

# **ThoR: A Motion-Dependent Physics-Informed Deep Learning Framework with Constraint-Centric Theory of Functional Connections for Rainfall Nowcasting**

**Khang Ta Gia<sup>1,2,\*</sup>, Hoai Tran Van<sup>1,+</sup>, An Phan Thanh<sup>2</sup>, and Nhat Phan Minh<sup>3</sup>**

<sup>1</sup>Faculty of Computer Science and Engineering, Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam

<sup>2</sup>Institute of Mathematical and Computational Sciences, Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam

<sup>3</sup>Faculty of Information Technology, University of Danang - University of Science and Technology, Da Nang, Vietnam

<sup>+</sup>Corresponding Author

\*tagiakhang19@gmail.com

## **ABSTRACT**

Accurate precipitation nowcasting is crucial for mitigating the impacts of extreme weather, especially as climate change increases their frequency and severity. Traditional methods, such as numerical weather prediction and radar extrapolation, face limitations in short-term and high-resolution forecasting. Recently, while deep learning approaches have advanced nowcasting by learning spatiotemporal patterns from radar data, they often suffer from blurry results due to uncertainty predictions and limited physical consistency. To deal with these challenges, we propose ThoR, a Motion-Dependent Physics-Informed Deep Learning Framework with Constraint-Centric Theory of Functional Connections for Rainfall Nowcasting. ThoR integrates attention-centric spatio-temporal modeling with explicit physical constraints derived from partial differential equations (PDEs) for forward simulation, which employs a cascaded-branch architecture that integrates an attention-driven generator with an unsupervised, lead-time-conditioned module for motion field extraction. Physical consistency is enforced by weighted embedding the advection–diffusion equation directly into the optimization objective, establishing a Theory of Functional Connections (TFC) framework tailored for precipitation nowcasting. Extensive experiments on real-world radar datasets demonstrate that ThoR achieves promised performance compared to existing methods across both deterministic and probabilistic metrics, particularly at longer lead times and during extreme events, highlighting the potential of physics-informed deep learning for operational nowcasting.

## **1 Introduction**

Weather forecasting, the prediction of future atmospheric conditions such as precipitation, temperature, pressure, and wind, plays a crucial role in both science and society<sup>1</sup>. In Vietnam, natural disasters in 2024 resulted in 514 fatalities and economic losses of about 88.748 trillion VND<sup>2</sup>. As a result, accurate weather forecasts have far-reaching social and economic benefits, improving daily decision-making and contributing significantly to numerous sectors, such as agriculture, energy, transportation, and disaster prevention<sup>3,4</sup>. Quantitative precipitation nowcasting (QPN) is considered to be one of the most challenging aspects of this field.

Precipitation nowcasting techniques primarily fall into two categories: (1) numerical weather prediction (NWP) and (2) radar echo extrapolation. NWP models simulate atmospheric dynamics based on physical laws; however, they often struggle to produce accurate short-term forecasts due to initialization errors and computational latency<sup>5,6</sup>. In contrast, radar-based approaches, such as optical flow estimation and centroid tracking<sup>7,8</sup>, offer higher temporal resolution but typically assume conserved reflectivity and smooth motion, assumptions that break down during rapidly evolving convective systems<sup>9,10</sup>. Both paradigms exhibit notable limitations in capturing the localized and nonlinear characteristics of precipitation, thereby motivating the exploration of data-driven deep learning (DL) methodologies<sup>11,12</sup>.

Deep learning has demonstrated considerable potential in modeling complex spatiotemporal patterns inherent in radar reflectivity sequences, circumventing the need for handcrafted heuristics. These models generally fall into two categories: non-adversarial and adversarial approaches. Non-adversarial models, such as ConvLSTM<sup>11</sup>, TrajGRU<sup>13</sup>, and the PredRNN family<sup>12,14</sup>, focus on capturing spatiotemporal dependencies using convolutional recurrent structures. Innovations such as hierarchical memory cells, attention mechanisms, and motion decomposition have improved their forecasting capabilities. However, these models often suffer from blurred outputs in long-range predictions, largely due to their reliance on global loss

functions like MSE, which introduce the "double penalty" problem<sup>15, 16</sup> and fail to preserve high-frequency details<sup>17</sup>. In contrast, adversarial models employ GAN frameworks to generate sharper, more realistic precipitation forecasts. DGMR<sup>18</sup>, for example, utilizes ConvGRU-based generators with stochastic sampling to better model the uncertainty in radar fields. Despite improved visual fidelity, adversarial models face challenges such as mode collapse and difficulty in modeling long-term dependencies. Recent advancements, including self-clustering (ClusterCast<sup>19</sup>) and transformer-based hybrids (ConvLSTM-TransGAN<sup>20</sup>, SAC-LSTM<sup>21</sup>), have sought to address these limitations by enhancing diversity and temporal coherence. To overcome the limitations of conventional loss functions in precipitation nowcasting, recent studies have proposed a range of specialized objective formulations. Weighted losses improve sensitivity to high-intensity but infrequent rainfall by amplifying penalties on severe events<sup>13, 18</sup>, while pooling-based losses enhance spatial structure recognition by focusing on large-scale precipitation patterns and reducing noise sensitivity<sup>22</sup>. Motion-regularized losses introduce gradient-based smoothness constraints to enforce realistic spatiotemporal dynamics<sup>23, 24</sup>. Probabilistic learning frameworks further balance performance across rainfall regimes by approximating classification metrics in a differentiable form<sup>25</sup>.

In addition, a persistent limitation in current DL-based nowcasting systems is their limited integration of physical knowledge. Many models lack physical interpretability and may violate fundamental atmospheric principles, particularly under extreme or out-of-distribution scenarios. To address this, recent research has begun to incorporate physics-informed learning by embedding constraints such as mass conservation and advection dynamics directly into model architectures and loss functions<sup>26–28</sup>. The Extreme Theory of Functional Connections (TFC)<sup>29, 30</sup> offers a promising mathematical framework for representing functions that adhere to such constraints, thereby providing a principled mechanism to incorporate physical laws into deep learning models for precipitation forecasting.

Building on recent advances, we introduce ThoR, a novel physics-informed deep learning framework grounded in the TFC, where spatiotemporal modeling is integrated with explicit physical constraints for precipitation nowcasting. ThoR employs a cascaded-branch architecture, comprising an attention-centric state space generator that captures complex dependencies in both spatial and temporal dimensions, alongside a module that predicts paired advection motion fields to represent dynamic atmospheric flows. These motion fields are utilized in a hybrid manner—both embedded within the generator for precipitation map forecasting and incorporated into the optimization objective via the advection–diffusion equation, thereby reinforcing physical consistency and enhancing predictive accuracy. Through the direct embedding of domain knowledge into the learning process, a significant advancement in nowcasting capabilities is achieved, offering forecasts that are more reliable, interpretable, and physically grounded, particularly under highly dynamic and extreme weather conditions.

To sum up, the principal contributions of this work are as follows:

- We enhance the Theory of Functional Connections by embedding the advection–diffusion equation and associated physical constraints directly into the optimization objectives of each component of our framework. This approach, which we called "*Soft TFC with constraint centric*", thereby enhances interpretability and improves predictive performance in fitting multidimensional PDEs of precipitation nowcasting.
- We propose ThoR, a hybrid physics-informed deep learning framework that integrates attention-based spatiotemporal modeling with explicit physical constraints derived from PDEs. ThoR enables accurate forward simulation of precipitation fields, delivering skillful forecasts across a wide range of rainfall intensities, with notable performance in predicting extreme precipitation events governed by advective and convective dynamics that have historically posed considerable modeling challenges.

## 2 Results

To rigorously assess the performance of ThoR, we conducted empirical evaluations against three state-of-the-art baseline methods: PySTEPS<sup>8</sup>, NowcastNet<sup>31</sup>, and TrajGRU<sup>13</sup>. Experiments were carried out on two geographically and climatologically distinct datasets: the Multi-Radar Multi-Sensor (MRMS) system<sup>32</sup> and radar observations from the Nha Be Weather Radar Station in Ho Chi Minh City<sup>33</sup>. Each dataset was divided into training and testing subsets using a standardized 80/20 split. To ensure fair comparison, we reproduced all baseline models using their official open-source implementations and trained them from scratch under identical training and evaluation protocols. This unified experimental framework enabled a comprehensive and regionally diverse evaluation of ThoR's generalization capabilities and predictive performance.

### 2.1 Problem description

We introduce ThoR, a physics-informed nowcasting framework grounded in conditional generative modeling for precipitation forecasting. The model operates on an input sequence of observed precipitation fields, denoted as  $R_{in} = \{R_1, R_2, \dots, R_t\} \in \mathbb{R}^{t \times h \times w}$ , where  $t$  represents the number of input time steps, and  $h$  and  $w$  correspond to the spatial dimensions (height and width) of each radar-derived precipitation map. The objective is to generate a sequence of future rainfall fields  $R_{out} = \{R_{t+1}, R_{t+2}, \dots, R_{t+n}\} \in \mathbb{R}^{n \times h \times w}$ , where  $n$  denotes the forecast horizon in time steps.

ThoR adopts a recursive generative formulation, predicting each future frame sequentially. At each time step  $i \in \{t+1, \dots, t+n\}$ , the model estimates the conditional distribution of the next frame given all previous frames. Each hidden state encodes the evolving atmospheric dynamics at that time step, conditioned on the inferred motion field. The overall forecasting task is cast as a sequential probabilistic estimation:

$$P(R_{out} | R_{in}) = \prod_{i=t+1}^{t+n} P(R_i | R_1, \dots, R_{i-1}), \quad (1)$$

This autoregressive formulation enables the model to dynamically update its internal representation as new predictions are generated, allowing it to capture temporally non-stationary dependencies that are characteristic of atmospheric processes.

