Rad-cGAN v1.0: Radar-based precipitation nowcasting model with conditional Generative Adversarial Networks for multiple domains

Suyeon Choi[1], Yeonjoo Kim[1]

[1]Department of Civil and Environmental Engineering, Yonsei University, Seoul 03722, Korea

5 *Correspondence to:* yeonjoo.kim@yonsei.ac.kr

Abstract. Numerical weather prediction models and probabilistic extrapolation methods using radar images have been widely used for precipitation nowcasting. Recently, machine-learning-based precipitation nowcasting models have also been actively developed for relatively short-term precipitation predictions. This study aimed to develop a radar-based precipitation
10 nowcasting model using an advanced machine learning technique, conditional generative adversarial network (cGAN), which shows high performance in image generation tasks. The cGAN-based precipitation nowcasting model, named Rad-cGAN, developed in this study was trained with a radar reflectivity map of the Soyang-gang Dam region in South Korea with a spatial domain of 128 × 128 km, spatial resolution of 1 km, and temporal resolution of 10 min. The model performance was evaluated using previously developed machine-learning-based precipitation nowcasting models, namely convolutional long short-term
15 memory (ConvLSTM) and U-Net, as well as the baseline Eulerian persistence model. We demonstrated that Rad-cGAN outperformed other models not only for the chosen site but also for the entire domain across the Soyang-gang Dam region. Additionally, the proposed model maintained good performance even with lead times up to 80 min based on the critical success index at the intensity threshold of 0.1 mm h^{-1}, while U-Net and ConvLSTM achieved lead times of 70 and 40 min, respectively. We also demonstrated the successful implementation of the transfer learning technique to efficiently train model with the data
20 from other dam regions in South Korea, such as the Andong and Chungju Dam regions. We used pre-trained model, which was completely trained in the Soyang-gang Dam region. This study confirms that Rad-cGAN can be successfully applied to precipitation nowcasting with longer lead times, and using the transfer learning approach it shows good performance in regions other than the originally trained region.

Geoscientific
Model Development
Discussions

1 Introduction

25

Nowcasting is defined as a description of the current weather and then forecasting within few hours and is generally applied to mesoscale and local scales. Due to the increasing number of disasters in small spatiotemporal scales, nowcasting plays an important role in risk management (WMO, 2017). Therefore, the need for accurate precipitation nowcasting for early warning systems of floods is increasing to reduce the damage caused by heavy rain, landslides, and flash floods.

30 Such systems mostly use the precipitation nowcasting data provided by numerical weather prediction (NWP), which are insufficient for forecasting at relatively short (0–2 h) lead times (Berenguer et al., 2012). Several studies have shown that radar-based models using the extrapolation method perform better than NWP, especially in the case of precipitation nowcasting at lead times of up to 6 h (Berenguer et al., 2012; Pierce et al., 2012; Renzullo et al., 2017; Imhoff et al., 2020). Additionally, the availability of high-resolution remote sensing observation data (e.g., radar) and computer resources has led to the
35 development of advanced data-driven precipitation prediction models. For example, Ayzel et al. (2019) developed an optical flow-based precipitation nowcasting model called rainymotion, and Pulkkinen et al. (2019) developed a deterministic and probabilistic nowcasting application called PySTEPS, which has potential in several countries (Finland, Switzerland, the United States, and Australia). Both models were written in an open-source Python library. Furthermore, the blending technique, which combines NWP and radar-based models, has improved the precipitation nowcasting performance for short-term flood
40 forecasting (Poletti et al., 2019; Hwang et al., 2020).

Recent availability of large amount of data and increased computational resources led to the development of radar-based models using machine learning techniques. Shi et al. (2015) developed a radar-based model with a convolutional long short-term memory (ConvLSTM) architecture that outperformed the optical flow-based model (real-time optical flow by variational methods for echoes of radar). They showed that ConvLSTM can capture the spatiotemporal correlation between input rainfall
45 image frames, which are recorded every 6 min across Hong Kong. Several studies have shown that the ConvLSTM architecture can be successfully applied to the precipitation nowcasting model (Kim et al., 2017; Moishin et al., 2021; Sønderby et al., 2020; Jeong et al., 2021). Although the convolution neural network (CNN) does not have a structure to conserve temporal information, Agrawal et al. (2019) showed that a fully connected CNN called U-Net can make better predictions than traditional NWP models. Further studies (e.g., Ayzel et al., 2020; Trebing et al., 2021) also confirmed that the U-Net
50 architecture can accurately predict precipitation.

In the field of computer science, the generative adversarial network (GAN) architecture (Goodfellow et al., 2014) showed remarkable performance in image-to-image tasks. Isola et al. (2017) demonstrated that the U-Net model with a conditional GAN (cGAN) approach called Pix2Pix can generate higher-quality images than the original U-Net model. Rüttgers et al. (2019) showed that typhoon tracks and cloud patterns over the Korean Peninsula could be successfully predicted using cGAN
55 architecture with satellite cloud images. Ravuri et al. (2021) developed a precipitation nowcasting model using a deep generative model inspired by the video GAN model (Clark et al., 2019). The video GAN model contains sequential components in the generator, while the discriminator uses a dual architecture that distinguishes the real and generated frames to ensure both temporal and spatial consistency. From the case study of connective cells over eastern Scotland, it was observed that using video GAN in the model significantly improved the quality of precipitation forecasts (Ravuri et al., 2021). These studies
60 indicate that the performance of precipitation nowcasting models can be improved by advanced machine learning techniques. However, because machine learning is a data-driven technique, it will perform effectively for only trained data domains. Generally, it is vital to train from the beginning to develop a model for a new domain, and computation costs will be high even if new data are similar to old data. Thus, the models trained for one domain will be limited in their applications for multiple regions.

65 In this study, we developed an advanced precipitation nowcasting model using a cGAN approach (Rad-cGAN) for multiple domains of the Soyang-gang, Andong, and Chungju Dam basins in South Korea. We trained the model using radar reflectivity data of the Soyang-gang Dam Basin for summer season during 2014–2017 (provided by the Korea Meteorological Administration, KMA), and evaluated the model performance by comparing it with reference models of ConvLSTM, U-Net, and Eulerian persistence using 2018 data. We proposed the transfer learning technique (Pan and Yang, 2009) that uses the
70 previously trained model with cost-effective computation to train the model for the other two abovementioned domains. Three transfer learning strategies were compared to evaluate which was most effective for the Andong and Chungju Dam basins.

2 Materials and methods

2.1 Study area and Radar reflectivity data

We developed a precipitation nowcasting model for dam basin areas where an accurate rainfall forecasting system is essential
75 for the estimation of urban water supply and flood prevention. The target domains were the Soyang-gang Dam Basin (D1), Chungju Dam Basin (D2), and Andong Dam basin (D3) areas. These dams are multi-purpose and are located upstream of the major rivers of South Korea (Fig. 1). The 1.5-km the constant altitude plan position indicator (CAPPI) radar reflectivity data, provided by KMA, were used as input data for training and evaluation of our model. The map product represents the quality-controlled radar reflectivity composite
80 (dBZ) of 11 weather radar stations across South Korea (Fig. 1a), with a size of 960 × 1200 pixels, spatial resolution of 1 × 1 km, and temporal resolution of 10 min.

The radar composite data were cropped to 128 × 128 pixels, covering three target basins. Figure. 1b shows the different topographical characteristics of each domain, suggesting their different rainfall patterns.

85 We used the available radar reflectivity data in summer (June–August, JJA) from 2014 to 2018. Data from 2014 to 2017 were used for training the model, and those from 2018 were used for evaluation. For quick and effective training, we converted the raw radar reflectivity data (dBZ) to grayscale (0–255) and scaled the data range to 0–1 via the min-max scaler. The predicted radar reflectivity data were converted into precipitation using the Z-R relationship (Eq. (1)) (Marshall and Palmer, 1948) to evaluate the rainfall prediction performance of the model.

$$Z = 200R^{1.6} \tag{1}$$

90 where Z is the radar reflectivity (dBZ) and R is the rainfall rate (mm h^{-1}).

