remote sensing

MDPI

Article

A Quality-Control Fusion Algorithm for Cloud-Radar Data in Complex Weather Scenarios Integrating LightGBM and Neighborhood Filtering

Chang Hou [1,2], Weihua Liu [3], Fa Tao [1,*] and Shuzhen Hu [1]

[1] Meteorological Observation Center of China Meteorological Administration, Beijing 100081, China
[2] Liaoning Meteorological Equipment Support Center, Shenyang 110166, China
[3] Dalian Meteorological Bureau, Dalian 116201, China
* Correspondence: taofa@asu.edu.pl; Tel.: +86-134-2625-2723

Highlights

This study focuses on the quality-control challenges of cloud-radar data under complex weather conditions, achieving three key breakthroughs in data, methodology and evaluation. In terms of data, based on the long-term Ka-band cloud radar observations from 49 stations across 24 provinces in China for the period of 2023–2024, we adopt a dual-mechanism approach of "manual labeling + automatic quality control verification" to construct a training dataset containing over 20 million groups of samples. This dataset, which covers scenarios such as snow, rain, fog, low clouds and dust as well as eight types of key features, overcomes the limitations of traditional single-station and short-time-series data. Methodologically, we innovate a two-step framework combining the Light Gradient Boosting Machine (LightGBM) with neighborhood filtering. Specifically, the LightGBM first performs initial clutter recognition, and then the neighbor frequency method is used to self-adaptively filter residual isolated clutter. This framework not only addresses the limitations of single machine learning but also prevents the loss of weak meteorological signals inherent in traditional filtering techniques. Through typical case studies and statistical data analysis, the proposed algorithm achieves better quality-control scores in comparison with the Random Forest and single LightGBM algorithms. For evaluation purposes, a "clutter rejection rate + meteorological echo retention rate" system is established. Verification results demonstrate that the proposed method achieves a weighted average clutter rejection rate of 98.47% (a 1.89% improvement over the single LightGBM method) and a weighted average comprehensive score of 92.13. Furthermore, the method proposed in this research is compatible with cross-regional and cross-device data, providing reliable support for meteorological operations.

What are the main findings?

- The quality-control method of "LightGBM + neighborhood filtering" proposed in this study outperforms single machine-learning algorithms, realizing an efficient balance between clutter rejection and meteorological echo retention. This method achieves a weighted average clutter-rejection rate of 98.47%, representing an improvement of 1.89% over the single LightGBM method (96.58%). In addition, its weighted average meteorological echo retention rate reaches 85.79%, which is only about 1% lower than that of the single LightGBM method (86.92%). The overall weighted average comprehensive score of the proposed method is 92.13%. Moreover, this method demonstrates robust and stable performance under complex meteorological conditions, with comprehensive scores of 90.73, 94.23, 96.49, 91.10 and 95.79 in snow, rain, low cloud, fog, and dust scenarios, respectively.

Check for updates

Academic Editors: Junqiang Yao, Jinfang Yin, Haoran Li, Chang Yi and Ismail Gultepe

Received: 21 December 2025
Revised: 13 February 2026
Accepted: 17 February 2026
Published: 26 February 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

- This approach demonstrates cross-regional and cross-device adaptability, while still requiring scenario-specific optimization. It is applicable to radar data from 49 cloud radar stations across China and supports four radar models: YLU3, HTKAAA, YWKAAA and HMBKPS. The method can efficiently eliminate near-surface isolated clutter in low-cloud scenarios and improve the clarity of fog echoes. However, to maintain the retention rate of snow echoes and the recognition accuracy of weak echoes at the edge of low clouds, this method still requires targeted parameter adjustments and supplementary features.

What are the implications of the main findings?

- The method provides reliable data support for key meteorological operations. The quality-controlled cloud radar base data enables clear precipitation retrieval, laying the foundation for subsequent quantitative estimation and particle phase identification. It overcomes the adaptability limitations of traditional quality-control methods under complex weather conditions, thereby improving the accuracy of data application in meteorological operations.
- This study offers a technical paradigm and data reference for research on the quality control of radar data. The construction of a long-term and wide-coverage training dataset (fusion of multi-source observation data with more than 20 million samples) provides a template for similar radar quality-control research based on machine learning. The two-step framework of "machine learning + physical filtering" can also provide a new approach to reconcile the trade-off between model generalization and data authenticity. Subsequently, this method can be further extended to quality control applications for X-band and C-band radar data.

Abstract

In order to address the challenges of limited accuracy in identifying non-meteorological clutter and the spatial overlap between meteorological and non-meteorological echoes in cloud radar observations under complex weather conditions, in this study, we propose a quality-control method for cloud-radar data, which integrates machine learning with neighborhood filtering, This quality-control method first uses the Light Gradient Boosting Machine (LightGBM) to initially identify clutter, then employs a customized neighborhood filtering module to optimize and eliminate residual isolated clutter. This two-stage framework combines the strengths of accurate machine-learning-based classification and physically motivated filtering optimization, enabling reliable discrimination between meteorological and non-meteorological echoes. Based on multi-region, long-term and multi-model radar baseline observations, which cover typical complex weather types such as snow, fog, rain, low clouds and dust, the refined manual labeling of meteorological and non-meteorological echoes is carried out, combined with multi-source ground observation data such as surface observations, temperature and humidity. Based on this, a feature training dataset for machine learning is constructed, which contains over 20 million samples. A multi-index evaluation system—including echo classification accuracy and non-meteorological clutter rejection rate—is used to quantitatively assess the quality-control performance of the method in different weather scenarios. The results indicate that the proposed method demonstrates stable performance in typical complex weather scenarios, with comprehensive scores of 90.73 (snow), 94.23 (rain), 96.49 (low clouds), 91.10 (fog) and 95.79 (dust) on a 100-point scale. Through typical case studies and statistical data analysis, the proposed algorithm achieves better quality-control scores in comparison with the Random Forest and single LightGBM algorithms. It provides a new technical approach for cloud-radar data quality control and also offers a theoretical basis for the feature selection

of machine-learning-based quality-control models, further enhancing the application value of cloud-radar data in refined meteorological observations.

