Geophysical Research Letters®

RESEARCH LETTER
10.1029/2024GL113699

Enhancing Nowcasting With Multi-Resolution Inputs Using Deep Learning: Exploring Model Decision Mechanisms

Yuan Cao[1], **Lei Chen**[1,2,3], **Junjing Wu**[3], **and Jie Feng**[4]

[1]Shanghai Central Meteorological Observatory, Shanghai, China, [2]Key Laboratory of High Impact Weather (Special), China Meteorological Administration, Changsha, China, [3]Shanghai Typhoon Institute, China Meteorological Administration, Shanghai, China, [4]Department of Atmospheric and Oceanic Sciences Fudan University, Shanghai, China

Key Points:

- We present a nowcasting model that is scalable and flexible, enabling it to incorporate heterogeneous inputs
- With multi-source data and pre-training, performance enhancement is achieved in terms of both general statistics and representative event
- The interpretability method reveals how the model generates predictions in a physically meaningful manner

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

L. Chen,
qqydss@163.com

Citation:

Cao, Y., Chen, L., Wu, J., & Feng, J. (2025). Enhancing nowcasting with multi-resolution inputs using deep learning: Exploring model decision mechanisms. *Geophysical Research Letters*, *52*, e2024GL113699. https://doi.org/10.1029/2024GL113699

Received 19 NOV 2024
Accepted 20 JAN 2025

Abstract

Nowcasting methods based on deep learning typically rely solely on radar data. However, effectively leveraging multi-source data with diverse spatio-temporal resolutions remains a significant challenge in the field. To address this challenge, we propose and validate a novel deep learning model for nowcasting, termed Nowcastformer. This model utilizes radar data and upper-air atmospheric variables, and has been pretrained on satellite data from non-target regions. Quantitative statistical assessments demonstrate that both the integration of multi-source data and the implementation of pre-training strategies enhance the model's performance. Additionally, we conduct a comprehensive analysis of predictor importance, revealing a trend where atmospheric variables become increasingly important as the forecast horizon increases. To illustrate the model's interpretability, we employ the integrated gradients method, which highlights critical areas in representative cases and provides insights into the model's decision-making process.

Plain Language Summary

As a sophisticated monitoring tool, weather radar occupies a pivotal position in convective nowcasting. While numerous contemporary deep learning approaches predominantly concentrate on refining network architectures using radar reflectivity as the sole input, the impact of atmospheric physical information on nowcasting remains underexplored. To incorporate the contextual backdrop of atmospheric states in nowcasting, we devise a comprehensive deep learning framework that integrates atmospheric variables across multiple levels. To enhance generalization, we employ a transfer learning strategy to extract generalized spatialtemporal features. Rather than emphasizing a specific network design, we underscore the advantages of harnessing multi-source data and the decision mechanism of the model. By fusing atmospheric variables and radar reflectivity, and adopting a pre-training and fine-tuning approach, we achieve more reliable and resilient nowcasting. Overall, our successful implementation of transfer learning within this multi-modal model offers promising insights for advancing the field of nowcasting.

1. Introduction

Nowcasting involves predicting immediate future events, ranging from a few minutes to several hours. It holds significant importance, particularly in preventing damage from severe precipitation over a short period. (Prudden et al., 2020). Despite its critical importance, nowcasting has long been a challenging issue due to the complexity and chaotic nature of the atmosphere (Ravuri et al., 2021). In recent years, the meteorological field has witnessed a remarkable surge in the application of deep learning (DL) models, primarily due to their rapid advancement. Specifically, these models have garnered significant attention in tasks encompassing lightning nowcasting (Leinonen et al., 2022; Lin et al., 2019; Zhou et al., 2020), rainfall estimation (H. Chen et al., 2019), and precipitation forecasts (Geng et al., 2022; Kaae Sønderby et al., 2020; Weyn et al., 2020). Among the various DL approaches, convolutional neural networks (CNNs) have been extensively utilized for spatial modeling in computer vision and geoscientific domains, while recurrent neural networks (RNNs) excel in handling time series data sets by recursively feeding their outputs as subsequent inputs. Shi et al. (2015) pioneered the integration of Convolutional Long Short-Term Memory (ConvLSTM) to combine the strengths of CNNs and RNNs for precipitation nowcasting in Hong Kong. This innovation entailed substituting convolutional operations for the fully connected operations in the basic LSTM architecture. Later, Shi et al. (2017) further enhanced this model by introducing optical flow into the Trajectory GRU (TrajGRU), enabling it to capture and process a more comprehensive range of feature information. Meanwhile, Wang et al. (2017) designed the Predictive Recurrent

Neural Network (PredRNN), which facilitates information transfer between adjacent layers in a zigzag fashion. The MotionRNN (Wu et al., 2021) model, on the other hand, improves performance by learning instantaneous changes and jointly accumulating motion patterns. In our prior research (L. Chen et al., 2020), we introduced a novel ConvLSTM architecture with a star-shaped bridge design and employed a specialized multi-sigmoid loss function, which we consider a differentiable critical success index (CSI). Some studies have predominantly employed Convolutional Neural Networks (CNNs) for nowcasting. Among these, U-Net (Ronneberger et al., 2015) exemplifies a seminal network architecture characterized by a pyramidal shape, featuring a symmetric pathway that comprises contracting and expanding paths of feature maps, bridged by skip connections. Given its paradigm to treat the prediction problem as an image-to-image translation task, the U-Net based model has gained widespread adoption in recent nowcasting studiess (Agrawal et al., 2019; Ayzel et al., 2020; Kim & Hong, 2021; Ko et al., 2022), owing to its simpler form compared to RNN-fused models and its effective multi-scale processing architecture.

However, the aforementioned research, which primarily relies on radar reflectivity or satellite images, still faces limitations in accurately forecasting convective storms. From an information theory standpoint, measurements from different modalities often offer complementary insights into various aspects of a specific objective entity. In atmospheric dynamics, the physical evolution is a complex interplay of multiple variables (Karpachev & Gasilov, 2001). Consequently, data-driven models that integrate information from multi-source data have the potential to yield more accurate inferences (Baltrušaitis et al., 2018). With the rapid increase in the availability of observations rich in spatial and temporal structures, manually extracting useful information becomes impracticable (Hou et al., 2014). In contrast, nowcasting techniques leveraging DL with inputs from diverse data sources, known as multi-modal learning (Baltrušaitis et al., 2018), can autonomously organize and learn abstract representations of the interactions between multiple inputs. These characteristics render DL based multi-modal learning an especially attractive approach for enhancing the performance of nowcasting systems (Reichstein et al., 2019; Rothfusz et al., 2018; Weyn et al., 2020).