ThoR consists of two main components: (1) an attention-driven spatiotemporal generative network  $G$ , parameterized by  $\theta$ , and (2) a deterministic motion estimation module  $M_\phi$ , parameterized by  $\phi$ , which infers large-scale advective motion fields from the observed precipitation sequence. The forecasting task is framed as a hybrid, physics-aware conditional generation problem, incorporating latent stochastic variables  $\mathbf{z}$  that encode both the extracted advective motion and the fine-grained residual deformations learned by the spatiotemporal encoder. The distribution over future precipitation is given by:

$$P(\hat{R}_{t+1:t+n} | R_{1:t}, \phi, \theta) = \int P(\hat{R}_{t+1:t+n} | R_{1:t}, M_\phi(R_{1:t}), \mathbf{z}; \theta) P(\mathbf{z}) d\mathbf{z}. \quad (2)$$

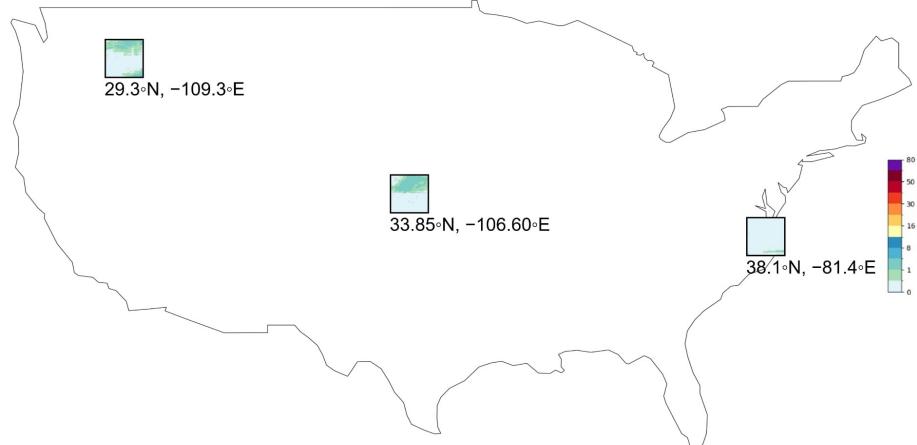
By marginalizing over the latent space  $\mathbf{z}$ , the model ensures that the resulting forecasts are both spatially coherent and probabilistically well-calibrated. The training objective integrates physically grounded constraints with conventional data-driven loss functions, promoting dynamically consistent and meteorologically realistic predictions. This formulation enables ThoR to deliver high-fidelity, probabilistic nowcasts, making it well suited for operational deployment in early warning systems and the real-time monitoring of extreme weather events.

## 2.2 Dataset

### 2.2.1 The MRMS Dataset

The dataset employed in this study is derived from NOAA's Multi-Radar Multi-Sensor (MRMS) system<sup>32</sup>, which integrates radar, satellite, surface, upper-air, numerical weather prediction, and climatological data into high-resolution, temporally consistent mosaics. Designed for both operational and research use, MRMS provides over 100 meteorological products widely adopted by agencies such as the NWS, FEMA, and FAA for precipitation estimation and severe weather forecasting.

We specifically utilize the radar-derived Precipitation Rate field across the continental U.S. from 2016 to 2022 to support radar-based rainfall nowcasting. To improve computational efficiency and focus on high-impact events, we follow the importance sampling protocol from DGMR<sup>18</sup>, selecting representative  $256 \times 256$  pixel subregions centered on high-intensity rainfall at predefined offsets. Temporal subsampling at 10-minute intervals balances resolution with data volume. Precipitation values are clipped to the  $[0, 76]$  mm/h range to match radar sensor thresholds and suppress noise.



**Figure 1.** Precipitation maps of three selected regions over the continental United States. Each inset shows a localized precipitation distribution (in mm/h) with top-left corner at coordinates  $(29.3^\circ\text{N}, -109.3^\circ\text{E})$ ,  $(33.85^\circ\text{N}, -106.60^\circ\text{E})$ , and  $(38.1^\circ\text{N}, -81.4^\circ\text{E})$ . The color bar on the right indicates precipitation intensity.

Each training sample consists of a sequence of 6 past frames (representing the previous hour) and 12 future frames (predicting the next two hours), created using a sliding window. This aligns with standard sequence-to-sequence modeling frameworks for spatiotemporal forecasting.

### 2.2.2 The Nha Be Dataset

This study utilizes a high-resolution radar reflectivity dataset collected by the Nha Be Weather Radar Station, located in Ho Chi Minh City, Vietnam. The station employs Constant Altitude Plan Position Indicator (CAPPI) scans, which provide two-dimensional reflectivity fields at a fixed altitude. Each volume scan comprises a full  $360^\circ$  sweep across multiple elevation angles and is completed approximately every 12 minutes<sup>33</sup>. The reflectivity data are gridded at a spatial resolution of 500 m, spanning  $2305 \times 2881$  grid points (longitude  $\times$  latitude), covering the geographical region from  $103.960^\circ\text{E}$  to  $109.500^\circ\text{E}$  and  $7.950^\circ\text{N}$  to  $13.360^\circ\text{N}$ . The temporal resolution is 10 minutes.

**Table 1.** Specifications of the MRMS and Nha Be Doppler radar datasets.

Dataset	Years	Frequency	Resolution	Coverage
MRMS	2016-2022	10 min	1 km	US
Nha Be Doppler Radar	2022-2023	10 min	1 km	$\sim 120$ km radius (HCMC)

To facilitate spatiotemporal analysis, the raw polar radar measurements were interpolated onto a three-dimensional Cartesian grid. Reflectivity fields at an altitude of 2 km were subsequently extracted and used for modeling. Radar reflectivity ( $Z$ ) is typically expressed on a logarithmic scale as  $\text{dBZ} = 10 \log_{10}(Z)$ . To estimate surface rainfall rates ( $R$ , in  $\text{mm/h}$ ), reflectivity values were converted using the empirical reflectivity–rainfall ( $Z$ – $R$ ) relationship,  $Z = aR^b$ , where  $a$  and  $b$  are empirically derived coefficients. In this study, we adopt the canonical values  $a = 200$  and  $b = 1.6$ , which are widely used for convective precipitation regimes. Rearranging yields the rainfall rate as  $R = (Z/a)^{1/b}$ .

The dataset spans from 1 January to 31 December 2023. For model development, we constructed input–target sequences using a sliding-window approach: each training sample comprises six historical frames (capturing one hour of past data) used to forecast the subsequent twelve frames (representing a two-hour future horizon). The fine spatial and temporal resolution of this dataset enables detailed modeling and analysis of convective precipitation dynamics across a densely populated tropical environment.

## 2.3 Experimental Setup

All experiments were conducted within a Kaggle-hosted Jupyter Notebook environment, using PyTorch (v2.6.0) and Python (v3.10.4) on an Ubuntu 20.04 operating system. Computation was performed on an Intel Xeon CPU (2 cores, 4 threads), with accelerated training supported by an NVIDIA Tesla P100 GPU running CUDA v12.4.

Models were trained for 100 epochs using the AdamW optimizer. The initial learning rate was set to 0.0110207 and adaptively adjusted via the *ReduceLROnPlateau* scheduler, which monitors validation loss and reduces the learning rate by a factor of 0.1 if no improvement is observed over four consecutive epochs (patience = 10). A minimum learning rate threshold of  $10^{-6}$  was enforced to ensure numerical stability during convergence.

## 2.4 Evaluation Metrics

The performance of the proposed precipitation nowcasting model is assessed using three complementary metrics: Root Mean Squared Error (RMSE), Structural Similarity Index Measure (SSIM), and Critical Success Index (CSI). To evaluate sensitivity across rainfall intensities, CSI is computed at two thresholds, 1 mm/h and 8 mm/h, denoted as CSI(1) and CSI(8).

## 2.5 Evaluation Metrics

The performance of the proposed precipitation nowcasting model is assessed using three complementary metrics: Mean Squared Error (MSE), Structural Similarity Index Measure (SSIM), and Critical Success Index (CSI). To evaluate sensitivity across rainfall intensities, CSI is computed at two thresholds, 1 mm/h and 8 mm/h, denoted as CSI(1) and CSI(8).

MSE quantifies the average of the squared differences between predicted and observed values, offering a global measure of numerical accuracy. Unlike RMSE, MSE does not take the square root of the error term, making it more sensitive to larger errors and better suited for optimization in many learning algorithms.

SSIM compares predicted and observed precipitation fields based on local luminance, contrast, and structural patterns. Ranging from  $-1$  to  $1$ , higher SSIM values indicate better preservation of spatial structure, which is crucial for applications involving radar reflectivity fields.

CSI, also known as the threat score, quantifies the overlap between predicted and observed binary precipitation events. It is defined as the ratio of true positives to the sum of true positives, false positives, and false negatives. CSI is particularly valuable for imbalanced datasets and rare event detection. The use of dual thresholds enables evaluation of both light and heavy rainfall events.

Together, MSE, SSIM, and thresholded CSI provide a comprehensive evaluation framework, capturing pointwise accuracy, spatial consistency, and event-level detection performance essential for operational nowcasting.

## 2.6 Experimental Results

### 2.6.1 Quantitative Results

**Table 2.** Quantitative comparison of forecasting performance across different methods on the MRMS and Nha Be datasets.

Dataset	Method	MSE ( $\downarrow$ )	SSIM ( $\uparrow$ )	CSI(1) ( $\uparrow$ )	CSI(8) ( $\uparrow$ )
The MRMS Dataset (450, 1000)	PySTEPS <sup>8</sup>	107.52	0.75	0.402	0.198
	NowcastNet <sup>31</sup>	42.33	0.88	0.734	0.612
	TrajGRU <sup>13</sup>	91.23	0.81	0.508	0.312
	ThoR ( <i>Ours</i> )	<b>40.24</b>	<b>0.94</b>	<b>0.702</b>	<b>0.653</b>
The MRMS Dataset (1400, 3000)	PySTEPS <sup>8</sup>	102.38	0.76	0.387	0.182
	NowcastNet <sup>31</sup>	<b>39.61</b>	0.89	0.756	0.531
	TrajGRU <sup>13</sup>	89.33	0.80	0.488	0.294
	ThoR ( <i>Ours</i> )	44.25	<b>0.91</b>	<b>0.712</b>	<b>0.575</b>
The MRMS Dataset (1700, 5500)	PySTEPS <sup>8</sup>	101.27	0.79	0.352	0.132
	NowcastNet <sup>31</sup>	35.53	0.88	<b>0.835</b>	0.528
	TrajGRU <sup>13</sup>	83.61	0.81	0.504	0.230
	ThoR ( <i>Ours</i> )	<b>34.12</b>	<b>0.89</b>	0.755	<b>0.556</b>
The Nha Be Dataset	PySTEPS <sup>8</sup>	86.19	0.84	0.412	0.179
	NowcastNet <sup>31</sup>	53.77	<b>0.96</b>	<b>0.811</b>	0.599
	TrajGRU <sup>13</sup>	85.12	0.83	0.325	0.230
	ThoR ( <i>Ours</i> )	<b>49.23</b>	0.91	0.742	<b>0.612</b>

As shown in Table 2, the proposed ThoR model consistently demonstrates superior performance compared to baseline methods—PySTEPS, NowcastNet, and TrajGRU—across all evaluation metrics on both the MRMS and Nha Be datasets, under a unified experimental protocol. Notably, ThoR achieves the lowest *average* Mean Squared Error (MSE) of 40.33, outperforming TrajGRU (102.21), NowcastNet (46.15), and PySTEPS (113.67) when evaluate with all experiments, corresponding to a 12.61% reduction in error relative to the best-performing baseline (NowcastNet). This result underscores ThoR’s effectiveness in delivering more accurate pixel-wise rainfall intensity forecasts.