2.2 Model architecture

2.2.1 Generative Adversarial Network (GAN) for image translation

GAN is a recently developed framework for training generative models (e.g., CNN encoder-decoder) via an adversarial process.
95 It consists of a generative model (G) that identifies the distribution of real data from random noise, and a discriminative model (D) that classifies whether the input sample is from the generative model or the original data distribution (Goodfellow et al., 2014). Furthermore, the cGAN framework uses additional conditions (e.g., input data of the generative model) for training and can generate targeted outputs that suit specific conditions (Mirza and Osindero, 2014). For image translation tasks, when G is trained to produce a targeted image (y) from input (x) with random noise (z), the objective of D will try to maximize the loss
100 function $\mathcal{L}_{cGAN}(G,D)$ while G will try to minimize $\log(1 - D(x, G(x, z)))$. This relation can be expressed as:

$$\min_G \max_D \mathcal{L}_{cGAN}(G,D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z)))], \tag{2}$$

where losses were calculated as expected ($\mathbb{E}$) values. After simultaneously training G and D, G was trained to generate an output that cannot be distinguished from real data (y) by D, which was trained in an adversarial manner to detect the fake image from G. Isola et al. (2017) showed that combining the traditional pixel-wise loss with cGAN loss can improve the quality
105 of output images. To generate sharp and realistic images, the $L1$ loss function $\mathcal{L}_{L1}(G)$ was used as the traditional loss (Eq. (3)).

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y}[\|y - G(x, z)\|_1] \tag{3}$$

By adding the traditional loss with a weight λ to the cGAN loss, the final objective was obtained (Eq. (4)).

$$G^* = arg \min_G \max_D \mathcal{L}_{cGAN}(G,D) + \lambda\mathcal{L}_{L1}(G) \tag{4}$$

In this study, we developed a radar-based precipitation nowcasting model using a cGAN framework. The structure of our
110 model was inspired by the Pix2Pix model, which shows outstanding performance in image-to-image translation tasks (Isola et al., 2017). We used the U-Net architecture as the generative model, and PatchGAN, which distinguishes images for each N × N patch (N can be smaller than the full size of the image), was used as the discriminative model for high-quality image generation. U-net-based precipitation nowcasting model has previously shown better performance than the traditional radar-based precipitation nowcasting model that uses optical flow (Ayzel et al., 2020). Therefore, we propose an improved
115 precipitation nowcasting model by applying the cGAN approach to the U-Net architecture.

2.2.2 Generative model

Figure 2a shows the generative model using U-Net architecture (a detailed description of U-Net is provided in Sect. 2.3.2). The model consists of nine convolutional layers, two max-pooling layers, two up-sampling layers, and an output convolutional
120 layer. Each convolutional layer, except for the output layer, is composed of the following operations: 3 × 3 2D convolution with zero-padding, batch normalization, and activation function of ReLU. In the contracting part of the generative model, a 2 × 2 2D max-pooling operation was used to down-sample the input images. To prevent model generalization (overfitting), a dropout layer with a rate of 0.5 was applied after the pooling and convolutional layers of the expanding part of the model (Srivastava et al., 2014). A 2 × 2 2D up-sampling operation was further applied in the expanding part after skip connection to
125 increase the resolution of featured images that contain both high- and low-level information. Finally, the output convolutional layer had a 1 × 1 2D convolution that used a linear function for activation to obtain future prediction of the radar reflectivity image.

2.2.3 Discriminative model

130 PatchGAN from the Pix2Pix model was used as the discriminative model (Fig. 2b). As in cGAN, the input pair of the discriminative model consists of historical radar reflectivity data (i.e., input of the generative model) and future radar reflectivity data. The discriminative model classifies real image pairs (input of generative model and ground truth image) as 1, and fake image pairs (input and generated image from generative model) as 0 (Mirza and Osindero, 2014). Particularly, PatchGAN only penalizes the structures at a certain scale of image patches; thus, this discriminative model classifies whether
135 the N × N patch in the input pair is real or fake. The size of the patch (N) was determined by the structure of the entire discriminative model, and it increased as the model became deeper. We constructed a discriminator model through optimization with a 34 × 34 patch size. The model consists of three convolutional layers and an output layer. The first two convolutional layers were composed of 4 × 4 2D convolution with strides of two and zero-padding, batch normalization, and ReLU activation function, which was leaky and had a 0.2 slope. The third convolutional layer had the same configuration as

140 the previous layers, except that its stride was 1. To distinguish the input pair in the image form, the output layer consisted of 4 × 4 2D convolution with zero-padding and sigmoid activation functions. Therefore, each pixel of the output referred to the probability that the discriminative model determines each patch of the input pair as the real one.

2.2.4 Optimization procedure

145 To optimize Rad-cGAN, we followed the training procedure suggested by Isola et al. (2017). First, we randomly selected samples that consisted of four consecutive radar reflectivity images (*t-30*, *t-20*, *t-10* min, and *t*) and the image at *t+10* min. Then, we created a training sample for the discriminative model by adding labels to classify whether the samples were real (image at *t+10* min from observation) or fake (*t+10* image from the generative model) pairs. Next, we updated the parameters of the discriminative model using the minibatch stochastic gradient descent (SGD) method for one step. Binary cross-entropy 150 was used as a loss function, and we applied the ADAM optimizer (Kingma and Ba, 2015) with a learning rate of 0.0002 and momentum parameters $\beta_1 = 0.5$ and $\beta_2 = 0.999$. Then, we trained the generative model for one step to optimize Eq. (4). We used binary cross-entropy as $\mathcal{L}_{cGAN}$ for the discriminator to classify the generated image into a real image. Additionally, λ of the traditional pixel-wise $L1$ loss was set to 100. The minibatch SGD and ADAM optimizer were applied to train the generative model with the same setting as the discriminative model. Both the procedures for updating the parameters of the discriminator 155 and generator were run simultaneously during one epoch. We trained our model using 600 epochs, with a batch size of 8. To effectively achieve the optimal model, we applied an early stopping technique that stops the training model when the loss stops improving. To monitor the loss, we set patience to 30 epochs and saved the model when the loss improved. The model architecture was written in Python (https://www.python.org/) using the Keras deep learning application (https://keras.io/). The entire procedure for training and all the experiments for evaluation were run on a computer with a single NVIDIA Tesla V100 160 GPU.

2.3 Reference models

The model performance of the Rad-cGAN model was compared and validated using reference models. We used the Eulerian persistence model (hereinafter referred to as persistence), a traditional radar-based rainfall prediction model, as the first 165 reference model. The persistence model assumes that rainfall prediction at any lead time will the same as rainfall in the forecast time. It is a simple but powerful model for predicting short-term precipitation. For comparison, we used ConvLSTM and U-Net, which are the common basic structures for machine-learning-based nowcasting models in several studies, as reference models.

2.3.1 ConvLSTM

170 LSTM is a special case of recurrent neural networks (RNNs) and is widely used in temporal sequence predictions (Hochreiter and Schmidhuber, 1997). Sutskever et al. (2014) proposed an LSTM encoder-decoder framework for sequence-to-sequence problems, which consists of concatenated LSTMs for the input and output sequences. Based on this model, Shi et al. (2015) developed a ConvLSTM network that can be applied to spatiotemporal sequence prediction, such as radar-based rainfall prediction. To handle spatiotemporal sequences, a convolution operator was used in state-to-state and input-to-state transitions. 175 The ConvLSTM model was shown to outperform the traditional optical flow-based precipitation nowcasting model. Recent studies have shown that the ConvLSTM model can be successfully applied to predict future radar-based precipitation (Kim et al., 2017; Moishin et al., 2021).

We designed a ConvLSTM model that uses four radar reflectivity image frames (*t-30*, *t-20*, *t-10* min, and *t*) as input to predict future frames at time *t+10* min, which is similar to input and output of Rad-cGAN. The model consists of three ConvLSTM 180 layers and an output layer. Each ConvLSTM layer contains 64 hidden states and 3 × 3 kernels. A 3D convolutional layer with a linear activation function was used as the output layer. To optimize the model, we used the mean squared error as the loss function and applied the ADAM optimizer (learning rate 0.002 and momentum parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$). We trained the model using 600 epochs (early stopping applied) with a batch size of 32.