Keywords: quality control of cloud radar data; LightGBM model; neighborhood filtering; complex weather; quality evaluation

1. Introduction

Cloud radars serve as key equipment for detecting cloud microphysical characteristics and dynamic processes, with the quality of their observational data directly determining the accuracy of critical professional meteorological operations such as atmospheric monitoring, numerical weather prediction and weather modification [1–8]. However, cloud-radar observations are susceptible to interference from near-surface dust, clear-sky clutter and ground echoes. These challenges are further amplified under complex weather conditions such as rain, snow, fog and low clouds, where traditional quality-control methods (e.g., threshold-based filtering and spatio-temporal smoothing) often encounter misidentification issues or insufficient preservation of valid signals [9–15], thereby severely constraining the operational effectiveness of cloud-radar data.

In recent years, machine learning has demonstrated broad application prospects in radar data quality control due to its powerful pattern recognition capabilities [16–20]. For instance, Zhang et al. (2025) employed a clustering algorithm to achieve efficient identification of abnormal signals [21]. Li et al. (2024) accurately distinguished meteorological and non-meteorological echoes using deep learning architectures [22]. These studies have validated the technical advantages of machine learning. In radar clutter classification tasks, model selection directly affects classification performance. For single machine-learning models, the support vector machine was an early mainstream method due to its good adaptability to small sample sizes, establishing optimal feature-space boundaries to complete echo classification [23–26]. Neural networks, especially convolutional neural networks, rely on their capabilities in image feature extraction to meet the complex pattern-recognition demands of echo images [27]. Ensemble models such as Random Forest, Gradient Boosting Decision Tree and the improved Light Gradient Boosting Machine (LightGBM) algorithms exhibit superior performance through multi-model integration [28–31]. Among them, LightGBM has emerged as a popular tool in radar data quality-control research due to its efficient gradient-boosting mechanism and strong processing capabilities for large-scale data [32,33].

With the rapid development of deep-learning techniques, Transformer-based models exemplified by the Swin Transformer have demonstrated superior performance in computer vision and remote-sensing applications [34,35]. These models are capable of effectively capturing the global dependency structures of radar echoes, which is beneficial for discriminating clutter from weak meteorological signals. However, such models feature complex architectures, and their global attention mechanisms may attenuate local feature representations in deep networks, potentially degrading the fine-scale identification of weak echoes associated with snowfall and fog. By contrast, the LightGBM framework provides a more favorable balance between classification accuracy and computational efficiency, making it better suited to the operational requirements of cloud-radar data quality control. Despite these advances, three major technical challenges persist in cloud-radar data quality control. First, the data support is inadequate. Most previous studies rely on single-station, short-duration data, lacking standardized datasets that involve diverse weather scenarios and comprehensive quality issues, markedly restricting model generalization [36]. Second, model efficacy is limited. Single models, constrained by the limited dimension of feature

learning, are prone to misclassification in isolated clutter recognition. Although ensemble models integrate multiple learners, they still struggle to delineate the ambiguous boundary between discrete noise and valid weak echoes in low-reflectivity scenes [37,38]. Third, physical authenticity is compromised. Although neighborhood filtering can selectively suppress discrete noise, its direct application may lead to the loss of key meteorological signals such as weak snowfall echoes, thereby compromising the physical authenticity of the data [39].

To address the aforementioned issues, we propose a collaborative innovation scheme encompassing both data and methodology in this study. The specific steps are as follows. Firstly, construct a multi-dimensional, wide-coverage cloud-radar feature dataset based on the one-year observations from 49 cloud-radar stations in 24 provinces. This cloud-radar feature dataset encompasses eight key features, such as basic reflectivity and radial velocity in typical scenarios including clear sky, precipitation and fog. A dual-labeling mechanism that combines manual labeling with automated quality-control verification is employed to refine the labels, eventually generating over 20 million pairs of valid range-bin samples. This dataset overcomes the limitations of traditional datasets in terms of geographical and meteorological coverage, providing comprehensive sample support for model training. Secondly, we propose a two-step quality-control framework of "LightGBM + neighborhood filtering", which integrates the pattern-recognition capability of machine learning with the physical constraints of neighborhood filtering, thereby ensuring clutter classification accuracy while avoiding the loss of critical meteorological signals. Finally, establish a quantitative evaluation system encompassing the indicators of accuracy rate and signal retention rate. This system aims to achieve a quantitative evaluation of quality-control performance across different weather scenarios.

This study aims to advance cloud-radar data quality control under complex weather conditions over China through the synergistic integration of data-driven support and methodological innovation, thereby providing more reliable data for meteorological operations.

2. Data and Methods

To address the technical challenges of misclassifying non-meteorological clutter and losing valid signals in cloud-radar data under complex weather conditions, we establish a highly reliable quality control system through dual dimensions of data support and methodological design. On one hand, a multi-region, long-term and multi-source-validated sample dataset is constructed to provide a comprehensive data foundation for model training. On the other hand, an integrated workflow, including feature extraction, model-based preliminary screening, post-processing optimization and quantitative evaluation, is proposed, which combines the pattern recognition capabilities of machine learning with the signal-preservation strengths of physical filtering, thereby providing a robust framework for the subsequent verification of quality control performance.

2.1. Equipment Models and Data

This study aims to break through the limitations of traditional single-station and short-term datasets and construct a multi-dimensional cloud-radar dataset that spans space, time, equipment models and weather scenarios, providing essential support for model generalization and quality-control reliability. The detailed characteristics of this dataset are as follows.