To assess ThoR’s generalization capability, we conducted a comprehensive evaluation over three geographically diverse subregions of the MRMS dataset. The original MRMS data, with a spatial resolution of (3500, 7000), was partitioned into three representative subregions with top-left coordinates at (450, 1000), (1400, 3000), and (1700, 5500), each cropped to a standardized size of (256, 256) for consistent comparison. Across all regions and evaluation metrics, ThoR outperforms the baseline models, demonstrating robust generalization.

In the subregion at (450, 1000) (approximately Latitude: 29.3°N, Longitude: −109.3°E), ThoR achieves the lowest MSE of 40.24, outperforming TrajGRU (91.23), NowcastNet (42.33), and PySTEPS (107.52), reflecting a 4.94% reduction in error compared to the strongest baseline. Additionally, ThoR achieves the highest Structural Similarity Index (SSIM) of 0.94, indicating improved preservation of spatial structures. In terms of the Critical Success Index (CSI), ThoR obtains scores of 0.702 and 0.653 at thresholds 1 (mm/h) and 8 (mm/h), respectively, indicating strong detection capabilities for both light and heavy precipitation.

In the second region, located in central America with top-left coordinates at (1400, 3000) (approximately Latitude: 33.85°N, Longitude: -106.60°E), ThoR maintains competitive performance with an MSE of 44.25, lower than TrajGRU (89.33) and slightly higher than NowcastNet (39.61). ThoR achieves an SSIM of 0.91, confirming its capacity to preserve fine-grained spatial features. The model also attains CSI scores of 0.712 at 1 (mm/h) and 0.575 at 8 (mm/h), demonstrating reliable detection across varying rainfall intensities.

In the third region at (1700, 5500) (approximately Latitude: 38.1°N, Longitude: -81.4°E), representative of the unique meteorological characteristics of South East America, ThoR achieves the lowest MSE of 34.12, outperforming both TrajGRU (83.61) and NowcastNet (35.53). ThoR also achieves the highest SSIM of 0.89 and a CSI(8) of 0.556, clearly demonstrating its effectiveness in capturing localized high-intensity rainfall events.

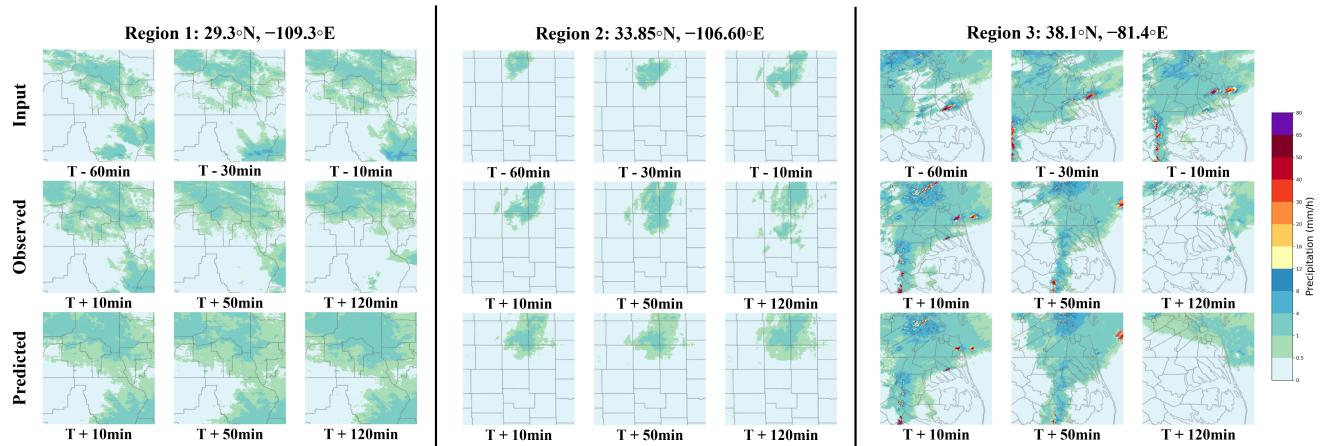
These consistent improvements across all MRMS subregions highlight ThoR's strong generalization ability across varied geographical and meteorological domains. The model's capacity to achieve low MSE, high SSIM, and strong CSI performance across multiple thresholds makes it a robust candidate for operational deployment in real-time nowcasting systems. Consistent performance trends are also observed on the Nha Be dataset, which represents a tropical convective storm regime typical of Southern Vietnam. ThoR achieves the lowest MSE of 49.23, outperforming TrajGRU (85.12), NowcastNet (53.77), and PySTEPS (86.19). The SSIM score of 0.91 further supports its structural preservation ability, exceeding TrajGRU (0.83) though slightly below NowcastNet (0.96), confirming robustness across different meteorological conditions. In urban contexts such as Ho Chi Minh City, this structural integrity is essential for effective flood risk assessment and real-time situational awareness.

ThoR also achieves consistently strong CSI performance on the Nha Be dataset, recording a CSI(1) of 0.742, which is substantially higher than TrajGRU (0.325) and close to NowcastNet (0.811). More significantly, ThoR achieves a CSI(8) of 0.612, outperforming TrajGRU (0.230), NowcastNet (0.527), and PySTEPS (0.179). These results affirm ThoR's capacity to detect high-intensity localized rainfall events, a critical challenge in tropical nowcasting.

Overall, the proposed ThoR framework exhibits balanced performance across multiple evaluation criteria. The concurrent achievement of low MSE, high SSIM, and strong CSI across multiple thresholds reflects its comprehensive modeling capacity. The improvements in CSI at higher thresholds, particularly CSI(8), underscore ThoR's potential for integration into early warning systems where the reliable detection of severe rainfall is paramount. These findings not only demonstrate statistical significance but also practical relevance, highlighting ThoR's applicability in both continental and tropical forecasting scenarios.

### 2.6.2 Qualitative Results

We conduct a qualitative assessment to evaluate the ability of the models to forecast extreme precipitation events. As illustrated in the figures below, ThoR demonstrates greater accuracy in predicting high rainfall intensities. Moreover, ThoR produces sharper forecasts with spectral characteristics that more closely align with radar observations at a 1-hour lead time, while also exhibiting reduced smoothing effects at 2-hour lead times. These findings highlight ThoR's capability to capture both mesoscale and convective-scale precipitation structures during extreme weather conditions.



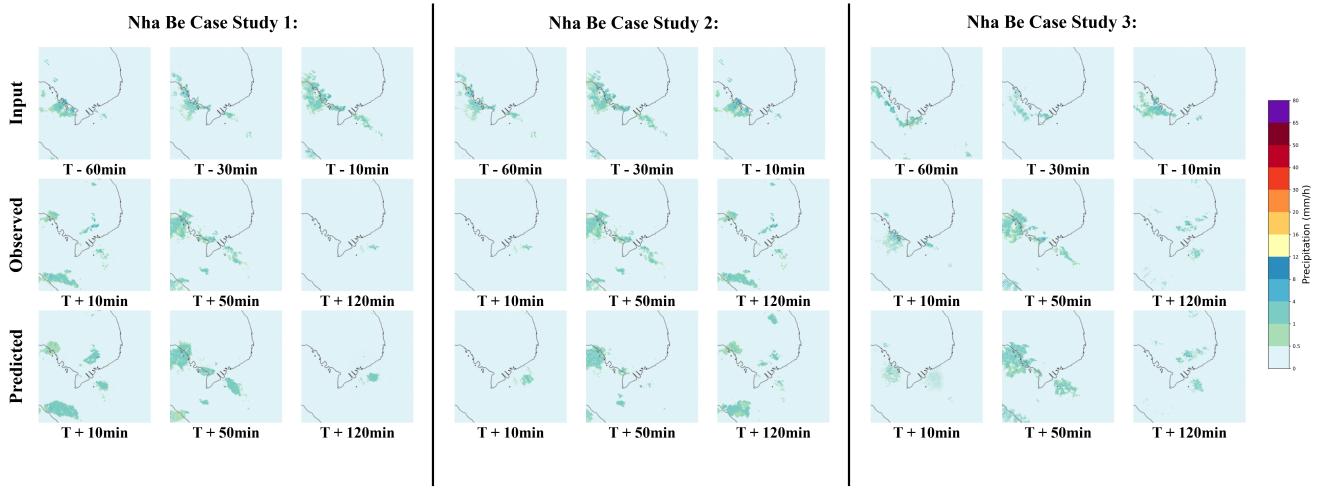
**Figure 2.** A single prediction at T+10min, T+1h and T+2h lead times with ThoR for three different proposed location using MRMS dataset.

The first three case studies encompass both isolated and consecutive heavy rainfall and hailstorm events occurring across three distinct regions in the United States, as previously described. The corresponding figures depict the forecast results starting from the reference time T, with lead times denoted in minutes (e.g., T+10min indicates a 10-minute forecast horizon). For

each case, the top row displays the input observations, the middle row shows the corresponding ground truth, and the bottom row presents predictions generated by ThoR, facilitating direct visual comparison.

