S., 2021. Skilful precipitation nowcasting using deep generative models of radar. *Nature* 597. <https://doi.org/10.1038/s41586-021-03854-z>.
- Roebber, P.J., 2009. Visualizing multiple measures of forecast quality. *Weather Forecast.* 24, 601–608.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. <https://arxiv.org/abs/1505.04597>.
- Shahrban, M., Walker, J., Wang, Q.J., Seed, A., Steinle, P., 2016. An evaluation of numerical weather prediction based on rainfall forecasts. *Hydrol. Sci. J.* 16, 2704–2717.
- Shin, H.C., Ha, J.H., Ahn, K.D., Lee, E.H., Kim, C.H., Lee, Y.H., Clayton, A., 2022. An overview of KMA's operational NWP data assimilation systems. In: Park, S.K., Xu, L. (Eds.), *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications*, vol. IV. Springer, Cham. [https://doi.org/10.1007/978-3-037722-7\\_26](https://doi.org/10.1007/978-3-037722-7_26).
- Shin, H.H., Hong, S., 2015. Representation of the subgrid-scale turbulent transport in convective boundary layers at gray-zone resolutions. *Mon. Weather Rev.* 143, 250–271.
- Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Liu, Z., Berner, J., Wang, W., Powers, J.G., Duda, M.G., Barker, D.M., Haung, X.Y., 2019. A Description of the Advanced Research WRF Model Version 4. NCAR Technical Notes. NCAR, Boulder, CO, USA.
- Song, H.J., Roh, S., 2023. Impact of horizontal resolution on the robustness of radiation emulators in a numerical weather prediction model. *Rem. Sens.* 15 <https://doi.org/10.3390/rs15102637>.
- Su, A., Li, H., Cui, L., Chen, Y., 2020. A convection nowcasting method based on machine learning. *Adv. Meteorol.* <https://doi.org/10.1155/2020/5124274>, 2020.
- Turner, B.J., Zawadzki, I., Germann, U., 2004. Predictability of precipitation from continental radar images. Part III: operational nowcasting implementation (MAPLE). *J. Appl. Meteorol.* 5, 231–248.
- Umer, M., Imtiaz, Z., Ullah, S., Mahmood, A., Choi, G.S., On, B.-W., 2020. Fake news stance detection using deep learning architecture (CNN-LSTM). *IEEE Access* 8, 156695–156706.
- WMO, 2021. In: Douris, J., Kim, G. (Eds.), *The Atlas of Mortality and Economic Losses from Weather, Climate and Water Extremes (1970–2019)*. World Meteorological Organization, p. pp90.
- Xiao, Y., Wu, J., Lin, Z., Zhao, X., 2018. A deep learning-based multi-model ensemble method for cancer prediction. *Comput. Methods Progr. Biomed.* 153, 1–9.
- Yen, M.H., Liu, D.W., Hsin, Y.C., Lin, C.E., Chen, C.C., 2019. Application of the deep learning for the precipitation of rainfall in Southern Taiwan. *Sci. Rep.* 9 <https://doi.org/10.1038/s41598-019-49242-6>.
- Zahumenský, I., 2004. Guidelines on quality control Procedures for Data from automatic weather stations. WMO CIMO/OPAG-SURFACE/ET-ST&MT-1/Doc 6.1 (2), p10.
- Zhang, F., Wang, X., Guan, J., Wu, M., Guo, L., 2021. RN-Net: a deep learning approach to 0-2 h rainfall nowcasting based on radar and automatic weather station data. *Sens* 21. <https://doi.org/10.3390/s21061981>.
- Zhang, W., Zhou, T., Zou, L., Zhang, L., Chen, X., 2018. Reduced exposure to extreme precipitation from 0.5°C less warming in global land monsoon regions. *Nat. Commun.* 9, 3153.
- Zhang, Y., Long, M., Chen, K., Xing, L., Jin, R., Jordan, M.I., Wang, J., 2023. Skillful nowcasting of extreme precipitation with NowcastNet. *Nature*. <https://doi.org/10.1038/s41586-23-06184-4>.
- Zhou, K., Zheng, Y., Li, B., Dong, W., Zhang, X., 2019. Forecasting different types of convective weather: a deep learning approach. *J. Meteorol. Res.* 33, 797–809.
- Zhou, T., Yu, R., Chen, H., Dai, A., Pan, Y., 2008. Summer precipitation frequency, intensity, and diurnal cycle over China: a comparison of satellite data with rain gauge observation. *J. Clim.* 21 (16), 3997–4010.