We selected Ka-band baseline observations from a total of 49 cloud radar stations across 24 provinces in China. The locations and elevations of these stations are shown in Figure 1. The dataset covers one year of observations (from 1 July 2023 to 30 June 2024) and

includes typical complex weather scenarios such as snow, rain, fog, and low clouds. The adopted cloud radars incorporate four manufacturer models YLU3, HTKAAA, YWKAAA and HMBKPS all operating under a unified technical architecture. These systems share a common core design featuring fully solid-state hardware, pulse compression and Doppler signal processing, ensuring the consistency and comparability of key observational variables such as radar reflectivity, radial velocity and spectral width. This uniformity provides a solid foundation for cross-model cloud-radar data quality control. All radars strictly comply with the China meteorological observation technical specifications and undergo standardized annual inspections and calibrations. Core calibration parameters include transmit power, receiver sensitivity, antenna-pointing accuracy and Doppler velocity measurement accuracy. These measures effectively safeguard the stability of radar observations and the reliability of the acquired data, ensuring the suitability of collected data for model training and operational meteorological applications.

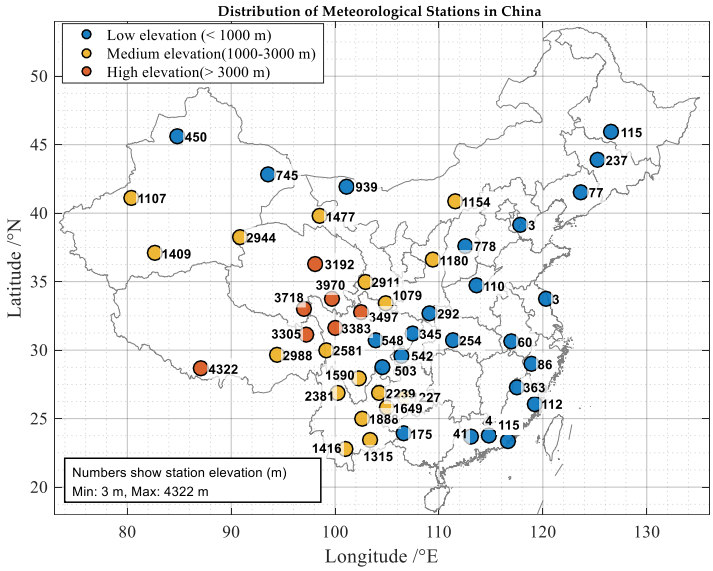


Figure 1. Distribution of cloud-radar stations.

The dataset includes the key parameters for both dual-polarization and single-polarization radars. Dual-polarization parameters include echo height (H), reflectivity (Z), linear depolarization ratio (LDR), radial velocity (V), spectrum width (W), and signal-to-noise ratios for horizontal (SNR1) and vertical channels (SNR2). Single-polarization radars lack the LDR and SNR2. The temporal resolution of the observations is one minute, and the spatial resolution is 30 m. Each station contains 631 range bins, with the first 400 range bins uniformly extracted for subsequent processing to ensure the parametric data of the cloud radar exhibit consistent dimensions in terms of altitudinal distribution and temporal resolution.

Training labels were generated through a combination of manual annotation and automated quality control rule verification. Three engineers with more than five years of experience in radar-echo interpretation independently annotated randomly selected samples following standardized guidelines based on Z–LDR distribution characteristics, spatial continuity and consistency with surface observations. Inter-annotator agreement was required, and any discrepancies were resolved by a senior analyst through coordinated assessment with weather situation. For inconsistencies between manual annotations and automated validation, a hierarchical strategy was adopted, including multi-source data cross-checking, priority assignment to manual labels in weak-echo scenarios, and majority voting for ambiguous samples. A binary classification labeling scheme was ultimately applied, categorizing echoes as “clutter” (label 0) or “precipitation echo” (label 1). Label

accuracy was ensured by a three-stage labeling procedure: traditional threshold-based preliminary screening, ground-based observation verification, manual correction. During model training, the cloud-radar echo samples spanning from July 2023 to July 2024 were adopted. For each station, two to three representative events were selected and strictly separated in chronological order. 70% of the earlier samples served as the training set, and 30% of later independent samples acted as the test set. The test set therefore consisted of previously unseen time-series data, effectively preventing data leakage caused by temporal overlap.

During model training, the dataset was further divided into training and validation subsets at a 7:3 ratio. Grid search was employed to optimize hyperparameters such as the number of decision trees, learning rate and maximum tree depth, in order to achieve an appropriate balance between model complexity and overfitting risk. This optimization ensures stable classification performance across diverse weather conditions and different regions. Ultimately, more than 20 million valid range-bin samples were constructed, providing a highly reliable sample base for model training.

2.2. Methods

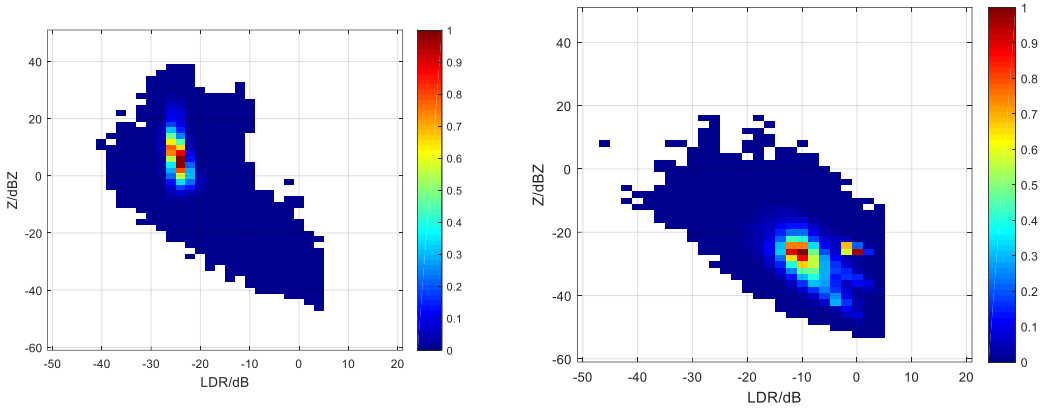
2.2.1. Recognition of Meteorological Echo Features