2.3.2 U-net

185

U-Net-based precipitation nowcasting models efficiently predict future precipitation using historical data, even though U-Net does not have a structure, such as RNN, that preserves temporal information (e.g., Ayzel et al., 2020; Trebing et al., 2021). U-Net was developed by modifying the fully convolutional network (Long et al., 2015), and performed well in image segmentation tasks (Ronneberger et al., 2015). This model architecture consists of two parts: a contracting network that 190 captures the context of the input images and an expanding network that increases the resolution of features from the contracting network.

The contracting network follows the usual CNN, which consists of convolution and max-pooling layers. Each convolution layer is composed of convolution, batch normalization, and activation operations. Batch normalization is used to prevent gradient vanishing or exploding problems and can effectively increase the convergence speed (Ioffe and Szegedy, 2015). The 195 max-pooling operation is applied for down-sampling after convolution of the input image. Through this process, the output of

the contracting network can incorporate the features of the input image. The expanding network consists of the up-sampling and convolution layers. Before applying the up-sampling operation, the skip connection is applied between each layer of the contracting network and the layer of the expanding network to prevent gradient vanishing and share the low-level information of the input data (Simonyan et al., 2015). The convolution layers of the expanding and contracting networks follow the same operation.

200 As the reference model, U-Net was designed to have the same structure as the proposed cGAN-based generative model in our study (Sect. 2.2.2). To optimize the model, $L1$ loss and ADAM optimizers were used similar to the way in ConvLSTM (Sect. 2.3.1). We trained the model using 600 epochs with early stopping and setting the batch size to 8.

205 **2.4 Experiments for evaluating model's prediction skills**

2.4.1 Performance evaluation

We trained the model using data from the summers (June–August) of 2014–2017 and assessed its precipitation nowcasting skill using data from the summer of 2018. To predict the 10 min ahead radar reflectivity data, four latest radar reflectivity data (*t-30*, *t-20*, *t-10* min, and *t* min; *t* being the forecast time) were used as input data. To predict beyond the 10 min lead time, we 210 used the prediction data at *t+10* min as the latest input data. Using this recursive process, predictions were obtained at a lead time of >10 min. Because the model predicts the radar reflectivity after 10 min using past consecutive radar images, we first evaluated the model performance at a lead time of 10 min. This allowed us to confirm the prediction tendency of our model and other reference models while performing precipitation nowcasting. Furthermore, to assess the applicability of our model to an actual early warning system that needs to ensure at least one hour of lead time, the predictive skill was evaluated for >10 215 min of lead time using the recursive process. We measured the verification metrics (see below) using rainfall prediction, converted from the radar reflectivity (Eq. (1)), at a lead time of up to 90 min to confirm the forecasting time in which the model ensured sufficient performance.

Furthermore, we evaluated the model performance at a specific dam site as well as for the entire dam domain (Fig. 1b). We chose the precipitation data at the dam site as it directly affects the dam water level, which is a major factor in the decision- 220 making process of dam management. To evaluate the entire domain, the verification metrics were calculated with increasing lead time for all pixels of the predicted image, and the model performance was represented by the 50th percentile (median) of the metrics. Additionally, to qualitatively evaluate the entire domain, we compared the resulting radar reflectivity images obtained using data at a certain forecast time. We set the forecast time at 23 August 2018, 17:50 UTC, when Typhoon Soulik, which landed on the Korean Peninsula from 23 August 2018, 12:00 UTC to 24 August 2018, 03:00 UTC, started affecting the 225 Soyang-gang Dam Basin.

Four metrics were used for model evaluation: Pearson correlation coefficient (R), root mean square error (RMSE), Nash–Sutcliffe efficiency (NSE), and critical success index (CSI). Since the precipitation prediction of the model was more accurate, the prediction and observation showed a strong positive linear relationship. Hence, we confirmed that the model is more accurate when R (Eq. (5)), calculated between the model prediction and observation, was closer to 1. To verify the precision 230 of the model, the RMSE (Eq. (6)) between prediction and observation was used. Since we propose a precipitation nowcasting model, NSE, which is widely used to assess hydrologic models, can be used as a goodness-of-fit index for it (McCuen et al., 2006) (Eq. (7)).

$$R = \frac{\sum_{i=1}^{N}(O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^{N}(O_i - \bar{O})^2 \sum_{i=1}^{N}(P_i - \bar{P})^2}} \qquad (5)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(O_i - P_i)^2}{N}} \qquad (6)$$

235 $$NSE = 1 - \frac{\sum_{i=1}^{N}(O_i - P_i)^2}{\sum_{i=1}^{N}(O_i - \bar{O})^2} \qquad (7)$$

where $\bar{O}$ and $\bar{P}$ are the means of observation and prediction, respectively, O_i and P_i are the observed and predicted precipitation, respectively, in the ith time of the data period, and N is the total data for the entire period.

Additionally, we used the CSI (Eq. (8)), which is a measure of categorical forecast performance, to verify the model accuracy for precipitation event detection. For this, we selected an intensity threshold of 0.1 mm h^{-1}, which is the criterion by which the 240 KMA determines the number of days of precipitation.

$$CSI = \frac{hits}{hits + false\ alarms + misses} \qquad (8)$$

where $hits$ (correct event forecasts), $false\ alarms$ (incorrect event forecasts), and $misses$ (missed events) are defined by a contingency table (Table 1).

245 **2.4.2 Experiments for transfer learning among different domains**

Since the machine learning model relies on input data as a data-driven model, training on the corresponding new data must be conducted from the beginning to develop a model for a new domain, which is also applicable for our precipitation nowcasting model for a new dam basin with different meteorological, environmental, and geographical characteristics (Fig. 1b). However, because this method is time-consuming and computationally expensive, we applied a transfer learning approach that can be 250 efficiently used to train models with multiple basin domains.

Transfer learning is a machine learning technique that uses knowledge and skills from pre-trained models to train a model for new datasets (Pan and Yang, 2009). This method is often used when the size of the provided dataset is insufficient for training and is also used to train the models for the new dataset due to its lower computational cost than that of training from scratch. The general training strategies of transfer learning are determined by the data size and similarity between the new and original data. For example, if the new dataset is similar to the dataset of the pre-trained model, the new model only fine-tunes for higher layers that learn specific features of the input data and freeze the lower layers that capture the general features. Fine-tuning uses only ~1/10th of the original learning rate and is one of the most effective ways to transfer knowledge. Several studies have shown that the transfer learning approach performs successfully well in image classification tasks (Krizhevsky et al., 2012; Simonyan and Zisserman, 2015; He et al., 2016). In the GAN approach, the discriminative model acts similar to the classifier of the image classification task. Wang et al. (2018) reported that fine-tuning both the generator and discriminator resulted in good performance, but the overfitting was a frequent issue that must be considered. Subsequently, Mo et al. (2020) proposed a strategy that works only on the discriminator called FreezeD, which freezes the lower layer of the discriminator and only fine-tunes the upper layers.

We used transfer learning to train our model for different dam basins, i.e., Andong and Chungju, with a pre-trained model that was completely trained by data from the Soyang-gang Dam Basin. The selected strategies were inspired by the previous approach of transferring GAN (Wang et al., 2018; Mo et al., 2020). We constructed three different strategies: Case 1 to train the entire model for the new domain, Case 2 to freeze the weights of the pre-trained generative model and train the discriminative model, and Case 3 to fine-tune the weights of the pre-trained generative model and train the discriminative model (Table 2a). We trained the model for the Chungju and Andong Dam domains separately using the three strategies (Table 2b). To determine the best strategy for training different dam domains, we estimated the performance at the 10-min lead time at each dam site (Fig. 1b). Additionally, we compared the predictive skill of each strategy at the lead time of up to 90 min by using the recursive process.

