https://doi.org/10.5194/egusphere-2026-590
Preprint. Discussion started: 10 June 2026

Deep learning for non-precipitation radar echo identification: Comparative evaluation of polarimetric, spatial, and temporal information

Rongze Yang[1,2,3], Chaoshi Wei[4], Xiang Pan[5], Kun Zhao[1,2,3,] Jie Ming[2,3], Chen Lu[1,2,3], Haotian Tan[1,2,3], Wenxuan Zhao[1,2,3], Hao Huang[1,2,3]

[1]State Key Laboratory of Severe Weather Meteorological Science and Technology, Nanjing University, Nanjing, 210023, China
[2]Key Laboratory of Mesoscale Severe Weather of MOE, Frontiers Science Center for Critical Earth Material Cycling, and School of Atmospheric Sciences, Nanjing University, Nanjing, 210023, China
[3]Key Laboratory of Radar Meteorology, China Meteorology Administration, Beijing, 100081, China
[4]Meteorological Center of East China Air Traffic Management Bureau, Shanghai, 200335, China
[5]Qinghai Provincial Key Laboratory of Plateau Climate Change and Corresponding Ecological and Environmental Effects, and School of Ecology and Environmental Science, Qinghai Institute of Technology, Xining, 810000, China

Correspondence to: Hao Huang, Jie Ming (huanghao@nju.edu.cn, jming@nju.edu.cn)

Abstract. Accurate identification of non-precipitation echoes (NPEs) in weather radar observations requires effective use of polarimetric signatures together with spatiotemporal structure. Here we present a unified deep-learning framework to quantify the independent and synergistic contributions of model architecture, dual-polarization variables, and short-term temporal evolution to NPE identification. Using data from the Guangzhou S-band dual-polarization radar, we conduct controlled comparative experiments with two representative architectures: a pointwise multilayer perceptron (MLP) and a Transformer-based Swin U-Net that explicitly learns spatial context. We further perform ablation experiments across single- versus dual-polarization inputs and single-volume versus two-volume inputs. Results show that architecture-driven spatial-context learning is the dominant factor: Swin U-Net consistently outperforms the pointwise MLP under all input settings. On a high-confidence test subset, for example, the Critical Success Index (CSI) increases from 0.887 for the dual-polarization MLP to 0.950 for the dual-polarization Swin U-Net. Dual-polarization variables provide essential microphysical constraints and substantially improve class separability, particularly for pointwise classifiers. Incorporating two consecutive volumes further improves performance by capturing short-term echo evolution, with larger gains for the MLP than for Swin U-Net. The best-performing configuration, combining Swin U-Net with dual-polarization and two-volume inputs, achieves a CSI of 0.953 on the high-confidence test subset. Notably, the Swin U-Net using only the reflectivity factor (Z_H) as input retains strong skill (CSI = 0.927), indicating that spatial-context learning can partially compensate for missing polarimetry and thus providing a practical pathway for quality control of legacy single-polarization archives.

https://doi.org/10.5194/egusphere-2026-590
Preprint. Discussion started: 10 June 2026

1 Introduction

Radar is one of the primary observational tools in modern weather monitoring and forecasting systems. It provides high-resolution observations for real-time monitoring of weather systems, nowcasting and early warning, and quantitative precipitation estimation (QPE). While detecting hydrometeors, radar electromagnetic waves inevitably receive returns from ground clutter, birds and insects, and anomalous propagation, giving rise to non-precipitation echoes (NPEs). On the one hand, NPEs contaminate precipitation echoes, degrade the accuracy of QPE, and lead to misidentification of weather systems and their features. On the other hand, NPEs such as insects carry information on the environmental wind field and convergence lines, which is beneficial for the early identification of convective initiation and for improving the accuracy of assimilation-based forecasts (Huang et al., 2023). Accurate identification of NPEs is therefore crucial for improving the quantitative identification, monitoring, and forecasting of weather systems.

However, efficiently and accurately identifying NPEs has long been one of the major challenges in radar data processing. The difficulty mainly stems from the highly nonlinear relationship between radar observables (and derived features) and NPEs, coupled with the large variability in their spatial morphology and temporal dynamics (Tang et al., 2014a; Zhao et al., 2025). There are different types of NPEs for ground-based polarimetric radars. Ground clutter is a major source of interference in precipitation measurements. Operational algorithms often use spectral separation or clutter filters to isolate precipitation echoes at a given range gate. When precipitation echoes are substantially weaker than the clutter, these methods may fail to fully suppress clutter, leading to contaminated radar variables and spurious or physically implausible echo patterns—a typical type of NPE contamination related to ground clutter. Anomalous propagation (AP) typically appears as elongated bands (Warde and Torres, 2014), and biological echoes generally have relatively low reflectivity factors. Co-frequency interference often appears as radial streaks confined within a narrow azimuth sector.

For polarimetric weather radars, variables such as differential reflectivity (Z_{DR}) and copolar correlation coefficient (ρ_{hv}) facilitate NPE identification, because they may provide fundamental discriminatory information related to scatterer shape, size, and microphysical (or scattering) properties. NPEs often depress ρ_{hv} due to the irregular shapes and heterogeneous compositions of the scatterers, whereas biological echoes from insects and birds tend to exhibit enhanced Z_{DR} owing to their preferential horizontal orientation (Tang et al., 2014b). Building on these signatures, researchers have developed rule-based polarimetric classification schemes, typically relying on fixed thresholds and fuzzy-logic membership functions. Such methods usually take Z_{DR} and ρ_{hv} as primary predictors, supplemented by reflectivity factor (Z_H), other polarimetric variables, and constraints based on texture measures and spatial-consistency checks (Ellis et al., 2009; Zhang et al., 2020). These algorithms have been standardized in operational systems such as CINRAD and have achieved relatively high skill scores (Tang et al., 2014b). However, rule-based approaches require expert-tuned thresholds, are mostly formulated using features at individual pixel/range gates, and have limited ability to fully exploit features from spatial structure and temporal evolution. As a result, their performance can degrade in complex biological echoes, mixed-echo scenes, and low signal-to-noise-ratio (SNR) conditions (Zou et al., 2018).

https://doi.org/10.5194/egusphere-2026-590
Preprint. Discussion started: 10 June 2026

From a spatiotemporal perspective, NPEs exhibit patterns that are distinct from meteorological precipitation echoes. Spatially, precipitation echoes typically form coherent and spatially continuous regions with relatively smooth textures, whereas NPEs often present fragmented structures and enhanced local variability. Temporally, precipitation echoes generally show strong continuity and coherent advection, while NPEs commonly manifest as either nearly stationary ground clutter or biological echoes (e.g., insects and birds) that drift with the ambient wind, whose signature is often captured in radial velocity, albeit with substantial measurement noise and aliasing. To reduce the instability inherent to pixel-wise classification, subsequent studies have introduced spatial and temporal context into the identification process. For conventional single-polarization radars, spatial textures of reflectivity factor and related variables are commonly used to link neighboring gates and mitigate misclassification of isolated NPEs (Lakshmanan et al., 2010; Lakshmanan et al., 2007). Temporal consistency and spectral characteristics (e.g., Doppler spectrum width and spectral asymmetry) have also been exploited to suppress transient NPEs, particularly chaff (Kim et al., 2013).

With the widespread deployment of polarimetric radars and advances in machine learning, NPE identification algorithms have further evolved. Various studies have engineered combinations of reflectivity, polarimetric variables, and their textures as inputs to machine-learning classifiers such as support vector machines, decision trees, and random forests. These approaches have improved the identification of specific targets, including ships and biological echoes from birds and insects (Chase et al., 2023; Jatau et al., 2021; Zhang et al., 2025). Huang et al. (2023), for example, combined typical polarimetric signatures of biological echoes with SNR for radar quality control, flagging pixels with anomalously high/low polarimetric values and large spatial variability as NPEs, and discarding pixels with SNR < 5 dB to reduce noise contamination. Nonetheless, these methods still rely heavily on manual feature engineering and empirical modeling, making it difficult to consistently learn the coupled relationships among temporal, spatial, and polarimetric information. Their performance is also sensitive to the statistical characteristics of the training samples, which limits their ability to generalize across regions.

More recently, rapid advances in deep learning have provided new possibilities for improving both the accuracy and computational efficiency of NPE identification (Chase et al., 2023). Convolutional neural networks (CNNs) and Transformer-based architectures can extract spatial textures and adaptively learn multi-scale information (Cuomo and Chandrasekar, 2021; Geng et al., 2024; Li et al., 2025). Sequence models such as ConvLSTM can encode the evolution of consecutive radar scans, thereby enabling joint ingestion of temporal evolution and polarimetric features (Li et al., 2024; Shi et al., 2015). Improved U-Net variants with multi-scale weighting and channel attention have demonstrated enhanced classification accuracy for boundaries and small targets, and have already shown promising performance in complex NPE scenarios (Cui et al., 2025). Collectively, these deep-learning approaches provide a unified framework to incorporate diverse sources of information, making it feasible to systematically assess the individual and combined contributions of polarimetric variables, spatial textures, and temporal evolution to NPE identification. Meanwhile, they reduce the reliance on expert-defined thresholds and mitigate inconsistencies that arise when different rule sets are used to compare methods or information sources. This, in turn, can deepen our understanding of NPE characteristics and support the development of robust, generalizable quality-control algorithms for operational applications.

https://doi.org/10.5194/egusphere-2026-590
Preprint. Discussion started: 10 June 2026

Motivated by these considerations, this study systematically evaluates the relative contributions of temporal evolution, spatial context, and polarimetric information for distinguishing NPEs from precipitation echoes in weather radar observations. To this end, we compare representative model architectures ranging from multilayer perceptrons (MLPs) to U-Net-type segmentation networks. Our objective is to develop echo-classification methods that are physically consistent, robust to diverse NPE scenarios, and computationally efficient for potential operational use, and to clarify, within a unified, interpretable, and operationally oriented framework, the relative value of different information sources for NPE identification. The rest of the paper is organized as follows. Section 2 describes the polarimetric radar data, labeling procedure, network architectures, and evaluation design. Section 3 presents the model comparison results. Section 4 provides a discussion of the results, and Section 5 draws a conclusion.