Significant progress has been achieved in generating gridded nowcasts from diverse input sources. For instance, Zhou et al. (2020) successfully integrated meteorological satellite observations, weather radar network data, and lightning location system information for lightning nowcasting. Pan et al. (2021) further explored the microphysics of convective storms by merging polarimetric radar variables, such as ZDR and KDP. Leinonen et al. (2022) introduced a deep learning model capable of utilizing multiple data sources and adapting to different hazard types. These models have shown relative success in nowcasting tasks, but they require all types of data to be spatio-temporally aligned. Due to the design of the model structure, they lack the ability to handle unaligned heterogeneous data, which is a common scenario in meteorological operations due to differences in data-generating infrastructure.

Furthermore, another issue with DL lies in its inherent opaque nature raises concerns regarding the explainability of the prediction processes. This lack of interpretability poses major challenges. First, it reduces trust from domain experts, such as meteorologists, who may be reluctant to rely on unexplained model outputs for high-stakes decision making. Second, it hinders further model refinement, as developers cannot easily diagnose errors or identify which relationships the models have captured. To address these limitations, explainable machine learning techniques have become essential to enhance trust in predictions, facilitate further model improvements, and uncover new meteorological insights. Drawing inspiration from the initial progress made in applying explainability techniques in weather and climate prediction applications (Deng et al., 2021; L. Yuan et al., 2022), our research introduces post-hoc explanations for trained models, aiming to provide insight into the fundamental atmospheric processes that lead to their predictions.

Against this backdrop of diverse spatio-temporal resolution inputs, this study aims to generalize the paradigm of multi-modal learning specifically tailored for next-two-hour nowcasting. Interpretable visualization methods are employed to facilitate the examination of the model's thermo-dynamic decision mechanisms. Meanwhile, we fully leverage the high flexibility offered by the Vision-Transformer (VIT) architecture (Liu et al., 2023; Lu et al., 2022; Yuan & Lin, 2020), which provides capabilities for pre-training and transfer learning (Andrychowicz et al., 2023; Nguyen et al., 2023). By incorporating satellite imagery located in the different region, our objective is to enhance the model's performance in nowcasting by learning from a more comprehensive array of meteorological information.

2. Data Sources

2.1. Satellite Data

In our pre-training efforts, we utilize satellite imagery from the 2023 edition of the Weather4cast competition (Gruca et al., 2022), administered and operated by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT). The data consists of 11 bands, including two visible (VIS) bands, two water vapor (WV) bands, and seven infrared (IR) bands. The WV and infrared data are scaled to the range [0, 1] by means of min-max normalization, with the minimum value being 170 K and the maximum value being 310 K. The training data set comprises 20,308 samples, the test data set 60 samples, and the validation data set, accounting for 20% of the training data set.

2.2. Radar Data

We primarily utilize the composite reflectivity variable due to its importance in nowcasting applications. The raw reflectivity is generated through volume-sampling at nine different elevation angles (VCP21 mode), with each volume-sampling occurring approximately every six minutes. To mitigate issues such as abnormal propagation, ground clutter, and particle clutter, the raw radar data undergo quality control procedures (Huang et al., 2018) before being interpolated to longitude and latitude coordinates, achieving a spatial resolution of 0.01° × 0.01°. Further preprocessing steps include clipping the composite reflectivity to the range [15, 70] and rescaling the data to the range [0, 1] using min-max normalization. This normalization step is crucial for ensuring that the data are suitable for input into machine learning models. After eliminating discontinuous frames resulting from radar malfunctions or maintenance, as well as excluding clear-sky images, we arrive at a data set of 34,745 sequential samples, each with a length of 15 frames. For training purposes, we utilize radar images from 2019 to 2022, reserving 20% of this training set for validation. Meanwhile, images from 2023 serve as our test set. Additionally, to facilitate the blending process in pySTEPS, the regional NWP rainfall data was converted to radar data using the local Z-R relationship, with coefficients referenced from (Wen et al., 2016). The NWP data was then regridded to match the radar data grid using the Kriging method. Some settings of our regional NWP can be found in Table S6 in Supporting Information S1.

2.3. Atmospheric Variables

To supply background physical information, we utilize ERA5 upper-air reanalysis data managed by the European Center for Medium-Range Weather Forecasting (ECMWF). These hourly variables encompass temperature (t), horizontal wind patterns (u,v), specific humidity (sh), and geopotential height (gh) measurements across four distinct pressure levels of 500, 700, 850, 925 hPa and have the horizontal resolution of 0.25° × 0.25°. All the atmospheric variables are scaled using z-score normalization.

2.4. Study Area

This study selected two regions (see Figure S1 in Supporting Information S1). Specifically, the outer region is roughly located in the East China region, while the inner area covers the Yangtze River Delta region around Shanghai, China. This inner area is located within the subtropical monsoon climate zone, where two major meteorological disasters prevail: hailstorms triggered by intense convective activity in spring, and typhoons during summer and autumn. Among them, the outer region serves as the area for atmospheric variable inputs, providing background information for nowcasting; while the inner region serves as the input and prediction area for radar reflectivity. Details of data and the relevant parameters (mean and standard deviation) can be found in Table S1–S3 in Supporting Information S1.

3. Methodology

3.1. Nowcastformer

Our goal is to establish a model that can utilize multi-modal heterogeneous data for pre-training and investigate its generalization ability under different regions and input conditions within the framework of transfer learning. The