3 Results and discussion

3.1 Model performance at the dam site

To evaluate the performance of our model, precipitation was predicted at a lead time of 10 min at the Soyang-gang Dam site during the summer of 2018 (Table 3). All models except for persistence performed extremely well as precipitation nowcasting models (R >0.8, NSE >0.5, CSI >0.5). Among them, Rad-cGAN significantly outperformed the other reference models (Fig. 3). The scatter plot of each model showed that the prediction of the proposed model strongly correlated with the observation (R = 0.86) compared to the reference models (average R = ~0.74). Particularly, Rad-cGAN showed improvements in R and NSE values by 2.63 % and 32.16 %, respectively, compared with the model results using U-Net, confirming that the cGAN approach could mitigate the tendency to underestimate precipitation. This improvement was also observed in the time series plot of the entire evaluation period, where our model performs better than U-Net in predicting peak precipitation. However, the prediction accuracy of the ConvLSTM model for maximum precipitation was higher than that of our model (Figs. 3a and c). Despite this, ConvLSTM predicts most low-intensity rainfall (<5 mm h^{-1}) as zero. Therefore, we can confirm that Rad-cGAN performs better than other refernce models in precipitation prediction.

We predicted the precipitation at a lead time of up to 90 min by using the recursive process to evaluate the model performance. (Fig. 4). Based on R, Rad-cGAN was the best model with an average improvement of 51.70 % and 25.02 % over ConvLSTM and U-Net, respectively, at overall lead times. Additionally, our model maintained good performance (R >0.5) at lead times of up to 50 min, while the U-Net and ConvLSTM models maintained it for up to 30 min and 20 min, respectively. Furthermore, our model maintained sufficient performance (CSI at the intensity of 0.1 mm h^{-1} >0.5) up to 80-min lead time, while ConvLSTM and U-Net preserved their performance for up to 30- and 70-min lead times, respectively. Although the NSE values converged to 0 with increasing lead times in all the models except persistence, our model showed the lowest performance degradation rate. However, with increasing lead time, the RMSE of our model increased more than that of the other models since our model generated more outliers than others when predicting rainfall for a lead time of 40 min or more. Thus, we observed a tendency of underestimation of the prediction of high-intensity precipitation in all the models, including Rad-cGAN (Figs. 3 and 4). This phenomenon was also reported by Kumar et al. (2020), wherein ConvLSTM made significant errors in predicting the precipitation >20 mm h^{-1}. The main reason was data imbalance, which is a common issue in machine learning studies (Wang et al., 2016). Data imbalance of our results occurred since unlike low-intensity precipitation (<10 mm h^{-1}), high-intensity precipitation rarely occurs during the training and testing periods.

3.2 Model performance for the dam domain

To apply our model in an early warning system, the prediction performance upstream of the dam should be sufficient. Hence, the verification metrics were calculated for each grid cell of the entire domain used to train the model, and their median values in all the grid cells were presented for each lead time in all the models (Fig. 5). Figure 5a shows an average increase of 9.02 % and 17.87 % for the R of Rad-cGAN at overall lead time compared to U-Net and ConvLSTM, respectively, which improved the precipitation prediction skill for the entire domain and was consistent with the results from the Soyang-gang Dam site (Fig. 4). Moreover, according to the CSI value, our model preserves its predictability performance (CSI at the intensity of 0.1 mm

310 h^{-1} >0.5) for the entire lead time, indicating that it can be applied to predict precipitation at lead times of >90 min. The lead time for the CSI at the intensity of 0.1 mm h^{-1} >0.5 was up to 40 min with ConvLSTM in this study while CSI at the intensity of 0.5 (not 0.1) mm h^{-1} >0.5 was up to 40 min with ConvLSTM-based nowcasting model for Hong Kong region (Shi et al., 2015). Ayzel et al. (2020) showed that the U-net-based model preserved performance (CSI at an intensity of 0.125 mm h^{-1} >0.5) at a lead time of >60 min in Germany, whereas the performance of our model with similar CSI (0.1 mm h^{-1} > 0.5) remained up to 80 min. Hence, we confirm that the reference model was sufficiently trained to be used for comparison with

315 our model. This result indicates that Rad-cGAN has a reliable performance for precipitation nowcasts. However, in the case of RMSE and NSE, Rad-cGAN performs worse than ConvLSTM, with an average increase of 1.90 % in the RMSE and a decrease of 7.67 % in the NSE over the entire lead time (Figs. 5b, d). This performance degradation shows that our model generates more outliers than ConvLSTM with increasing prediction lead time. However, the CSI of ConvLSTM recorded a markedly low compared to the persistence model, suggesting that the precipitation event detection performance deteriorated

320 rapidly with increased lead time (Fig. 5d).

To better understand the model performance with increasing lead time, we predicted the radar reflectivity results for lead times of 10, 30, 60, and 90 min for a specific forecast time of 23 August 2018, 17:50 UTC when Typhoon Soulik began affecting the Soyang-gang Dam Basin. We observed that the model performance degraded due to the blurring effect of the predicted image by increasing the lead time, which is an issue of previous machine-learning-based nowcasting models (Ayzel et al.,

325 2020; Shi et al., 2015). Despite this smoothing trend, Rad-cGAN produced qualitatively better results than the other reference models (Fig. 6). The prediction bias of Rad-cGAN at 90-min lead time ranged from -35.00 to 18.43 (median = -4.98 dBZ), indicating that our model alleviates the underestimation of radar reflectivity compared to U-Net, whose bias ranged from -43.00 to 15.24 (median = -11.67 dBZ). This result supports the improvement in Rad-cGAN's verification metrics compared to the U-Net (Fig. 5). Furthermore, in the case of ConvLSTM, the median bias of -9.71 dBZ at 90-min lead time prediction

330 showed that ConvLSTM was less prone to underestimation than U-Net. However, the bias of ConvLSTM was recorded to be approximately -20 to -40 dBZ in areas with an observation of ~20–40 dBZ, indicating that it predicted radar reflectivity to be close to zero in most areas with increasing lead times. This result confirms the cause of the remarkably low CSI of ConvLSTM (Fig. 5). Therefore, through the overall verification metrics, we conclude that Rad-cGAN shows the best prediction skills for nowcasting and spatial patterns of precipitation movement.

335

3.3 Performance with transfer learning at different dam sites

To develop a precipitation nowcasting model for multiple dam basins (Andong and Chungju Dam Basins) other than Soyang-gang Dam Basin, we proposed to not only retrain our model with data from new dam basins (Case 1) but also apply efficient transfer learning methodology (Cases 2 and 3; Sect. 2.4.2).

340 First, we measured the model performance of three cases with model predicted precipitation at a 10-min lead time at the Andong Dam site (Table 4a). The results showed that most of the verification metrics were similar in Case 3 compared to Case 1, and Case 2 performed better. In Case 2, which uses all the parameters of the generative model from the pre-trained model, an NSE of 0.72 was achieved, which is close to the NSE of the pre-trained model (0.73) with data from the Soyang-gang Dam area. This phenomenon was also consistent in the verification metrics results at lead times of up to 90 min (Fig. 7a). Through

345 the R and CSI (at the intensity of 0.1 mm h^{-1}) results, it was observed that Case 2 maintained the prediction performance (R, CSI > 0.5) for up to 80-min lead time, even though hyperparameter tuning was not performed to match the model with the Andong Dam data. However, in Case 3, the performance was lower than that of the other strategies. This was because of performance degradation due to overfitting during fine-tuning the pre-trained parameters. High similarity between the two datasets of the Andong and Soyang-gang dams may be the reason for a major performance degradation of the transfer learning

350 using the fine-tuning method (Wang et al., 2018).

For the Chungju Dam Basin, we trained the model using the methodology used for the Andong Dam Basin. The three cases performed similarly for predicting the precipitation at 10-min lead time at the Chungju Dam site (Table 4b). Although Case 1 showed the highest R of 0.86, by comparing other metrics (RMSE, NSE, CSI), transfer learning cases (Case 2 and Case 3) performed better than Case 1. Additionally, when the lead time was increased by up to 90 min, Cases 2 and 3 showed better

355 performance than Case 1 for the entire lead time (Fig. 7b). Case 3 preserved sufficient performance (CSI >0.5) at a lead time of up to 80 min, while Case 2 maintained it for up to 70 min. However, Case 3 showed a less improved average CSI value of ~1.44 % compared to Case 2 for the overall lead time, indicating that the difference in rainfall prediction performance between the two transfer learning strategies was insignificant.