2 Data And Methods

To quantify how different information sources and model choices affect NPE identification, we design a suite of controlled comparative experiments along three axes. First, to assess the role of spatial context, we compare two architectures with markedly different spatial-representation capacity: a multilayer perceptron (MLP) and a Swin U-Net. Second, to evaluate the added value of polarimetry, we test whether incorporating polarimetric variables beyond reflectivity improves classification performance. Third, to examine temporal information, we contrast models that use a single radar volume with those that ingest two consecutive scans, thereby quantifying the benefit of short-term echo evolution for classification.

2.1 Model Architecture for NPEs Identification

2.1.1 MLP Architecture

MLP is a classical neural network that is used as a pointwise classifier that predicts the echo type for each radar pixel independently. The input typically consists of radar reflectivity factor and (when available) polarimetric variables, and the network learns a nonlinear mapping from these predictors to echo labels through stacked fully connected layers. In this study, the MLP comprises four hidden layers, containing 512, 256, 128, and 64 neurons, respectively. Each hidden block follows the sequence: fully connected layer, normalization, LeakyReLU activation, and dropout (rate = 0.3). The fully connected layers perform feature transformation (and, where applicable, reduce feature dimensionality), the BatchNorm1d normalization layer stabilizes optimization, LeakyReLU improves nonlinear representation, and dropout helps mitigate overfitting and improves generalization. The output layer has two units and produces class probabilities for precipitation and NPEs. The model is trained using a binary classification objective based on the predicted NPE probability (elaborated on later).

Depending on the specific experimental objective, different combinations of predictors are provided to the MLP input layer (Table 1). The pixel-wise MLP using polarimetric radar variables (Z_H, Z_{DR}, ρ_{hv}, and the moving standard deviation of the

https://doi.org/10.5194/egusphere-2026-590
Preprint. Discussion started: 10 June 2026

differential phase Φ_{DP}, denoted as $\sigma_{\Phi_{DP}}$) is adopted as the benchmark (MLP-Model-1). Note that Φ_{DP} contains contributions from both the propagation phase and backscatter differential phase; the latter exhibits strong gate-to-gate variability in echoes from insects, birds, and ground clutter. Although $\sigma_{\Phi_{DP}}$ is a local texture-like measure and thus implicitly encodes limited spatial variability, it has been used for separating precipitation from non-precipitation echoes (Schuur et al., 2003). Therefore, it is included in this benchmark model.

To assess NPE-identification performance with and without polarimetric information, we further construct a single-polarization counterpart that uses only the horizontal reflectivity factor Z_H (MLP-Model-3). Furthermore, to account for the temporal evolution of the echoes, we design two additional experiments (MLP-Model-2 and MLP-Model-4) in which the inputs include not only the predictors from the current radar volume (t) but also the corresponding predictors from the previous volume $(t-1)$. Specifically, MLP-Model-2 extends MLP-Model-1 by concatenating the polarimetric predictors at $t-1$, while MLP-Model-4 extends MLP-Model-3 by concatenating $Z_H(t-1)$.

2.1.2 Swin U-Net Architecture

To explicitly account for spatial structure in echo classification, we adopt a Swin U-Net model. This architecture follows the U-Net encoder–decoder framework but replaces the standard convolutional blocks in both the encoder and decoder with Swin Transformer blocks, such that the core representation learning is driven by self-attention rather than conventional convolutions [22]. In Swin Transformers, self-attention is computed within non-overlapping local windows, and the windows are shifted between successive layers to enable cross-window information exchange. This shifted-window design captures long-range dependencies and multi-scale features with reduced computational cost, which is particularly beneficial for exploiting spatial context when identifying NPEs with diverse spatial extents. In addition, the patch-merging operations in Swin U-Net produce hierarchical feature maps, allowing the network to learn echo characteristics at multiple resolutions [23].

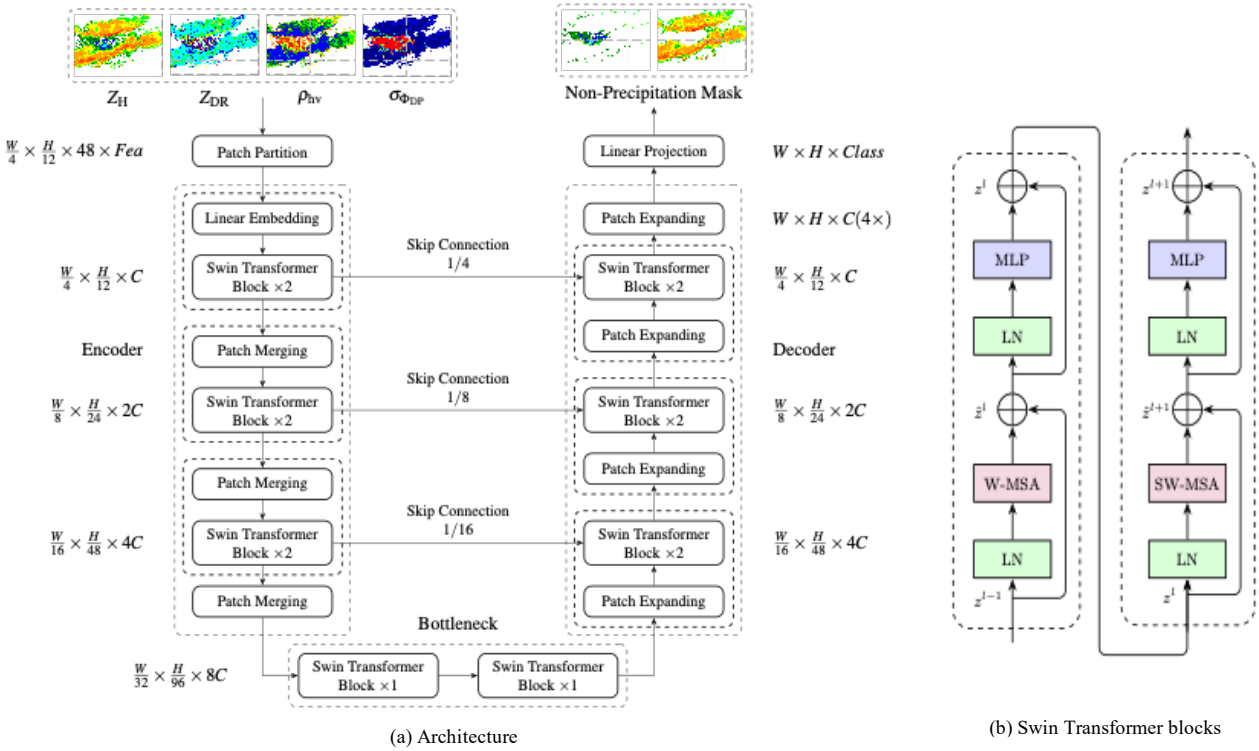
To tailor the Swin U-Net to our radar application, we configure the key architectural parameters as illustrated in Figure 1. The radar data are organized as PPI images with a size of 64 (azimuth angles) × 960 (range gates), consistent with the typical structure of a radar scan. The input image is first partitioned into non-overlapping patches of size 4 (azimuth) × 12 (range), which are then embedded and fed into the Transformer encoder. This patch design reflects both the range–azimuth resolution of the PPI images and the spatial characteristics of radar echoes, providing a practical balance between efficiency and representativeness.

For the backbone, we follow the configuration of Swin-Tiny. The initial embedding dimension is set to 96, and the network depth across the four Swin Transformer stages is $[2, 2, 6, 2]$ (i.e., 2, 2, 6, and 2 Transformer blocks per stage). The numbers of attention heads are $[3, 6, 12, 24]$ across the four stages, enabling the model to capture spatial relationships among radar variables at different scales. The local self-attention window size is fixed at 8×8, which provides a compromise between computational cost and the ability to represent local echo structure. The model further employs feedforward networks with

https://doi.org/10.5194/egusphere-2026-590
Preprint. Discussion started: 10 June 2026

an MLP expansion ratio of 4, uses a learnable bias in the qkv projections (qkv_bias = True), and adopts a drop-path rate of 0.1 to mitigate overfitting. Other settings include disabling absolute positional encoding (ape = False) and enabling patch normalization (patch_norm = True), which together ensure stable training while controlling memory usage. Overall, this configuration yields a small-to-medium–sized model that maintains relatively low computational complexity while retaining sufficient capacity to model radar features across spatial scales.

To isolate the effect of spatial information and facilitate comparison with the MLP, we design four Swin U-Net experiments, Swin-Model-1 through Swin-Model-4, in direct analogy to MLP-Model-1 through MLP-Model-4 (Table 1). Swin-Model-1 and Swin-Model-3 mainly examine the impact of including polarimetric variables in addition to reflectivity, while Swin-Model-2 and Swin-Model-4 further incorporate two consecutive radar volumes to capture short-term displacement and structural evolution. This design enables a controlled evaluation of temporal information under both single- and dual-polarization settings.



(a) Architecture

(b) Swin Transformer blocks

Figure 1: Architecture of the proposed S-band dual-polarization radar NPE identification model based on Swin U-Net (a), and the structure of a Swin Transformer block (b).

https://doi.org/10.5194/egusphere-2026-590
Preprint. Discussion started: 10 June 2026

Table 1: For the MLP model and Swin U-Net comparative experiment, the input time steps, the included radar variables, and the total number of input variables. Note that, in the MLP models, the observation variables are provided as pixel-wise inputs, whereas in the Swin U-Net, they are fed in as 2D image-like fields.