In machine-learning-based clutter recognition, accurately distinguishing between meteorological and non-meteorological echoes is essential for extracting valid meteorological information [40]. Z and LDR, as the key observation parameters of cloud radars exhibit joint distribution characteristics that reflect the scattering physics of different echo targets [41]. To clearly illustrate the differences in Z–LDR distributions between meteorological and non-meteorological echoes, echo identification is first conducted based on a combination of observational labeling and physical characteristics. Meteorological echoes, including rainfall, snowfall and fog, are identified using synchronous ground-based observations and their dynamic evolution features, while non-meteorological echoes are determined based on static spatial patterns, the absence of physically consistent meteorological evolution, and manual verification. All samples of these two echo categories undergo strict quality-control procedures, which include eliminating outliers and observations with data gaps, thereby ensuring the reliability of the Z–LDR distribution analysis. The joint Z–LDR distribution of meteorological echoes (such as precipitation) and non-meteorological echoes (such as clear-sky clutter) is statistically analyzed and compared. The objective is to find the inherent physical signatures of non-meteorological echoes, thereby supporting the quality control of cloud-radar data and accurate meteorological information retrieval.

Through statistically analyzing the correlation of echoes with multi-source data (i.e., ground-based observations, all-sky cloud images, and near-surface temperature and humidity measurements), an accurate correspondence between weather types and echo features is established, providing an interpretable feature guidance for model training.

Reflectivity factor (Z) represents the backscattering intensity of precipitation particles per unit volume; higher values indicate more particles, larger diameters, and stronger precipitation. Linear depolarization ratio (LDR) is defined as the ratio of the vertically polarized received power to the horizontally polarized received power when the radar transmits horizontally polarized waves. LDR quantifies the shape characteristics of detected targets, lower LDR values (e.g., below −20 dB) imply weaker depolarization effects, corresponding to regularly shaped, uniformly oriented particles (such as small raindrops); higher LDR values (e.g., above −15 dB) indicate stronger depolarization effects, associated with irregularly shaped, randomly oriented targets commonly found in non-meteorological clutter [42]. Comparing the joint Z–LDR distributions for meteorological and non-meteorological echoes (Figure 2), we find marked differences between them. Specifically, the precipitation echo

shows a compact distribution, with strong reflectivity and low depolarization, owing to the uniform scattering of near-spherical raindrops. In contrast, the clear-sky clutter shows a highly dispersed distribution, with weak reflectivity (Z centered between −40 dBz and −20 dBz) and high depolarization (LDR spanning from −15 dB to 0 dB), reflecting the complex scattering associated with atmospheric turbulence and ground clutter. These distinctions provide essential physical criteria for clutter recognition.



(**a**) (**b**)

Figure 2. Composite reflectivity–linear depolarization ratio distributions of (**a**) meteorological (rainfall) and (**b**) non-meteorological (clear-sky clutter) echoes.

Combined with the ground-based meteorological observation, precise identification criteria for snow, rain, fog and low clouds are further established as follows. Snowfall requires air temperature ≤ 0 °C, with Z typically ranging from −20 to 10 dBz and LDR between −25 and −15 dB [43]. Fog corresponds to surface relative humidity $\geq 85\%$, visibility ≤ 1 km and Z ≤ 5 dBz, and its identification is aided by cloud imagery captured by all-sky cloud imagers. Low-cloud systems are characterized by echo-top height ≤ 2 km, Z in the range of 5–20 dBz, and LDR between −25 and −15 dB. Rainfall typically exhibits Z between 10 dBz and 40 dBz, LDR in the range of −30–−20 dB [44,45], and V varying from -5 ms^{-1} to 10 ms^{-1}. By establishing the physical correlation between surface conditions and cloud-radar parameters, misidentification of weather types is minimized, which provides accurate scenario labels for the model.

2.2.2. Data Quality Control Algorithm Based on LightGBM Combined with Neighborhood Filtering

Based on the efficient fitting ability of the LightGBM algorithm for high-dimensional features and its advantage in capturing nonlinear relationships, a clutter-classification model is constructed following the workflow of feature engineering, label definition and training optimization. Feature selection takes both physical attributes and spatial distribution into account. Specifically, physical attribute characteristics include the key radar parameters (Z, LDR, V, W, SNR1, SNR2 and H), capturing differences in motion and scattering between precipitation particles and clutter. In addition, spatial correlation features are innovatively introduced to characterize the contrast between the spatial continuity of meteorological echoes and the discreteness of clutter. The spatial correlation feature is calculated based on a 3×3 spatiotemporal neighborhood frequency method applied to the cloud radar's height-time two-dimensional grid points. Here, i denotes the range bin and j the time step of a cloud-radar echo grid point, with each pair (i, j) uniquely corresponding to a target spatiotemporal echo grid point. Centered on this target grid point, a 3×3 spatiotemporal neighborhood window is constructed. This window includes the target point itself, along with eight neighboring points—the upper and lower adjacent

range bins $(i \pm 1)$ and the left and right adjacent scan times $(j \pm 1)$, resulting in a total of nine spatiotemporal grid points. The calculation method is as follows.

$$P(i,j) = \frac{N_{met}(W_{3\times3}(i,j))}{N_{total}(W_{3\times3}(i,j))} \quad (1)$$

$N_{met}(W_{3\times3}(i,j))$ refers to the number of valid echo points with non-null values in the window $P(i,j)$ at (i,j), relative to the total number of data points in the window. The value ranges from 0 to 1, and its magnitude directly quantifies the continuity characteristic of cloud-radar echoes in the temporal-two-dimensional space. When $P(i,j)$ approaches 1, it indicates that the spatiotemporal continuity of echoes in the target grid is strong and conforms to the natural distribution characteristics of meteorological echoes such as snow, fog, low clouds, and rainfall. When $P(i,j)$ approaches 0, it means there are almost no valid meteorological echoes around the target grid point, the echoes exhibit obvious spatial discreteness, and the probability of non-meteorological echoes is high.

