Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Precip-UNetGAN: An Attention-Guided Multi-Scale GAN for Radar-Based Precipitation Nowcasting with Metric-Oriented Optimization

**Juncong Shao**
Henan University
**Yanhong Liu**
yanhongliu@vip.henu.edu.cn
Henan University

**Research Article**

**Keywords:** precipitation nowcasting, radar echo extrapolation, generative adversarial network (GAN), multi-scale attention, metric-oriented learning, Soft-CSI loss

**Posted Date:** November 4th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-10727857/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

# Precip-UNetGAN: An Attention-Guided Multi-Scale GAN for Radar-Based Precipitation Nowcasting with Metric-Oriented Optimization

**Juncong Shao, Yanhong Liu***

School of Software, Henan University, Kaifeng 475004, China

## Abstract

Radar-based precipitation nowcasting must preserve the location, intensity, and spatial organization of rapidly evolving precipitation while maintaining skill at operationally relevant thresholds. This study introduces Precip-UNetGAN, a 0–30 min radar nowcasting framework designed to reduce smoothing of localized high-intensity echoes and the mismatch between pixel-wise training objectives and event-based verification metrics. The generator combines a UNet encoder–decoder with multi-scale spatial attention, while adversarial learning constrains the spatial realism of predicted precipitation structures. A metric-oriented objective combines reflectivity-weighted mean squared error with a time-weighted differentiable critical success index (Soft-CSI). On the Royal Netherlands Meteorological Institute dataset, Precip-UNetGAN achieved the highest critical success index and Heidke skill score among the evaluated baselines at thresholds of 0.5, 2, 5, and 10 mm h(^{-1}), with the clearest gains for moderate-to-heavy precipitation. Experiments on the Shanghai radar dataset reproduced the same overall ranking after dataset-specific training, indicating cross-dataset reproducibility while confirming the need for local calibration. Overall, combining scale-aware feature selection with threshold-oriented optimization improves the retention of intense precipitation structures and balances detection against false alarms, providing a practical basis for further development of short-term radar nowcasting。

**Keywords:** precipitation nowcasting; radar echo extrapolation; generative adversarial network (GAN); multi-scale attention; metric-oriented learning; Soft-CSI loss

*Corresponding author: yanhongliu@vip.henu.edu.cn

## 1. Introduction

Precipitation nowcasting supports flash-flood warning, urban drainage management, aviation operations, and short-fuse severe-weather alerts. The central scientific and operational difficulty is not simply to extrapolate radar echoes, but to represent the coupled displacement, deformation, growth, and decay of precipitation systems over short lead times. This difficulty is particularly acute for localized convective cells, whose high-reflectivity cores occupy a small fraction of the radar domain but dominate hydrometeorological impacts. Optical-flow and advection-based methods represent translation efficiently, yet their underlying persistence assumptions limit their ability to describe rapid intensification, splitting, merging, and dissipation (Germann and Zawadzki 2002; Sakaino 2013; Sun et al. 2014; Woo and Wong 2014).

Deep-learning models can learn nonlinear mappings from recent radar sequences to future precipitation fields, but three limitations remain important for atmospheric applications. First, pixel-wise losses tend to favor the dominant no-rain and weak-rain pixels, producing smooth forecasts and underestimating compact high-intensity echoes. Second, precipitation systems span multiple spatial scales: broad stratiform regions, embedded convective cores, and narrow boundaries must be represented simultaneously. Third, minimizing mean error does not necessarily improve event-based scores used in warning operations, including the critical success index (CSI), which measures the overlap of forecast and observed events; probability of detection (POD), which measures the fraction of observed events that are detected; precision, which measures the fraction of forecast events that are observed; and the Heidke skill score (HSS), which measures categorical forecast skill relative to

random agreement. Consequently, a model may produce visually plausible fields while still missing threshold exceedances that matter for warnings, or it may increase detection at the cost of excessive false alarms.

This study therefore examines whether scale-aware spatial attention, adversarial structural constraints, and direct optimization of differentiable event scores can jointly improve the representation of precipitation intensity and morphology. Precip-UNetGAN uses a UNet generator to retain fine and coarse spatial information, a multi-scale attention module to emphasize spatially sparse precipitation structures, and a discriminator to discourage unrealistic smoothing. Its loss function gives greater weight to moderate and heavy precipitation in reflectivity space and directly optimizes a time-weighted Soft-CSI across multiple rain-rate thresholds. The analysis focuses not only on aggregate performance, but also on how forecast skill changes with precipitation intensity and lead time, what type of error is reduced, and whether the behavior transfers between the Netherlands and Shanghai radar datasets.

1. The study formulates radar nowcasting as a joint problem of preserving multi-scale precipitation structure and detecting threshold exceedances, rather than as pixel-wise image reconstruction alone.

2. A reflectivity-weighted reconstruction loss and a time-weighted Soft-CSI objective are introduced to counter class imbalance and align training with operational verification at 0.5, 2, 5, and 10 mm/h.

3. Experiments are interpreted by precipitation intensity, lead time, error type, and region, allowing the physical and operational implications—and the limits of transferability—to be assessed.

## 2. Related Work

Radar-based precipitation nowcasting has gradually evolved from motion-based extrapolation to data-driven spatiotemporal prediction. Optical-flow and advection-based methods efficiently describe the displacement of existing radar echoes, but their persistence assumptions limit their ability to represent precipitation growth,

63 decay, splitting, and merging (Germann and Zawadzki 2002; Sakaino 2013; Sun et al. 2014; Woo and Wong
64 2014). Deep-learning methods attempt to overcome these limitations by learning nonlinear mappings from
65 recent radar observations to future precipitation fields.

66 Recurrent spatiotemporal models form an important class of deep-learning nowcasting methods. ConvLSTM
67 introduced convolutional state transitions to jointly represent spatial and temporal dependencies (Shi et al. 2015),
68 while TrajGRU, PredRNN, and Memory-in-Memory further improved nonlinear motion modeling and temporal
69 memory (Shi et al. 2017; Wang et al. 2019, 2023a). Recent studies have incorporated attention, multi-scale
70 processing, and lightweight recurrent units to improve local feature retention and computational efficiency (Liu
71 et al. 2025; Wang et al. 2025; Cheng et al. 2025; Wu et al. 2025). Nevertheless, recurrent forecasts often become
72 progressively smoother with increasing lead time, resulting in the attenuation of localized high-intensity echoes.

73 Convolutional encoder–decoder models provide an alternative approach to preserving precipitation structures at
74 multiple spatial resolutions. UNet-based architectures use hierarchical feature extraction and skip connections to
75 combine broad precipitation organization with local spatial detail (Ronneberger et al. 2015). RainNet,
76 Broad-UNet, and SmaAt-UNet have demonstrated the effectiveness of fully convolutional, multi-scale, and
77 attention-based designs for radar nowcasting (Ayzel et al. 2020; Fernández and Mehrkanoon 2021; Trebing et al.
78 2021). However, multi-scale feature extraction alone does not resolve the strong imbalance between widespread
79 weak precipitation and spatially sparse intense echoes. When conventional pixel-wise losses dominate training,
80 predicted fields may still exhibit weakened precipitation cores and blurred boundaries.

81 Generative adversarial learning has therefore been introduced to reduce the excessive smoothing associated with
82 deterministic regression. By constraining generated fields to resemble the spatial distribution of observations,
83 GAN-based methods can improve the continuity and sharpness of predicted radar echoes (Goodfellow et al.

2014; Isola et al. 2017; Ravuri et al. 2021; Wang et al. 2023b). Attention- and transformer-based models, including Rainformer and AA-TransUNet, further enhance global dependency modeling, while NowcastNet, SFTFormer, and MDTNet incorporate physical evolution information or frequency-domain representations (Bai et al. 2022; Yang and Mehrkanoon 2022; Zhang et al. 2023; Xu et al. 2024; Zhao et al. 2024). However, visually realistic or structurally sharp forecasts do not necessarily achieve accurate intensity calibration or improved threshold-based verification scores.

Despite these advances, three related challenges remain. First, rare moderate-to-heavy precipitation pixels are easily suppressed by the dominant no-rain and weak-rain background. Second, precipitation systems contain broad rain areas, compact convective cores, and sharp boundaries that must be represented at different spatial scales. Third, commonly used pixel-wise objectives are not directly aligned with categorical verification metrics such as the critical success index, probability of detection, precision, and Heidke skill score. Precip-UNetGAN addresses these challenges by combining multi-scale spatial attention, adversarial structural constraints, reflectivity-weighted reconstruction, and differentiable metric-oriented optimization. Its performance is evaluated across precipitation intensity, forecast lead time, error trade-offs, and two regional radar datasets.