Model	Time steps	Radar variables	Variable numbers
MLP-Model-1/ Swin-Model-1	single	Z_H, Z_{DR}, ρ_{hv}, and $\sigma_{\Phi_{\mathrm{DP}}}$	4
MLP-Model-2/ Swin-Model-2	double	Z_H, Z_{DR}, ρ_{hv}, and $\sigma_{\Phi_{\mathrm{DP}}}$	8
MLP-Model-3/ Swin-Model-3	single	Z_H	1
MLP-Model-4/ Swin-Model-4	double	Z_H	2

2.2 Dataset for training and evaluation

A large volume of accurate training data is essential for improving the performance of deep learning models, yet obtaining such high-quality labels is often the most difficult part of the workflow. This creates an inherent dilemma: deep learning critically depends on accurate labels, but generating those labels is precisely the major bottleneck in practice. When a mature method already exists for producing reliable training data, the added value of deep learning can be diminished. Fortunately, for polarimetric weather radars, identification of potential NPEs based on the ρ_{hv} is well established and physically justified. In this study, we therefore use a ρ_{hv}-based screening strategy to produce initial labels, and then combine it with manual inspection and correction to construct a high-quality reference dataset. This procedure improves label accuracy and consistency for subsequent model training and evaluation. More importantly, it provides a trustworthy basis for interpreting the roles of different model designs and information sources in NPE identification.

In this study, we use data from the Guangzhou S-band dual-polarization Doppler radar as a representative example. This radar has been employed in numerous previous studies, demonstrating robust performance after quality control. NPEs are typically more pronounced at lower elevation angles and strongly affect applications such as QPE and convection initiation detection. Therefore, we construct the dataset using the 0.5° elevation scans. The identification of NPEs follows Huang et al. (2023) and Huang et al. (2018). Specifically, echoes with ρ_{hv} lower than 0.85 are initially flagged as potential NPEs. If an NPE region embedded within precipitation echoes contains fewer than 20 connected gates, it is reclassified as precipitation. Conversely, some NPEs may exhibit relatively high ρ_{hv} values (> 0.85) due to random measurement errors. In such cases, regions with fewer than 20 connected gates are detected using a de-speckle filter and classified as NPEs.

Figure 2 illustrates representative examples from the dataset, demonstrating the effectiveness of the rule-based labeling scheme in identifying diverse types of NPEs. Figures 2a1-2d1 show convection interspersed with multiple types of NPE, including residual ground clutter, biological echoes, and co-frequency interference. The identified NPEs (used as labels; Figure 2e1) successfully capture regions of low ρ_{hv}, even when such features are embedded within strong convective cells or located in the gaps between weather systems. Figures 2a2–f2 present the case with NPEs from co-frequency interference or

https://doi.org/10.5194/egusphere-2026-590
Preprint. Discussion started: 10 June 2026

AP-induced radial streaks. These NPEs exhibit irregular structures and, in many areas, relatively high Z_H values (Figure 2d2), making them difficult to detect using threshold-based methods on any single variable. Nevertheless, the labels accurately isolate these radial, non-precipitation artifacts while preserving the real precipitation echoes in the southern sector. The ability to distinguish among small speckle-like NPEs, elongated radial artifacts, and true precipitation structures highlights the robustness of the labeling process across a variety of environmental conditions and echo morphologies.

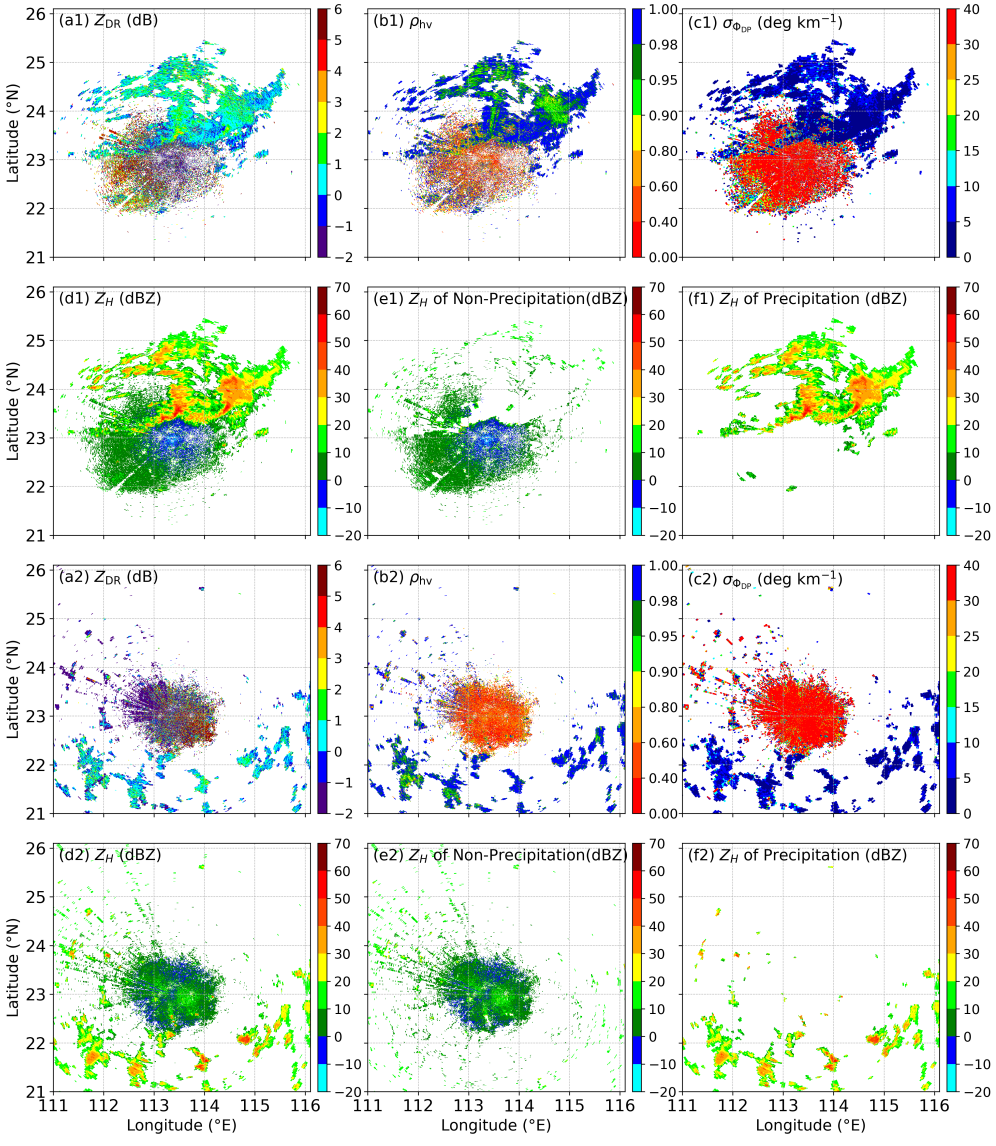


Figure 2: Example samples from the Guangzhou S-band weather-radar dataset. Dual-polarization radar fields and the corresponding labels at the 0.5° elevation angle are shown for 15:00 UTC on 5 March 2019 (a1–f1) and 03:00 UTC on 29 July 2019 (a2–f2), including Z_{DR} (a; dB), ρ_{hv} (b), $\sigma_{\Phi_{DP}}$ (c; degree), and Z_H (d; dBZ). Also shown are the Z_H classified as NPEs with the threshold-based method (e), and the Z_H of precipitation echoes with NPEs removed (f).

https://doi.org/10.5194/egusphere-2026-590
Preprint. Discussion started: 10 June 2026

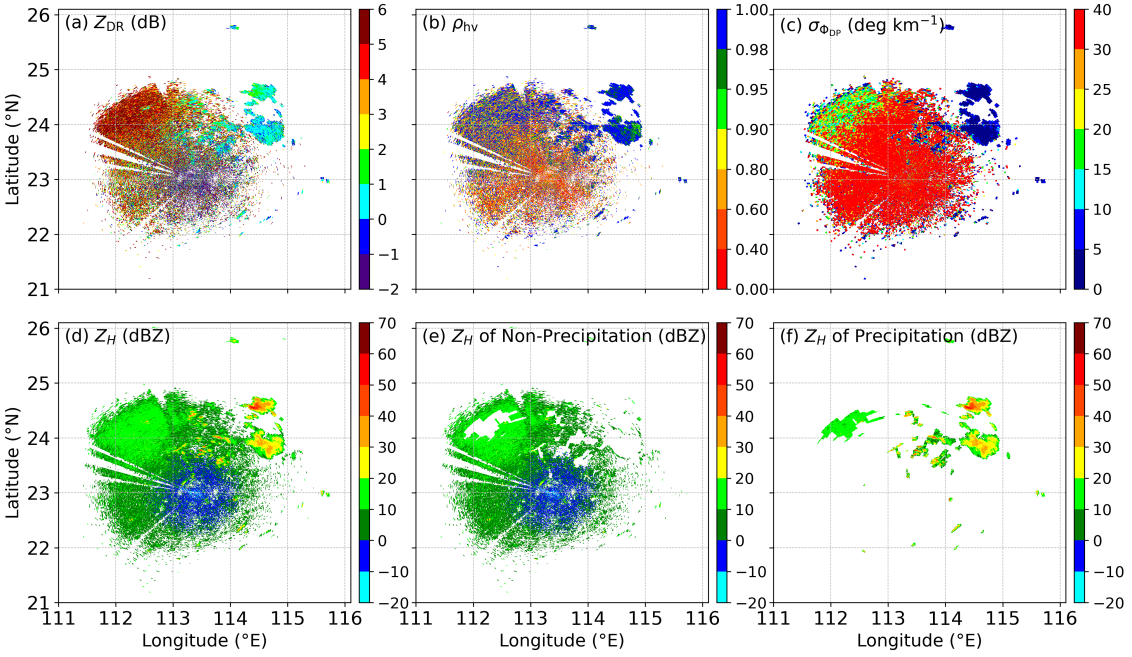


Figure 3: An example for accurate data representation & universal readability of figures. Example from the Guangzhou S-band weather-radar dataset illustrating a misclassification that can occur when labels are generated using the rule-based method. The radar fields and corresponding labels are the same as in Figures 2a1-f1, but for the case at 17:00 UTC on 15 May 2019.