To address the limitation that the LightGBM model is prone to retaining isolated clutter in complex weather conditions, neighborhood filtering with physical constraints is innovatively introduced. This technology leverages the spatial continuity characteristics of meteorological echoes to achieve a balance between clutter rejection and preservation of weak meteorological signals, thereby avoiding the loss of weak echoes such as snow and fog caused by the "one-size-fits-all" approach of traditional filtering.

Weather-adaptive window design: The weather-adaptive window design is to dynamically adjust the neighborhood size according to spatial scale differences in different weather echoes. To validate the rationality of this design, this study compared the baseline LightGBM method with the use of 3×3, 5×5 and 7×7 neighborhood filtering windows, based on 9600 base data samples (5760 for snowfall scenarios and 3840 for rainfall scenarios) (Figure 3). The results indicate that in snowfall scenarios, the 5×5 window achieved the highest comprehensive score (CS = 94.14), as its larger size better captures the spatial continuity of loose snow echoes. In rainfall scenarios, the 3×3 window performed best (CS = 95.33), since the smaller window can finely characterize the local features of compact rain echoes. The experiments demonstrate that selecting window sizes based on echo structure (5×5 for snow, 3×3 for rain) effectively enhances the balance between clutter suppression and signal retention.

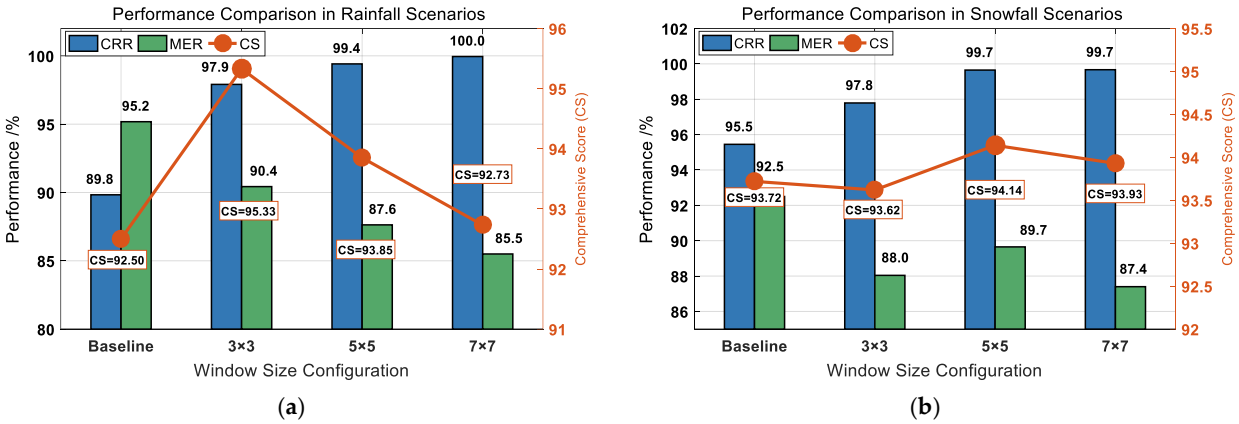


Figure 3. Performance comparison of different window sizes (**a**) rainfall scenarios; (**b**) snowfall scenarios.

Quantitative criteria of isolation degree: Define echo isolation degree as Isolation $= \frac{\text{Number of nonnull points in the neighborhood}}{\text{Total number of points}}$. Figure 4 presents the receiver operating characteristic analysis of isolation-degree thresholds. Following machine-learning-based

processing, an isolation-degree threshold of approximately 60% is identified to yield an optimal trade-off between clutter removal and signal retention. A threshold exceeding 60% is prone to excessive elimination of weak meteorological echoes, whereas a lower threshold leads to insufficient clutter suppression. Based on both experimental results and prior experience, a 60% threshold is employed in this study [4]. When the isolation degree < 60%, it is determined as residual clutter (marked as an invalid value), and when the isolation degree $\geq$ 60%, it is retained as a valid meteorological echo. This method shows high calculation efficiency and strong adaptability, complementing the LightGBM model.

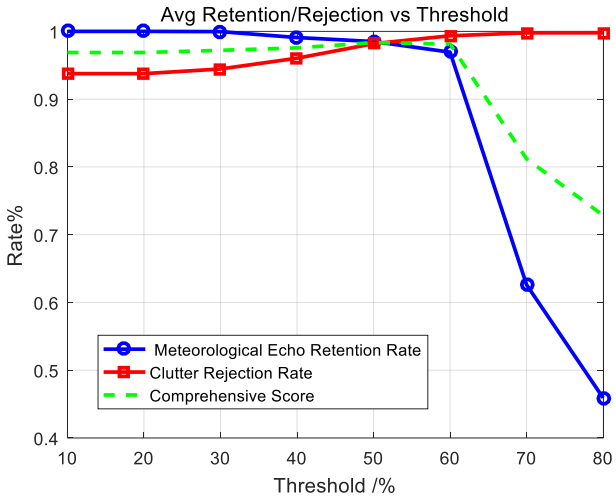


Figure 4. Performance metrics at different isolation-degree thresholds based on receiver operating characteristic analysis.

2.2.3. Performance Evaluation Metrics

In this study, we establish two key indicators, namely clutter rejection rate and meteorological-echo retention rate, to measure the performance of quality control. Clutter rejection rate (the proportion of correctly recognized noise to total noise) reveals the noise-suppression capability of the model. Meteorological-echo retention rate (the proportion of correctly preserved meteorological echoes to total valid meteorological echoes) reveals the signal-preservation capability of the model. The performance of quality control is comprehensively judged by the average score of the two indicators. This evaluation framework avoids the limitations of relying on a single indicator and provides a unified and objective standard for the subsequent assessment of quality-control performance across multiple weather types and stations. The proposed algorithm was deployed on a general-purpose hardware platform equipped with 16 GB of Random Access Memory and 8-core virtual Central Processing Units. Under this configuration, the algorithm is capable of completing real-time quality control of approximately 1.6 GB of hourly cloud-radar raw data from 20 stations within 3 min.