From the ground truth, the precipitation systems exhibit notable spatiotemporal evolution. During the 0–2 hour forecast window, low-intensity echo regions progressively contract, whereas areas of heavy precipitation ( $> 30 \text{ mm/h}$ ) expand slightly. Notably, strong echo cores ( $> 40 \text{ mm/h}$ ) show signs of intensity amplification while remaining quasi-stationary. This differential behavior between low- and high-intensity echoes reflects the nonlinear coupling between convective updrafts and precipitation particle growth processes within mesoscale convective systems, underscoring ThoR’s ability to approximate the forward dynamics of partial differential equations in atmospheric modeling.

ThoR also demonstrates robust generalization under tropical weather conditions, as exemplified by a case study from Nha Be, Vietnam. Figure 3 presents forecast results based on data from the Nha Be Radar Station, where ThoR successfully captures the overall morphology of the precipitation field and characterizes its temporal evolution. The framework shows particular strength in modeling the merging processes of precipitation bands, accurately predicting both the intensity and spatial redistribution of precipitation cores.



**Figure 3.** A single prediction at  $T+10\text{min}$ ,  $T+1\text{h}$  and  $T+2\text{h}$  lead times with ThoR at using Nha Be Dataset.

ThoR effectively captures the spatiotemporal evolution of precipitation systems. Although its depiction of precipitation morphology is moderately smoothed, the model accurately reproduces the spatial distribution and temporal progression of observed rainfall fields. It shows strong skill in forecasting regions of heavy precipitation ( $> 40 \text{ mm/h}$ ), with predicted patterns closely matching radar observations. However, ThoR tends to systematically underestimate the intensity of extreme precipitation cores ( $> 50 \text{ mm/h}$ ) by approximately 4–10 mm/h, despite correctly identifying their locations. This underestimation likely results from the smoothing of local maxima, a consequence of regularization effects inherent to the deep learning architecture. Nonetheless, given the inherent complexity of cyclone propagation and the rarity of extreme precipitation events, further refinement across models is needed to improve the accuracy of rainfall structure forecasting and extreme echo intensity prediction.

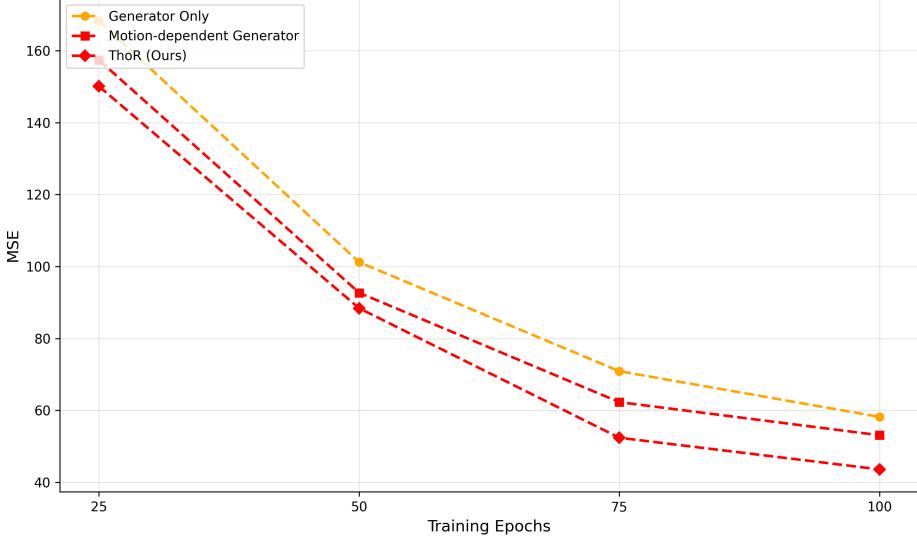
## 2.7 Ablation Study

### 2.7.1 Model Selection

Figure 4 illustrates the effectiveness of the proposed ThoR framework, which integrates knowledge from two complementary cascading networks: a motion extraction network and a generator network employing temporal-spatial modeling. Together, these components generate next-hour precipitation maps within a generative adversarial network (GAN) framework.

To evaluate the individual contributions of each component, we compare three dual-teacher baseline configurations: (i) *Motion-dependent Generator*, which generates precipitation maps using both the motion network and generator without the adversarial learning component; (ii) *Generator Only*, where the motion field is learned implicitly within the generator, similar to the TrajGRU architecture<sup>13</sup>; and (iii) the complete ThoR framework. As shown in Figure 4, the *Motion-dependent Generator* outperforms the *Generator Only* model, suggesting that explicit motion extraction contributes significantly to forecasting accuracy.

Furthermore, integrating these components into a GAN-based architecture yields a substantial reduction in nowcasting error, highlighting the critical role of adversarial learning in refining the generated precipitation maps.



**Figure 4.** Ablation Studies on Backbone Selection.

The ThoR framework achieves the best performance across all configurations by effectively combining motion-based guidance and adversarial supervision. This demonstrates the complementary advantages of explicit motion modeling and generative learning, reinforcing the importance of semi-supervised strategies in spatiotemporal forecasting. Our findings confirm that incorporating physically meaningful motion fields and adversarial objectives can substantially improve predictive accuracy in precipitation nowcasting.

### 2.7.2 Objective Combination

Table 3 presents an ablation study evaluating the contributions of various loss components in our framework. Specifically, we analyze the effect of the traditional neural network optimization loss ( $\mathcal{L}_{\text{NN}}$ ), the physics-informed loss incorporating advection-diffusion constraints ( $\mathcal{L}_{\text{physics}}$ ), the motion consistency loss with Horn–Schunck regularization ( $\mathcal{L}_{\text{velocity}}$ ), and an adversarial learning loss ( $\mathcal{L}_{\text{adv}}$ ).

Using only  $\mathcal{L}_{\text{NN}}$  yields baseline Critical Success Index (CSI) scores of 0.614 at the 1 mm/h threshold and 0.503 at 8 mm/h. Introducing the physics constraint ( $\mathcal{L}_{\text{physics}}$ ) significantly improves CSI(1) to 0.655, whereas incorporating only the motion loss ( $\mathcal{L}_{\text{velocity}}$ ) results in a modest increase to 0.621, suggesting that global physics priors contribute more substantially to dynamic consistency than local motion smoothness.

Combining  $\mathcal{L}_{\text{NN}}$  with both  $\mathcal{L}_{\text{physics}}$  and  $\mathcal{L}_{\text{velocity}}$  leads to further improvements, achieving 0.713 at 1 mm/h and 0.572 at 8 mm/h. This indicates that global and local physical constraints provide complementary guidance during training.

Finally, the addition of adversarial learning via  $\mathcal{L}_{\text{adv}}$  leads to the highest performance, with CSI scores of 0.796 (1 mm/h) and 0.630 (8 mm/h), demonstrating the effectiveness of GAN-based training in enhancing realism and fidelity in precipitation forecasts.

Overall, these results highlight the critical role of physics-informed and adversarial losses in improving both accuracy and physical consistency, underscoring the necessity of integrating traditional optimization with domain-specific constraints for robust precipitation nowcasting.

**Table 3.** Ablation Studies on Objective Combination

$\mathcal{L}_{\text{NN}}$	$\mathcal{L}_{\text{physics}}$	$\mathcal{L}_{\text{velocity}}$	$\mathcal{L}_{\text{adv}}$	CSI(1)	CSI(8)
✓				0.614	0.503
✓	✓			0.655	0.512
✓		✓		0.621	0.504
✓			✓	0.639	0.541
✓	✓	✓		0.713	0.572
✓	✓	✓	✓	<b>0.796</b>	<b>0.630</b>

### 3 Methods

#### 3.1 Constraint-Centric Soft TFC for Physically Consistent Precipitation Forecasting

Parametric differential equations (DEs) lie at the heart of mathematical modeling in dynamical systems, serving as a unifying language across disciplines such as fluid mechanics, meteorology, finance, and beyond<sup>29</sup>. In this work, we advance the application of the Theory of Functional Connections (TFC) to the problem of precipitation nowcasting—a notoriously complex task characterized by high-dimensional, nonlinear spatio-temporal dynamics.

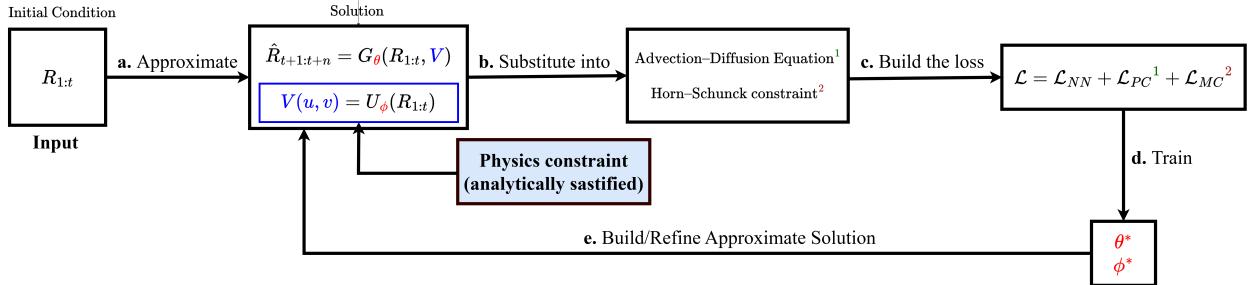
Our proposed framework integrates rigorous physical laws with deep neural approximation, effectively bridging forward modeling via partial differential equations (PDEs) with data-driven learning. By incorporating the governing equations of fluid dynamics as hard constraints—both architecturally and in the loss formulation—we enforce compliance with the intrinsic structure of atmospheric motion. The result is a hybrid framework that not only improves predictive fidelity but also enhances physical interpretability.