When comparing Cases 2 and 3, the Andong and Chungju dams showed different best cases. However, considering that the

360 performance of both cases at Chungju Dam was almost similar, we decided that the Case 2 method, with the lowest computational cost, was the appropriate strategy for training the model with a new dam basin. Thus, we infer that by using transfer learning, a model can be successfully developed with different domains, although it does not optimize the hyperparameter to fit the model with the new domain.

365 ### 4 Conclusions

In this study, we proposed a radar-based precipitation nowcasting model using the cGAN approach. The model architecture was inspired by the image-to-image translation model called Pix2Pix, which consists of U-Net as the generative model, and

PatchGAN as the discriminative model (Isola et al., 2017). For 10-min lead time precipitation prediction at the Soyang-gang Dam site, our model outperformed the other reference models. Additionally, when we applied the recursive process to predict

370 precipitation at lead times up to 90 min, our model achieved sufficient performance (CSI >0.5) with lead times up to 80 min, which was an improvement over ConvLSTM (up to 30 min) and U-Net (up to 70 min). For the median values of R and CSI of the entire domain, our model generated precipitation prediction more accurately at the overall lead times compared to the reference models. Although our model tends to underestimate strong precipitation, the qualitative evaluation of the Typhoon Soulik confirmed that our model can capture the spatiotemporal change in the extent of radar reflectivity closest to the ground

375 truth. Therefore, we conclude that our Rad-cGAN model is the most advanced precipitation nowcasting model compared to the other reference models.

To develop the precipitation nowcasting model for different dam basins (Andong and Chungju dam basins), we proposed different transfer learning strategies by using the previously trained model with data from the Soyang-gang Dam Basin. In the case of the Andong Dam Basin, the Case 2 strategy, which uses all pre-trained parameters of the generative model and re-

380 trains the discriminative model, showed the highest performance. However, in the case of the Chungju Dam Basin, the Case 3 strategy, which fine-tuned the pre-trained parameters of the generative model and re-trained the discriminative model, showed almost the same but slightly better performance than Case 2. Therefore, we conclude that the transfer learning approach of using all pre-trained parameters of the generative model and re-training discriminative model is the most suitable strategy to develop the precipitation nowcasting model for new dam basins because of its computational efficiency.

385 We confirmed that the proposed precipitation nowcasting model demonstrated improved performance over conventional machine learning-based models (U-Net and ConvLSTM) and showed that transfer learning strategies could be effectively applied to develop models for other domains with summer precipitation in South Korea. However, there are remaining issues that must be considered to ensure auditability of our model for real problems, such as predicting heavy precipitation events and flash flood forecasting. First, the tendency of the model to underestimate precipitation is a major issue. The decisive cause

390 of this issue is data imbalance, as mentioned in general machine-learning tasks (Wang et al., 2016). To address this issue, further studies need to be conducted to improve the predictive performance of extreme precipitation events by extending the duration of training data and assigning weights to the extreme or other events. Additionally, adding information about domain characteristics, such as the digital elevation model and land cover map is expected to improve the precipitation nowcasting model. Another issue is that we trained models for different domains using basic transfer learning strategies, and evaluated the

395 performance only for the new domains, which are not sufficient to develop models for multiple domains that can be used in early warning systems. To overcome this issue, for example, Wang et al. (2020) presented a new transfer learning approach that simultaneously mined the knowledge of multiple pre-trained generators. Therefore, it is expected that further research using more advanced transfer learning strategies will help to develop a precipitation nowcasting model with good performance in all different domains to enhance practicality.

400

Code and data availability

The source code of the model architecture is available at the GitHub repository https://github.com/SuyeonC/Rad-cGAN (last access: 11 November 2021). The radar reflectivity composite data samples provided by the Korea Meteorological Administration (KMA) are available at the public service: https://data.kma.go.kr/resources/html/en/ncdci.html (last access: 405 11 November 2021). The dataset for the entire period can be obtained through a separate request to the KMA.

Author contributions

SC and YK designed the study, and SC performed the model development, simulations, and result analysis under the supervision of YK. SC wrote the original manuscript, and YK reviewed and edited the manuscript.

Competing interests

410 The authors declare that they have no conflict of interest.

References

Agrawal, S., Barrington, L., Bromberg, C., Burge, J., Gazen, C., and Hickey, J.: Machine learning for precipitation nowcasting from radar images, arXiv [preprint], arXiv:1912.12132, 2019.

420   Ayzel, G., Heistermann, M., and Winterrath, T.: Optical flow models as an open benchmark for radar-based precipitation nowcasting (rainymotion v0.1), Geosci. Model Dev., 12, 1387-1402, 10.5194/gmd-12-1387-2019, 2019.

Ayzel, G., Scheffer, T., and Heistermann, M.: RainNet v1.0: a convolutional neural network for radar-based precipitation nowcasting, Geosci. Model Dev., 13, 2631-2644, 10.5194/gmd-13-2631-2020, 2020.

Clark, A., Donahue, J., and Simonyan, K.: Adversarial video generation on complex datasets, arXiv [preprint], 425   arXiv:1907.06571, 2019.

Clive, P., Alan, S., Sue, B., David, S., and Zhihong, L.: Nowcasting, in: Doppler Radar Observations, edited by: Joan, B., and Jorge Luis, C., IntechOpen, Rijeka, 10.5772/39054, 2012.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y.: Generative adversarial networks, Commun. ACM, 63, 139–144, 10.1145/3422622, 2020.

430   Graves, A.: Generating sequences with recurrent neural networks, arXiv [preprint], arXiv:1308.0850, 2013.

He, K., Zhang, X., Ren, S., and Sun, J.: Deep Residual Learning for Image Recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 27-30 June 2016, 770-778, 10.1109/CVPR.2016.90, 2016

Hochreiter, S. and Schmidhuber, J.: Long Short-Term Memory, Neural Computation, 9, 1735-1780, 10.1162/neco.1997.9.8.1735, 1997.

435   Hwang, S., Yoon, J., Kang, N., and Lee, D.-R.: Development of flood forecasting system on city·mountains·small river area in Korea and assessment of forecast accuracy, Journal of Korea Water Resources Association, 53, 225-236, 10.3741/JKWRA.2020.53.3.225, 2020.

Imhoff, R. O., Brauer, C. C., Overeem, A., Weerts, A. H., and Uijlenhoet, R.: Spatial and Temporal Evaluation of Radar Rainfall Nowcasting Techniques on 1,533 Events, Water Resources Research, 56, e2019WR026723, 440   https://doi.org/10.1029/2019WR026723, 2020.

Ioffe, S. and Szegedy, C.: Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, in: Proceedings of the 32nd International Conference on Machine Learning, Proceedings of Machine Learning Research, 448-456, 2015.

Isola, P., Zhu, J., Zhou, T., and Efros, A. A.: Image-to-Image Translation with Conditional Adversarial Networks, in: 2017 445   IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 21-26 July 2017, 5967-5976, 10.1109/CVPR.2017.632, 2017.

Jeong, C. H., Kim, W., Joo, W., Jang, D., and Yi, M. Y.: Enhancing the Encoding-Forecasting Model for Precipitation Nowcasting by Putting High Emphasis on the Latest Data of the Time Step, Atmosphere, 12, 261, 2021.

Kim, S., Hong, S., Joh, M., and Song, S.-k.: Deeprain: Convlstm network for precipitation prediction using multichannel radar 450   data, in: 7th International Workshop on Climate Informatics, 20-22 September 2017, 89-92, 2017.

Kingma, D. P. and Ba, J.: Adam: A method for stochastic optimization, in: Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015), 2015.

Krizhevsky, A., Sutskever, I., and Hinton, G. E.: ImageNet classification with deep convolutional neural networks, in: Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, Lake Tahoe, 455   Nevada, 2012.

Kumar, A., Islam, T., Sekimoto, Y., Mattmann, C., and Wilson, B.: Convcast: An embedded convolutional LSTM based architecture for precipitation nowcasting using satellite data, PLOS ONE, 15, e0230114, 10.1371/journal.pone.0230114, 2020.

Long, J., Shelhamer, E., and Darrell, T.: Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE 460   Conference on Computer Vision and Pattern Recognition (CVPR), 3431-3440, 2015.