## 3. Model

In this section, a precipitation field denotes a two-dimensional grid of radar-derived surface rain rate. Rain rate is denoted by R or r and is expressed in millimetres per hour (mm/h). A forecast lead time is the elapsed time between the latest input observation and the predicted field; the model uses 12 input fields separated by 5 min and predicts six fields at lead times of 5, 10, 15, 20, 25, and 30 min. The generator G maps the historical sequence to the forecast sequence, whereas the discriminator D returns the probability that an input field is observational rather than generated.

Precip-UNetGAN maps 12 consecutive radar-derived precipitation fields to six forecasts at 5-min intervals. The framework contains a UNet generator, multi-scale spatial attention, and a convolutional discriminator (Fig. 1). The generator represents the evolving precipitation field at several spatial scales; attention increases the contribution of spatially sparse but meteorologically important echo regions; and adversarial training penalizes structurally implausible outputs. The joint objective combines intensity-sensitive reconstruction and threshold-based event detection. Detailed architectural specifications are provided in Supplementary Figs. S1–S3 and Supplementary Table S3, while the main scientific role of each component is summarized below.

The model is designed around three forecast-error mechanisms. Multi-scale attention addresses the coexistence of broad precipitation shields and compact convective cores. The discriminator addresses the tendency of regression models to average uncertain futures and blur echo boundaries. The dBZ-weighted and Soft-CSI terms address the under-representation of rare intense pixels and the mismatch between continuous reconstruction error and categorical warning metrics. The method remains data-driven and does not explicitly solve atmospheric governing equations; therefore, "physical plausibility" in this study refers to consistency with observed radar morphology and intensity distributions rather than explicit dynamical conservation.

### 3.1 Multi-Scale Attention-Guided Generative Adversarial Extrapolation Network

Radar-based precipitation nowcasting requires the simultaneous representation of broad precipitation systems, compact convective cores, and sharp echo boundaries. These structures occur at different spatial scales and evolve differently with forecast lead time. Pixel-wise regression models often average over uncertain future states, which can attenuate high-intensity echoes and blur precipitation boundaries. To reduce these errors, the proposed multi-scale attention-guided generative adversarial network combines a UNet generator, a multi-scale spatial attention module, and a convolutional discriminator.

Let X ∈ R^(T_in×H×W) denote the historical precipitation sequence, where T_in = 12 is the number of input times and H and W are the grid height and width, respectively. The generator G maps X to the predicted sequence

$$\hat{Y} = G(X) \quad (1)$$

where $\hat{Y} \in R^{(T_{out} \times H \times W)}$ contains T_out = 6 forecast fields at lead times of 5, 10, 15, 20, 25, and 30 min. The corresponding observed sequence is denoted by Y.

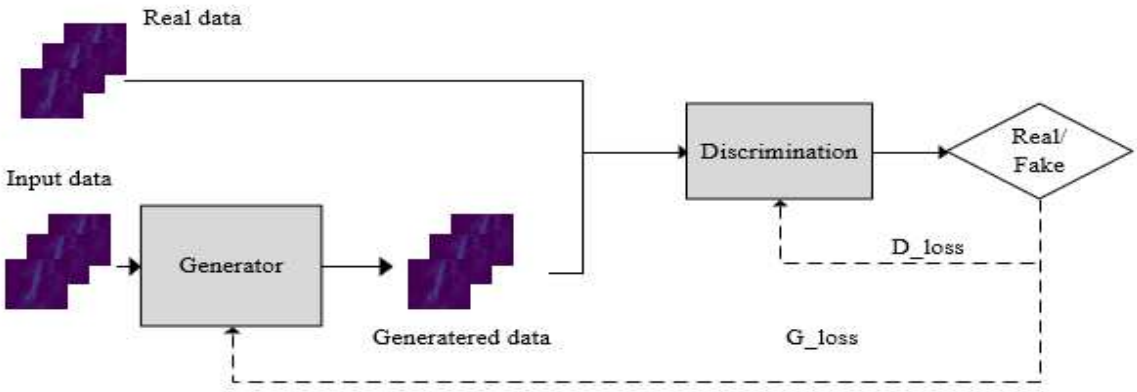



**Fig. 1** Overview of Precip-UNetGAN. Twelve historical precipitation fields are used to predict six future fields at 5–30 min lead times, with adversarial training providing a structural constraint.

The generator follows a UNet encoder-decoder structure. The encoder progressively reduces spatial resolution while extracting precipitation features at increasingly broad spatial scales. The decoder restores the original grid resolution and combines deep features with encoder features through skip connections. This design allows large-scale precipitation organization to be represented together with local gradients and compact echo structures.

To increase the contribution of spatially sparse but meteorologically important precipitation regions, a multi-scale spatial attention module is applied to intermediate generator features. Let U ∈ R^(C×H_s×W_s) denote an input feature map at a given network resolution, where C is the number of feature channels and H_s and W_s are the

corresponding spatial dimensions. Channel-wise average pooling and channel-wise maximum pooling are first applied to U, producing two spatial descriptors:

$$U_avg = AvgPool_c(U) \quad (2)$$

$$U_max = MaxPool_c(U) \quad (3)$$

$$S = [U_avg ; U_max] \quad (4)$$

where AvgPool_c and MaxPool_c denote pooling along the channel dimension, and [U_avg ; U_max] denotes concatenation along the channel dimension.

The spatial descriptor S is then processed by three convolution branches with kernel sizes 3 × 3, 5 × 5, and 7 × 7. Their outputs are denoted by $F_k$ , where $k \in \{1, 2, 3\}$ identifies the spatial scale:

$$F_k = \sigma(BN(Conv_k(S))) \quad (5)$$

$$\alpha_k = \frac{\exp(w_k)}{\sum_{i=1}^{N} \exp(w_i)}, k \in \{1, ..., N\}, 0 \leq \alpha_k \leq 1, \quad \sum_{k=1}^{N} \alpha_k = 1 \quad (6)$$

$$A = \sum_{k=1}^{N} \alpha_k \cdot F_k \quad (7)$$

$$\tilde{U} = U \odot A \quad (8)$$

Here, N = 3 is the number of convolution branches, i is the summation index, and exp(·) denotes the exponential function. The fused spatial attention map is denoted by A. The attention-enhanced feature map $\tilde{U}$ is obtained by element-wise multiplication, where ⊙ denotes the Hadamard product. Through this operation, the network can increase the influence of compact heavy-precipitation cores and strong spatial gradients without discarding the broader precipitation background. The attention module therefore addresses the coexistence of precipitation structures at different spatial scales rather than merely increasing network complexity.

A convolutional discriminator D is introduced to reduce the excessive smoothing commonly produced by regression-only training. The discriminator receives either an observed precipitation field Y or a generated field $\hat{Y}$ and returns the probability that the input originates from the observational distribution:

$$D(Y) \in [0, 1] (9)$$

$$D(\hat{Y}) \in [0, 1] (10)$$

Values close to 1 indicate that the discriminator considers the input observational, whereas values close to 0 indicate that it considers the field generated. During training, the discriminator is optimized to distinguish observed and generated fields, while the generator is optimized to produce forecasts that are difficult to distinguish from observations. The adversarial constraint complements the intensity-based reconstruction losses by penalizing unrealistic spatial discontinuities, diffuse boundaries, and implausible echo textures.

The scientific roles of the three components are distinct. The UNet architecture retains information across coarse and fine spatial resolutions; the multi-scale attention module emphasizes sparse precipitation structures whose contribution may otherwise be suppressed by the dominant weak-rain background; and the discriminator discourages spatial smoothing caused by averaging uncertain future states. The framework remains data-driven and does not explicitly represent atmospheric dynamics or microphysical processes. Consequently, improved structural realism refers to closer agreement with observed precipitation morphology and intensity organization, rather than explicit physical conservation or causal representation of storm development.