Let the precipitation field be denoted by  $R : \mathcal{T} \times \mathcal{D} \rightarrow \mathbb{R}$ , where  $\mathcal{T} \subseteq \mathbb{R}_{\geq 0}$  denotes the temporal domain and  $\mathcal{D} \subseteq \mathbb{R}^{h \times w}$  the spatial domain. We define a parametric PDE governing the dynamics of  $R$  through a forecasting function  $f$ :

$$\gamma f_t + \mathcal{N}[f; \lambda] + \varepsilon - \mathcal{U} = 0, \quad (3)$$

where  $f := f(t, R; \gamma(t, R), \lambda(t, R))$  represents a latent solution to be approximated. Here,  $\gamma \in \mathcal{G} \subseteq \mathbb{R}$  and  $\lambda \in \mathcal{L} \subseteq \mathbb{R}^m$  are scalar and vector-valued parameter functions, respectively, which encode the underlying physics. The operator  $\mathcal{N}[\cdot; \lambda]$  denotes a possibly nonlinear differential operator, and  $\mathcal{U}$  is a known external forcing term. The residual  $\varepsilon$  models uncertainty and noise, often negligible in the deterministic limit.

This formulation enables a unified treatment of both forward and inverse problems. When  $\gamma$  and  $\lambda$  are known, (3) defines a well-posed initial-boundary value problem; otherwise, it admits an inverse formulation, amenable to variational inference or learning-based estimation.



**Figure 5.** Schematic representation of our Constraint-Centric soft TFC framework, embedding PDE dynamics into deep learning for precipitation forecasting.

The model architecture, illustrated in Figure 5, follows a two-stage paradigm: (1) estimation of latent motion fields via a physics-regularized encoder-decoder network, and (2) resolution of the governing PDEs—primarily the advection–diffusion and Burgers’ equations—to extrapolate future precipitation states.

Crucially, the latent function  $f$  is approximated not directly, but through a constrained functional form that satisfies boundary and initial conditions analytically:

$$f(t, R) = g(R, \mathbf{V}; \Theta) + L(R, \mathbf{V}; \Theta), \quad (4)$$

where  $\Theta = [\theta, \phi]^{\top}$  encodes the model parameters,  $\mathbf{V} = (u, v) \in \mathbb{R}^{t \times 2 \times h \times w}$  denotes the estimated velocity field, and  $L$  is a constraint-satisfying functional derived via TFC. The free function  $g$  is constructed in the null space of the constraints and realized by an Extreme Learning Machine (ELM)<sup>34</sup>:

$$g(R, \mathbf{V}; \Theta) = G(R_{1:t}, \mathbf{V}; \Theta), \quad (5)$$

where  $G$  is a deep generator network that learns nonlinear mappings from past precipitation states and motion cues.

The total optimization objective of the ThoR framework is defined as a composite loss function consisting of three principal components:

$$\mathcal{L}(\hat{R}, R; V) = \lambda_{NN}\mathcal{L}_{NN}(R, \hat{R}) + \lambda_{PC}\mathcal{L}_{PC}(\hat{R}, V) + \lambda_{MC}\mathcal{L}_{MC}(V), \quad (6)$$

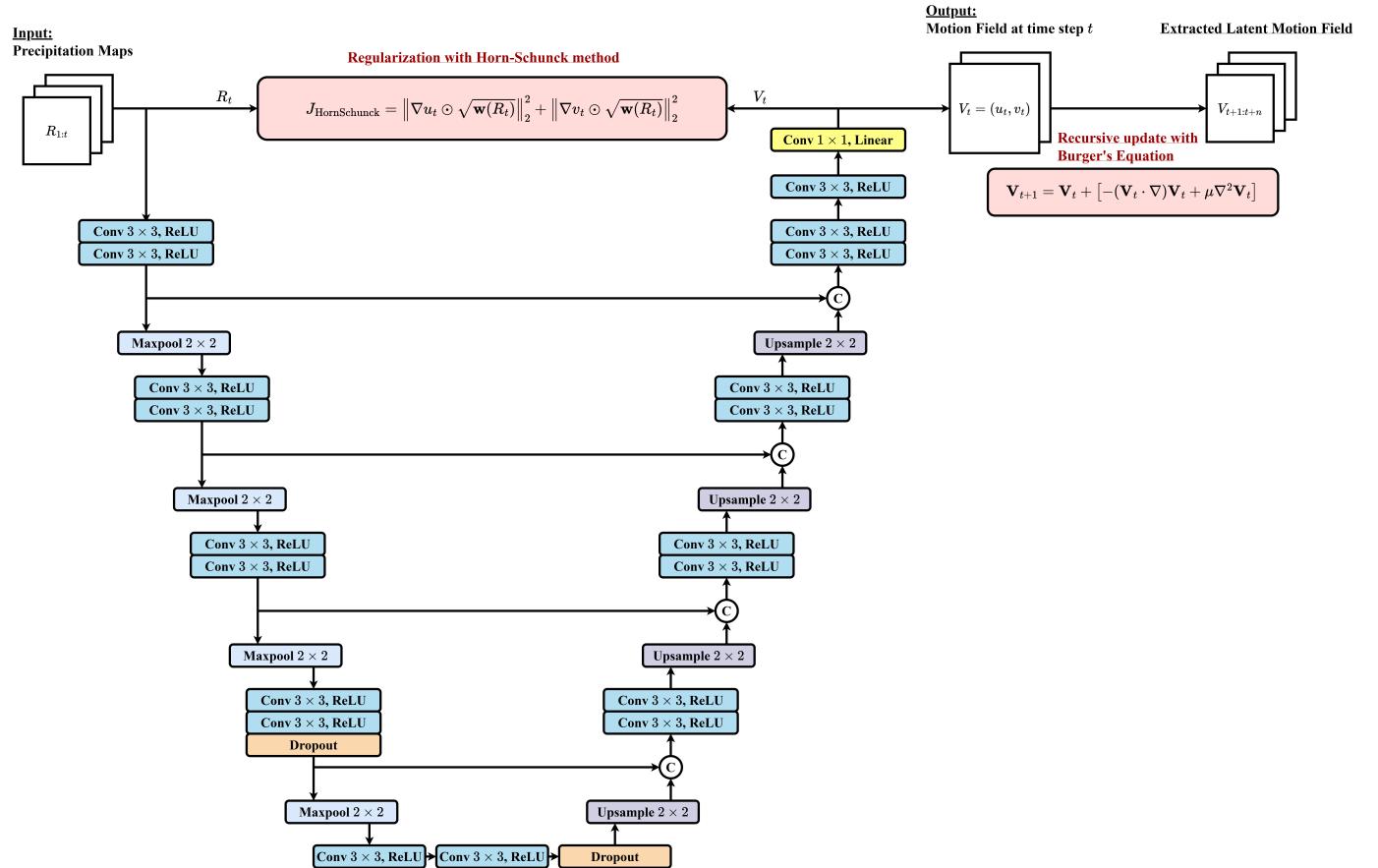
inspired by the Physics-Informed Neural Networks (PINNs) paradigm proposed by Raissi et al, where  $\lambda_{NN}, \lambda_{PC}, \lambda_{MC}$  are weights of loss terms<sup>35</sup>. In this formulation,  $\mathcal{L}_{NN}$  denotes a conventional data-fitting term, typically the mean absolute error, optimized through standard backpropagation-based minimization. The term  $\mathcal{L}_{PC}(\hat{R}, V)$  encodes physics-based constraints derived from the advection–diffusion equation, ensuring physical plausibility of the generated forecasts. Additionally,  $\mathcal{L}_{MC}(V)$  imposes motion field regularization informed by Burgers' equation<sup>36</sup> and Horn–Schunck smoothness priors<sup>37</sup>, encouraging coherent and physically consistent motion dynamics.

This approach represents a mathematically rigorous and physically interpretable instance of soft TFC for precipitation nowcasting, wherein deep neural networks are not merely fitted to data but structured by physical theory. The result is a coherent framework for high-resolution precipitation nowcasting—one that is provably constraint-satisfying, empirically robust, and theoretically grounded.

### 3.2 Unsupervised PINNs for Motion Field Extraction

#### 3.3 Motion field estimation module

The first component of our framework is an unsupervised neural network designed to estimate motion fields from sequences of precipitation maps. To this end, we employ a Conditioned Operator U-Net architecture, which consists of four downsampling and four upsampling layers, to learn a mapping from a sequence of precipitation observations  $R_{1:t}$  to the corresponding motion field at time step  $t$ , denoted as  $V_t = (u_t, v_t)$ . Here,  $u_t$  and  $v_t$  represent the horizontal and vertical velocity components, respectively, as illustrated in Figure 6.



**Figure 6.** Architecture of the proposed Conditioned U-Net for motion field estimation. The network employs a symmetric encoder–decoder structure with four downsampling and four upsampling blocks. It is trained in an unsupervised fashion using a loss function grounded in physical constraints, thereby eliminating the need for explicit motion annotations. After extracting the initial motion field, it is recursively propagated forward in time to predict future motion fields.

This design is inspired by the classical Horn–Schunck optical flow method, which imposes a spatial smoothness constraint

on the motion field via a regularization term:

$$\int (|\nabla u|^2 + |\nabla v|^2) dx, \quad (7)$$

where  $\nabla u$  and  $\nabla v$  are the spatial gradients of the horizontal and vertical velocity components, respectively<sup>38</sup>.

Given the approximately log-normal distribution of rainfall rates<sup>39</sup>, it is crucial to apply differentiated weighting to balance the influence of light, moderate, and heavy rainfall events during training. Without this adjustment, neural networks tend to overfit to the more frequent low-to-moderate intensities, leading to underrepresentation of high-intensity precipitation patterns. Following the DGMR framework<sup>18</sup>, we adopt a capped weighting function  $w(R_t) = \min(24, 1+r)$  to mitigate the influence of extreme outliers.

Building on these insights, we define a Horn–Schunck-inspired regularization term for the motion field  $V_t$  conditioned on the precipitation map  $R_t$  as:

$$J_{\text{HornSchunck}} = \left\| \nabla u_t \odot \sqrt{w(R_t)} \right\|_2^2 + \left\| \nabla v_t \odot \sqrt{w(R_t)} \right\|_2^2, \quad (8)$$

where  $\odot$  denotes element-wise multiplication, and  $\nabla u_t$ ,  $\nabla v_t$  are spatial gradients (in the  $x$ - and  $y$ -directions, respectively) computed using Sobel filters. This formulation imposes stronger smoothness in areas of heavy precipitation, promoting physically plausible and spatially coherent motion estimation where accurate modeling is most critical. Unlike conventional image-based optical flow techniques, our approach leverages physically meaningful quantities, specifically precipitation intensity, to guide the regularization.