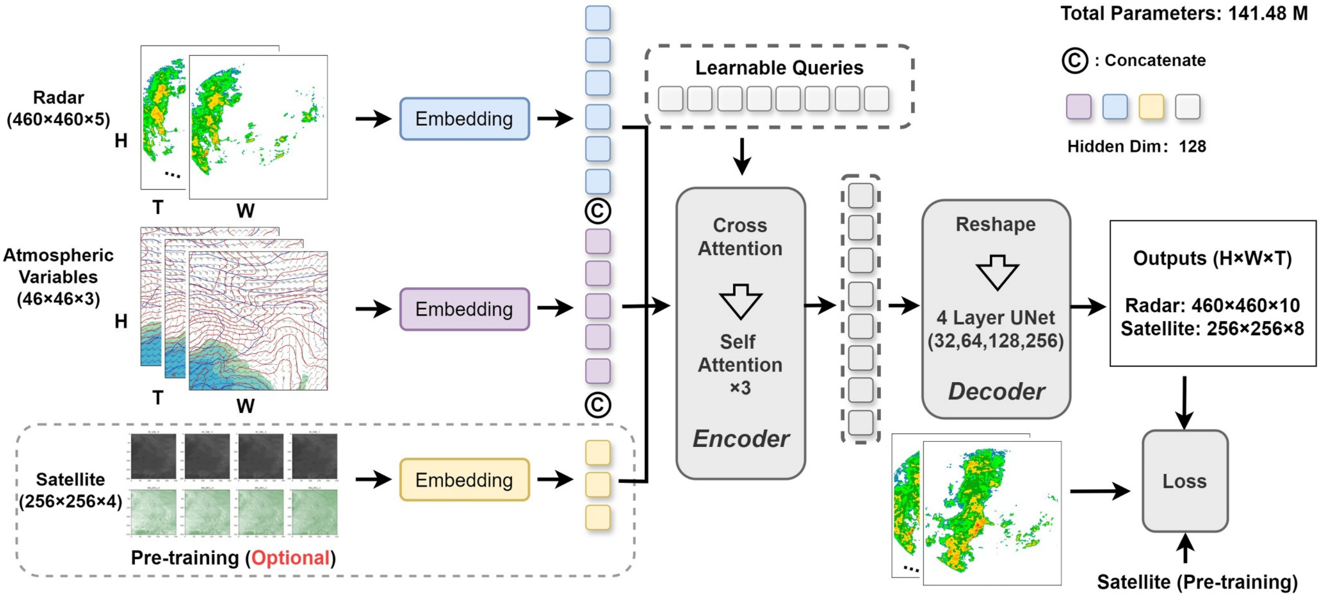

Figure 1. The main pipeline and schematic depiction of Nowcastingformer encapsulate three essential components: embedding, encoder, and decoder. Diverse inputs are individually embedded through distinct branches and subsequently fused by a cross-attention block, resulting in compressed tensor representations. These tensors undergo refinement through multiple stacked self-attention blocks and are ultimately fed into the decoder to generate predictions.

architecture was built upon VIT with three major components, namely variable embedding, encoder and decoder (Figure 1). Due to the heterogeneity of multi-modal input data, embedding operations are initially carried out separately for each input type. Subsequently, an encoding block is employed to derive a concise representation from the variable tokens generated by the embedding blocks. The decoder then receives the encoder's outputs and utilizes a U-Net to generate the final prediction. Detailed hyperparameter configurations and input/hidden tensor/output shapes are provided in Table S4 in Supporting Information S1. The forward propagation process and encoder-decoder architecture are detailed in Text S1, and Figure S2 in Supporting Information S1.

In the pre-training phase, we utilize the basic Mean Squared Error (MSE,Equation 1) as the training loss function. During the fine-tuning phase, due to the significantly uneven distribution of radar reflectivity intensity, we adopt the balanced MSE(BMSE,Equation 2) the training loss, with the weights being determined through a simple search. The fine-tuning process consists of two steps: Partial fine-tuning and full fine-tuning (see Text S2 in Supporting Information S1 for details).

$$\mathrm{MSE} = \frac{\sum_{t \in T_t} \sum_{i=1}^{N} (F_{t,i} - O_{t,i})^2}{N}, \tag{1}$$

$$\mathrm{BMSE} = \frac{\sum_{t \in T_t} \sum_{i=1}^{N} w(O_{t,i})\,(F_{t,i} - O_{t,i})^2}{N}, \tag{2}$$

$$w(O_{t,i}) = \begin{cases} 1, & O_{t,i} < 25 \\ 10, & 25 \le O_{t,i} < 35 \\ 30, & 35 \le O_{t,i} < 45 \\ 50, & 45 \le O_{t,i} < 55 \\ 80, & O_{t,i} \ge 55 \end{cases} \tag{3}$$

3.2. Interpretability Methods

To enhance our understanding of the model's decision-making mechanisms and analyze how input features impact the overall performance of the model, this study introduces two interpretability methods: Permutation Method (*PM*)(Rasp & Lerch, 2018) and Integrated Gradients (*IG*)(Sundararajan et al., 2017). These approaches facilitate a deeper comprehension of the model at both the statistical and individual case levels.

The *PM* (Equation 4) is used to determine the relative importance of each predictor. The contribution of a specific predictor is measured by temporarily shuffling its values while keeping other predictors unchanged. Then, the performance decrease due to this permutation is examined. Permutation importance has two variations: single-pass and multi-pass. Since the single-pass *PM* has limitations in recognizing information redundancy among predictors, this study opts for the multi-pass version, whose algorithm is detailed in Text S3 of Supporting Information S1.

IG(Equation 5) employs backpropagation to calculate gradients and integrates them along a defined path from a baseline (often inputs filled with zeros or averages) to the actual input data. Through this approach, the *IG* algorithm comprehensively evaluates the significance of each spatial location within a specific input sample to the model's output.

$$PM(k) = \frac{M(base) - M(k)}{M(k)} \tag{4}$$

$$\begin{cases} IG(x_i) = \left(x_i - x_i^{'}\right) \times \int_{\alpha=0}^{1} \frac{\partial f\left(x_i^{\alpha}\right)}{\partial x_i^{\alpha}} d\alpha, \\ x_i^{\alpha} = x_i^{'} + \alpha\left(x_i - x_i^{'}\right) \end{cases} \tag{5}$$

4. Results and Discussion

4.1. Overall Performance Evaluation

To verify the effectiveness of multi-source data inputs and pre-training scheme, we conducted performance comparisons between various models, including PySteps (Pulkkinen et al., 2019), LDCast (Leinonen et al., 2023), PredRNN, U-Net, and Nowcastformer (with and without the use of multi-source data and pre-training.). We report CSI, POD, and Fractional Skill Score (FSS) on the test set in Figure 2. Additional continuous metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) were introduced to evaluate reflectivity intensity. For calculation formulas of these metrics, see Text S4 in Supporting Information S1. The pySTEPS blending is distinguished by an augmented Lagrangian extrapolation framework, which incoporated regional NWP precipitation forecasts. Other models are trainable architectures.