Marshall, J. S. and Palmer, W. M. K.: THE DISTRIBUTION OF RAINDROPS WITH SIZE, Journal of Atmospheric Sciences, 5, 165-166, 10.1175/1520-0469(1948)005<0165:Tdorws>2.0.Co;2, 1948.

Mathieu, M., Couprie, C., and LeCun, Y.: Deep multi-scale video prediction beyond mean square error, in: 4th International Conference on Learning Representations (ICLR 2016), 2016.

465   McCuen, R. H., Knight, Z., and Cutter, A. G.: Evaluation of the Nash-Sutcliffe Efficiency Index, Journal of Hydrologic Engineering, 11, 597-602, doi:10.1061/(ASCE)1084-0699(2006)11:6(597), 2006.

Mirza, M. and Osindero, S.: Conditional generative adversarial nets, arXiv [preprint], arXiv:1411.1784, 2014.

Mo, S., Cho, M., and Shin, J.: Freeze the discriminator: a simple baseline for fine-tuning gans, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition AI for Content Creation Workshop (CVPRW), 2020.

470   Moishin, M., Deo, R. C., Prasad, R., Raj, N., and Abdulla, S.: Designing Deep-Based Learning Flood Forecast Model With ConvLSTM Hybrid Algorithm, IEEE Access, 9, 50982-50993, 10.1109/ACCESS.2021.3065939, 2021.

Pan, S. J. and Yang, Q.: A Survey on Transfer Learning, IEEE Transactions on Knowledge and Data Engineering, 22, 1345-1359, 10.1109/TKDE.2009.191, 2010.

Poletti, M. L., Silvestro, F., Davolio, S., Pignone, F., and Rebora, N.: Using nowcasting technique and data assimilation in a 475   meteorological model to improve very short range hydrological forecasts, Hydrol. Earth Syst. Sci., 23, 3823-3841, 10.5194/hess-23-3823-2019, 2019.

Pulkkinen, S., Nerini, D., Pérez Hortal, A. A., Velasco-Forero, C., Seed, A., Germann, U., and Foresti, L.: Pysteps: an open-source Python library for probabilistic precipitation nowcasting (v1.0), Geosci. Model Dev., 12, 4185–4219, https://doi.org/10.5194/gmd-12-4185-2019, 2019.

480 Ranzato, M., Szlam, A., Bruna, J., Mathieu, M., Collobert, R., and Chopra, S.: Video (language) modeling: a baseline for generative models of natural videos, arXiv [preprint], arXiv:1412.6604, 2014.

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., and Mohamed, S.: Skilful precipitation nowcasting using deep generative models of radar, Nature, 597, 672-677,

485 10.1038/s41586-021-03854-z, 2021.

Rawat, W. and Wang, Z.: Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review, Neural Computation, 29, 2352-2449, 10.1162/neco_a_00990, 2017.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat: Deep learning and process understanding for data-driven Earth system science, Nature, 566, 195-204, 10.1038/s41586-019-0912-1, 2019.

490 Renzullo, L., Velasco-Forero, C., and Seed, A.: Blending radar, NWP and satellite data for real-time ensemble rainfall analysis: a scale-dependent method, CSIRO, https://doi.org/10.4225/08/594eb78c96025, 2017.

Ronneberger, O., Fischer, P., and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, edited by: Navab, N., Hornegger, J., Wells, W. M., and Frangi, A. F., Springer International Publishing, Cham, pp. 234–241, https://doi.org/10.1007/978-3-319-24574-4_28,

495 2015.

Rüttgers, M., Lee, S., Jeon, S., and You, D.: Prediction of a typhoon track using a generative adversarial network and satellite images, Scientific Reports, 9, 6057, 10.1038/s41598-019-42339-y, 2019.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-k., and Woo, W.-c.: Convolutional LSTM Network: a machine learning approach for precipitation nowcasting, in: Proceedings of the 28th International Conference on Neural Information

500 Processing Systems-Volume 1, 802-810, 2015.

Shi, X., Gao, Z., Lausen, L., Wang, H., Yeung, D.-Y., Wong, W.-k., and WOO, W.-c.: Deep Learning for Precipitation Nowcasting: A Benchmark and A New Model, in: Advances in Neural Information Processing Systems, 30, 5617-5627, 2017.

Simonyan, K. and Zisserman, A.: Very deep convolutional networks for large-scale image recognition, , in: Proceedings of the

505 3rd International Conference on Learning Representations (ICLR 2015), 2015.

Sønderby, C. K., Espeholt, L., Heek, J., Dehghani, M., Oliver, A., Salimans, T., Agrawal, S., Hickey, J., and Kalchbrenner, N.: Metnet: A neural weather model for precipitation forecasting, arXiv [preprint], arXiv:2003.12140, 2020.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting, The journal of machine learning research, 15, 1929-1958, 2014.

510 Sutskever, I., Vinyals, O., and Le, Q. V.: Sequence to sequence learning with neural networks, in: Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2, Montreal, Canada, 2014.

Trebing, K., Staǹczyk, T., and Mehrkanoon, S.: SmaAt-UNet: Precipitation nowcasting using a small attention-UNet architecture, Pattern Recognition Letters, 145, 178-186, https://doi.org/10.1016/j.patrec.2021.01.036, 2021.

Wang, S., Liu, W., Wu, J., Cao, L., Meng, Q., and Kennedy, P. J.: Training deep neural networks on imbalanced data sets, in:

515 2016 International Joint Conference on Neural Networks (IJCNN), 24-29 July 2016, 4368-4374, 10.1109/IJCNN.2016.7727770, 2016.

Wang, Y., Coning, E., Harou, A., Jacobs, W., Joe, P., Nikitina, L., Roberts, R., Wang, J., and Wilson, J.: Guidelines for nowcasting techniques, WMO publication, published online: https://library. wmo. int/opac/doc_num. php, 2017.

Wang, Y., Gonzalez-Garcia, A., Berga, D., Herranz, L., Khan, F. S., and Weijer, J. v. d.: MineGAN: Effective Knowledge

520 Transfer From GANs to Target Domains With Few Images, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 13-19 June 2020, 9329-9338, 10.1109/CVPR42600.2020.00935, 2020.

Wang, Y., Wu, C., Herranz, L., van de Weijer, J., Gonzalez-Garcia, A., and Raducanu, B.: Transferring gans: generating images from limited data, in: Proceedings of the European Conference on Computer Vision (ECCV), 218-234, 2018.

Wang, Y., Yao, Q., Kwok, J., and Ni, L.: Generalizing from a Few Examples: A Survey on Few-shot Learning, ACM

525 Computing Surveys, 53, 1-34, 10.1145/3386252, 2020.

Yosinski, J., Clune, J., Bengio, Y., and Lipson, H.: How transferable are features in deep neural networks?, in: Advances in Neural Information Processing Systems, 27, 3320-3328, 2014.
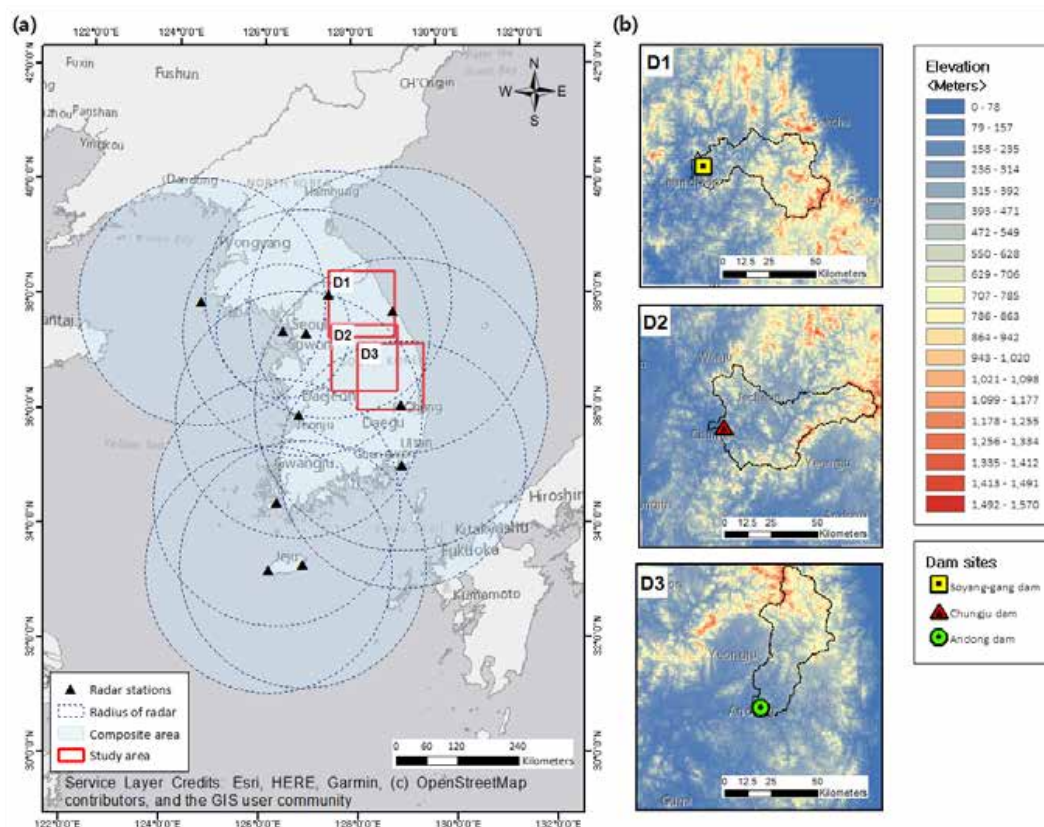
530