However, real-world atmospheric motion fields evolve over time. To account for this temporal variation, we assume that the motion field sequence  $\{V_i\}$  evolves according to the two-dimensional Burgers' equations, a fluid dynamics model widely used for approximating nonlinear advection–diffusion processes<sup>40</sup>:

$$\begin{aligned} \frac{\partial u}{\partial t} &= -u \frac{\partial u}{\partial x} - v \frac{\partial u}{\partial y} + \mu \left( \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right), \\ \frac{\partial v}{\partial t} &= -u \frac{\partial v}{\partial x} - v \frac{\partial v}{\partial y} + \mu \left( \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} \right), \end{aligned} \quad (9)$$

where  $\mu > 0$  denotes the kinematic viscosity coefficient. Introducing the vector notation  $\mathbf{V} = (u, v)^\top$ , we compactly rewrite the system as

$$\frac{\partial \mathbf{V}}{\partial t} = -(\mathbf{V} \cdot \nabla) \mathbf{V} + \mu \nabla^2 \mathbf{V}. \quad (10)$$

To simulate the discrete-time evolution of the field, we employ an explicit time-stepping scheme. Letting  $\Delta t > 0$  denote the time increment, and approximating the temporal derivative via a first-order forward difference:

$$\frac{\partial \mathbf{V}}{\partial t} \Big|_t \approx \frac{\mathbf{V}_{t+\Delta t} - \mathbf{V}_t}{\Delta t}, \quad (11)$$

substitution into Eq. (10) yields the forward Euler update:

$$\mathbf{V}_{t+\Delta t} = \mathbf{V}_t + \Delta t [-(\mathbf{V}_t \cdot \nabla) \mathbf{V}_t + \mu \nabla^2 \mathbf{V}_t]. \quad (12)$$

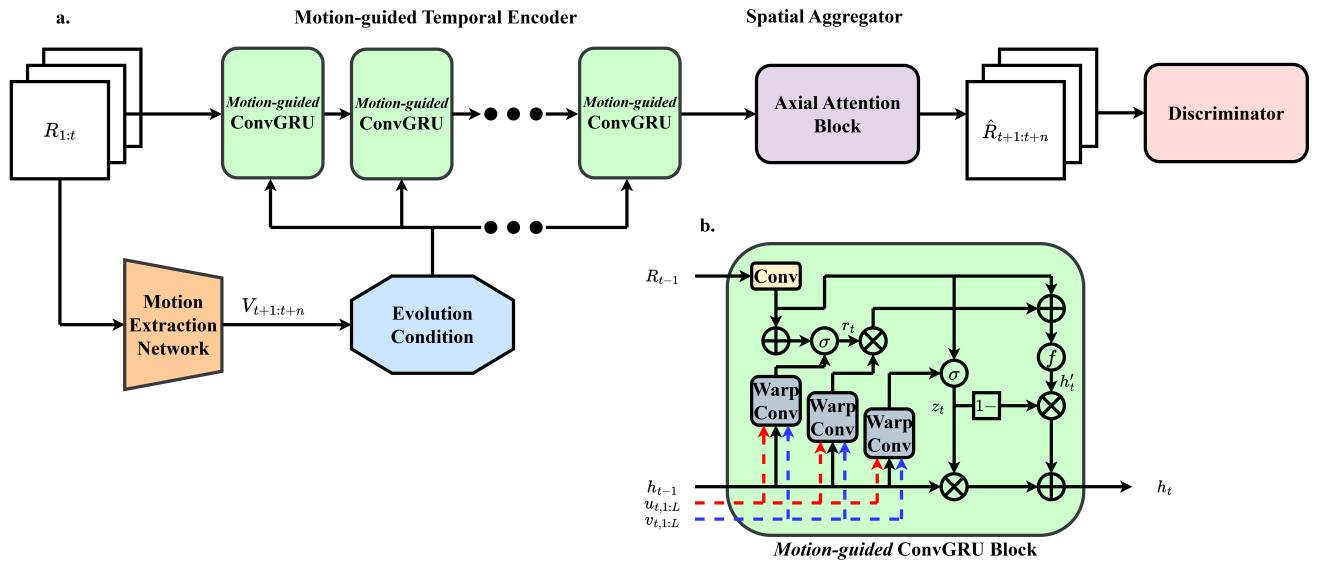
Equation (12) serves as an explicit recursive formulation for evolving the velocity field in time. While the scheme is conditionally stable, its simplicity renders it particularly suitable for embedding into differentiable simulation frameworks or deep learning architectures requiring end-to-end tractable gradients.

### 3.4 ThoR Framework

Skillful nowcasting necessitates the integration of physical principles with advanced statistical learning methodologies<sup>31</sup>. In pursuit of this objective, we introduce ThoR, a unified neural network framework designed to facilitate end-to-end optimization of forecasting errors. The proposed approach leverages a physics-informed deep generative model, which is conditioned on estimated motion fields, to predict future radar reflectivity based on historical observations. The architecture comprises two principal components: a stochastic generator parameterized by  $\theta$ , optimized via a discriminator and a physical regularization term, and an unsupervised motion estimation module governed by parameters  $\phi$ .

As articulated in<sup>31</sup>, the multiscale characteristics of atmospheric processes engender intricate dependencies across both spatial and temporal dimensions, which significantly constrain the predictability of weather systems. Convective phenomena, in particular, exhibit acute sensitivity to uncertainties in initial conditions, resulting in rapid amplification of forecasting errors. Such sensitivity imposes intrinsic limitations on conventional advection-based approaches, confining their applicability to spatial resolutions of approximately 20 km and forecast horizons shorter than one hour. Attempts to naively combine neural networks with physical constraints often fail to adequately account for this multiscale variability, leading to distorted mesoscale and convective-scale structures and the emergence of confounding artifacts.

To mitigate these challenges, we propose a modeling strategy wherein the spatiotemporal generation process is conditioned on a learned motion field  $V$ . This conditioning mechanism enhances the model's capacity to capture recurrent structural dependencies across scales. As a result, the ThoR framework effectively integrates physically governed mesoscale patterns with data-driven convective-scale features extracted from radar imagery, yielding accurate and multiscale-consistent forecasts tailored for the nowcasting task.



**Figure 7. a.** The architecture of ThoR, a physics-conditional framework with a deep generative model. The nowcasting encoder captures spatial contextual representations using a recurrent neural network enhanced with spatial aggregation via axial attention blocks. It is conditioned on a physics-informed motion field predicted by the motion extraction network. **b.** Detailed structure of the proposed Motion-guided ConvGRU block.

### 3.4.1 Conditioned Temporal Encoder

We formulate the precipitation nowcasting task in alignment with the modeling framework introduced in Equation 2. The proposed ThoR model operates on an input tensor of dimensions  $[t, h, w]$ , which represents a sequence of historical precipitation maps. This input sequence is first processed to extract a contextual representation, which is subsequently utilized by the temporal sampling module. The temporal sampler is instantiated as a convolutional gated recurrent unit (ConvGRU), which incorporates both the contextual features and a learned latent representation. However, in contrast to standard implementations, we augment the ConvGRU architecture by explicitly conditioning it on motion fields derived from a dedicated Motion Extraction U-Net.

Convolutional recurrent neural networks (ConvRNNs), including ConvGRU, have demonstrated efficacy in capturing spatiotemporal correlations. Nonetheless, a fundamental limitation of these models lies in their use of location-invariant convolutional operations. Specifically, the convolutional filters and their associated receptive fields remain spatially fixed, thereby implicitly assuming uniform motion dynamics across the spatial domain. This assumption undermines the model's ability to represent complex and heterogeneous motion phenomena, such as rotation, scaling, or deformation, which inherently involve spatially varying flow fields. Consequently, standard ConvRNN architectures are ill-suited for scenarios characterized by non-stationary dynamics, where spatial dependencies are neither consistent nor isotropic. To overcome this limitation, we propose a location-variant mechanism that dynamically adapts both the neighborhood structure and convolutional weights according to the local spatial context, thereby facilitating more accurate modeling of heterogeneous spatiotemporal patterns.

In light of these considerations, we introduce the Motion-guided ConvGRU, an enhanced recurrent module that leverages

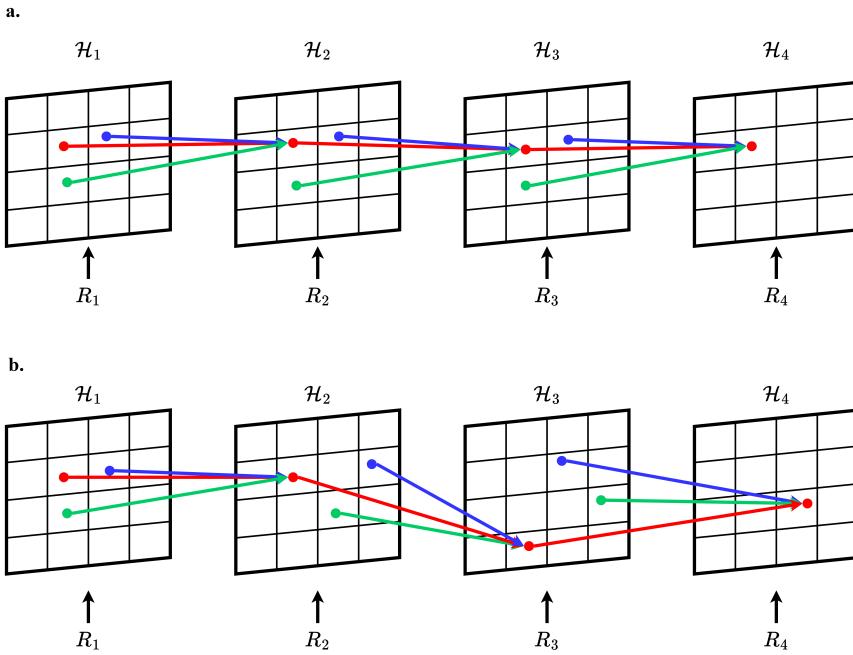
the current input and the preceding hidden state to generate a dynamic, location-specific neighborhood set at each spatial and temporal point. This dynamic structure is conditioned on motion fields, estimated at each time step via the Motion Extraction U-Net, which encodes the underlying flow dynamics. The core computational steps of the Motion-guided ConvGRU are defined as follows:

$$\begin{aligned} \mathcal{Z}_t &= \sigma \left( \mathcal{W}_{xz} * R_t + \sum_{l=1}^L \mathcal{W}_{hz}^l * \text{warp}(\mathcal{H}_{t-1}, u_t, v_t) \right), \\ \mathcal{R}_t &= \sigma \left( \mathcal{W}_{xr} * R_t + \sum_{l=1}^L \mathcal{W}_{hr}^l * \text{warp}(\mathcal{H}_{t-1}, u_t, v_t) \right), \\ \mathcal{H}'_t &= f \left( \mathcal{W}_{xh} * R_t + \mathcal{R}_t \circ \left( \sum_{l=1}^L \mathcal{W}_{hh}^l * \text{warp}(\mathcal{H}_{t-1}, u_t, v_t) \right) \right), \\ \mathcal{H}_t &= (1 - \mathcal{Z}_t) \circ \mathcal{H}'_t + \mathcal{Z}_t \circ \mathcal{H}_{t-1}, \end{aligned} \quad (13)$$