### *3.2. dBZ-Weighted Mean Squared Error Loss（DBL）*

The conventional mean squared error (MSE), defined as the mean of squared differences between predicted and

observed values, is easily dominated by the large proportion of no-rain and weak-precipitation pixels. To increase sensitivity to different precipitation intensities, this study introduces the dBZ-weighted MSE loss (DBL). Here, R denotes surface rain rate in mm /h, which represents the depth of liquid precipitation that would accumulate in one hour if the instantaneous rate remained constant. Z denotes the radar reflectivity factor in $mm^6$ /$m^3$, a quantity proportional to the sixth moment of the hydrometeor diameter distribution under Rayleigh-scattering assumptions. Radar reflectivity is reported on the logarithmic decibel scale dBZ, defined relative to 1 $mm^6/m^3$; larger dBZ values generally indicate stronger radar echoes. The empirical Marshall–Palmer Z–R relation uses the positive coefficients a and b to convert R to Z.

With a denoting the multiplicative coefficient and b the exponent in the empirical Z–R relation, the reflectivity factor is calculated as:

$$Z=aR^b \quad (11)$$

For the Marshall–Palmer relation used here, a = 200 and b = 1.6. These coefficients provide a common conversion for stratiform rain, although their optimal values can vary with precipitation type, drop-size distribution, and region.

Subsequently, the reflectivity factor is converted into the dBZ representation through a logarithmic transformation, which can be written as:

The logarithmic reflectivity value dBZ is then obtained from Z, where log10(·) denotes the base-10 logarithm:

$$dBZ=10\log_{10}(Z) \quad (12)$$

By substituting Equation (11) into Equation (12), the complete mapping from mm /h to dBZ can be obtained as:

Substituting a = 200 and b = 1.6 into the preceding two equations gives the direct mapping from rain rate R to dBZ:

$$dBZ=10\log_{10}(200\cdot R^{1.6})=10\log_{10}(200)+16\log_{10}(R) \quad (13)$$

After the mapping from mm/h to dBZ is completed, the error term is weighted according to the intensity interval to which the ground-truth dBZ value belongs. Specifically, the dBZ range is divided into several intervals, and adjustable weights W1, W2, W3, and W4 are assigned to different intervals, thereby constructing the weighted MSE loss as follows:

Let $Z$ denote the observed (ground-truth) reflectivity in dBZ and let $\hat{Z}$ denote the predicted reflectivity in dBZ. Let W(Z) be a dimensionless piecewise weight determined by the observed intensity interval, E[·] the mean over the samples and pixels included in the loss, and $\mathscr{L}_{DW-MSE}$ the resulting weighted reconstruction loss. It is defined as:

$$\mathscr{L}_{DW-MSE}=E\left[W(Z)\cdot(\hat{Z}-Z)^2\right] \quad (14)$$

The weighting function uses four positive dimensionless coefficients W1, W2, W3, and W4. Larger coefficients assign a greater training contribution to rarer, stronger echoes.

The four reflectivity intervals are separated at 25, 40, and 60 dBZ. With Z representing the observed reflectivity, the weighting function is:

$$W(Z)=\begin{cases} W1, & Z<25\ \text{dBZ}, \\ W2, & 25\ \text{dBZ}\le Z<40\ \text{dBZ}, \\ W3, & 40\ \text{dBZ}\le Z<60\ \text{dBZ}, \\ W4, & Z\ge 60\ \text{dBZ}. \end{cases} \quad (15)$$

*3.3. Time-Weighted Soft-CSI Loss*

The critical success index (CSI) is a dimensionless categorical verification score that measures the overlap between forecast and observed precipitation events while penalizing both false alarms and misses. Conventional CSI uses a hard rain-rate threshold and is non-differentiable. The proposed time-weighted Soft-CSI replaces the hard event indicator with a differentiable approximation and assigns explicit weights to forecast lead times.

Let $\hat{Y}t$ denote the predicted rain-rate field and Yt the observed rain-rate field at forecast time index t, both in mm/h. Let θ be a rain-rate threshold in mm/h, $\sigma(x) = \frac{1}{(1 + \exp(-x))}$ the Sigmoid function, and $\tau > 0$ a temperature parameter with the same unit as rain rate so that $\frac{Y_t - \theta}{\tau}$ is dimensionless. The soft forecast-event indicator $p_{t,\theta}$ and soft observed-event indicator $g_{t,\theta}$ take values in [0, 1] and approximate whether the corresponding field exceeds θ:

$$p_{t,\theta} = \sigma\left(\frac{\hat{Y}_t - \theta}{\tau}\right) \quad (16)$$

$$g_{t,\theta} = \sigma\left(\frac{Y_t - \theta}{\tau}\right) \quad (17)$$

A smaller τ produces a sharper transition that more closely approximates hard thresholding, whereas a larger τ produces a smoother transition.

Using $p_{t,\theta}$ and $g_{t,\theta}$, define the soft true-positive amount $TP_{t,\theta}$ soft false-positive amount $FP_{t,\theta}$ and soft false-negative amount $FN_{t,\theta}$. TP represents forecast–observation agreement above the threshold, FP represents forecast exceedance without observed exceedance, and FN represents observed exceedance without forecast exceedance. E[·] denotes averaging over the batch and spatial dimensions:

$$TP_{t,\theta} = E\left[p_{t,\theta} \cdot g_{t,\theta}\right] \quad (18)$$

$$FP_{t,\theta} = E\left[p_{t,\theta} \cdot (1 - g_{t,\theta})\right] \quad (19)$$

$$FN_{t,\theta} = E\left[(1 - p_{t,\theta}) \cdot g_{t,\theta}\right] \quad (20)$$

where *E*[.] denotes the averaging operation over both the batch dimension and the spatial dimensions, that is, the mean value over all pixel locations of all samples in a batch, which can be expressed as:

For an arbitrary field f, let B be the batch size, H and W the grid height and width in pixels, b the sample index, and i and j the row and column indices. The value at location (i, j) of sample b is fb,i,j. The averaging operator

is:

$$E[f]=\frac{1}{BHW}\sum_{b=1}^{B}\sum_{i=1}^{H}\sum_{j=1}^{W} f_{b,i,j} \quad (21)$$

This normalization makes TP, FP, and FN comparable across batches and grid sizes.

Based on the above definitions, the differentiable Soft-CSI metric can be further formulated as:

Let $\varepsilon > 0$ be a small dimensionless numerical constant that prevents division by zero. The differentiable Soft-CSI at time t and threshold $\theta$ is:

$$CSI_{t,\theta}=\frac{TP_{t,\theta}}{TP_{t,\theta}+FP_{t,\theta}+FN_{t,\theta}+\varepsilon} \quad (22)$$

The resulting Soft-CSI is dimensionless and approaches the conventional CSI as the soft indicators approach binary values.

Let T denote the number of forecast times, $t \in \{1, \ldots, T\}$ the forecast-time index, and $\alpha \geq 0$ a dimensionless exponent controlling how strongly later lead times are emphasized. The normalized temporal weight assigned to time t is denoted by $w_t$:

$$w_t=\frac{t^{\alpha}}{\frac{1}{T}\sum_{k=1}^{T} k^{\alpha}} \quad (23)$$

In this study $\alpha = 0.5$. The denominator normalizes the weights so that their mean over the T forecast times equals 1, preventing a change in the overall loss magnitude.

Let $\Theta$ denote the set of rain-rate thresholds and $|\Theta|$ its number of elements.

With $\Theta = \{0.5, 2, 5, 10\}$ mm /h, the time-weighted Soft-CSI loss $\mathscr{L}_{TW-SoftCSI}$ is the average of $1 - CSI_{t,\theta}$ over all forecast times and thresholds:

$$\mathscr{L}_{TW-SoftCSI}=\frac{1}{T|\Theta|}\sum_{t=1}^{T}\sum_{\theta \square \Theta} w_t\,(1-CSI_{t,\theta}) \quad (24)$$

Let $\mathscr{L}_G$ denote the total generator objective, $\mathscr{L}_{TW-SoftCSI}$ the dBZ-weighted reconstruction loss, $\mathscr{L}_{TW-SoftCSI}$ the time-weighted event loss, and $\mathscr{L}$_GAN the adversarial generator loss supplied by the discriminator. The

dimensionless coefficients $\lambda_{DW}$, $\lambda_{CSI}$and $\lambda_{adv}$ control the relative contributions of these three terms. The total objective is:

$$\mathcal{L}_G=\lambda_{DW}\mathcal{L}_{DW-MSE}+\lambda_{CSI}\mathcal{L}_{TW-SoftCSI}+\lambda_{adv}\mathcal{L}_{GAN} \quad (25)$$

*3.4. Training Procedure*

During the training stage, the maximum number of epochs was set to 200, the batch size was set to 8, and the Adam optimizer was employed for parameter updating. The model was trained end-to-end under a generative adversarial learning framework, in which the parameters of the generator and discriminator were updated alternately. The generator takes historical radar echo sequences as input and outputs precipitation forecasts for multiple future time steps. The discriminator takes both real echoes and generated echoes as input and determines their authenticity, thereby providing adversarial constraint signals to the generator. During training, the parameters of the generator were first fixed while the discriminator was updated to improve its ability to distinguish real samples from generated ones. Subsequently, the discriminator parameters were fixed, and the generator was updated by minimizing the generator loss so that the forecasting quality could be progressively improved through the adversarial game.