Figure 1: (a) Radar reflectivity composite map and location of the domain; (b) selected domains over the dam basin. The boxes of D1, D2, and D3 present the domains of Soyang-gang, Chungju, and Andong dam basins, respectively. Maps were created using ArcGIS software by Esri; Base-map source: Esri, HERE, Garmin, © OpenStreetMap contributors, and the GIS User Community.
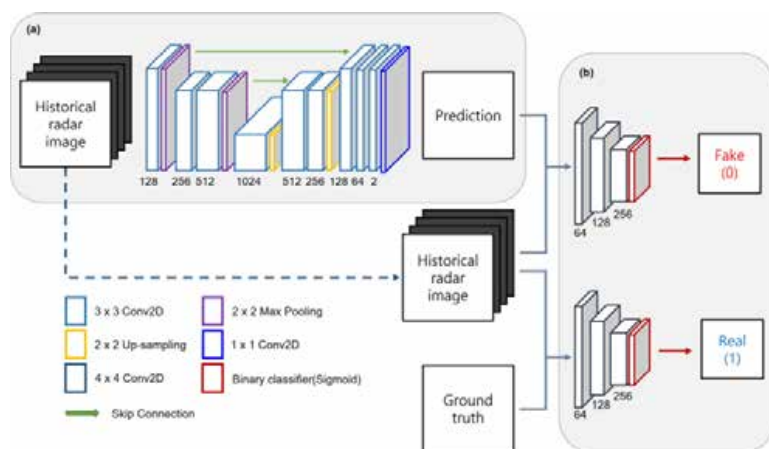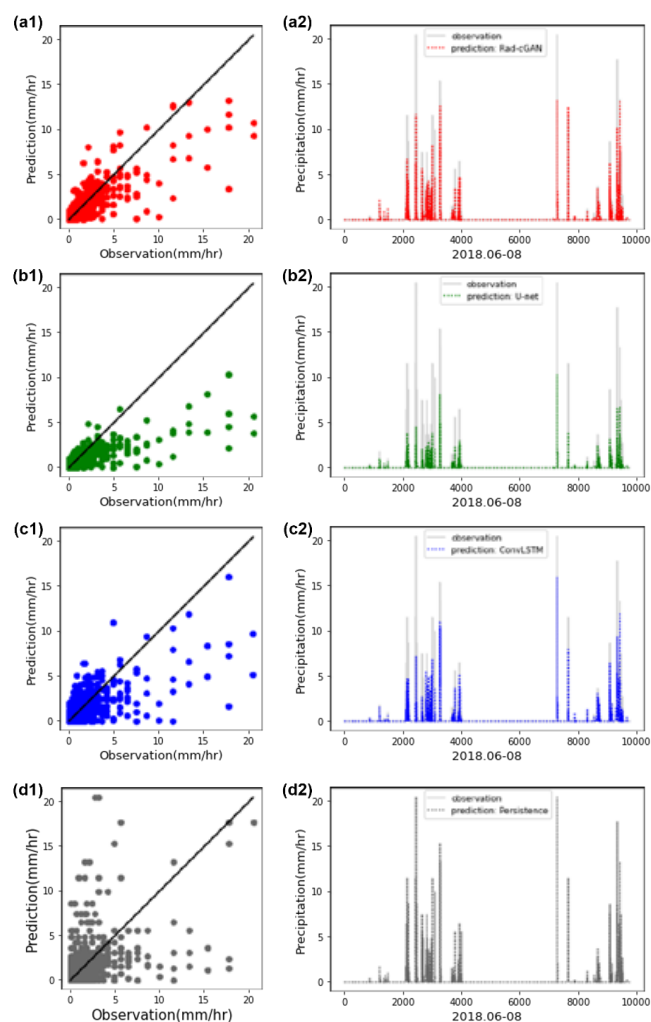
535

Figure 2: Model architecture consists of a (a) generative and (b) discriminative model.

540 **Figure 3: Evaluation of model performance for 10-min lead time prediction at Soyang-gang Dam site with the scatter plot (left column) and the time series (right column) based on (a) Rad-cGAN model, (b) U-net based model, (c) ConvLSTM model, and (d) Eulerian Persistence model.**
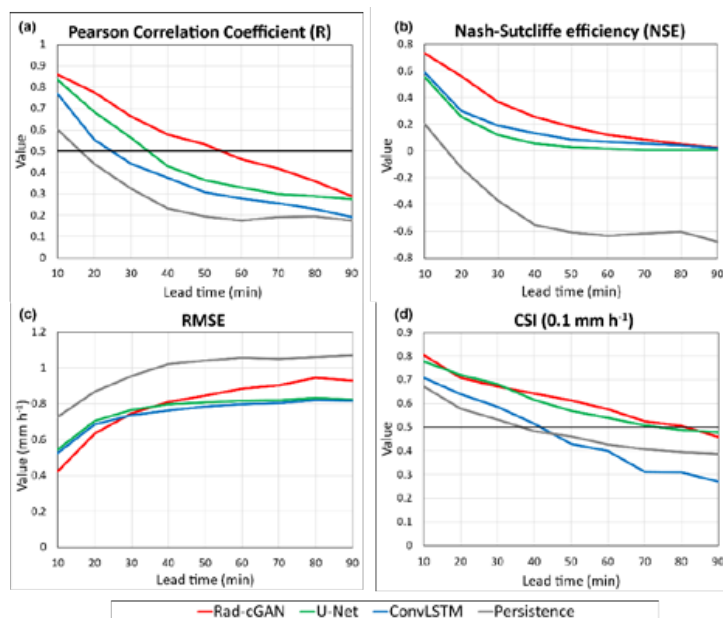
545