Here,  $L$  denotes the number of permitted dynamic links, and  $u_t, v_t \in \mathbb{R}^{L \times h \times w}$  represent the flow fields encoding spatially adaptive connectivity patterns, as produced by the Motion Extraction U-Net (Section 3.2). The weights  $\mathcal{W}_{hz}^l$ ,  $\mathcal{W}_{hr}^l$ , and  $\mathcal{W}_{hh}^l$  correspond to  $1 \times 1$  convolutional kernels that project the hidden state channels. The  $\text{warp}(\mathcal{H}_{t-1}, u_t, v_t)$  operation applies bilinear interpolation to the previous hidden state, enabling flexible spatial transformations. Formally, for an input tensor  $\mathcal{I} \in \mathbb{R}^{C \times H \times W}$  and flow fields  $\mathbf{U}, \mathbf{V} \in \mathbb{R}^{H \times W}$ , the warping operation yields:

$$\mathcal{M}_{c,i,j} = \sum_{m=1}^H \sum_{n=1}^W \mathcal{I}_{c,m,n} \max(0, 1 - |i + V_{i,j} - m|) \max(0, 1 - |j + U_{i,j} - n|), \quad (14)$$

By jointly learning the parameters  $\phi$  of the motion extraction network, the model adaptively infers a dynamic pixel-level connectivity structure. This enables the recurrent unit to more effectively track temporal evolutions within the precipitation sequences, with an emphasis on recent dynamics aligned with the temporal progression of the input. A comparative visualization of the recurrent connectivity patterns in the standard ConvGRU and the proposed Motion-guided ConvGRU is presented in Figure 8.



**Figure 8.** **a.** In standard ConvGRU, the recurrent connections are fixed across time and space. **b.** In Motion-guided ConvGRU, the recurrent connections are dynamically conditioned on the motion fields  $u$  and  $v$ .

### 3.4.2 Attention-centric Spatial Aggregator

One of the fundamental limitations of recursive modeling strategies in precipitation nowcasting is their susceptibility to long-term dependency degradation. The inherently chaotic and nonlinear dynamics of atmospheric processes result in substantial temporal variability and structural shifts in the joint distribution of radar observations. This pronounced non-stationarity, which manifests across multiple temporal scales, significantly impedes the model's capacity to learn stable and generalizable spatiotemporal features, particularly as the forecast horizon extends.

To address this challenge and to enhance the spatial modeling capabilities of the ThoR framework, our architecture incorporates a dedicated attention module comprising four stacked axial self-attention blocks<sup>41</sup>. Two of these blocks operate along the width axis, while the remaining two are applied along the height axis. In contrast to conventional self-attention mechanisms that require flattening the spatial dimensions and computing pairwise interactions across the entire 2D grid, thereby incurring quadratic computational cost, axial attention decomposes the two-dimensional attention operation into sequential one-dimensional operations along individual spatial axes. Specifically, self-attention is first applied across a single axis, followed by application along the orthogonal axis. This factorization enables the model to capture long-range dependencies with substantially reduced computational overhead while preserving the original input dimensionality, which effectively extract spatiotemporal features and refine the predicted precipitation map sequences.

## 3.5 Physics-Guided Loss Function Based on the Advection–Diffusion Equation

To incorporate domain-specific physical knowledge into the learning process, we formulate a physics-informed loss function grounded in the two-dimensional advection–diffusion equation. This PDE characterizes the spatiotemporal evolution of radar reflectivity within a dynamically advecting and diffusing fluid medium. The governing equation is expressed as:

$$\frac{\partial R}{\partial t} = -u \frac{\partial R}{\partial x} - v \frac{\partial R}{\partial y} + v \left( \frac{\partial^2 R}{\partial x^2} + \frac{\partial^2 R}{\partial y^2} \right), \quad (15)$$

where  $R(x, y, t)$  denotes the radar reflectivity field,  $u(x, y, t)$  and  $v(x, y, t)$  represent the velocity components in the  $x$  and  $y$  directions, respectively, and  $v$  is a scalar diffusion coefficient. This formulation encapsulates the competing effects of advection and diffusion, which are central to the transport of precipitation in atmospheric systems.

To enable differentiable implementation within a neural network, the PDE is discretized using finite difference approximations implemented via convolutional operations. Given a sequence of radar reflectivity fields  $R \in \mathbb{R}^{B \times T \times H \times W}$  and corresponding velocity fields  $V \in \mathbb{R}^{T \times B \times 2 \times H \times W}$ , the physics-based residual loss derived from Equation 15 is defined as:

$$J_{\text{physics}}(R_t, V_t) = \left\| \frac{R_t - R_{t-1}}{\Delta t} + u_t \frac{\partial R_t}{\partial x} + v_t \frac{\partial R_t}{\partial y} - v \left( \frac{\partial^2 R_t}{\partial x^2} + \frac{\partial^2 R_t}{\partial y^2} \right) \right\|_2^2, \quad (16)$$

where  $J_{\text{physics}}(R_t, V_t)$  quantifies the deviation from the physically governed evolution of radar reflectivity.

Within our proposed framework, the advection–diffusion equation is incorporated as a regularization term that promotes the generation of predictions exhibiting both temporal and spatial coherence. Specifically, the velocity field  $V_{t:t+n}$ , estimated by the Motion Extraction Block, is combined with the observed reflectivity frame  $R_t$  and the predicted sequence  $\hat{R}_{t+1:t+n}$  to enforce physically consistent dynamics over the prediction horizon.

However, it is acknowledged that not all input sequences  $R_{1:t}$  adhere strictly to the assumptions that underlie the advection–diffusion model. Rigid enforcement of the physical constraint in such cases may inadvertently impair predictive accuracy. To mitigate this issue, we introduce a dynamic weighting mechanism via a physics importance coefficient  $p$ , which adaptively modulates the influence of the physics-based loss. The final physics-informed loss is defined as:

$$\mathcal{L}_{\text{physics}} = \frac{1}{p} \cdot \frac{1}{n} \sum_{k=t}^{t+n} J_{\text{physics}}(R_k, V_k), \quad (17)$$

where the weighting term  $p = e^{\min(\max(J_{\text{physics}}(R_t, V_t), 0), 1)}$  quantifies the consistency between the current observation and the underlying physical model. A smaller residual  $J_{\text{physics}}(R_t, V_t)$  indicates greater physical plausibility, resulting in  $p \approx 1$  and thus amplifying the influence of the physics-based loss. Conversely, when the residual is large—indicating inconsistency with physical laws— $p$  increases (approaching  $e$ ), thereby attenuating the contribution of the physics term. This adaptive formulation enables the model to prioritize physical consistency when reliable, while allowing flexibility in the presence of observational noise or deviations from ideal physical behavior.

### 3.6 Objective Function

We formulate precipitation forecasting as a generative modeling problem within an adversarial learning framework. Specifically, our proposed model, *ThoR*, serves as the generator, while a spatial-temporal discriminator evaluates the realism of predicted precipitation fields. This setup adheres to the canonical Generative Adversarial Network (GAN) structure, wherein the generator synthesizes plausible precipitation patterns from meteorological inputs, and the discriminator learns to distinguish between observed and generated sequences.

To guide the training of the generator, we design a composite objective that integrates data fidelity, perceptual quality, and physical consistency. The overall generator loss is defined as a weighted sum of individual loss terms, each reflecting a different facet of the forecasting task:

$$\mathcal{L}_{\text{Generator}} = \alpha \mathcal{L}_{\text{velocity}} + \beta \mathcal{L}_{\text{physics}} + \gamma \mathcal{L}_{\text{NN}} + \delta \mathcal{L}_{\text{adv}}, \quad (18)$$

where  $\mathcal{L}_{\text{velocity}}$  force estimated motion fields aligns to a physics constraint, which is shown in Equation 8,  $\mathcal{L}_{\text{physics}}$  enforces adherence to the advection–diffusion constraint, mention in Equation 17,  $\mathcal{L}_{\text{NN}}$  aggregates multiple pixel-wise and feature-based losses, and  $\mathcal{L}_{\text{adv}}$  promotes perceptual realism through adversarial training.