The training objective of the generator consists of multiple loss terms, including the dBZ-weighted mean squared error loss, the time-weighted Soft-CSI loss and the adversarial loss. Among them, the time-weighted Soft-CSI loss directly optimizes event-level forecasting performance by using a Sigmoid-based soft-threshold approximation. The threshold set was defined as $\Theta = \{0.5, 2, 5, 10\}$, and the exponent parameter for temporal weighting was set to $\alpha = 0.5$ to balance the relative importance of different forecast lead times. The temperature parameter $\tau = 0.5$ was used to control the smoothness of the soft-threshold transition. The dBZ-weighted mean squared error loss maps the precipitation rate into the dBZ space through the Z-R relationship and applies

piecewise weights W1–W4 to strengthen the constraints on moderate-to-heavy precipitation regions. These loss terms are linearly combined through weighting coefficients so as to achieve a balance among numerical accuracy, spatial structure, and event-level forecasting performance. Overall, through the joint optimization of multiple loss terms and the adversarial training mechanism, the model is able to learn the spatiotemporal evolution patterns of radar echoes and generate precipitation forecasts that perform favorably in terms of intensity distribution, spatial structure, and operational evaluation metrics.For the rare $r \geq 10$ mm/h experiment, Precip-UNetGAN was trained separately using a stronger dBZ-weighting configuration to increase the contribution of high-intensity precipitation pixels. The corresponding weighting coefficients are provided in Supplementary Table S2.

## 4. Experiments

### *4.1. Datasets*

The primary dataset was obtained from the Royal Netherlands Meteorological Institute (KNMI) and follows the data configuration used by Trebing et al. (2021). Here, a 5-min precipitation accumulation is the liquid-water depth accumulated during the preceding 5 min, expressed in millimetres (mm). It contains 5-min radar-derived precipitation accumulations over the Netherlands and surrounding areas from 2016–2019. The composite is based on two C-band Doppler radars, where C-band denotes microwave wavelengths of approximately 4–8 cm, at De Bilt and Den Helder and is gauge-adjusted. Each original grid contains 765 × 700 pixels at 1-km spacing; pixel values represent precipitation accumulated during the preceding 5 min in units of 0.01 mm. Thus, the model predicts radar-derived surface precipitation estimates rather than raw reflectivity volumes.

Images were screened for invalid regions, center-cropped to 288 × 288 pixels, and normalized for model training. Each sample contains 12 input fields (60 min of history) and six target fields (30 min of forecast) at

5-min intervals. Data from 2016–2018 were used for training (314,940 images), and 2019 was held out for testing (105,003 images). The temporal split prevents individual storm sequences from appearing in both training and test data.

Transferability was evaluated with a Shanghai radar dataset previously used by Wang et al. (2025). The domain covers 501 km × 501 km around Shanghai and contains more localized and intense short-duration precipitation than the KNMI dataset. The fields were resampled to 288 × 288 pixels and divided into 4,795 training, 792 validation, and 1,048 test sequences. Because spatial resolution, temporal resolution, radar characteristics, and precipitation climatology differ between the two datasets, this experiment evaluates cross-dataset reproducibility after dataset-specific training rather than zero-shot transferability.

The KNMI data are publicly available through the KNMI open-data platform under the stated data license. Access to the Shanghai dataset is governed by the Shanghai Central Meteorological Observatory and the conditions described in the original dataset publication.

### *4.2. Evaluation Metrics and Threshold Definitions*

Forecast verification was performed at rain-rate thresholds r of 0.5, 2, 5 and 10 mm/h, where r denotes radar-derived surface precipitation rate. At a selected threshold, each grid cell is classified as an event when r is greater than or equal to the threshold. TP is the number of hits (forecast event and observed event), FN the number of misses (no forecast event but an observed event), FP the number of false alarms (forecast event but no observed event), and TN the number of correct negatives (neither forecast nor observed event). All four quantities are non-negative counts. The critical success index (CSI), probability of detection (POD), precision, and Heidke skill score (HSS) are dimensionless. CSI measures event-area overlap, POD measures the detected

fraction of observed events, precision measures the observed fraction of forecast events, and HSS measures categorical skill after removing agreement expected by chance.

Using the above contingency-table counts, the categorical metrics are defined below. Larger CSI, POD, precision, and HSS generally indicate better performance, although POD and precision should be interpreted together because increased forecast area can raise POD while reducing precision.

CSI is the ratio of hits to the union of forecast and observed events:

$$\mathrm{CSI}=\frac{\mathrm{TP}}{\mathrm{TP+FP+FN}} \quad (26)$$

HSS compares the number of correct categorical forecasts with the number expected from random agreement:

$$\mathrm{HSS}=\frac{2(\mathrm{TP}\cdot\mathrm{TN}-\mathrm{FP}\cdot\mathrm{FN})}{(\mathrm{TP+FN})(\mathrm{FN+TN})+(\mathrm{TP+FP})(\mathrm{FP+TN})} \quad (27)$$

POD is the fraction of observed events that are successfully detected:

$$\mathrm{POD}=\frac{\mathrm{TP}}{\mathrm{TP+FN}} \quad (28)$$

Precision is the fraction of forecast events that are observed:

$$\mathrm{Precision}=\frac{\mathrm{TP}}{\mathrm{TP+FP}} \quad (29)$$

The structural similarity index measure (SSIM) is a dimensionless image-comparison metric that evaluates agreement in mean intensity, contrast, and local spatial structure. Let $\hat{Y}$ denote the continuous predicted precipitation field and Y the corresponding observed field. Let $\mu_{\hat{Y}}$ and $\mu_Y$ be their mean values, $\sigma_{\hat{Y}}^2$ and $\sigma_Y^2$ their variances, and $\sigma_{\hat{Y}Y}$ their covariance. C1 and C2 are small positive stabilization constants with units chosen to match the squared field values. SSIM is calculated as:

$$\mathrm{SSIM}(\hat{Y},Y)=\frac{(2\mu_{\hat{Y}}\mu_Y+C_1)(2\sigma_{\hat{Y}Y}+C_2)}{(\mu_{\hat{Y}}^2+\mu_Y^2+C_1)(\sigma_{\hat{Y}}^2+\sigma_Y^2+C_2)} \quad (30)$$

SSIM is typically bounded above by 1, with 1 indicating identical fields. In this study it complements the threshold-based categorical metrics by assessing the preservation of precipitation morphology and intensity organization.

*4.3. Forecast Skill across Intensity, Lead Time, and Region*

Tables 1–4 summarize the mean categorical scores on the KNMI test set. The results at 0.5, 2, and 5 mm/h use the standard dBZ-weighting configuration, whereas the r ≥ 10 mm/h results are obtained from a separately trained configuration with stronger weighting of high-intensity echoes, as detailed in Supplementary Table S2.