Figure 4: Verification metrics of model predictions at the lead time up to 90 min at the Soyang-gang Dam site. (a) R, (b) NSE, (c) RMSE, and (d) CSI at intensity threshold of 0.1 mm h⁻¹.
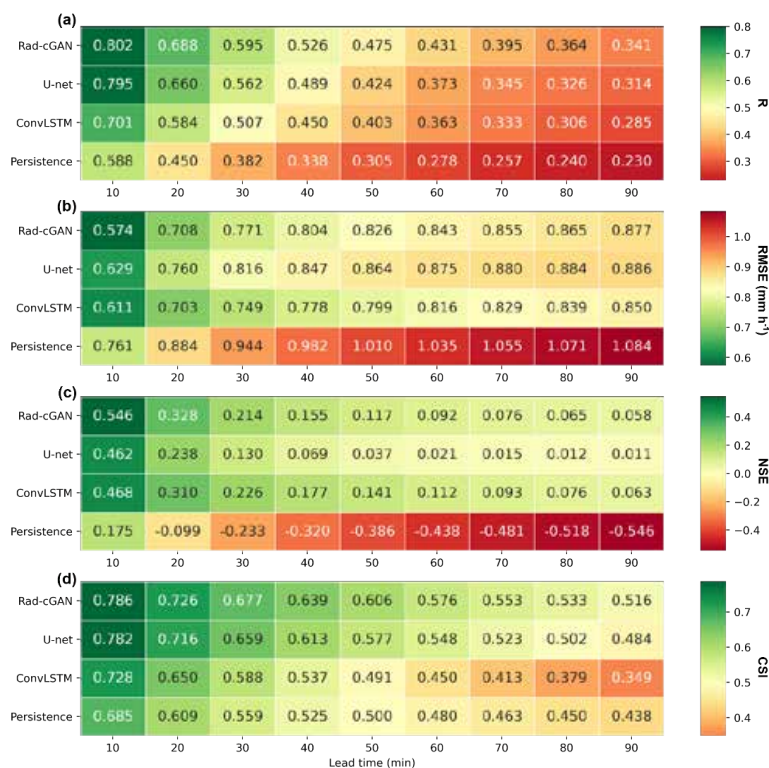
Figure 5: Median values of verification metrics of model predictions at the lead time up to 90 min over all grid cells from the Soyang-gang Dam region. Panels from top to bottom represent R, RMSE, NSE, and CSI at intensity threshold of 0.1 mm h^{-1}.
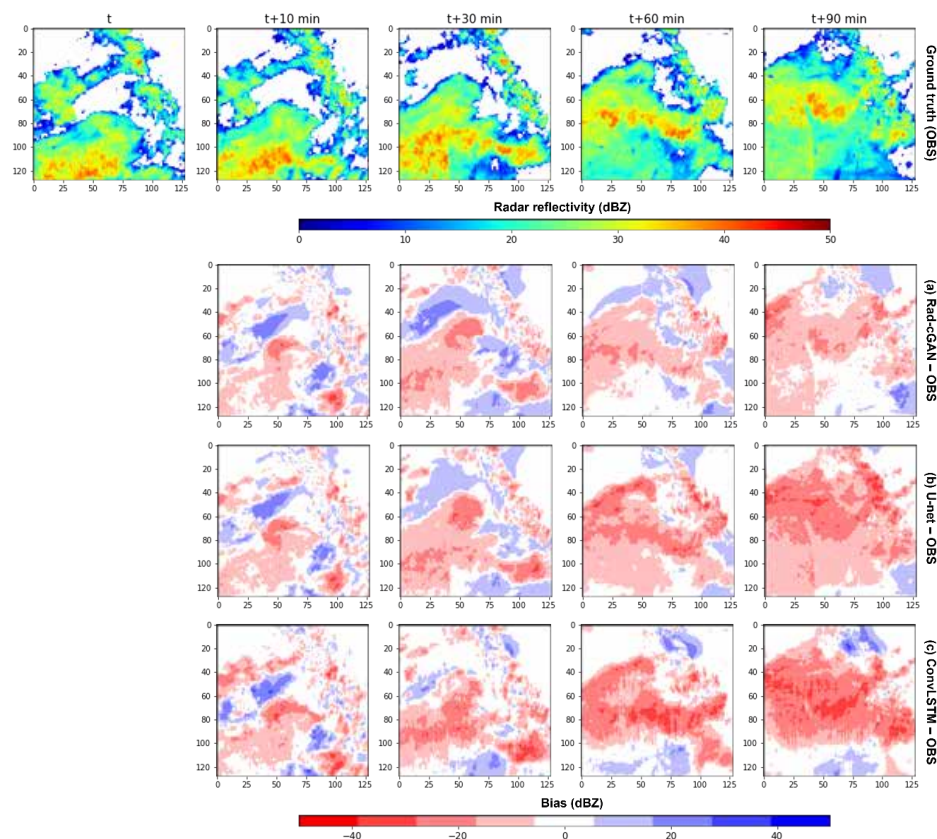
Figure 6: Radar reflectivity (dBZ) observation example at forecasting time _t_ = 23 August 2018, 17:50 UTC, and the bias that represents difference between model prediction and ground truth (OBS). Panels from top to bottom express ground truth: (a) difference of Rad-cGAN model, (b) difference of U-net based model, and (c) difference of ConvLSTM.
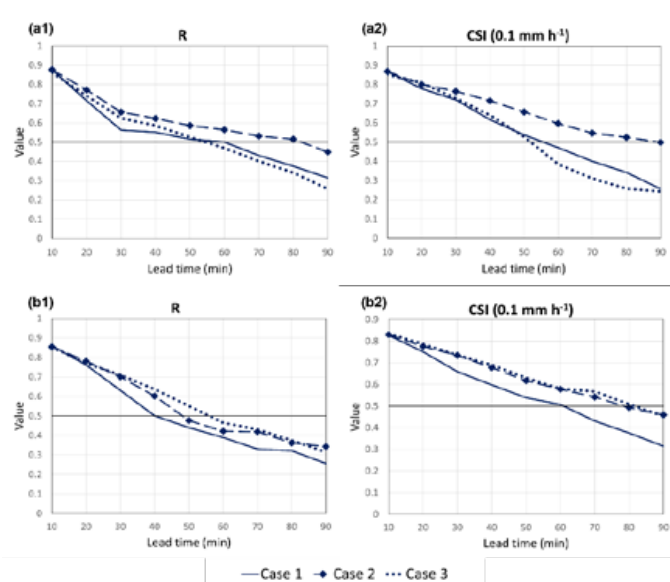
Figure 7: Verification metrics of model predictions at lead time up to 90 min at each domain at (a) Andong Dam site and (b) Chungju Dam site with R (left column) and CSI at intensity threshold of 0.1 mm h⁻¹ (right column).

Table 1: Contingency table for the categorical scores.

		Observation	
		Event detected	Event not detected
Prediction	Event detected	Hit	False alarm
	Event not detected	Miss	Correct non-event

565

Table 2: Experimental design for transfer learning strategies to train model with different domain. (a) Detailed training procedure of each strategy, (b) data used to train the model according to each strategy.

(a) Training strategies		
No.	Generator	Discriminator
Case 1	Train from the scratch	Train from the scratch
Case 2	Use pre-trained parameters	Train from the scratch
Case 3	Fine tuning pre-trained parameters[*]	Train from the scratch

(b) Training dataset			
	Pre-trained domain	Andong Dam domain	Chungju Dam domain
Case 1	-	2014–2017 (JJA) at Andong Dam domain	2014–2017 (JJA) at Chungju Dam domain
Case 2	2014-2017 (JJA) at Soyang-gang Dam domain	2014–2017 (JJA) at Andong Dam domain	2014–2017 (JJA) at Chungju Dam domain
Case 3	2014-2017 (JJA) at Soyang-gang Dam domain	2014–2017 (JJA) at Andong Dam domain	2014–2017 (JJA) at Chungju Dam domain

[*] Use 1/10th of original learning rate

570 **Table 3: Comparison of the performance scores for 10-min precipitation prediction of different models at the Soyang-gang Dam site during summer season (June–August) of 2018.**

	R	RMSE (mm h^{-1})	NSE	CSI
Rad-cGAN	**0.8587**	**0.4248**	**0.7277**	**0.8065**
U-Net	0.8367	0.5457	0.5506	0.7781
ConvLSTM	0.7716	0.5241	0.5855	0.7086
Persistence (baseline)	0.6023	0.7287	0.1989	0.6716

575

580

Table 4: Comparison of the 10-min precipitation prediction performance scores for three different model using different transfer learning strategies for the (a) Andong and (b) Chungju Dam sites in the summer season (June–August) of 2018.

(a) Andong Dam site				
	R	RMSE (mm h^{-1})	NSE	CSI
Case 1	0.8597	0.5338	0.5778	0.8541
Case 2	0.8760	0.4919	0.7152	0.8683
Case 3	0.8611	0.5999	0.5762	0.8522
(b) Chungju Dam site				
	R	RMSE (mm h^{-1})	NSE	CSI
Case 1	0.8603	0.8562	0.6134	0.8271
Case 2	0.8539	0.7328	0.7168	0.8301
Case 3	0.8526	0.7937	0.6677	0.8344

585

590

595