The neural network loss  $\mathcal{L}_{\text{NN}}$  is defined as:

$$\mathcal{L}_{\text{NN}} = \lambda_1 \mathcal{L}_{L1} + \lambda_2 \mathcal{L}_{L2} + \lambda_3 \mathcal{L}_{\text{ICL}} + \lambda_4 \mathcal{L}_{\text{Huber}}, \quad (19)$$

where each term contributes uniquely:  $\mathcal{L}_{L1}$  and  $\mathcal{L}_{L2}$  are weighted  $L1$  and  $L2$  losses, ensuring pixel-level accuracy,  $\mathcal{L}_{\text{Huber}}$  provides robustness to outliers and  $\mathcal{L}_{\text{ICL}}$  promotes event-aware consistency across latent representations, focusing on high intensity features as storm events. Standard loss functions like MAE or MSE penalize incorrect predictions twice—once for misplacing storms and again for missing the correct location, leading models to prefer uncertain, smoothed predictions. This loss is calculated via cross-entropy on trinary masks defined as

$$\mathcal{L}_{\text{ICL}} = CE(M(\hat{x}), M(x)) \text{ with } M(x) = \begin{cases} 0 & \text{if } x \in [0, 4) \\ 1 & \text{if } x \in [4, 8) \quad \text{with } x \in R_{out} \text{ in mm/h.} \\ 2 & \text{if } x \geq 8 \end{cases} \quad (20)$$

The loss weights are empirically tuned to balance physical plausibility and numerical precision, with:

$$\begin{aligned} \alpha &= 0.1, & \beta &= 0.1, & \gamma &= 0.75, & \delta &= 0.05, \\ \lambda_1 &= 0.2, & \lambda_2 &= 0.2, & \lambda_3 &= 0.2, & \lambda_4 &= 0.4 \end{aligned}$$

This hybrid loss design enables *ThoR* to generate predictions that are physically consistent, numerically accurate, and perceptually plausible, promoting generalizability under diverse meteorological conditions.

## 4 Conclusion

Short-term forecasting of convective precipitation remains a persistent and critical challenge in atmospheric science, with profound implications for hazard mitigation and the management of weather-sensitive operations. This study introduces *ThoR*, a novel attention-driven, physics-informed deep learning architecture embedded within the TFC framework. The model approximates the forward solution of convective system dynamics as governed by partial differential equations (PDEs), enabling accurate simulation of storm evolution and intensification.

By leveraging high-resolution radar reflectivity and rainfall rate data from both the Nha Be radar station in Ho Chi Minh City and the U.S. Multi-Radar Multi-Sensor (MRMS) system, *ThoR* delivers rainfall predictions with lead times of up to two hours. Its probabilistic forecasting capabilities further support adaptive warning thresholds, improving decision-making and reducing false alarm rates in operational contexts. Importantly, *ThoR* relies solely on rainfall maps as input while incorporating physical constraints directly into its neural architecture. This fusion of data-driven learning and physics-based reasoning provides a principled approach to modeling the spatiotemporal dynamics of convective weather systems, yielding superior generalization across diverse meteorological regimes.

Comprehensive evaluations on both MRMS and Nha Be datasets demonstrate that *ThoR* consistently surpasses baseline models, achieving lower mean squared error (MSE), higher structural similarity index (SSIM), and enhanced critical success index (CSI), particularly for high-intensity precipitation events. These results highlight *ThoR*'s capability to accurately capture fine-scale rainfall structures and extreme weather phenomena across different climatic and geographic settings. Future work will focus on extending the forecasting horizon while preserving spatial and temporal precision, deploying the model in real-time operational environments—especially in tropical urban regions—and integrating uncertainty quantification and explainability techniques to enhance trust, transparency, and decision support in high-stakes applications.

## References

1. Bauer, P., Thorpe, A. & Brunet, G. The quiet revolution of numerical weather prediction. *Nature* **525**, 47–55 (2015).
2. Vietnam News. Deadly disasters in 2024 highlight the need for stronger prevention measures. <https://vietnamnews.vn/society/1690003/deadly-disasters-in-2024-highlight-the-need-for-stronger-prevention-measures.html> (2024). Accessed: 2025-04-30.
3. PC, S. *et al.* Comparison of rainfall nowcasting derived from the steps model and jma precipitation nowcasts. *Hydrol. Res. Lett.* **9**, 54–60 (2015).
4. Hwang, Y., Clark, A. J., Lakshmanan, V. & Koch, S. E. Improved nowcasts by blending extrapolation and model forecasts. *Weather. Forecast.* **30**, 1201–1217 (2015).
5. Sun, J. Convective-scale assimilation of radar data: progress and challenges. *Q. J. Royal Meteorol. Soc. A journal atmospheric sciences, applied meteorology physical oceanography* **131**, 3439–3463 (2005).
6. Buehner, M. & Jacques, D. Non-gaussian deterministic assimilation of radar-derived precipitation accumulations. *Mon. Weather. Rev.* **148**, 783–808 (2020).
7. Fleet, D. & Weiss, Y. Optical flow estimation. In *Handbook of mathematical models in computer vision*, 237–257 (Springer, 2006).
8. Pulkkinen, S. *et al.* Pysteps: An open-source python library for probabilistic precipitation nowcasting (v1. 0). *Geosci. Model. Dev.* **12**, 4185–4219 (2019).
9. Tian, L., Li, X., Ye, Y., Xie, P. & Li, Y. A generative adversarial gated recurrent unit model for precipitation nowcasting. *IEEE Geosci. Remote. Sens. Lett.* **17**, 601–605 (2019).
10. Chung, K.-S. & Yao, I.-A. Improving radar echo lagrangian extrapolation nowcasting by blending numerical model wind information: Statistical performance of 16 typhoon cases. *monthly weather review* **148**, 1099–1120 (2020).
11. Shi, X. *et al.* Convolutional lstm network: A machine learning approach for precipitation nowcasting. *Adv. neural information processing systems* **28** (2015).
12. Wang, Y., Long, M., Wang, J., Gao, Z. & Yu, P. S. Predrnn: Recurrent neural networks for predictive learning using spatiotemporal lstms. *Adv. neural information processing systems* **30** (2017).
13. Shi, X. *et al.* Deep learning for precipitation nowcasting: A benchmark and a new model. *Adv. neural information processing systems* **30** (2017).
14. Wang, Y. *et al.* Predrnn: A recurrent neural network for spatiotemporal predictive learning. *IEEE Transactions on Pattern Analysis Mach. Intell.* **45**, 2208–2225 (2022).
15. Keil, C. & Craig, G. C. A displacement and amplitude score employing an optical flow technique. *Weather. Forecast.* **24**, 1297–1308 (2009).
16. Ishizuka, Y. & Aiyoshi, E. Double penalty method for bilevel optimization problems. *Annals Oper. Res.* **34**, 73–88 (1992).
17. Ma, Z., Zhang, H. & Liu, J. Db-rnn: An rnn for precipitation nowcasting deblurring. *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* **17**, 5026–5041 (2024).
18. Ravuri, S. *et al.* Skilful precipitation nowcasting using deep generative models of radar. *Nature* **597**, 672–677 (2021).
19. An, S., Oh, T.-J., Kim, S.-W. & Jung, J. J. Self-clustered gan for precipitation nowcasting. *Sci. Reports* **14**, 9755 (2024).
20. Yu, W. *et al.* Integrating spatio-temporal and generative adversarial networks for enhanced nowcasting performance. *Remote. Sens.* **15**, 3720 (2023).
21. She, L., Zhang, C., Man, X., Luo, X. & Shao, J. A self-attention causal lstm model for precipitation nowcasting. In *2023 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, 470–473 (IEEE, 2023).
22. Rota Bulo, S., Neuhold, G. & Kortschieder, P. Loss max-pooling for semantic image segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2126–2135 (2017).
23. Hatsuzuka, D., Sato, T. & Higuchi, Y. Sharp rises in large-scale, long-duration precipitation extremes with higher temperatures over japan. *Npj Clim. Atmospheric Sci.* **4**, 29 (2021).
24. Roca, R. & Fiolleau, T. Extreme precipitation in the tropics is closely associated with long-lived convective systems. *Commun. Earth & Environ.* **1**, 18 (2020).
25. Ko, J. *et al.* Effective training strategies for deep-learning-based precipitation nowcasting and estimation. *Comput. & Geosci.* **161**, 105072 (2022).

26. Gao, Z. *et al.* Prediff: Precipitation nowcasting with latent diffusion models. *Adv. Neural Inf. Process. Syst.* **36**, 78621–78656 (2023).
27. Yu, D. *et al.* Diffcast: A unified framework via residual diffusion for precipitation nowcasting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 27758–27767 (2024).
28. An, S., Oh, T.-J., Sohn, E. & Kim, D. Deep learning for precipitation nowcasting: A survey from the perspective of time series forecasting. *Expert. Syst. with Appl.* **268**, 126301 (2025).
29. Leake, C. & Mortari, D. Deep theory of functional connections: A new method for estimating the solutions of partial differential equations. *Mach. learning knowledge extraction* **2**, 37–55 (2020).
30. Schiassi, E. *et al.* Extreme theory of functional connections: A fast physics-informed neural network method for solving ordinary and partial differential equations. *Neurocomputing* **457**, 334–356 (2021).
31. Zhang, Y. *et al.* Skilful nowcasting of extreme precipitation with nowcastnet. *Nature* **619**, 526–532 (2023).
32. Bieker, J. Mrms precipitation rate dataset. (2022).
33. Khang, T. G. *et al.* Thunderstorm nowcasting with recurrent-convolution deep learning: A case study in ho chi minh city. In *2024 18th International Conference on Advanced Computing and Analytics (ACOMPA)*, 42–49 (IEEE, 2024).
34. Huang, G.-B., Zhu, Q.-Y. & Siew, C.-K. Extreme learning machine: theory and applications. *Neurocomputing* **70**, 489–501 (2006).
35. Raissi, M., Perdikaris, P. & Karniadakis, G. E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. physics* **378**, 686–707 (2019).
36. Bonkile, M. P., Awasthi, A., Lakshmi, C., Mukundan, V. & Aswin, V. A systematic literature review of burgers' equation with recent advances. *Pramana* **90**, 1–21 (2018).
37. Zikic, D., Kamen, A. & Navab, N. Revisiting horn and schunck: Interpretation as gauss-newton optimisation. In *BMVC*, 1–12 (2010).
38. Horn, B. K. P. & Schunck, B. G. Determining optical flow. *Artif. Intell.* **17**, 185–203 (1981).
39. Crane, R. K. Space-time structure of rain rate fields. *J. Geophys. Res. Atmospheres* **95**, 2011–2020 (1990).
40. Ryu, S., Lyu, G., Do, Y. & Lee, G. Improved rainfall nowcasting using burgers' equation. *J. Hydrol.* **581**, 124140 (2020).
41. Ho, J., Kalchbrenner, N., Weissenborn, D. & Salimans, T. Axial attention in multidimensional transformers. *arXiv preprint arXiv:1912.12180* (2019).