Here, Vanilla denotes Nowcastformer trained from scratch using only radar reflectivity. It serves as our baseline architecture to investigate the impacts of pre-training and multi-source inputs. Its fundamental performance is crucial, and therefore, we compared it with pySTEPS blending and three unimodal deep learning models. In terms of the FSS20 and FSS30 metrics, pySTEPS blending outperformed PredRNN for the one hour forecast, and was on par with PredRNN regarding the CSI20 metric. However, it did not demonstrate competitiveness when compared to multi-modal deep learning methods. The performance of pySTEPS blending is influenced by multiple factors. The NWP bias, stemming from the direct resolution of convective processes without convective parameterization in the 3 km resolution and the insufficient spin-up time, along with the intermediate handcrafted feature engineering, all contribute to the performance gap. Vanilla, which augments the U-Net with a transformer structure as the encoder, demonstrates improvements over U-Net in terms of weak reflectivity (CSI20) and moderate-intensity reflectivity (CSI30). This enhancement may be attributed to the transformer's ability to capture global contextual information and establish long-range dependencies. However, The performance advantage of Vanilla disappears for strong reflectivity (CSI40), potentially due to CNNs inherently possessing more suitable inductive biases for gridded data, while transformers, lacking structural awareness, may require significantly more data or robust data augmentation (Arkin et al., 2023). Additionally, we evaluated the FSS metric with a radius of 5 km, which can be considered as a softened version of CSI, accommodating a certain degree of spatial

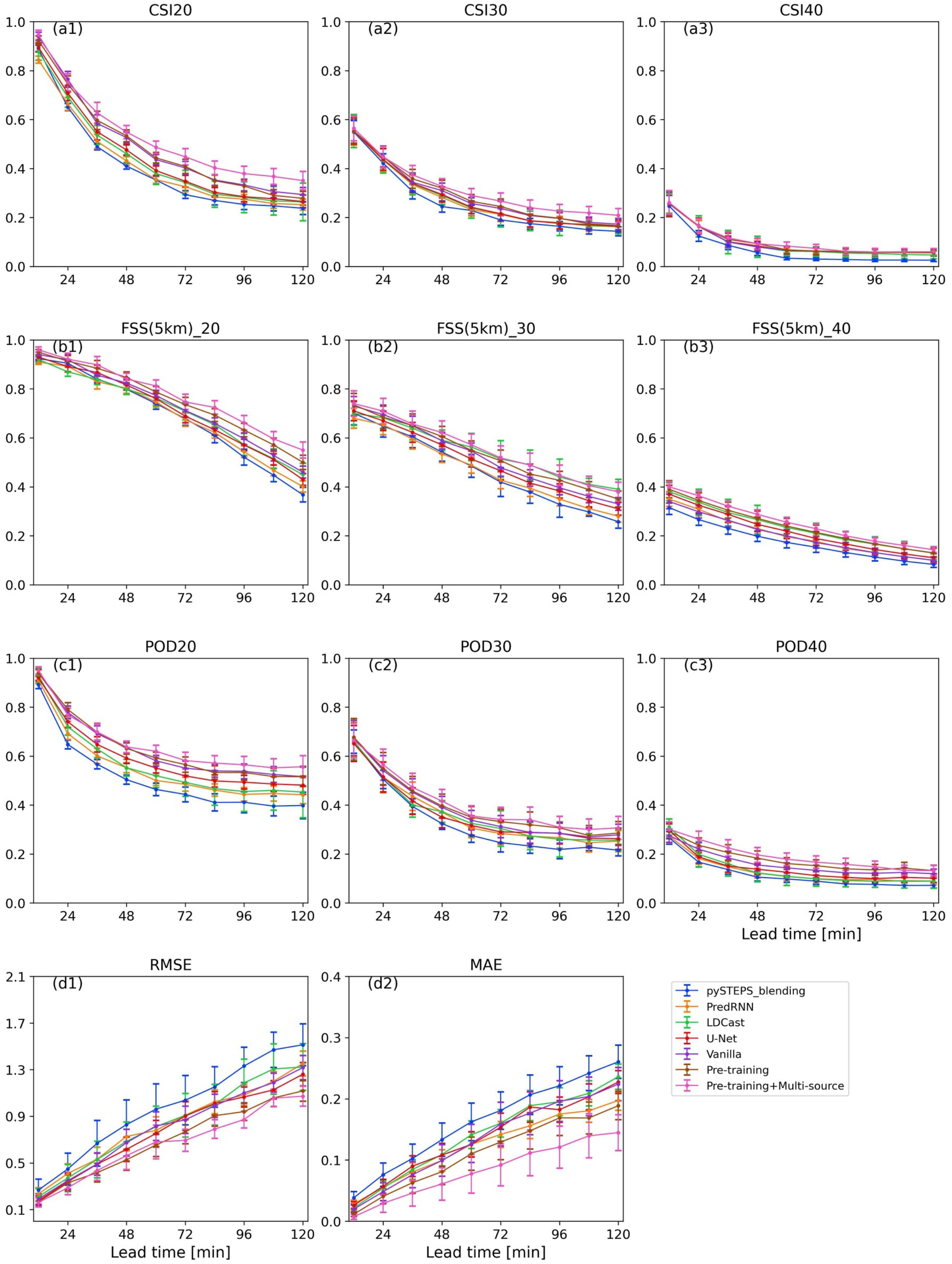

Figure 2. Quantitative assessment of the performance of various models for two-hour nowcasting with thresholds of 20, 30, and 40 dBZ, depicted as functions of lead time, with error bars representing the standard deviation (std). Panels (a1–a3) display the critical success index with corresponding error bars. Panels (b1–b3) present the Fractional Skill Score with a radius of 5 km, also accompanied by error bars. Panels (c1–c3) showcase the Probability of Detection with std error bars. Panels (d1–d2) are the Root Mean Squared Error and Mean Absolute Error, respectively, both including error bars to indicate the standard deviation.

displacement in prediction. For FSS20, Vanilla outperformed all unimodal models. Nevertheless, for FSS30 and FSS40, LDCast, a generative model, demonstrated a more pronounced advantage. This superiority could be attributed to LDCast's utilization of direct sampling, which mitigates blurring effect, whereas FSS do not penalize minor pixel-level inconsistencies. However, for FSS40, the performance of LDCast did not surpass that of the multimodal model, indicating that, given the scarcity of strong reflectivity training samples, relying solely on model architecture without additional input information is insufficient to enhance the prediction performance of strong reflectivity. For the POD metric, similar to the CSI metric, the Vanilla model achieved the best performance among the unimodal models. In terms of the RMSE metric, the Vanilla model was comparable to the PredRNN model; whereas for the MAE metric, the Vanilla model was on par with the U-Net model. Taken together, these results indicate that our architecture, with comparable performance to other traditional DL models, is adequate as a starting point for the research focus of this study.