3. Evaluation Results

To comprehensively validate the performance of the "LightGBM + neighborhood filtering" method on data quality control in complex weather scenarios, This study first analyzes representative cases under various weather conditions (snowfall, rainfall, fog, low-level clouds and dust), and compares the cloud-radar data quality control performance of the Random Forest, the LightGBM algorithm and the proposed "LightGBM + neighborhood filtering" method. Using surface observations as references, the original echoes and the quality-controlled results from the three methods are systematically compared. The results indicate that although the Random Forest algorithm achieves basic clutter sup-

pression and meteorological echo retention, its clutter rejection rate is relatively low under complex weather conditions, and its ability to identify and preserve weak meteorological echoes (e.g., snowfall and fog) is limited, resulting in moderate overall comprehensive quality-control performance.

In addition, statistical analyses are conducted to calculate weighted scores of the quality control algorithms under different weather scenarios based on sample duration, which enables a quantitative evaluation of the performance differences among quality control methods. These analyses ultimately verify the effectiveness, stability and cross-scenario adaptability of the proposed method.

In the snowfall scenarios, cloud radar echoes are susceptible to interference from near-surface discrete clutter. The clutter reflectivity is primarily concentrated between −40 dBz and −30 dBz, spatially overlapping with snowfall echoes. Traditional methods often suffer from "incomplete clutter removal" or "loss of weak snowfall echoes". In this research, we selected two typical snowfall cases that occurred at Shenyang station (No. 54340) from 0000 UTC to 1200 UTC on 30 January 2024 (Figure 5a) and at Anqing station (No. 58424) from 0000 UTC to 2300 UTC on 18 December 2024 (Figure 5b).

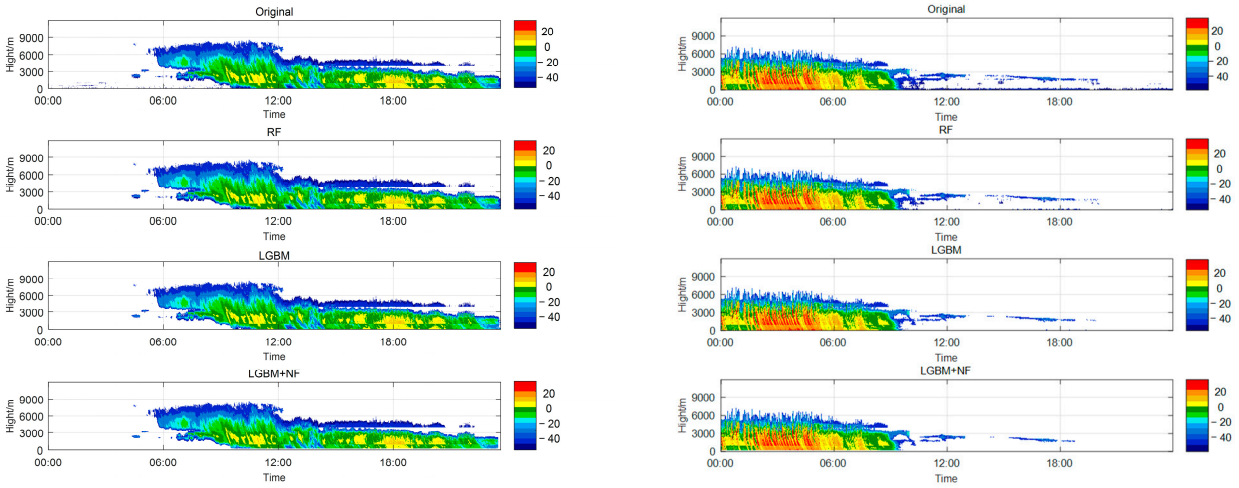


Figure 5. Echoes of the snowfall cases (**a**) at Shenyang station in Liaoning Province from 0000 UTC to 1200 UTC on 30 January 2024; (**b**) at Anqing station in Anhui Province from 0000 UTC to 2300 UTC on 18 December 2024.

From Figure 5, we find that the original echoes exhibit substantial discrete clutter below 1 km near the surface. The vertical structure and intensity distribution of snow echoes are obscured. Specifically, although the Random Forest algorithm removes a portion of clutter, it yields an average clutter rejection rate of 84.71%, a meteorological echo retention rate of 89.08% and a comprehensive score of 86.89. In particular, in the Harbin station case (25 December 2023), its clutter rejection rate drops to 67.75%, which indicates that the Random Forest algorithm exhibits limited suppression capability under dense clutter conditions. After LightGBM-only quality control, although most clutter is preliminarily filtered out, some isolated noise points remain. After processing with the "LightGBM + neighborhood filtering" method, the residual clutter can be precisely recognized and eliminated, based on the spatial continuity of snowfall echoes. Specifically, in the case at Shenyang station, the time-height distribution features of snowfall echoes are relatively clearer. In the case at Anqing station, the clutter-induced fuzzy boundaries are fully eliminated, while the intensity gradient of the snow echo is preserved. The assessment results (Table 1) indicate that in the snowfall scenarios, the average clutter rejection rate of the

"LightGBM + neighborhood filtering" method reaches 98.14%. This represents an improvement of 1.99% over the LightGBM algorithm (96.15%) and is substantially higher than that of the Random Forest algorithm (84.71%). The clutter rejection rate increases from 99.28% to 100% at Harbin station (No. 50953) and rises from 99.59% to 99.88% at Shenyang station (No. 54340). Although the average meteorological echo retention rate decreases slightly from 91.51% to 90.18%, the comprehensive score reaches 94.16, higher than the scores of the LightGBM algorithm (93.85) and Random Forest algorithm (86.89). These results demonstrate the ability of the proposed method to balance noise suppression and preservation of weak snowfall echoes.

Table 1. Comparison of the scores of quality control results in the snowfall scenario.