However, some NPEs may exhibit signatures that closely resemble those of precipitation echoes in some radar variables, causing them to be mistakenly classified as precipitation echoes by the rule-based method. Figure 3 shows a representative misclassification case. In Figure 3f, the narrow band of echoes within the marked blue rectangle would be labeled as precipitation because it has a relatively high ρ_{hv}. Nevertheless, its low Z_H and anomalously large Z_{DR}, together with pronounced spatial fluctuations in both Z_{DR} and ρ_{hv}, indicate that it is in fact a non-precipitation echo—possibly insects covered with dew under humid nighttime conditions.

In this context, we manually collected high-quality samples from the initial rule-based NPE identification results to build the training dataset. To ensure robust performance across precipitation regimes, we selected representative cases from 2016–2018, including squall lines, isolated convective cells, typhoon events, and stratiform precipitation. We also included clear-air cases to help the model learn characteristic non-precipitation signatures. In addition, the dataset spans multiple seasons and different times of day to improve generalization. To facilitate effective training and alleviate class imbalance, we collected 1,000 clear-air cases, 3,000 strong-precipitation cases, 500 weak-precipitation cases, and 500 typhoon cases. This sampling strategy provides sufficient coverage of typical cases while maintaining manageable model complexity and training cost, and yields a more balanced distribution of samples. During training, 20% of the dataset was held out as a validation set.

https://doi.org/10.5194/egusphere-2026-590
Preprint. Discussion started: 10 June 2026

Following the same principles, we constructed an independent test set using Doppler radar observations over Guangzhou in 2019. To enable a more rigorous assessment of model robustness, we deliberately retained a subset of samples with imperfect (potentially erroneous) labels in the test set, thereby evaluating the sensitivity of the deep-learning models to label noise. In total, 1500 samples were selected. Among them, 1,000 samples have labels judged to be correct and physically consistent, forming Test Dataset A. Note that the cases shown in Figure 2 are in the Test Dataset A. The remaining 500 samples correspond to situations in which the rule-based labeling is prone to misclassification under specific conditions (e.g., atypical biological echoes or mixed scenes), forming Test Dataset B.

Because Swin U-Net requires inputs with fixed spatial dimensions, we applied the following preprocessing steps. To balance GPU memory consumption and model capacity, each full 360°−PPI (with ~1° azimuthal resolution and 1920 gates) was then partitioned into 6 azimuthal sectors using a sliding window with overlap (each with about 60° per sector). Along the range dimension, we further downsampled the data by retaining every other range gate, thereby reducing the input size and computational cost. After these steps, each sector sample has a fixed size of 64 azimuth bins × 960 range gates. For a fair comparison, the MLP-based models were trained and evaluated using the same preprocessed inputs.

2.3 Model training and performance evaluation

We formulate NPE identification as a binary classification task. During training, we use the binary cross-entropy (BCE) loss to quantify the discrepancy between model predictions (i.e., the predicted probability of NPE) and the ground-truth labels. Models are optimized with Adam, and mixed-precision training is enabled to improve computational efficiency and reduce GPU memory usage.

We adopt an early-stopping strategy to mitigate overfitting: training is terminated if the validation loss does not decrease for 20 consecutive epochs. The checkpoint that achieves the highest $F1$ score(Rosenhoover et al., 2025) on the validation set is selected as the final model. To account for differences in model size and convergence behavior, the batch size is set to 24 for the lightweight MLP and 8 for the larger Swin U-Net.

The learning rate is controlled by a scheduler with a warm-up phase followed by polynomial decay. The base learning rate is set to $lr_{\text{init}} = 1 \times 10^{-3}$. During warm-up (the first epoch), the learning rate increases linearly from $warmup\_factor \times lr_{\text{init}}$ to lr_{init} (with $warmup_{factor} = 10^{-3}$), which helps stabilize optimization at the start of training. After warm-up, the learning rate decays according to a polynomial schedule ($power = 0.1$), following Chen et al. (2018), balancing convergence speed and final performance. All training and testing are performed in PyTorch with CUDA acceleration on an NVIDIA L40S GPU.

For model performance evaluation, we compute verification statistics at the grid-point level. Following Xingjian (2015), we use Critical Success Index (CSI), Probability of Detection (POD) and False Alarm Ratio (FAR) to quantitatively compare different models. CSI is the fraction of correctly identified NPEs considering both misses and false alarms; POD is the detection ability for true NPEs; and FAR is the proportion of precipitation echoes incorrectly labeled as NPEs.

https://doi.org/10.5194/egusphere-2026-590
Preprint. Discussion started: 10 June 2026

3 Results

3.1 Model Architecture for NPEs identification

To provide a qualitative verification of the proposed models, we first examine two representative cases shown in Figure 2 and compare the corresponding single-volume predictions in Figure 4. Figures 4a1–b1 show the results at 15:00 UTC on 5 March 2019. Using the dual-polarization inputs (Z_H, Z_{DR}, ρ_{hv}, and $\sigma_{\Phi_{\mathrm{DP}}}$), MLP-Model-1 successfully removes most NPEs while preserving the overall structure of meteorological precipitation echoes, achieving a CSI of 0.894. Nevertheless, some NPEs—particularly biological echoes (e.g., insects or birds) and residual ground clutter contamination—remain and are not completely removed (Figure 4a1). A major reason is that these echoes can exhibit large spatial fluctuations in polarimetric variables; locally, ρ_{hv} may become relatively high, making their feature signatures partially overlap with those of light rain or melting-layer echoes. Without explicit spatial-texture or spatial-consistency constraints, a pointwise MLP classifier is prone to such residual misclassification.

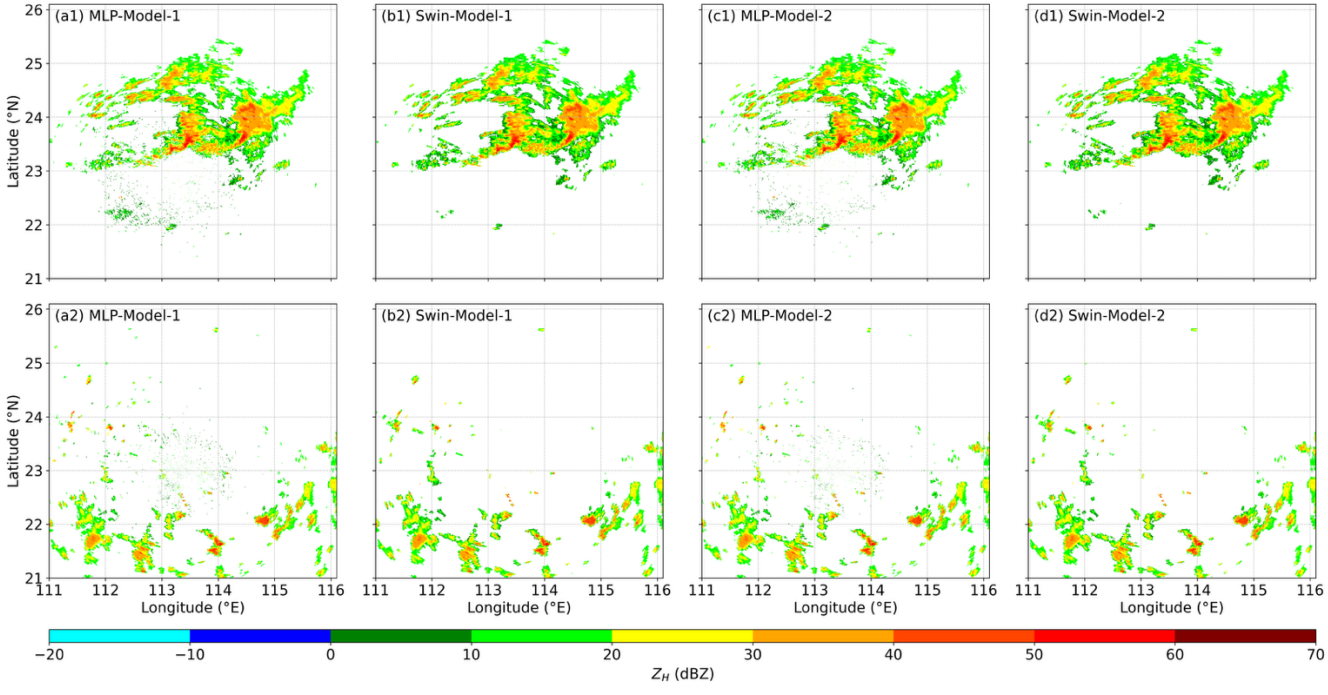


Figure 4: Predicted Z_H of precipitation echoes with NPEs removed at the 0.5° elevation angle from the Guangzhou S-band radar for 15:00 UTC on 5 March 2019 (a1, b1, c1, d1) and 03:00 UTC on 29 July 2019 (a2, b2, c2, d2), produced by MLP-Model-1 (a1, a2), Swin-Model-1 (b1, b2), MLP-Model-2 (c1, c2), and Swin-Model-2 (d1, d2).

In contrast, Swin-Model-1 explicitly leverages spatial context through self-attention in Swin Transformer blocks, which enables the network to model spatial coherence and neighborhood dependencies. As a result, Swin-Model-1 more effectively

https://doi.org/10.5194/egusphere-2026-590
Preprint. Discussion started: 10 June 2026

handles these confusing NPE regions associated with strong local variability, producing cleaner precipitation fields with fewer residual artifacts (Figure 4b1). The CSI improves to 0.944, substantially higher than that of MLP-Model-1 under the same single-volume dual-polarization setting.

Figures 4a2–b2 present the second case at 03:00 UTC on 29 July 2019, characterized by smaller-scale convection and more scattered, irregularly distributed NPEs. MLP-Model-1 already performs well in this case (CSI = 0.952), removing most NPE contamination. However, some isolated speckles remain, consistent with occasional local increases in ρ_{hv} and other variability-driven effects. Swin-Model-1 further suppresses these residual scattered pixels and yields a visually cleaner and more spatially coherent precipitation field (Figure 4b2), achieving an even higher CSI of 0.975.