Building upon the Vanilla model, we further investigated the benefits of pre-training and multimodal input using the CSI with thresholds of 30 and 40 dBZ as examples. For predictions within 24 min, the performance of Nowcastformer's three configurations is indistinguishable. An intuitive explanation is that for forecasts very close to the initial time, relying solely on input radar reflectivity provides sufficient information. However, as the forecast time extends, the effects of pre-training and multi-modal information begin to emerge. The improvement from pre-training is mainly observed from 24 to 84 min for medium-intensity reflectivity and from 24 to 60 min for strong reflectivity. On the basis of pre-training, further introduction of atmospheric variables mainly enhances the performance from 24 to 120 min for medium-intensity reflectivity and from 36 to 84 min for strong reflectivity. It is noteworthy that pre-training alone lead to moderate improvement. This may be attributed to differences in regions and sensors between the pre-training data and the nowcasting downstream task data. Nevertheless, the model still extracts transferable information between the two data sets, which encourages us to subsequently seek data with greater similarity (such as satellite or radar data from adjacent regions) to more effectively enhance the model's generalization performance. Atmospheric context information provides additional thermodynamic and dynamic information, inherently reducing the uncertainty between input and output, thus offering greater potential for improvement. Notably, for strong reflectivity, predictions beyond 84 min do not show significant improvement even with the inclusion of atmospheric information. This may be because strong reflectivity samples are too scarce and often associated with small-scale convective phenomena, requiring more detailed thermodynamic and dynamic information. This motivates us to refine the training process (such as improving the loss function or sample distribution) and introduce high-resolution data (such as outputs from regional rapid update assimilation systems). For the metrics of POD, FSS, MAE, and RMSE, we also obtained the optimal values in the end. Overall, the performance consistently improved as we incrementally incorporated the pre-training strategy and atmospheric variables. In consideration of operational settings, we further conducted fine-tuning using limited archived NWP prediction data (Zhang et al., 2021). It was observed that the performance remained stable. This may be due to the continuous distribution of atmospheric variables, which exhibits good predictability, allowing the model pre-trained on ERA5 to adapt well to NWP inputs. This supplementary experiment can also be viewed as assimilating direct observational data into numerical models, demonstrating the initial potential of Deep Learning in data assimilation, albeit further extensive validation and exploration are required. The detailed findings of our regional NWP can be found in Figure S3 in Supporting Information S1.

4.2. Feature Importance Analyses

Figure 3 displays the importance ranking using the permutation method for two lead times of 60 and 120 min (the result including radar data is presented in Figure S4 in Supporting Information S1). The selection of the 30 dBZ threshold was based on Empirical evidence from extensive field observations in the Yangtze River Delta region. Specifically, most of severe weather events in this area were characterized by radar echoes reaching or exceeding 30 dBZ. This threshold also enables us to issue warnings with a relatively long lead time, averaging around 40 min, which is critical for effective disaster prevention and mitigation. At a lead time of 1 hr, even when all atmospheric variables are shuffled, the model maintains 85% of its performance, indicating that the 1-hr forecast skills primarily derive from the radar reflectivity inputs. For the 2-hr lead time, the importance scores of atmospheric variables become notably higher, suggesting that forecast skills are associated with atmospheric backgrounds at longer lead times. Specifically, *u* and *v* provide the most benefit for 2-hr

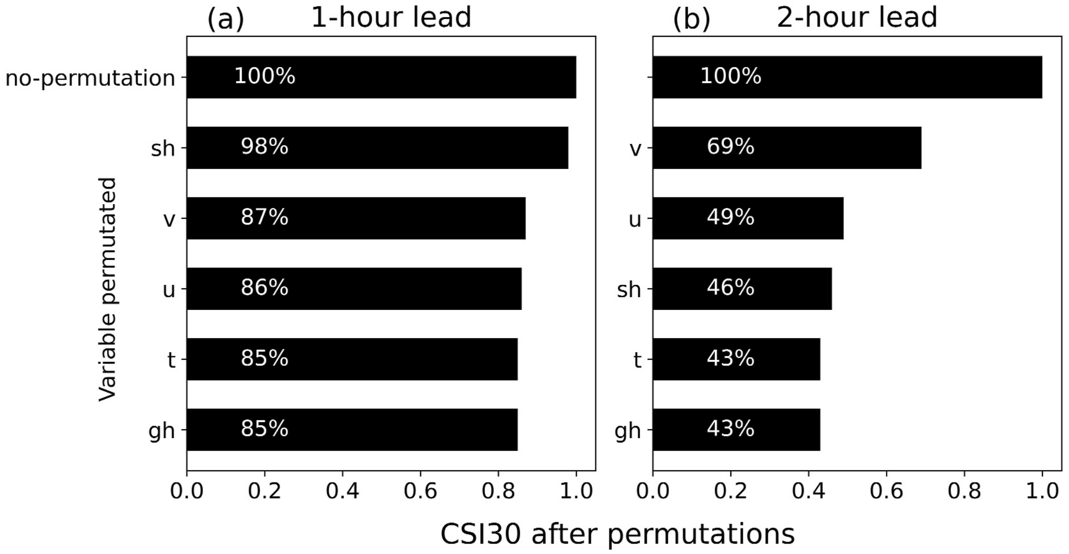

Figure 3. Bar chart summarizing the distributions of importance metrics for atmospheric variables at lead times of 60 and 120 min, with results from the permutation experiment normalized by the no-permutation metric.

forecast, while gh and *t* have minimal impacts on both 1-hr and 2-hr forecasts. The importance of *u* and *v* in improving the 2-hr forecast can be attributed to their direct impact on the movement and development of convection systems. Precipitation particles move with the wind, leading to changes in the location of radar reflectivity corresponding to these movements. Additionally, horizontal convergence and divergence influence air motion, causing upward or downward movement that can enhance or weaken cloud development. Vertical wind shear, by separating ascending and descending air currents, supports the maintenance and intensification of storms.