**Table 1**

Average values of different evaluation metrics for different models on the KNMI dataset averaged over the six 5-min lead times for r ≥ 0.5 mm/h. CSI, POD, HSS, and precision are dimensionless; higher values are better.

| Algorithm | CSI↑ | POD↑ | HSS↑ | Precision↑ |
|---|---|---|---|---|
| Precip-UNetGAN | **0.721** | 0.851 | **0.720** | **0.840** |
| SmatUnet | 0.687 | 0.869 | 0.695 | 0.764 |
| Predrnn | 0.605 | 0.692 | 0.629 | 0.822 |
| ConvLstm | 0.617 | **0.921** | 0.617 | 0.650 |
| Unet | 0.637 | 0.868 | 0.636 | 0.753 |

**Table 2**

Average values of different evaluation metrics for different models on the KNMI dataset averaged over the six

5-min lead times for r ≥ 2 mm/h. CSI, POD, HSS, and precision are dimensionless; higher values are better.

| Algorithm | CSI↑ | POD↑ | HSS↑ | Precision↑ |
|---|---|---|---|---|
| Precip-UNetGAN | **0. 515** | **0. 701** | **0. 633** | **0. 780** |
| SmatUnet | 0. 481 | 0. 592 | 0. 605 | 0. 708 |
| Predrnn | 0. 413 | 0. 508 | 0. 534 | 0. 675 |
| ConvLstm | 0. 453 | 0. 601 | 0. 569 | 0. 632 |
| Unet | 0. 419 | 0. 635 | 0. 541 | 0. 674 |

**Table 3**

Average values of different evaluation metrics for different models on the KNMI dataset averaged over the six 5-min lead times for r ≥ 5 mm/h. CSI, POD, HSS, and precision are dimensionless; higher values are better.

| Algorithm | CSI↑ | POD↑ | HSS↑ | Precision↑ |
|---|---|---|---|---|
| Precip-UNetGAN | **0. 296** | **0. 575** | **0. 441** | **0. 655** |
| SmatUnet | 0. 275 | 0. 327 | 0. 413 | 0. 619 |
| Predrnn | 0. 266 | 0. 339 | 0. 401 | 0. 544 |
| ConvLstm | 0. 246 | 0. 293 | 0. 371 | 0. 597 |
| Unet | 0. 266 | 0. 381 | 0. 404 | 0. 598 |

**Table 4**

Average values of different evaluation metrics for different models on the KNMI dataset averaged over the six

5-min lead times for r ≥ 10 mm/h. CSI, POD, HSS, and precision are dimensionless; higher values are better.

| Algorithm | CSI↑ | POD↑ | HSS↑ | Precision↑ |
|---|---|---|---|---|
| Precip-UNetGAN | **0. 162** | **0. 219** | **0. 268** | 0. 461 |
| SmatUnet | 0. 157 | 0. 195 | 0. 260 | 0. 427 |
| Predrnn | 0. 112 | 0. 142 | 0. 187 | 0. 383 |
| ConvLstm | 0. 088 | 0. 102 | 0. 147 | **0. 469** |
| Unet | 0. 140 | 0. 209 | 0. 237 | 0. 420 |

**Table 5**

Proportions of precipitation samples under different thresholds

| r >=0.5 mm/h | r >= 2 mm/h | r >=5mm/h | r >=10 mm/h |
|---|---|---|---|
| 36.24% | 10.62% | 1.84% | 0.27% |

Tables 1–4 present the average CSI, POD, HSS, and Precision values of Precip-UNetGAN and the comparison models under four precipitation thresholds. Overall, Precip-UNetGAN achieves the best or near-best comprehensive performance across different threshold conditions, indicating that the proposed method has strong adaptability under precipitation scenarios of varying intensities. To further provide an intuitive illustration of the performance variation of different models as the forecast lead time increases, the changing curves of CSI, POD, HSS, Precision, and SSIM over six time steps are also presented in Figures 5–9.

Table 5 represents fraction of KNMI pixels meeting or exceeding each rain-rate threshold. The strong decrease with threshold quantifies the class imbalance faced during training.

Across the KNMI thresholds, Precip-UNetGAN produced the highest CSI and HSS and maintained a more balanced relationship between POD and precision than the comparison models. At 0.5 mm/h, ConvLSTM

obtained a higher POD but substantially lower precision, indicating that its greater detection rate was accompanied by more false alarms. The proposed model therefore improved event overlap not by maximizing detection alone, but by retaining precipitation objects while limiting spurious expansion. The advantage persisted at 2 and 5 mm/h, where compact high-intensity regions are increasingly sparse and conventional pixel-wise training is more prone to attenuation.

At 10 mm/h, all models showed a marked reduction in CSI and HSS. This decline is consistent with both sample rarity and the limited predictability of rapidly evolving convective cores. Precip-UNetGAN remained best in CSI, HSS, and POD, but the absolute scores indicate that the model should not be interpreted as resolving convective initiation or decay reliably. Rather, the principal benefit is improved short-term persistence of already observed intense structures. This distinction is important for operational use: the method can support warning continuity, but it does not replace process-aware diagnosis or numerical guidance for newly developing convection.

The lead-time curves in Figs. 5–9 show that skill decreases for every model as forecast uncertainty accumulates. Precip-UNetGAN generally degrades more slowly, especially at the 0.5–5 mm/h thresholds. The simultaneous retention of CSI, precision, and SSIM suggests that the improvement is associated with better preservation of object location and boundary structure rather than indiscriminate enlargement of rainy areas. At the highest threshold, model differences narrow by the final forecast times, indicating that loss reweighting cannot fully compensate for uncertainty in convective growth and decay.

**Table 6**

Average values of different evaluation metrics for different models on the Shanghai dataset averaged over the six 5-min lead times for r ≥ 0.5mm/h. CSI, POD, HSS, and precision are dimensionless; higher values are better.

| Algorithm | CSI↑ | POD↑ | HSS↑ | Precision↑ |
|---|---|---|---|---|
| Precip-UNetGAN | **0.738** | **0.905** | **0.403** | **0.842** |
| SmatUnet | 0.708 | 0.872 | 0.390 | 0.801 |
| Predrnn | 0.722 | 0.889 | 0.396 | 0.828 |
| ConvLstm | 0.723 | 0.893 | 0.397 | 0.812 |
| Unet | 0.731 | 0.901 | 0.401 | 0.835 |

**Table 7**

Average values of different evaluation metrics for different models on the Shanghai dataset averaged over the six 5-min lead times for r ≥ 2 mm/h. CSI, POD, HSS, and precision are dimensionless; higher values are better.

| Algorithm | CSI↑ | POD↑ | HSS↑ | Precision↑ |
|---|---|---|---|---|
| Precip-UNetGAN | **0.652** | **0.879** | **0.378** | **0.793** |
| SmatUnet | 0.623 | 0.814 | 0.365 | 0.751 |
| Predrnn | 0.641 | 0.855 | 0.372 | 0.774 |
| ConvLstm | 0.640 | 0.853 | 0.371 | 0.768 |
| Unet | 0.643 | 0.858 | 0.374 | 0.768 |

**Table 8**

Average values of different evaluation metrics for different models on the Shanghai dataset averaged over the six 5-min lead times for r ≥ 5 mm/h. CSI, POD, HSS, and precision are dimensionless; higher values are better.

| Algorithm | CSI↑ | POD↑ | HSS↑ | Precision↑ |
|---|---|---|---|---|
| Precip-UNetGAN | **0. 497** | 0. 575 | **0. 443** | **0. 662** |
| SmatUnet | 0. 468 | 0. 540 | 0. 298 | 0. 608 |
| Predrnn | 0. 485 | **0. 591** | 0. 308 | 0. 626 |
| ConvLstm | 0. 472 | 0. 554 | 0. 301 | 0. 618 |
| Unet | 0. 484 | 0. 581 | 0. 304 | 0. 632 |

**Table 9**

Average values of different evaluation metrics for different models on the Shanghai dataset averaged over the six 5-min lead times for r ≥ 10 mm/h. CSI, POD, HSS, and precision are dimensionless; higher values are better.

| Algorithm | CSI↑ | POD↑ | HSS↑ | Precision↑ |
|---|---|---|---|---|
| Precip-UNetGAN | **0. 292** | **0. 576** | **0. 207** | **0. 468** |
| SmatUnet | 0. 283 | 0. 540 | 0. 201 | 0. 441 |
| Predrnn | 0. 288 | 0. 564 | 0. 205 | 0. 452 |
| ConvLstm | 0. 276 | 0. 551 | 0. 198 | 0. 446 |
| Unet | 0. 277 | 0. 553 | 0. 195 | 0. 454 |

On the Shanghai dataset (Tables 6–9), Precip-UNetGAN again achieved the highest CSI at all thresholds and retained favorable POD and precision. The consistency of the ranking across a maritime mid-latitude radar domain and a more convectively active East Asian domain suggests that the combination of multi-scale attention and metric-oriented optimization is robust to different precipitation distributions.However, because the model

was trained separately on the Shanghai dataset, these experiments demonstrate cross-dataset reproducibility after dataset-specific training rather than direct portability. Operational transfer to a new radar network would still require quality control, unit harmonization, threshold calibration, and local validation.