Figures 4c1–d1 and 4c2–d2 further compare models that ingest two consecutive radar volumes (MLP-Model-2 and Swin-Model-2). Across both cases, adding temporal information clearly benefits the MLP. By incorporating the previous-volume observations as short-term history, MLP-Model-2 can better distinguish persistent precipitation structures from transient or sporadic NPE artifacts, leading to a noticeable reduction of scattered, speckle-like non-precipitation echoes (Figures 4c1 and 4c2). Accordingly, the CSI increases markedly to 0.911 and 0.970 for the two cases, respectively. Even so, its performance remains lower than that of Swin-Model-1 (CSI = 0.944/0.975), underscoring that model architecture and explicit spatial-context learning can be as important as temporal cues for NPE identification.

For the Swin U-Net, the single-volume dual-polarization setting (Swin-Model-1) already achieves high skill by leveraging spatial coherence and morphology, leaving less room for improvement from additional temporal inputs. After incorporating echo evolution, Swin-Model-2 shows essentially unchanged performance in the first case (Figure 4d1; CSI = 0.942), but a modest improvement in the second case (from 0.975 to 0.983; Figure 4d2). This contrast indicates that the incremental value of temporal information is scene dependent: when polarimetric signatures and spatial structure are already sufficiently distinctive, temporal evolution adds limited benefit; when echoes are more fragmented or ambiguous, short-term temporal consistency can provide complementary constraints and further refine the classification.

To enable a more complete comparison, Table 2 summarizes CSI, POD, and FAR on both the Validation Dataset and Test Dataset A. Although the absolute values of these metrics differ between the two datasets, the relative ranking of models is consistent across them and is also in agreement with the case-study results. This consistency suggests that the training and evaluation are stable (i.e., no obvious overfitting or dataset-specific artifacts), and therefore we focus the subsequent quantitative discussion primarily on Test Dataset A, which provides the most reliable assessment under high-confidence labels.

First, the benefit of explicit spatial-context learning is evident when comparing MLP and Swin U-Net. Using the same single-volume polarimetric inputs, Swin-Model-1 achieves a substantially higher CSI than MLP-Model-1 on Test Dataset A (0.950 vs. 0.887) and simultaneously reduces FAR (0.032 vs. 0.072). This indicates that leveraging spatial structure and neighborhood coherence is highly effective for suppressing NPE artifacts that are difficult to resolve via pixel-wise decision rules.

https://doi.org/10.5194/egusphere-2026-590
Preprint. Discussion started: 10 June 2026

Table 2: Quantitative performance of the MLP and Swin U-Net variants on the validation set and the high-confidence test subset (Test Dataset A). For each model, we report the Critical Success Index (CSI), Probability of Detection (POD), and False Alarm Ratio (FAR). Higher CSI/POD and lower FAR indicate better performance.

Model	Validation Dataset			Test Dataset A		
	CSI	POD	FAR	CSI	POD	FAR
MLP-Model-1	0.860	0.942	0.092	0.887	0.951	0.072
MLP-Model-2	0.899	0.963	0.069	0.913	0.966	0.064
MLP-Model-3	0.577	0.877	0.372	0.604	0.873	0.324
MLP-Model-4	0.677	0.845	0.228	0.749	0.864	0.137
Swin-Model-1	0.959	0.980	0.022	0.950	0.973	0.032
Swin-Model-2	0.962	0.977	0.015	0.953	0.967	0.024
Swin-Model-3	0.932	0.963	0.034	0.927	0.955	0.042
Swin-Model-4	0.940	0.967	0.029	0.929	0.951	0.040

Meanwhile, temporal information improves performance, especially for the MLP. Adding the previous-volume predictors increases CSI from 0.887 (MLP-Model-1) to 0.913 (MLP-Model-2), accompanied by a higher POD (0.951 to 0.966) and a lower FAR (0.072 to 0.064), suggesting that short-term echo evolution provides useful constraints for removing scattered or transient NPE residuals. For Swin U-Net, however, temporal gains are comparatively modest: CSI changes from 0.950 (Swin-Model-1) to 0.953 (Swin-Model-2), consistent with the observation that once strong spatial modeling is available, the incremental value of two-volume temporal cues can become scene dependent and closer to saturation.

3.2 Models trained only using reflectivity factors

Although dual-polarization weather radars are increasingly available, a large fraction of operational radars and many historical archives still provide single-polarization observations only. Therefore, evaluating deep-learning performance under Z_H-only conditions is practically important for broad deployability and for reprocessing legacy datasets. Figure 5 (a1–d1, a2–d2) shows the results for the two cases in Figure 2 using the single-polarization settings: MLP-Model-3 / Swin-Model-3 (single volume) and MLP-Model-4 / Swin-Model-4 (two consecutive volumes). When only the pointwise reflectivity factor is provided, MLP-Model-3 exhibits a pronounced performance degradation as expected. Lacking polarimetric microphysical discrimination and any explicit spatial-context modeling, the pointwise MLP struggles to separate precipitation and NPEs with similar reflectivity magnitudes, and tends to retain only the most prominent precipitation cores and remove widespread precipitation echoes (Figures 5a1 and 5a2). Consistent with this visual impression, both cases show high FAR (0.367/0.234) and low CSI (0.625/0.711). In contrast, Swin-Model-3 maintains strong classification skills even with only Z_H as the input (Figures 5b1 and 5b2), achieving CSI values of 0.906 and 0.961 for the two cases, respectively. This demonstrates that Swin

https://doi.org/10.5194/egusphere-2026-590
Preprint. Discussion started: 10 June 2026

U-Net can effectively compensate for reduced input information by learning discriminative spatial morphology and texture directly from the reflectivity field, whereas the pointwise MLP lacks the mechanism to exploit such spatial cues.

Models MLP-Model-4 and Swin-Model-4 further incorporate temporal evolution by using two consecutive volumes. In Case 1 (Figures 4c1 and 5d1), adding temporal context partially mitigates the limitations of the MLP under Z_H-only inputs, leaving more precipitation echoes and improving CSI from 0.625 to 0.669, and in Case 2, the CSI score increases from 0.711 to 0.883. These results suggest that short-term echo continuity and displacement provide additional constraints for suppressing isolated or transient non-precipitation artifacts. Swin-Model-4 also shows a slight improvement (CSI = 0.910 in Case 1 and 0.970 in Case 2), but the total gain is smaller because Swin U-Net already captures rich spatial structure in the single-volume setting. Overall, these comparisons indicate that temporal evolution provides useful motion-related information, but its incremental benefit is typically smaller than that provided by dual-polarization variables. In particular, when anomalous NPEs persist coherently over time, temporal cues alone may become less discriminative and cannot substitute for the microphysical information encoded in polarimetric measurements.

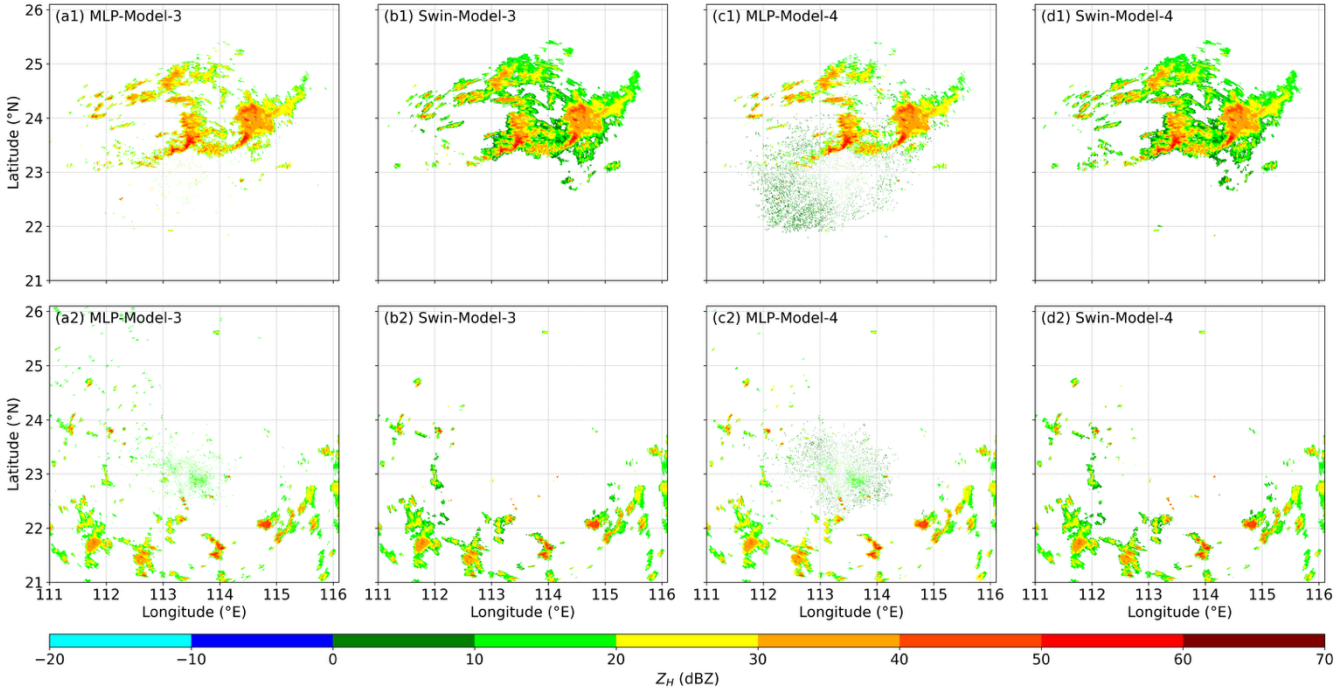


Figure 5: Similar to Figure 4, but for the results given by MLP-Model-3 (a1, a2), Swin-Model-3 (b1, b2), MLP-Model-4 (c1, c2), and Swin-Model-4 (d1, d2).