4.3. Representative Cases

An example of thunderstorms occurred on 15 April 2023, is presented in Figure 4 (an additional case is described in Text S5, Figure S5 and S6 in Supporting Information S1). The storm was characterized by rapid movement toward the east-southeast and significant deformation. Ground truth indicates that Area A evolved into a linear convective system. Given that there were no significant differences among models for forecasts within 60 min, we focused on the 96-min predictions. When using only reflectivity data, The pySTEPS blending fails to predict the deformation of the storm and significantly overestimated the overall storm intensity, forecasting most storm areas to exceed 40 dBZ at the 96-min mark. Similarly, the Vanilla model, LDCast, PredRNN, and U-Net also struggled with predicting the storm's shape characteristics when relying solely on reflectivity data. Specifically, with the exception of LDCast, these models exhibited weak intensity predictions, and notably, the Vanilla model tended to overestimate the storm's extent, leading to substantial false alarms. PredRNN, and U-Net, on the other hand, underestimated the storm's range, decaying more rapidly and resulting in higher miss rates. Notably, although LDCast maintained a better estimation of the storm's extent and preserved some high-frequency features, it also produced numerous spurious reflectivity values beyond the main echo region. When incorporating a pre-training strategy, there was some improvement in shape prediction, with linear features emerging, and the reflectivity evolution in Area B was partially predicted. However, the overall prediction remained notably underestimated. Experiments utilizing both pre-training and multi-source data successfully predicted conditions in Areas A and B. Our successful predictions in Areas A and B can be attributed to the collective consequence of both architecture design and the diverse training data corpus. However, it was observed that as time progressed, our model produced blurry nowcasting images, potentially excluding small-scale weather patterns. This is a common issue in data-driven deep learning models for video prediction.

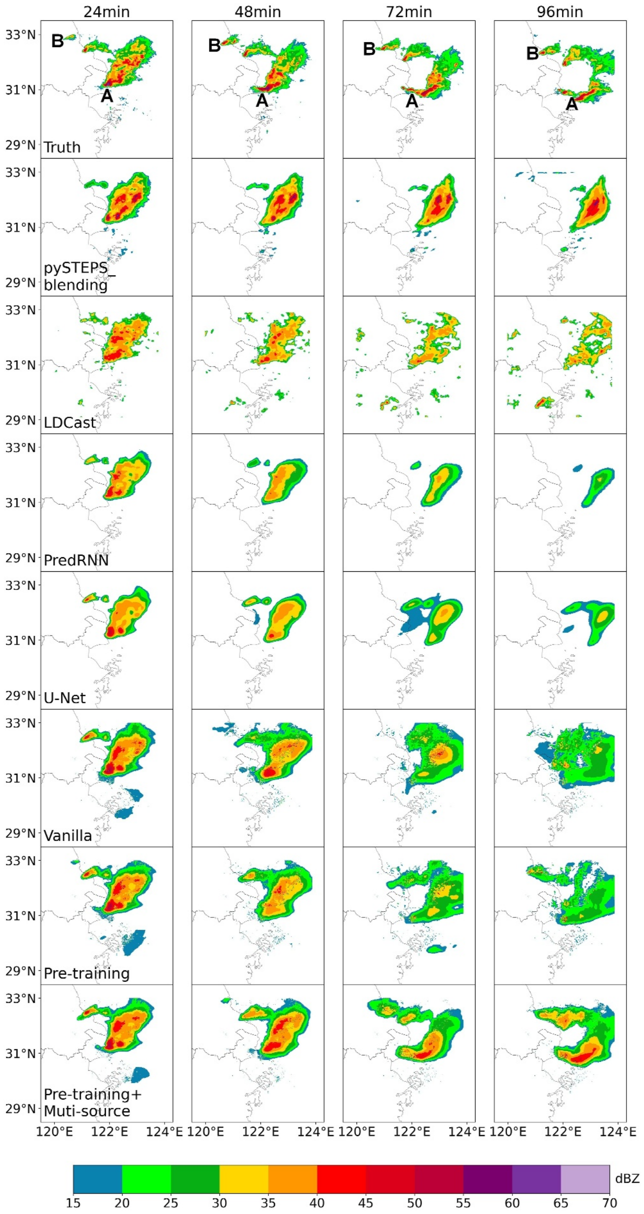

Figure 4. A comparison between models is shown at four future time steps: the first row presents the observation, the second row displays nowcasting from pySTEPS_blending, nowcasting from the DL model, utilizing only radar reflectivity shown from the third row to the sixth row, and the seventh and eighth rows correspond to nowcasting from experiments with pre-training and pre-training + multi-source, respectively. Area A and B are investigated areas.

The *IG* maps for predicting Area A are presented in Figure 5. Given the context of a cold front and upper-level trough, our analysis focuses solely on temperature at the 850 hPa and winds at the 500 hPa. The patterns observed in these heatmaps are meteorologically meaningful. Specifically, examining the area enclosed by the rectangle in the left panel of Figure 5, there is evident activation within the low-temperature trough, where the lifting effect of cold air intrusion at mid-to-low levels favors the development and maintenance of storms. Furthermore, examining the area denoted by the arrows and mark "+" in the right panel of Figure 5, there is notable activation within the positive vorticity region, with positive vorticity advection downstream in its wind field. According to the geopotential height tendency equation, positive vorticity advection promotes ascending motion at lower levels.

5. Summary and Future Work

In this paper, we have presented a deep learning model, Nowcastformer, capable of utilizing multi-modal data with diverse spatio-temporal resolutions for nowcasting tasks. This architecture is characterized by its flexibility in handling spatio-temporally unaligned heterogeneous data. This attribute stems from the fundamental design principles that underpin its ability to encode diverse data types, exemplified by image pixels, into token representations analogous to those in natural language. Consequently, the model transcends task-specificity, thereby accommodating a wide array of input features. We use satellite data sets from the Weather4cast competition for pre-training the model. The encoder allows our model to accommodate data from various sensors, enabling us to utilize satellite data for pre-training even when the target task involves radar data. Despite originating from different sensors and observation principles, both satellite and radar data capture information about rain clouds, thereby sharing relevant information. When direct augmentation of the task-specific training data set is not feasible, we adopt pre-training followed by task transfer as an alternative approach. This strategy effectively enhances the information capacity of the task-specific training samples by leveraging data from other regions, thus improving our model's performance in terms of both general statistics and representative events.

The predictor importance analysis revealed that in Supporting Information S1 on atmospheric variables did not significantly enhance performance within 1-hr leading time. However, as the lead time extended to 2 hr, the contribution from atmospheric variables, especially horizontal winds, rise rapidly. The representative nowcasting cases demonstrated that the model's decision mechanism is consistent with prior meteorological principles. From this perspective, we reproduce the thermal-dynamical relationships of storm development and associated atmospheric variables to some extent.

Despite the innovations in investigating the effectiveness of pre-training and multi-source data for enhancing nowcasting, the continued exploration is necessary. Future research directions include investigating more advanced loss functions, leveraging high spatio-temporal resolution atmospheric data from rapid-update NWP models, incorporating physical priors to guide model design, and exploring the potential of ensemble methods. Ultimately, we believe that improving the performance of DL-based nowcasting models involves numerous factors, warranting further scientific and engineering efforts.