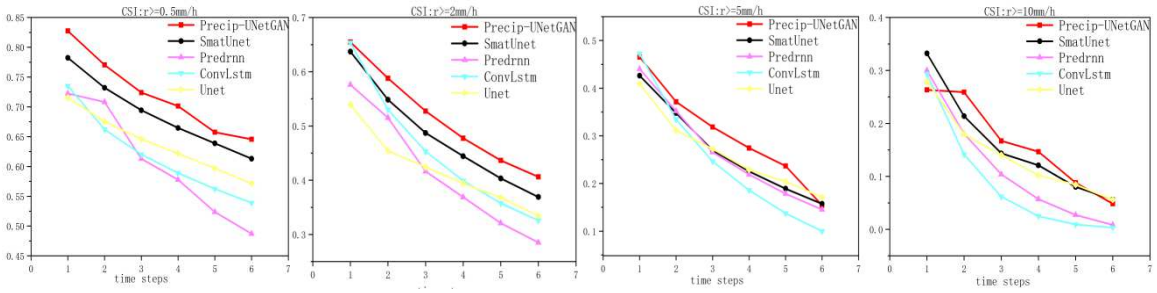



**Fig. 5.** Variation of CSI over six time steps under different precipitation thresholds.

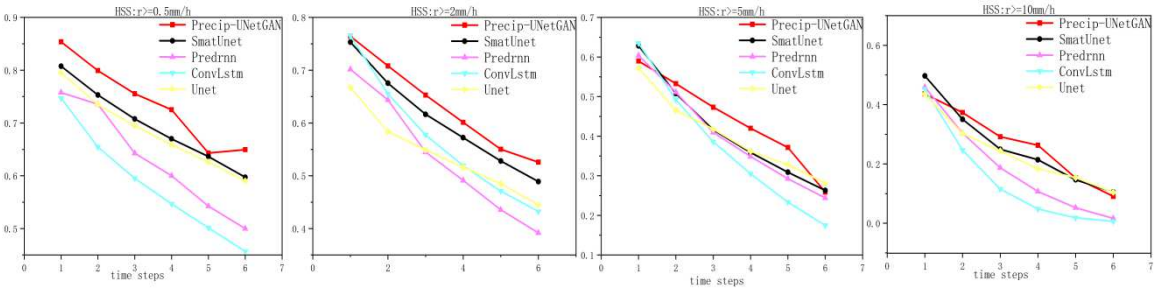



**Fig. 6.** Variation of HSS over six time steps under different precipitation thresholds.

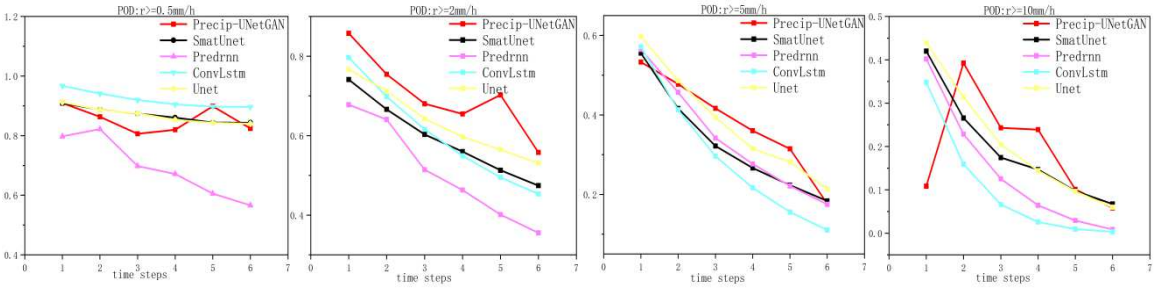



**Fig. 7.** Variation of POD over six time steps under different precipitation thresholds.

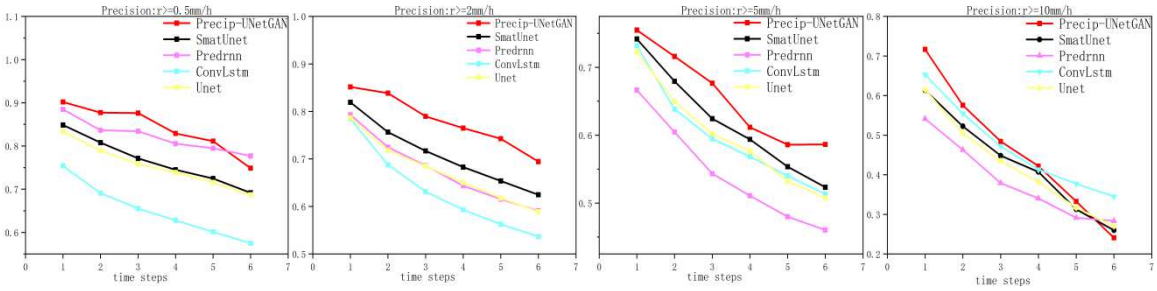



**Fig. 8.** Variation of Precision over six time steps under different precipitation thresholds.

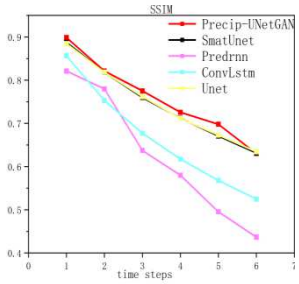



**Fig. 9.** Variation of SSIM over six time steps under different precipitation thresholds.

Figures 5–8 clarify the error trade-offs behind the mean scores. The proposed model does not uniformly maximize POD; instead, it generally improves CSI and HSS while keeping precision high. This behavior is meteorologically useful because excessive areal growth of predicted echoes can inflate detection but produce false alarms over adjacent catchments or urban areas. The metric-oriented objective therefore appears to improve the spatial agreement of threshold exceedances rather than simply increasing forecast intensity everywhere.

The comparison also reveals different failure modes. ConvLSTM tends toward high detection at low thresholds but weaker precision, consistent with broader predicted rain areas. UNet and PredRNN lose skill more rapidly for moderate and heavy precipitation, consistent with smoothing or attenuation of compact cores. SmaAt-UNet remains competitive for weak precipitation but shows less benefit as the threshold increases. These patterns support the interpretation that multi-scale attention is most useful when intense echoes occupy a small area within a larger weak-precipitation background.

SSIM (Fig. 9) provides complementary evidence: higher structural similarity indicates that improvements in categorical scores are accompanied by better preservation of precipitation morphology. Nevertheless, SSIM does not measure displacement error or hydrological impact directly, and visually sharper GAN output is not necessarily more quantitatively calibrated. The categorical and structural scores should therefore be considered

jointly.Taken together, the results indicate that Precip-UNetGAN primarily reduces three errors over the 0–30 min window: attenuation of moderate-to-heavy echo cores, loss of object boundaries, and mismatch between detected and observed threshold-exceedance areas. The experiments do not establish improved representation of unobserved convective initiation, and they do not demonstrate dynamical causality. These limits motivate the discussion in Sect. 5.

### *4.4. Ablation Study*

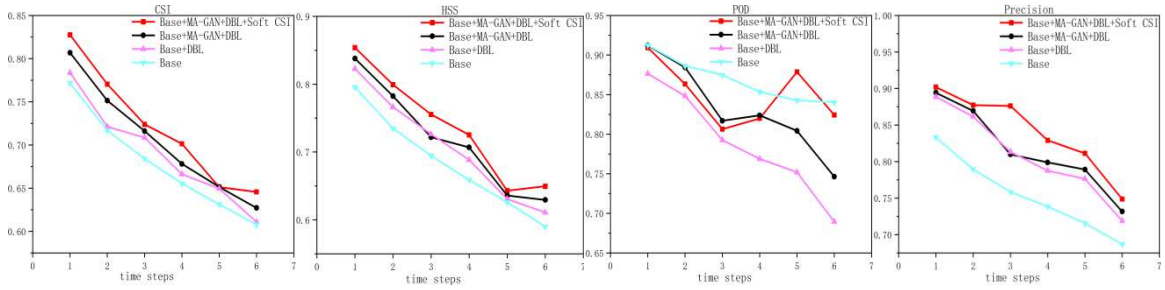



**Fig. 10.** The change process of CSI, HSS, POD, and Precision over 6 time steps.

The ablation experiment isolates the contribution of the model components at $r \geq 0.5$ mm/h (Fig. 10). Adding dBZ-weighted reconstruction produces the largest broad improvement, consistent with the severe pixel imbalance documented in Table 5. Soft-CSI further improves event-overlap measures by penalizing misses and false alarms in a form aligned with operational verification. Multi-scale attention and adversarial learning contribute mainly to structural retention and boundary clarity. The fact that the highest POD is not always obtained by the full model is not a deficiency by itself: POD can be increased by predicting excessively large rain areas. The higher CSI and precision of the complete model indicate a more favorable balance between missed events and false alarms. A limitation of the current ablation is that it is reported for one threshold; threshold-specific ablations would strengthen attribution of the heavy-rain improvements.