Table 2 further quantifies performance under the single-polarization (Z_H-only) setting and enables a controlled comparison of temporal and spatial contributions when polarimetric information is absent. Within the MLP family, the Z_H-only, single-

https://doi.org/10.5194/egusphere-2026-590
Preprint. Discussion started: 10 June 2026

volume configuration (MLP-Model-3) shows a pronounced drop in skill on Test Dataset A (CSI = 0.604, POD = 0.873, FAR = 0.324), indicating that a pointwise classifier struggles to separate meteorological and non-meteorological echoes when only reflectivity intensity is available. Introducing short-term temporal context substantially improves the MLP results: MLP-Model-4 (two volumes, Z_H-only) raises CSI to 0.749 and reduces FAR to 0.137 (POD = 0.864), demonstrating that echo continuity and displacement cues provide strong additional constraints for suppressing false alarms in the Z_H-only regime.

In contrast, Swin U-Net remains highly effective even with reflectivity as the sole input. Swin-Model-3 (single volume, Z_H-only) achieves CSI = 0.927, POD = 0.955, and FAR = 0.042, and adding the previous volume yields only a marginal change (Swin-Model-4: CSI = 0.929, POD = 0.951, FAR = 0.040). Comparing architectures under the same Z_H-only inputs highlight the importance of spatial-context learning: Swin-Model-3 exceeds MLP-Model-3 by 0.323 CSI (0.927 vs. 0.604) and reduces FAR by 0.282 (0.042 vs. 0.324), showing that spatial morphology and texture learned from the Z_H are highly discriminative for NPE identification. Moreover, Swin-Model-3 (single-volume, Z_H-only) even outperforms the dual-polarization pointwise baseline MLP-Model-1 (CSI = 0.927 vs. 0.887), indicating that strong spatial modeling can partially compensate for missing polarimetric microphysical information. Overall, in the single-polarization setting, temporal information provides a meaningful gain for the MLP, whereas the dominant improvement comes from spatial-context modeling in Swin U-Net.

4 Discussion

4.1 Effects of input information and model architecture on NPE identification

Overall, Swin U-Net consistently outperforms the pointwise MLP across all input configurations, indicating that explicitly modeling spatial context is a dominant factor for NPE identification in this dataset. Specifically, Swin U-Net achieves CSI values of 0.927–0.953 (Swin-Model-3/4 to Swin-Model-2), whereas the MLP achieves 0.604–0.913 (MLP-Model-3 to MLP-Model-2). This gap is also reflected in FAR: for the matched settings, Swin U-Net yields substantially lower false alarms (e.g., 0.032 vs. 0.072 for the dual-pol single-volume setting, and 0.042 vs. 0.324 for the single-pol single-volume setting). These results are consistent with the case studies: Swin U-Net benefits from learning echo morphology, texture, and spatial coherence, while a pointwise MLP cannot directly exploit such spatial structure.

Polarimetric information provides a clear benefit within each architecture, but its impact depends on whether the model can already leverage spatial cues. For the MLP, adding dual-polarization variables leads to a major improvement: CSI increases from 0.604 (MLP-Model-3, Z_H only) to 0.887 (MLP-Model-1, dual-pol) and FAR drops sharply from 0.324 to 0.072, highlighting the importance of polarimetric microphysical signatures for pixel-wise discrimination. For Swin U-Net, polarimetry still improves performance (CSI 0.927 to 0.950, FAR 0.042 to 0.032 from Swin-Model-3 to Swin-Model-1), but the gain is comparatively smaller because spatial context already provides strong discrimination even under Z_H-only inputs.

https://doi.org/10.5194/egusphere-2026-590
Preprint. Discussion started: 10 June 2026

Temporal evolution acts as a complementary cue, with a noticeably larger contribution for the MLP than for the Swin U-Net. When two consecutive volumes are used, the MLP improves from 0.887 to 0.913 in the dual-pol setting (MLP-Model-1 vs. MLP-Model-2) and from 0.604 to 0.749 in the single-pol setting (MLP-Model-3 vs. MLP-Model-4), accompanied by consistent FAR reductions (e.g., 0.324 to 0.137 for Z_H-only). In contrast, Swin U-Net exhibits only marginal gains from adding the previous volume (CSI 0.950 to 0.953 for dual-pol; 0.927 to 0.929 for single-pol), suggesting that once spatial morphology is well captured, the incremental value of short-term temporal cues can be limited and scene dependent.

Taken together, these results indicate that spatial-context modeling is the primary driver of performance, while polarimetric variables provide strong additional microphysical discrimination—especially for pointwise classifiers—and temporal information offers useful supplementary constraints, most notably when spatial modeling or input richness is limited.

4.2 Generalization across different types of cases

To further examine the robustness of the proposed approach under diverse observational regimes, we evaluate the best-performing configuration, Swin-Model-2 (dual-polarization, two consecutive volumes), in different types of cases. Results for organized convection and small-scale convection have been presented in Section 3.

4.2.1 Stratiform precipitation case

Stratiform precipitation is typically characterized by broad, spatially continuous echo coverage, relatively small spatial gradients, and weak to moderate reflectivity. In this regime, separating precipitation from NPEs is challenging because weak precipitation can overlap with clutter (after filtering) or other artifacts in reflectivity magnitude, making intensity-based screening unreliable.

Figure 6 illustrates a representative stratiform event. The Z_H observations (Figure 6a) show a widespread precipitation region, but it is contaminated by non-meteorological patterns, including banded/radial streak-like structures and scattered speckles that are inconsistent with the smooth morphology expected for stratiform precipitation. The corresponding ρ_{hv} field (Figure 6b) exhibits pronounced spatial variability, including areas of reduced ρ_{hv} within the broader echo region, which can cause ambiguity for fixed-threshold labeling.

https://doi.org/10.5194/egusphere-2026-590
Preprint. Discussion started: 10 June 2026

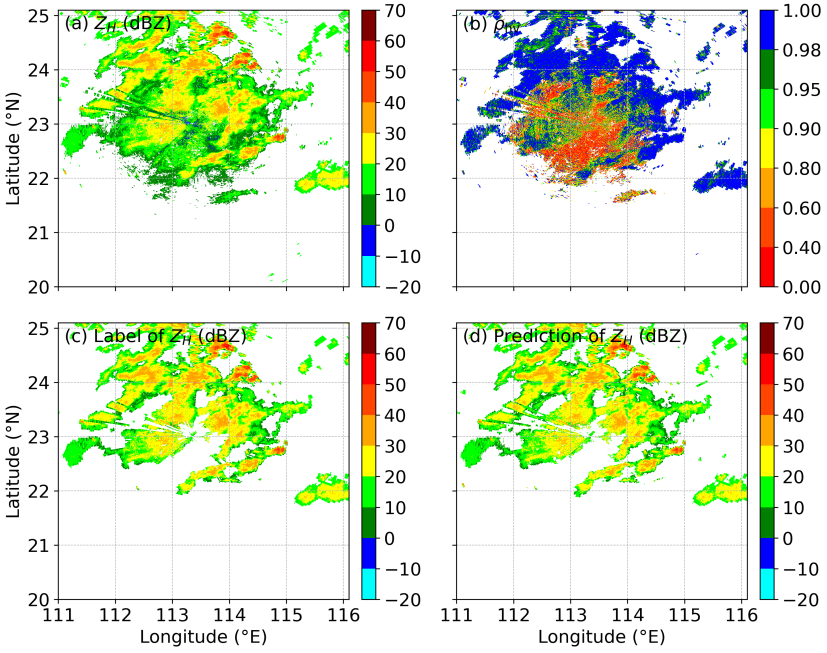


Figure 6: Example of NPE identification for a stratiform-precipitation case using Swin-Model-2. (a) and (b) are observed Z_H and ρ_{hv}. (c) is the Z_H of precipitation echoes labeled by the rule-based method. (d) is the corresponding results given by the Swin-Model-2 using dual-polarization variables and two consecutive volumes.

The threshold-based meteorological field (Figure 6c) suppresses most NPE contamination and provides a generally reliable reference for this case. The Swin-Model-2 prediction (Figure 6d) also removes NPEs while retaining the overall morphology of the stratiform precipitation band. Notably, the rule-based procedure can introduce minor discontinuities in the precipitation region and may remove portions of stratiform echoes when polarimetric signatures deviate from typical "clean precipitation" values. Although manual screening was applied to curate high-quality labels, it should be acknowledged that the resulting labels are not perfect and may still contain residual errors or inconsistencies. In comparison, Swin-Model-2 tends to further suppress streak-like and speckle-like artifacts and produces a more spatially continuous precipitation field. This observation highlights an important practical point: because the ground-truth labels are imperfect, evaluation metrics such as CSI are not expected to reach 1.0, and small discrepancies between model output and labels may sometimes reflect limitations of the labeling procedure rather than model failure.

4.2.2 Clear-air echo case

Under clear-air conditions, radars frequently observe weak echoes induced by biological scatterers (e.g., insects and birds) and residual ground clutter. Such echoes are typically spatially sparse, scattered, or beam-aligned streaky and are generally characterized by low reflectivity.

https://doi.org/10.5194/egusphere-2026-590
Preprint. Discussion started: 10 June 2026

Figure 7 shows a representative clear-air echo case. The Z_H filed (Figure 7a) is dominated by weak and discontinuous echoes with banded echoes aligned along the radar beams, lacking the smooth and spatial continuity expected for precipitation. The corresponding ρ_{hv} field (Figure 7b) exhibits pronounced spatial variability, showing correlation patterns that are inconsistent with homogeneous hydrometeor Scatterers, thereby enabling identification based on polarimetric information. The threshold-based meteorological mask (Figure 7c) suppresses almost all echoes and yields an approximately empty reference field, consistent with the no-precipitation scenario. The Swin-Model-2 prediction (Figure 7d) also produces a near-empty reflectivity field, indicating that the model can effectively suppress non-precipitation echoes under clear-air conditions while avoiding noticeable false alarms. This case further demonstrates the robustness and generalization of the proposed approach in non-precipitation conditions.