Station/No.	Time	RF			LGBM			LGBM + NF		
		CRR (%)	MER (%)	CS	CRR (%)	MER (%)	CS	CRR (%)	MER (%)	CS
Harbin/50953	25 December 2023	67.75	96.85	82.30	99.28	95.67	97.47	100.00	95.17	97.59
Harbin/50953	27 February 2024	71.54	93.00	82.27	94.88	91.39	93.14	97.05	90.18	93.61
Anqing/58424	18 December 2024	99.54	68.81	84.17	90.91	85.77	88.40	95.63	84.36	89.99
Shenyang/54340	30 January 2024	100.00	97.67	98.83	99.59	93.21	96.40	99.88	90.99	95.43
Average Score		84.71	89.08	86.89	96.17	91.51	93.85	98.14	90.18	94.16

Note: RF = Random Forest; LGBM = Light Gradient Boosting Machine; NF = Neighborhood Filtering; LGBM + NF = Light Gradient Boosting Machine with neighborhood filtering; CRR = Clutter Rejection Rate; MER = Meteorological Echo Retention Rate; CS = Comprehensive Score.

In rainfall scenarios, heavy precipitation echoes are prone to overlap with near-surface dense clutter, masking the spatial structure of valid signals. In this study, we select two typical cases for analysis, which occurred at Zhengzhou station (No. 57083) from 0000 UTC to 2000 UTC on 19 April 2024 (Figure 6a) and at Shantou station (No. 59316) from 0200 UTC to 0800 UTC on 12 May 2024 (Figure 6b). For the original echoes at Zhengzhou station, the strongest reflectivity exceeded 20 dBz, with extensive discrete clutter below 3 km near the surface and at the echo edge. From 1500 UTC to 1800 UTC, during the short-term rainfall episode, the precipitation echo and the clutter exhibited complete overlap. After the initial clutter rejection by the LightGBM, the isolated clutter still remains at the edge. After adding neighborhood filtering, the isolated clutter is eliminated, and the precipitation echo exhibits clear and continuous boundaries. In addition, the intensity gradient and temporal-spatial distribution of the precipitation echo are completely retained, and there is no misidentification of valid signals. The precipitation echo was weak at Shantou station, and thus the single LightGBM method tends to miss clutter. The joint method, through neighborhood frequency verification (when the isolation degree < 60%, it is identified as clutter), not only eliminates the surface clutter but also completely retains the weak precipitation echo.

As shown in Table 2, in the rainfall scenarios, the average clutter rejection rate of the "LightGBM + neighborhood filtering" reaches 96.37%, representing a 2.33% improvement over the LightGBM algorithm 94.04% and a marked enhancement compared with the Random Forest (85.16%). The clutter rejection rate increases from 98.40% to 99.66% at Shantou station (No. 59316) and rises from 93.72% to 97.22% at Zhengzhou station (No. 57083). The comprehensive score is 95.13, highlighting its strong adaptability; the proposed method not only improves clutter-rejection efficiency but also minimizes the loss of valid precipitation signals.

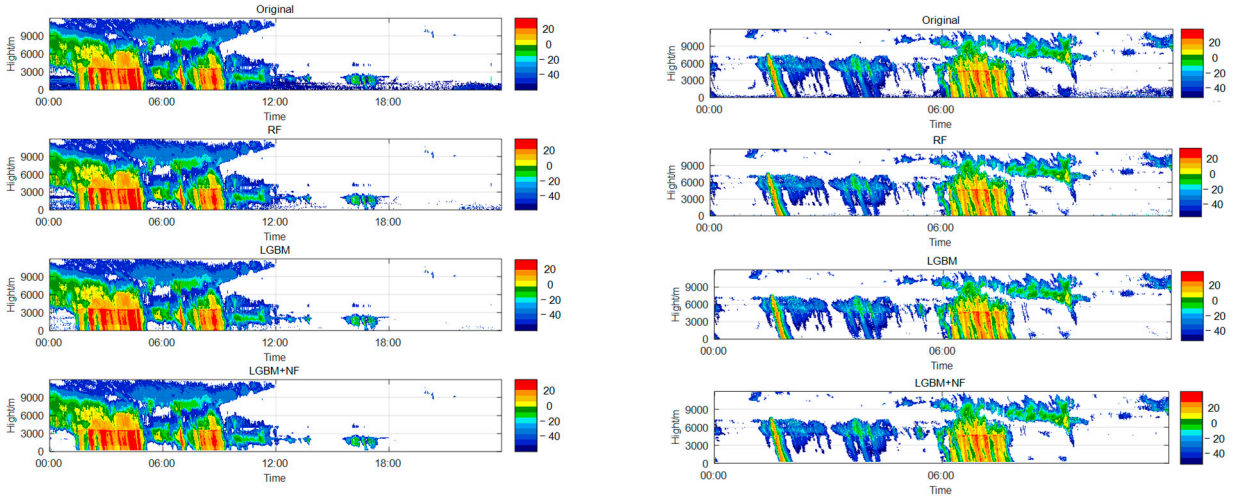


Figure 6. Same as Figure 5, but for precipitation cases (**a**) at Zhengzhou station in Henan Province from 0000 UTC to 2000 UTC on 19 April 2024; (**b**) Shantou station in Guangzhou Province from 0200 UTC to 0800 UTC on 12 May 2024.

Table 2. Comparison of the scores of quality control result scores in the precipitation scenario.

Station/No.	Time	RF			LGBM			LGBM + NF		
		CRR (%)	MER (%)	CS	CRR (%)	MER (%)	CS	CRR (%)	MER (%)	CS
Zhengzhou/57083	19 April 2024	73.08	97.64	85.36	93.72	95.62	94.66	97.22	95.33	96.28
Qingyuan/59280	8 May 2024	79.30	98.26	88.78	86.91	93.12	90.01	89.94	92.51	91.22
Ankang/57245	6 June 2024	93.42	97.06	95.23	97.13	93.80	95.47	98.64	92.80	95.72
Shantou/59316	12 May 2024	94.84	93.16	94.00	98.40	95.72	97.06	99.66	94.92	97.29
Average Score		85.16	96.53	90.84	94.04	94.57	94.30	96.37	93.89	95.13

Note: RF = Random Forest; LGBM = Light Gradient Boosting Machine; NF = Neighborhood Filtering; LGBM + NF = Light Gradient Boosting Machine with neighborhood filtering; CRR = Clutter Rejection Rate; MER = Meteorological Echo Retention Rate; CS = Comprehensive Score.