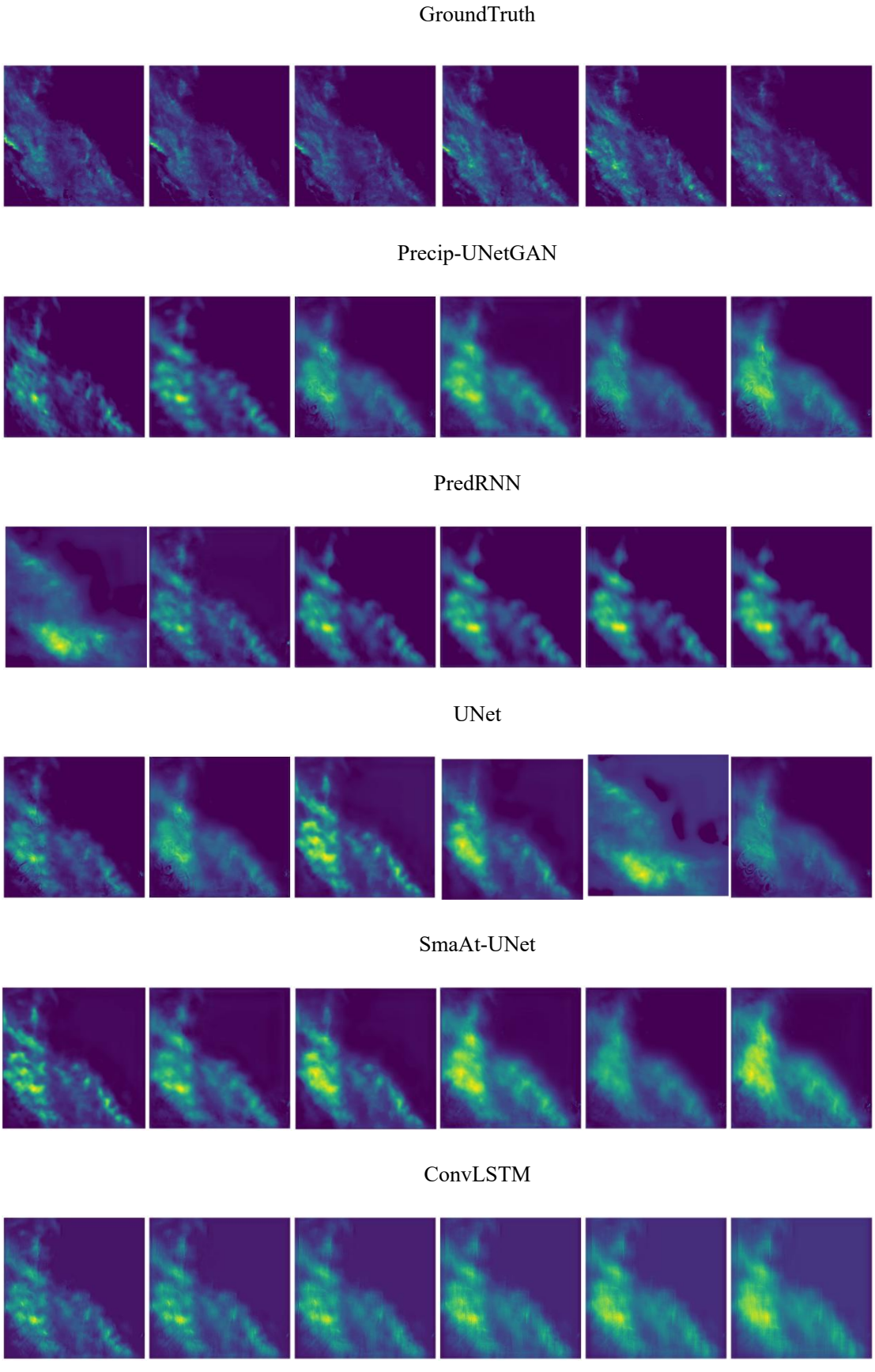



**Fig. 11.** Comparison between real images and the prediction results of five models

### *4.5. Case-Based Interpretation of Spatial Errors*

Figure 11 compares a representative sequence in terms of four physically relevant forecast attributes: displacement of the precipitation area, retention of peak intensity, continuity of echo boundaries, and generation of spurious cells. Precip-UNetGAN retains a more coherent precipitation object and sharper gradients than the deterministic baselines. PredRNN and UNet show stronger smoothing at the object edge and greater attenuation of local maxima. SmaAt-UNet retains weak-precipitation texture but introduces more small-scale noise, while ConvLSTM tends to broaden the rainy area. These visual differences are consistent with the POD–precision trade-offs in the quantitative results. Because one case cannot establish general behavior, the figure is used only to illustrate the error mechanisms suggested by the aggregate scores, not as independent evidence of superiority.

## 5. Discussion

The experiments show that the largest practical gain of Precip-UNetGAN occurs when precipitation occupies a small fraction of the domain and forecast quality depends on retaining compact structures. This behavior is consistent with the scale separation of precipitation fields: broad weak-rain regions are comparatively easy to reconstruct, whereas embedded convective cores and sharp gradients are vulnerable to averaging by pixel-wise losses. Multi-scale attention increases the influence of these sparse structures, while dBZ weighting counteracts their low frequency in the training distribution. Soft-CSI then aligns the optimization with the areal overlap of threshold exceedances. The combined effect explains why CSI and precision improve together in many experiments, rather than through an indiscriminate increase in POD.

The atmospheric interpretation should nevertheless remain limited. The model learns statistical evolution from radar histories and does not explicitly represent moisture convergence, instability, cold-pool dynamics, or microphysical processes. Improved morphology therefore indicates better reproduction of observed echo

evolution, not discovery of a new physical mechanism. The rapid decline at 10 mm/h and at later lead times also shows that the method remains constrained by the intrinsic unpredictability of convective growth and decay. Incorporating environmental fields from numerical weather prediction, satellite observations, lightning, or radar-derived motion and growth diagnostics may help distinguish advection from genuine development.

Transferability has two components.The similar ranking on the KNMI and Shanghai datasets supports methodological reproducibility across contrasting precipitation regimes after dataset-specific training. Direct operational transfer, however, is affected by radar wavelength, calibration, clutter filtering, spatial and temporal resolution, Z–R conversion, terrain blockage, and local rain-rate climatology. The present results therefore justify retraining and local calibration rather than universal deployment without adaptation. Future work should include leave-one-region-out tests, zero-shot or few-shot adaptation, reliability analysis, object-based verification, and evaluation on independent severe-weather seasons.

For operational nowcasting, the model's balanced CSI, POD, and precision is more informative than visual sharpness alone. A useful deployment would provide deterministic fields together with uncertainty estimates and quality flags, particularly because adversarial models may generate plausible-looking but poorly calibrated details. Computational cost and training stability must also be compared with the latency constraints of warning systems. These considerations define the main path from the present methodological study toward a robust meteorological application.

## 6. Conclusions

This study evaluated whether scale-aware representation and metric-oriented optimization can improve 0–30 min radar precipitation nowcasting. Three conclusions emerge. First, Precip-UNetGAN consistently improved CSI and HSS across 0.5–10 mm/h thresholds on the KNMI dataset, with the clearest relative gains for moderate-to-heavy precipitation. Second, the joint behavior of CSI, precision, POD, and SSIM indicates that the

improvement is associated with better retention of compact precipitation structures and boundaries, rather than simple expansion of predicted rain areas. Third, the same overall ranking was reproduced on the Shanghai dataset after dataset-specific training, supporting cross-dataset reproducibility while also showing that local calibration remains necessary.

The principal take-home message is that precipitation nowcasting benefits when the training objective reflects both the multi-scale organization of radar echoes and the categorical thresholds used in warning practice. The model is most appropriately viewed as an improved short-term extrapolation tool for observed precipitation systems, not as a solution to convective initiation or a physically explicit atmospheric model. Future development should extend evaluation beyond 30 min, quantify uncertainty and calibration, use object-based and hydrological verification, test direct transfer across radar networks, and incorporate environmental predictors that describe precipitation growth and decay. These steps are necessary before the method can be considered for real-time operational deployment.

*Disclosures*

The authors declare that there are no financial interests, commercial affiliations, or other potential conflicts of interest that could have influenced the objectivity of this research or the writing of this paper.