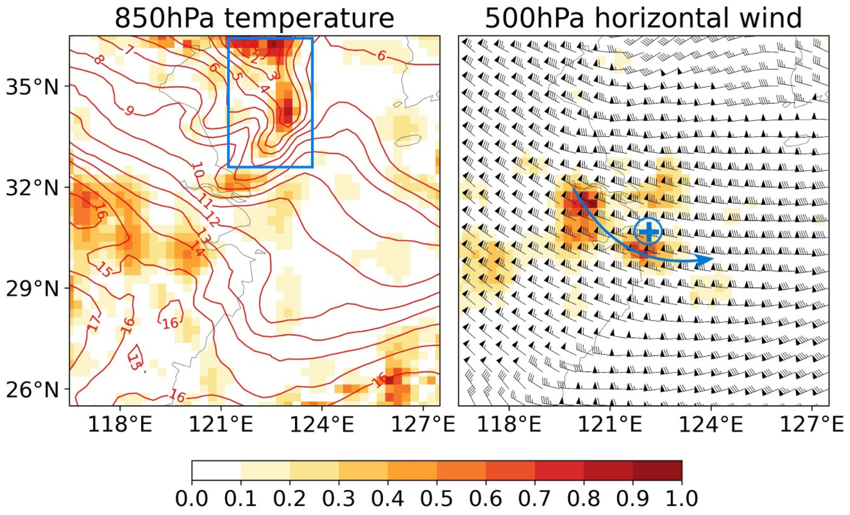

Figure 5. Heatmaps obtained using *IG*. The heatmap of the temperature (°C) at 850 hPa level can be found in left with the trough area enclosed by a blue rectangle. The heatmap illustrating the horizontal winds at the 500 hPa level is presented in right, where blue arrows and "+" Marks indicate positive curvature.

Data Availability Statement

The radar dataset is available on the data repository Zenodo (L. Chen, 2024). The ERA5 data are available at Copernicus Climate Data Store (Hersbach et al., 2023). The satellite data utilized in the pre-training phase can be accessed via SFTP at the following address: ftp://w4c@ala.boku.ac.at/. The access phrase required is "Weather4cast23!" (including the exclamation mark). For further details on data access and usage guidelines, kindly visit the official website: https://weather4cast.net/neurips2024/.

Acknowledgments
This work was jointly supported by the National Natural Science Foundation of China(Grant 42105001) and Open Project of Key Laboratory of High Impact Weather(special), China Meteorological Administration.

References

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

Andrychowicz, M., Espeholt, L., Li, D., Merchant, S., Merose, A., Zyda, F., et al. (2023). Deep learning for day forecasts from sparse observations. *arXiv preprint arXiv:2306.06079*.

Arkin, E., Yadikar, N., Xu, X., Aysa, A., & Ubul, K. (2023). A survey: Object detection methods from CNN to transformer. *Multimedia Tools and Applications*, *82*(14), 21353–21383. https://doi.org/10.1007/s11042-022-13801-3

Ayzel, G., Scheffer, T., & Heistermann, M. (2020). Rainnet v1. 0: A convolutional neural network for radar-based precipitation nowcasting. *Geoscientific Model Development*, *13*(6), 2631–2644. https://doi.org/10.5194/gmd-13-2631-2020

Baltrušaitis, T., Ahuja, C., & Morency, L.-P. (2018). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(2), 423–443. https://doi.org/10.1109/tpami.2018.2798607

Chen, H., Chandrasekar, V., Tan, H., & Cifelli, R. (2019). Rainfall estimation from ground radar and TRMM precipitation radar using hybrid deep neural networks. *Geophysical Research Letters*, *46*(17–18), 10669–10678. https://doi.org/10.1029/2019gl084771

Chen, L. (2024). Radar reflectivity [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.12749010

Chen, L., Cao, Y., Ma, L., & Zhang, J. (2020). A deep learning-based methodology for precipitation nowcasting with radar. *Earth and Space Science*, *7*(2), e2019EA000812. https://doi.org/10.1029/2019ea000812

Deng, H., Zou, N., Du, M., Chen, W., Feng, G., & Hu, X. (2021). A unified Taylor framework for revisiting attribution methods. *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 35(13), 11462–11469). https://doi.org/10.1609/aaai.v35i13.17365

Geng, L., Geng, H., Min, J., Zhuang, X., & Zheng, Y. (2022). Af-srnet: Quantitative precipitation forecasting model based on attention fusion mechanism and residual spatiotemporal feature extraction. *Remote Sensing*, *14*(20), 5106. https://doi.org/10.3390/rs14205106

Gruca, A., Serva, F., Lliso, L., Rípodas, P., Calbet, X., Herruzo, P., et al. (2022). Weather4cast at neurips 2022: Super-resolution rain movie prediction under spatio-temporal shifts. In M. Ciccone, G. Stolovitzky, & J. Albrecht (Eds.), *Proceedings of the neurips 2022 competitions track* (Vol. 220, pp. 292–313). PMLR. Retrieved from https://proceedings.mlr.press/v220/gruca22a.html

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., & Muñoz Sabater, J., (2023). Era5 hourly data on single levels from 1940 to present. 2023 [Dataset]. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*. https://doi.org/10.24381/cds.adbb2d47

Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., et al. (2014). The global precipitation measurement mission. *Bulletin of the American Meteorological Society*, *95*(5), 701–722. https://doi.org/10.1175/bams-d-13-00164.1

Huang, H., Zhao, K., Zhang, G., Lin, Q., Wen, L., Chen, G., et al. (2018). Quantitative precipitation estimation with operational polarimetric radar measurements in southern China: A differential phase–based variational approach. *Journal of Atmospheric and Oceanic Technology*, *35*(6), 1253–1271. https://doi.org/10.1175/jtech-d-17-0142.1

Kaae Sønderby, C., Espeholt, L., Heek, J., Dehghani, M., Oliver, A., Salimans, T., et al. (2020). Metnet: A neural weather model for precipitation forecasting. *arXiv e-prints, arXiv–2003*.