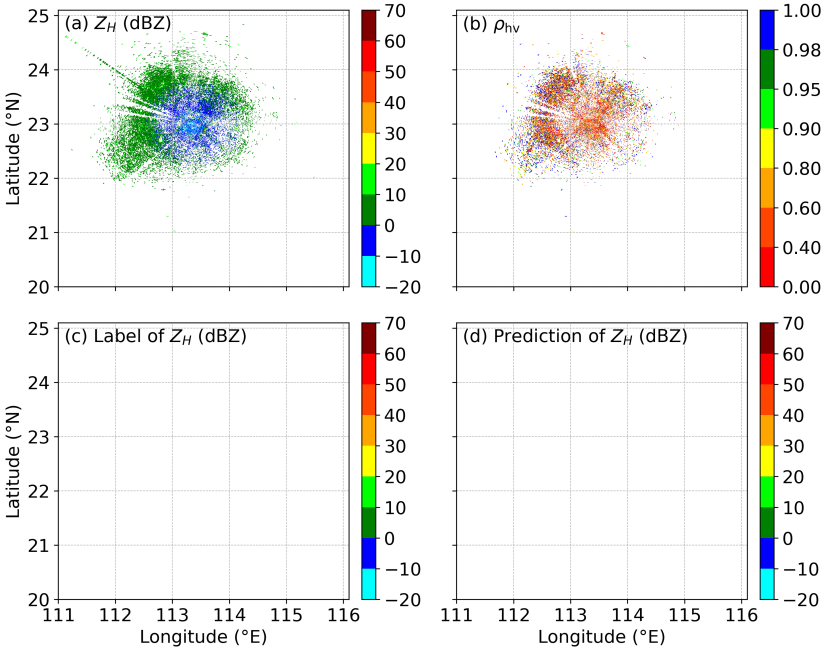


Figure 7: Similar to Figure 6, but for the example of NPE identification for a clear-air echo case.

4.3 Model capability to correct mislabeled regions

Rule-based (threshold) labeling can fail under complex conditions, such as enhanced nocturnal biological echoes, residual ground clutter, and AP, where NPEs may partially overlap with precipitation in polarimetric signatures. To evaluate robustness to such label errors, we use Test Dataset B, which intentionally contains samples where the rule-based labels are more likely to be incorrect. In this setting, CSI/POD/FAR quantify agreement with an imperfect reference, so small metric differences may reflect not only model skill but also whether the model is "following" mislabeled regions.

https://doi.org/10.5194/egusphere-2026-590
Preprint. Discussion started: 10 June 2026

Figure 8 provides a representative example. Based on the polarimetric diagnosis for this event (Figure 3), the suspicious sector (highlighted region) exhibits low and spatially noisy ρ_{hv}, anomalously large Z_{DR}, and clear scan-line/streak artifacts in Z_H, which is inconsistent with typical precipitation structure and suggests that this area is a mislabeled NPE region. Among the Swin variants, Swin-Model-2 (dual-polarization, two-volume input) most effectively suppresses this artifact while preserving the physically plausible precipitation structure and retaining coherent high-ρ_{hv} precipitation regions, indicating a stronger ability to override erroneous labels using consistent spatial–temporal–polarimetric cues. In contrast, the other configurations (Swin-Model-1/3/4) leave more residual contamination or remove precipitation less cleanly, consistent with their missing input components (temporal context and/or polarimetric variables).

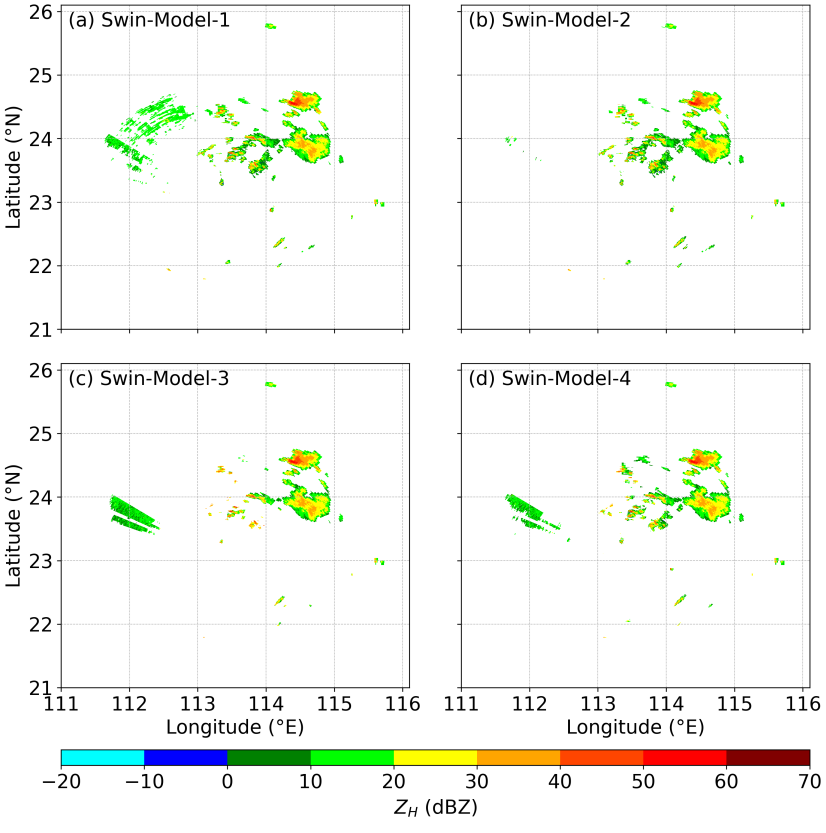


Figure 8: Comparison of predicted Z_H (dBZ) of precipitation echoes for a convective case over Guangzhou at 17:00 UTC on 15 May 2019, for which the reference labels contain a localized mislabeling region. Panels show results from (a) Swin-Model-1, (b) Swin-Model-2, (c) Swin-Model-3, and (d) Swin-Model-4. The potential mislabeled region is highlighted by the blue rectangle.

These qualitative findings help interpret the quantitative behavior in Table 3. Overall, Swin U-Net models remain the top performers on Test Dataset B (CSI = 0.935–0.952), indicating strong robustness even under label noise. Richer information generally improves performance: dual-polarization Swin models (Swin-Model-1/2: CSI = 0.952/0.951) exceed the Z_H-only

https://doi.org/10.5194/egusphere-2026-590
Preprint. Discussion started: 10 June 2026

Swin models (Swin-Model-3/4: CSI = 0.939/0.935), and a similar pattern holds for MLP (MLP- Model-1/2 > MLP-Model-3/4). Temporal information provides clear gains for MLP (e.g., CSI 0.866 to 0.906 for MLP-Model-1 to MLP-Model-2; 0.629 to 0.720 for MLP-Model-3 to MLP-Model-4), while changes among Swin temporal variants are smaller—partly because Swin already captures strong spatial structure, and partly because on Test B improved "correction" of mislabeled regions may reduce strict agreement with noisy labels. This is consistent with Figure 8, where Swin-Model-2 appears to remove a region that the reference labels would otherwise treat as precipitation, illustrating why metrics on Test Dataset B should be interpreted as agreement with imperfect labels rather than absolute physical correctness.

Table 3: Quantitative metrics (CSI, POD, and FAR) of the MLP and Swin U-Net variants on Test Dataset B, which contains samples with partially mislabeled regions due to failure cases of the rule-based (threshold) labeling procedure. Note that metrics on this set primarily reflect agreement with an imperfect reference, and thus are used to assess model robustness under label-noise conditions.

Model	CSI	POD	FAR
MLP-Model-1	0.866	0.892	0.032
MLP-Model-2	0.906	0.939	0.037
MLP-Model-3	0.629	0.761	0.211
MLP-Model-4	0.720	0.752	0.048
Swin-Model-1	0.952	0.972	0.021
Swin-Model-2	0.951	0.980	0.030
Swin-Model-3	0.939	0.984	0.047
Swin-Model-4	0.935	0.980	0.047

5 Conclusions

This study investigated how polarimetric information, spatial context, and short-term temporal evolution contribute to NPE identification for an S-band weather radar over Guangzhou. Using a controlled experimental design, we compared a pointwise MLP with a spatial-context Swin U-Net, and evaluated single-volume vs. two-volume inputs under both dual-polarization and single-polarization (Z_H-only) configurations. Performance was assessed with CSI, POD, and FAR, complemented by case studies and additional scenario tests.

It is found that model architecture (spatial-context learning) is the dominant driver of performance. Across all input configurations, Swin U-Net consistently outperformed the pointwise MLP, demonstrating that learning echo morphology, texture, and spatial coherence is crucial for separating precipitation from NPEs. On the high-confidence test subset (Test Dataset A), under identical single-volume dual-polarization inputs, Swin U-Net achieved substantially higher skill than MLP

https://doi.org/10.5194/egusphere-2026-590
Preprint. Discussion started: 10 June 2026

(e.g., CSI 0.950 vs. 0.887, with a lower FAR 0.032 vs. 0.072), consistent with the qualitative improvements seen in case studies.

Within the same architecture, adding dual-polarization predictors improved discrimination between precipitation and NPEs by reducing ambiguity that cannot be resolved from reflectivity magnitude alone. This effect was particularly pronounced for the MLP, where moving from Z_H-only to dual-polarization inputs yielded large gains in CSI and substantial reductions in FAR (Test Dataset A), indicating that polarimetric signatures play a critical role when spatial context is not explicitly modeled.

Meanwhile, Temporal evolution provides complementary information, with larger gains for the MLP than for the Swin U-Net. Incorporating the previous radar volume improved MLP performance in both dual-pol and single-pol settings (e.g., Test Dataset A, CSI 0.887 to 0.913 for dual-pol; 0.604 to 0.749 for Z_H-only), suggesting that short-term continuity and displacement help suppress scattered/transient artifacts. In contrast, Swin U-Net showed smaller incremental gains from two-volume inputs (e.g., Test Dataset A, CSI 0.950 to 0.953 for dual-pol; 0.927 to 0.929 for Z_H-only), consistent with spatial modeling already capturing much of the discriminative structure, making additional temporal cues more scene dependent.