*Code Availability*

The source code generated during this study is available from the corresponding author upon reasonable request via email at 2710766198@qq.com.

549 Germann U, Zawadzki I. “Scale-dependence of the predictability of precipitation from
550 continental radar images. Part I: Description of the methodology,” Mon. Weather Rev.,
551 130(12), 2859–2873, (2002).

552 Sakaino H. “Spatio-temporal image pattern prediction method based on a physical model
553 with time-varying optical flow,” IEEE Trans. Geosci. Remote Sens., 51(5), 3023–3036,
554 (2013).

555 Sun J, Xue M, Wilson JW, Zawadzki I, Ballard SP, Onvlee-Hooimeyer J, Joe P, Barker DM,
556 Li PW, Golding B, Xu M, Pinto J. “Use of NWP for nowcasting convective precipitation:
557 Recent progress and challenges,” Bull. Am. Meteorol. Soc., 95(3), 409–426, (2014).

558 Woo WC, Wong WK. “Application of optical flow techniques to rainfall nowcasting,” in
559 27th Conf. Severe Local Storms, (2014).

560 Shi X, Chen Z, Wang H, Yeung DY, Wong WK, Woo WC. “Convolutional LSTM network:
561 A machine learning approach for precipitation nowcasting,” in Adv. Neural Inf. Process.
562 Syst., Cambridge, MA: MIT Press, (2015), pp. 802–810.

563 Wang Y, Wu H, Zhang J, Gao Z, Wang J, Yu PS, Long M. “PredRNN: A recurrent neural
564 network for spatiotemporal predictive learning,” IEEE Trans. Pattern Anal. Mach. Intell.,
565 45(2), 2208–2225, (2023).

566 Ronneberger O, Fischer P, Brox T. “U-Net: Convolutional networks for biomedical image
567 segmentation,” in Int. Conf. Med. Image Comput. Comput.-Assist. Interv., Cham:
568 Springer,(2015), pp. 234–241.

569 Ravuri S, Lenc K, Willson M, Kang S, Lam R, Mirowski P, Mohamed S. “Skillful
570 precipitation nowcasting using deep generative models of radar,” Nature, 597(7878), 672–
571 677, (2021).

572 Shi X, Gao Z, Lausen L, Wang H, Yeung DY, Wong WK, Woo WC. “Deep learning for
573 precipitation nowcasting: A benchmark and a new model,” in Adv. Neural Inf. Process. Syst.,
574 (2017), pp. 5617–5629.

575 Reyniers M. Quantitative Precipitation Forecasts Based on Radar Observations: Principles,
576 Algorithms and Operational Systems, Institut Royal Météorologique de Belgique, (2008).

577 Shi X, Chen Z, Wang H, Yeung DY, Wong WK, Woo WC. “Convolutional LSTM network:
578 A machine learning approach for precipitation nowcasting,” in Adv. Neural Inf. Process.
579 Syst., Cambridge, MA: MIT Press, (2015), pp. 802–810.

580 Wang Y, Long M, Wang J, Gao Z, Yu PS. “Memory in memory: A predictive neural network
581 for learning higher-order non-stationarity from spatiotemporal dynamics,” in Proc. IEEE
582 Conf. Comput. Vis. Pattern Recognit., Long Beach, CA: IEEE, (2019), pp. 9154–9162.

583 Srivastava N, Mansimov E, Salakhudinov R. “Unsupervised learning of video representations
584 using LSTMs,” in Int. Conf. Mach. Learn.,(2015), pp. 843–852.

585 Mathieu M, Couprie C, LeCun Y. “Deep multi-scale video prediction beyond mean square
586 error,” in Int. Conf. Learn. Represent., (2016).

587 Finn C, Goodfellow I, Levine S. “Unsupervised learning for physical interaction through
588 video prediction,” in Adv. Neural Inf. Process. Syst., (2016), pp. 64–72.

589 Lotter W, Kreiman G, Cox D. “Deep predictive coding networks for video prediction and
590 unsupervised learning,” in Int. Conf. Learn. Represent., (2017).

591 Liu C, Wang H, Quan W, Si J, Yan Y, Chen Z, Lu Y. “A precipitation nowcasting method
592 using radar echo data based on attentive ConvLSTM and three-channel UNet,” Signal, Image
593 Video Process., 19, 985, (2025).

594 Wang Y, Yao L, Jiang H, Liu T, Lu Y, Zhou C. “Precipitation nowcasting based on radar echo
595 images via multi-scale spatiotemporal LSTM,” IEEE Trans. Geosci. Remote Sens., (2025).
596 DOI: 10.1109/TGRS.2025.3584824.

597 Cheng Q, Su Y, Yang H, Wu Y, Liu F, Rao Y, Chao Y, Wang K, Liu Z, Liu J, Yao C.
598 “Enhanced radar echo extrapolation for precipitation nowcasting quality using the
599 convolutional Kolmogorov–Arnold networks,” J. Hydrol., 663(Part A), 134134, (2025).

600 Wu D, Zhang S, Zhao G, Feng Y, Ma Y, Zhang Y, et al. “Short-term rainfall prediction based
601 on radar echo using an efficient spatio-temporal recurrent unit,” Sci. Rep., 15, 28620, (2025).

602 Agrawal S, Barrington L, Bromberg C, Burge J, Gazen C, Hickey J. “Machine learning for
603 precipitation nowcasting from radar images,” arXiv preprint arXiv:1912.12132, (2019).

604 Ayzel G, Scheffer T, Heistermann M. “RainNet v1.0: All convolutional neural network for
605 radar-based precipitation nowcasting,” Geosci. Model Dev., 13(6), 2631–2644, (2020).

606 Han L, Liang H, Chen H, Zhang W, Ge Y. “Convective precipitation nowcasting using U-Net
607 model,” IEEE Trans. Geosci. Remote Sens., 60, 1–8, (2021).

608 Fernández JG, Mehrkanoon S. “Broad-UNet: Multi-scale feature learning for nowcasting

609 tasks," Neural Networks, 144, 419–427, (2021).

610 Trebing K, Stanczyk T, Mehrkanoon S. "SmaAt-UNet: Precipitation nowcasting using a
611 small attention-UNet architecture," Pattern Recognit. Lett., 145, 178–186, (2021).

612 Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A,
613 Bengio Y. "Generative adversarial nets," in Adv. Neural Inf. Process. Syst., Cambridge, MA:
614 MIT Press, (2014), pp. 2672–2680.

615 Isola P, Zhu JY, Zhou T, Efros AA. "Image-to-image translation with conditional adversarial
616 networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Honolulu, HI: IEEE, (2017),
617 pp. 1125–1134.

618 Wang R, Su L, Wong WK, Lau AKH, Fung JCH. "Skillful radar-based heavy rainfall
619 nowcasting using task-segmented generative adversarial network," IEEE Trans. Geosci.
620 Remote Sens., 61, 1–13, (2023).

621 Bai C, Sun F, Zhang J, Song Y, Chen S. "Rainformer: Features extraction balanced network
622 for radar-based precipitation nowcasting," IEEE Geosci. Remote Sens. Lett., 19, 1–5, (2022).

623 Yang Y, Mehrkanoon S. "AA-TransUNet: Attention augmented TransUNet for nowcasting
624 tasks," in Int. Joint Conf. Neural Networks, (2022), pp. 1–8.

625 Zhang Y, Long M, Chen K, Xing L, Jin R, Jordan MI, Wang J. "Skilful nowcasting of
626 extreme precipitation with NowcastNet," Nature, 619(7970), 526–532, (2023).

627 Xu L, Lu W, Yu H, Yao F, Sun X, Fu K. "SFTFormer: A spatial-frequency-temporal
628 correlation-decoupling transformer for radar echo extrapolation," IEEE Trans. Geosci.

Remote Sens., 62, 1–15, (2024).

Zhao Z, Dong X, Wang Y, Wang J, Chen Y, Hu C. “MDTNet: Multiscale deformable transformer network with Fourier space losses toward fine-scale spatiotemporal precipitation nowcasting,” IEEE Trans. Geosci. Remote Sens., 62, 1–17, (2024).

Wang L, Wang Z, Hu W, Bai C. “RainHCNet: Hybrid High-Low Frequency and Cross-Scale Network for Precipitation Nowcasting,” IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens., (2025).

Zhang J, Zheng Y, Qi D. “Deep spatio-temporal residual networks for citywide crowd flows prediction,” in AAAI Conf. Artif. Intell., (2017), pp. 1655–1661.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementary.pdf