Karpachev, A., & Gasilov, N. (2001). Zonal and meridional wind components derived from intercosmos-19 hmf2 measurements. *Advances in Space Research*, *27*(6–7), 1245–1252. https://doi.org/10.1016/s0273-1177(01)00170-3

Kim, Y., & Hong, S. (2021). Very short-term rainfall prediction using ground radar observations and conditional generative adversarial networks. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1–8. https://doi.org/10.1109/tgrs.2021.3108812

Ko, J., Lee, K., Hwang, H., Oh, S.-G., Son, S.-W., & Shin, K. (2022). Effective training strategies for deep-learning-based precipitation nowcasting and estimation. *Computers and Geosciences*, *161*, 105072. https://doi.org/10.1016/j.cageo.2022.105072

Leinonen, J., Hamann, U., & Germann, U. (2022). Seamless lightning nowcasting with recurrent-convolutional deep learning. *Artificial Intelligence for the Earth Systems*, *1*(4), e220043. https://doi.org/10.1175/aies-d-22-0043.1

Leinonen, J., Hamann, U., Nerini, D., Germann, U., & Franch, G. (2023). Latent diffusion models for generative precipitation nowcasting with accurate uncertainty quantification. *arXiv preprint arXiv:2304.12891*.

Lin, T., Li, Q., Geng, Y.-A., Jiang, L., Xu, L., Zheng, D., et al. (2019). Attention-based dual-source spatiotemporal neural network for lightning forecast. *IEEE Access*, *7*, 158296–158307. https://doi.org/10.1109/access.2019.2950328

Liu, Y., Hu, T., Zhang, H., Wu, H., Wang, S., Ma, L., & Long, M. (2023). itransformer: Inverted transformers are effective for time series forecasting. *arXiv preprint arXiv:2310.06625*.

Lu, K., Grover, A., Abbeel, P., & Mordatch, I. (2022). Frozen pretrained transformers as universal computation engines. *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 36(7), 7628–7636). https://doi.org/10.1609/aaai.v36i7.20729

Nguyen, T., Brandstetter, J., Kapoor, A., Gupta, J. K., & Grover, A. (2023). Climax: A foundation model for weather and climate. *arXiv preprint arXiv:2301.10343*.

Pan, X., Lu, Y., Zhao, K., Huang, H., Wang, M., & Chen, H. (2021). Improving nowcasting of convective development by incorporating polarimetric radar variables into a deep-learning model. *Geophysical Research Letters*, *48*(21), e2021GL095302. https://doi.org/10.1029/2021gl095302

Prudden, R., Adams, S., Kangin, D., Robinson, N., Ravuri, S., Mohamed, S., & Arribas, A. (2020). A review of radar-based nowcasting of precipitation and applicable machine learning techniques. *arXiv preprint arXiv:2005.04988*.

Pulkkinen, S., Nerini, D., Pérez Hortal, A. A., Velasco-Forero, C., Seed, A., Germann, U., & Foresti, L. (2019). Pysteps: An open-source python library for probabilistic precipitation nowcasting (v1. 0). *Geoscientific Model Development*, *12*(10), 4185–4219. https://doi.org/10.5194/gmd-12-4185-2019

Rasp, S., & Lerch, S. (2018). Neural networks for postprocessing ensemble weather forecasts. *Monthly Weather Review*, *146*(11), 3885–3900. Retrieved from https://doi.org/10.1175/MWR-D-18-0187.1

Ravuri, S., Lenc, K., Willson, M., Kangin, D., Lam, R., Mirowski, P., et al. (2021). Skilful precipitation nowcasting using deep generative models of radar. *Nature*, *597*(7878), 672–677. https://doi.org/10.1038/s41586-021-03854-z

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat (2019). Deep learning and process understanding for data-driven earth system science. *Nature*, *566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Medical image computing and computer-assisted intervention–miccai 2015: 18th international conference, munich, germany, october 5-9, 2015, proceedings, Part III* (Vol. 18, 234–241). https://doi.org/10.1007/978-3-319-24574-4_28

Rothfusz, L. P., Schneider, R., Novak, D., Klockow-McClain, K., Gerard, A. E., Karstens, C., et al. (2018). Facets: A proposed next-generation paradigm for high-impact weather forecasting. *Bulletin of the American Meteorological Society*, *99*(10), 2025–2043. https://doi.org/10.1175/bams-d-16-0100.1

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., & Woo, W.-c. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in Neural Information Processing Systems*, *28*.

Shi, X., Gao, Z., Lausen, L., Wang, H., Yeung, D.-Y., Wong, W.-k., & Woo, W.-c. (2017). Deep learning for precipitation nowcasting: A benchmark and a new model. *Advances in Neural Information Processing Systems* (Vol. 30).

Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic attribution for deep networks. In *International conference on machine learning* (pp. 3319–3328).

Wang, Y., Long, M., Wang, J., Gao, Z., & Yu, P. S. (2017). Predrnn: Recurrent neural networks for predictive learning using spatiotemporal LSTMS. *Advances in Neural Information Processing Systems*, *30*.

Wen, L., Zhao, K., Zhang, G., Xue, M., Zhou, B., Liu, S., & Chen, X. (2016). Statistical characteristics of raindrop size distributions observed in east China during the Asian summer monsoon season using 2-d video disdrometer and micro rain radar data. *Journal of Geophysical Research: Atmospheres*, *121*(5), 2265–2282. https://doi.org/10.1002/2015jd024160

Weyn, J. A., Durran, D. R., & Caruana, R. (2020). Improving data-driven global weather prediction using deep convolutional neural networks on a cubed sphere. *Journal of Advances in Modeling Earth Systems*, *12*(9), e2020MS002109. https://doi.org/10.1029/2020ms002109

Wu, H., Yao, Z., Wang, J., & Long, M. (2021). Motionrnn: A flexible model for video prediction with spacetime-varying motions. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (pp. 15435–15444).

Yuan, L., Park, H. S., & Lejeune, E. (2022). Towards out of distribution generalization for problems in mechanics. *Computer Methods in Applied Mechanics and Engineering*, *400*, 115569. https://doi.org/10.1016/j.cma.2022.115569

Yuan, Y., & Lin, L. (2020). Self-supervised pretraining of transformers for satellite image time series classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 474–487. https://doi.org/10.1109/jstars.2020.3036602

Zhang, X., Yang, Y., Chen, B., & Huang, W. (2021). Operational precipitation forecast over China using the Weather Research and Forecasting (WRF) model at a gray-zone resolution: Impact of convection parameterization. *Weather and Forecasting*, *36*(3), 915–928. https://doi.org/10.1175/waf-d-20-0210.1

Zhou, K., Zheng, Y., Dong, W., & Wang, T. (2020). A deep learning network for cloud-to-ground lightning nowcasting with multisource data. *Journal of Atmospheric and Oceanic Technology*, *37*(5), 927–942. https://doi.org/10.1175/jtech-d-19-0146.1