Spatial perception models can partially compensate for missing polarimetry, supporting legacy single-pol applications. A notable outcome is that Swin U-Net maintained high skill even with Z_H alone (Test Dataset A CSI 0.927), reaching more than 97% of the dual-pol Swin performance (CSI 0.950). Moreover, the Z_H-only Swin configuration exceeded the dual-pol pointwise MLP baseline (CSI 0.927 vs. 0.887), highlighting the practical value of spatial-context architectures for operational radars and historical archives where dual-polarization data are unavailable.

Robustness under label noise is supported by both quantitative results and qualitative correction behavior. On Test Dataset B (containing samples with increased label imperfection), Swin U-Net models remained highly competitive (CSI ≈ 0.935–0.952). Because metrics on Test B measure agreement with an imperfect reference, small differences among high-performing models should be interpreted cautiously; they may reflect a model's tendency to deviate from mislabeled regions rather than degraded physical discrimination, consistent with the observed correction of localized label-error regions in representative cases.

Despite strong performance and consistent trends, several limitations remain. First, the study is centered on data from a single radar site; broader multi-region validation is needed because terrain, climate, and local NPE composition can differ substantially. Second, the training set may not fully cover rare but operationally important NPE types (e.g., chaff), which can limit generalization under previously unseen conditions. Third, temporal modeling here is limited to two consecutive volumes; longer sequences or more explicit motion modeling may further improve robustness in challenging scenes. Finally, operational deployment would benefit from exploring lighter architectures or acceleration techniques (e.g., pruning/distillation) and from improving label quality via scalable strategies (e.g., active learning or semi-supervised refinement).

Overall, the results indicate that the best performance is achieved by combining architectures that effectively exploit spatial context (Swin U-Net) with richer information sources (dual-polarization and short-term temporal evolution), while also

https://doi.org/10.5194/egusphere-2026-590
Preprint. Discussion started: 10 June 2026

showing that spatial modeling alone can recover much of the discrimination capability when polarimetric data are unavailable—an encouraging pathway for operational quality control and legacy data reprocessing.

Code availability

Codes are available from the authors upon request.

Author contributions

RY and HH conceived the study and designed the experiments; RY developed the model code, conducted the experiments and simulations, and analyzed the results; CW and XP contributed to data collection, preprocessing, and case selection; KZ, JM, and CL contributed to the interpretation of the radar-physics and polarimetric signatures and provided critical scientific feedback; HT and WZ assisted with visualization, validation, and result checking; RY prepared the manuscript draft; HH and JM supervised the research and reviewed and edited the manuscript with contributions from all co-authors.

Competing interests

The contact author has declared that none of the authors has any competing interests.

Acknowledgements

The authors would like to thank the reviewers and editors for their careful evaluation of the manuscript, and the engineers at Guangzhou Radar for their support.

Financial support

This work is supported by the National Natural Science Foundation of China (Grants 42422501 and 42475006), the Jianghuai Meteorological Joint Project of Anhui Natural Science Foundation (Grant 2408055UQ002), the AI & AI for Science Project of Nanjing University (Grant 020714380231), and the Major Science and Technology Demonstration Project for Social Development in Jiangsu Province (Grant BE2023766).

References

Chase, R. J., Harrison, D. R., Lackmann, G. M., and McGovern, A.: A Machine Learning Tutorial for Operational Meteorology. Part II: Neural Networks and Deep Learning, Weather and Forecasting, 38, 1271–1293, 10.1175/waf-d-22-0187.1, 2023.

https://doi.org/10.5194/egusphere-2026-590
Preprint. Discussion started: 10 June 2026

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation, Proceedings of the European conference on computer vision (ECCV), 801–818,

Cui, M., Zeng, C., Xu, X., Bilal, M., and Xia, X.: MW-UNet: Multi-Scale Weighted Connection UNet for Identification and Classification of Non-Meteorological Clutter over Big Radar Data, Big Data Mining and Analytics, 8, 65–77, 10.26599/bdma.2024.9020032, 2025.

Cuomo, J. and Chandrasekar, V.: Use of Deep Learning for Weather Radar Nowcasting, Journal of Atmospheric and Oceanic Technology, 10.1175/jtech-d-21-0012.1, 2021.

Ellis, S. M., Dixon, M., and Hubbert, J. C.: Weather Radar Ground Clutter. Part II: Real-Time Identification and Filtering, Journal of Atmospheric and Oceanic Technology, 26, 1181–1197, 10.1175/2009jtecha1160.1, 2009.

Geng, H., Wu, F., Zhuang, X., Geng, L., Xie, B., and Shi, Z.: The MS-RadarFormer: A Transformer-Based Multi-Scale Deep Learning Model for Radar Echo Extrapolation, Remote Sensing, 16, 10.3390/rs16020274, 2024.

Huang, H., Zhao, K., Zhang, G., Lin, Q., Wen, L., Chen, G., Yang, Z., Wang, M., and Hu, D.: Quantitative Precipitation Estimation with Operational Polarimetric Radar Measurements in Southern China: A Differential Phase–Based Variational Approach, Journal of Atmospheric and Oceanic Technology, 35, 1253–1271, 10.1175/JTECH-D-17-0142.1, 2018.

Huang, K., Huang, H., Gu, T., Wang, B., and Lou, C.: Polarimetric Radar Observations of Biological Scatterers in the Eye of Typhoon Lekima (2019), Journal of Geophysical Research-Biogeosciences, 128, 10.1029/2023jg007533, 2023.

Jatau, P., Melnikov, V., and Yu, T.-Y.: A machine learning approach for classifying bird and insect radar echoes with S-band Polarimetric Weather Radar, Journal of Atmospheric and Oceanic Technology, 10.1175/jtech-d-20-0180.1, 2021.

Kim, S., Kim, Y. H., Han, H.-Y., Heo, B.-H., and You, C.-H.: Real-Time Detection and Filtering of Chaff Clutter from Single-Polarization Doppler Radar Data, Journal of Atmospheric and Oceanic Technology, 30, 873–895, 10.1175/jtech-d-12-00158.1, 2013.

Lakshmanan, V., Zhang, J., and Howard, K.: A Technique to Censor Biological Echoes in Radar Reflectivity Data, Journal of Applied Meteorology and Climatology, 49, 453–462, 10.1175/2009jamc2255.1, 2010.

Lakshmanan, V., Fritz, A., Smith, T., Hondl, K., and Stumpf, G.: An Automated Technique to Quality Control Radar Reflectivity Data, Journal of Applied Meteorology and Climatology, 46, 288–305, 10.1175/jam2460.1, 2007.

Li, J., Shi, Y., Zhang, T., Li, Z., Wang, C., and Liu, J.: Radar precipitation nowcasting based on ConvLSTM model in a small watershed in north China, Natural Hazards, 120, 63–85, 10.1007/s11069-023-06193-6, 2024.

Li, Z., Lu, Z., Li, Y., and Liu, X.: SwinNowcast: A Swin Transformer-Based Model for Radar-Based Precipitation Nowcasting, Remote Sensing, 17, 10.3390/rs17091550, 2025.

Rosenhoover, M., Rushing, J., Beck, J., White, K., and Graves, S.: Improving Doppler Radar Precipitation Prediction with Citizen Science Rain Gauges and Deep Learning, Sensors (Basel), 25, 10.3390/s25123719, 2025.

Schuur, T., Ryzhkov, A., Heinselman, P., Zrnic, D., Burgess, D., and Scharfenberg, K.: Observations and classification of echoes with the polarimetric WSR-88D radar, Report of the National Severe Storms Laboratory, Norman, OK, 73069, 46, 2003.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., and Woo, W.-c.: Convolutional LSTM network: A machine learning approach for precipitation nowcasting, Advances in neural information processing systems, 28, 2015.

Tang, L., Krause, J., Karstens, C., and Lakshmanan, V.: Quality Control of Weather Radar Data Using Polarimetric Variables, Journal of Atmospheric and Oceanic Technology, 31, 1234–1249, 10.1175/jtech-d-13-00073.1, 2014a.

Tang, L., Zhang, J., Langston, C., Krause, J., Howard, K., and Lakshmanan, V.: A Physically Based Precipitation–Nonprecipitation Radar Echo Classifier Using Polarimetric and Environmental Data in a Real-Time National System, Weather and Forecasting, 29, 1106–1119, 10.1175/waf-d-13-00072.1, 2014b.

Warde, D. A. and Torres, S. M.: Ground Clutter Mitigation for Weather Radars Using the Autocorrelation Spectral Density, Journal of Atmospheric and Oceanic Technology, 31, 2049–2066, 10.1175/jtech-d-13-00117.1, 2014.

Xingjian, S.: Convolutional lstm network: A machine learning approach for precipitation nowcasting, Advances in neural information processing systems, 28, 1, 2015.

Zhang, S., Li, H., Moisseev, D., and Leskinen, M.: A hybrid algorithm for ship clutter identification in pulse compression polarimetric radar observations, Atmospheric Measurement Techniques, 18, 4839–4855, 10.5194/amt-18-4839-2025, 2025.

Zhang, S., Huang, X., Min, J., Chu, Z., Zhuang, X., and Zhang, H.: Improved fuzzy logic method to distinguish between meteorological and non-meteorological echoes using C-band polarimetric radar data, Atmospheric Measurement Techniques, 13, 537–551, 10.5194/amt-13-537-2020, 2020.

https://doi.org/10.5194/egusphere-2026-590
Preprint. Discussion started: 10 June 2026

Zhao, Z., Wen, H., Wu, L., Li, R., Zhuang, T., and Zhang, Y.: Non-Precipitation Echo Identification in X-Band Dual-
605 Polarization Weather Radar, Remote Sensing, 17, 10.3390/rs17173023, 2025.
Zou, H., Zhang, S., Liang, X., and Yi, X.: Improved Algorithms for Removing Isolated Non-Meteorological Echoes and Ground Clutters in CINRAD, Journal of Meteorological Research, 32, 584–597, 10.1007/s13351-018-7176-9, 2018.