In the scenarios of low clouds, the main quality-control challenge lies in low echo-top height ($\leq$2 km) and weak intensity on echo edges, and thus, low-cloud echoes are easily confused with clutter. In this study, we select the low cloud cases for analysis, which occurred at Taiyuan station (No. 53772) in Shanxi Province from 0000 UTC to 1600 UTC on 11 June 2024 (Figure 7a) and at Karamay station (No. 51243) in Xinjiang from 1000 UTC to 1400 UTC on 14 June 2024 (Figure 7b). The original echo at Taiyuan station showed that the bottom edge of the low cloud echo was at an altitude of about 1 km, and its intensity varied from -30 dBz to -15 dBz. The near-surface isolated clutter overlapped with the weak echo at the edge. The strong clutter can be filtered by the LightGBM alone, but this method misclassifies part of the echo at the edge with a weak intensity. The combined neighborhood filtering thoroughly eliminates the near-surface clutter and fully retains the weak echo of 5–10 dBz at the edge. As a result, the vertical structure of the echo is clearer, and the bottom edge of the low clouds at Karamay station is at an altitude of about 3 km. The non-meteorological clutter is concentrated below 1 km. After the "LightGBM + neighborhood filtering" processing, the residual clutter below 1 km can be basically eliminated, and the intensity and spatial extent of the low-cloud echo remain unaffected.

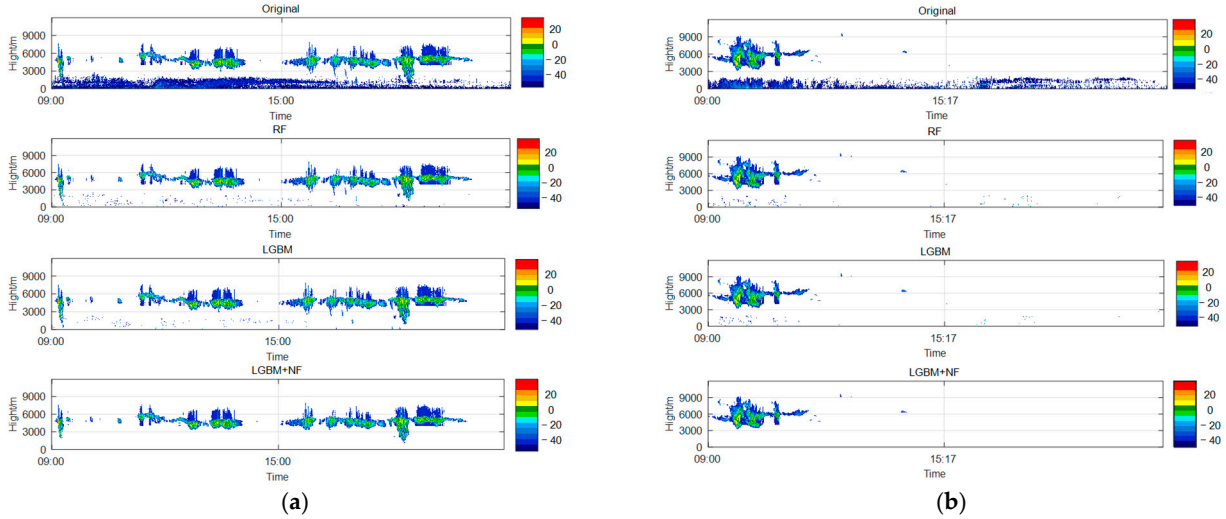


Figure 7. Same as Figure 5, but for low clouds (**a**) at Taiyuan station in Shanxi Province from 0000 UTC to 1600 UTC on 11 June 2024; (**b**) Karamay station in Xinjiang from 1000 UTC to 1400 UTC on 14 June 2024.

According to Table 3, in the low cloud scenarios, the performance of the Random Forest is also inferior to that of the LightGBM and the "LightGBM + neighborhood filtering" method in low cloud scenarios. At Karamay station, the Random Forest clutter rejection rate is 96.44%, and at Taiyuan station, it reaches 98.11%; both substantially lower than the 99.83% achieved by the "LightGBM + neighborhood filtering" method. Although the Random Forest algorithm attains a relatively high meteorological echo retention rate, these results still expose a critical drawback of its insufficient clutter suppression capability. The average clutter rejection rate of the "LightGBM + neighborhood filtering" is 97.35%, 1.42% higher than that of the single machine learning algorithm 95.93%. The clutter rejection rate increases from 98.56% to 99.98% at Karamay station (No. 51243) and rises from 98.11% to 99.83% at Taiyuan station (No. 53772). The comprehensive score is 96.49, fully demonstrating the excellent performance of the proposed method in low cloud scenarios.

Table 3. Comparison of the scores of quality control results across different stations in the low cloud scenario.

Station/No.	Time	RF			LGBM			LGBM + NF		
		CRR (%)	MER (%)	CS	CRR (%)	MER (%)	CS	CRR (%)	MER (%)	CS
Wenjiang/56187	22 January 2024	82.78	99.46	91.12	97.96	89.55	93.76	98.37	88.34	93.36
Taiyuan/53772	11 June 2024	96.85	98.52	97.68	98.11	99.92	99.02	99.83	98.73	99.28
Simao/56964	13 November 2023	85.35	91.49	88.42	89.08	97.67	93.38	91.20	95.68	93.44
Karamay/51243	14 June 2024	96.44	99.90	98.17	98.56	100.00	99.28	99.98	99.79	99.89
Average Score		90.35	97.34	93.85	95.93	96.79	96.36	97.35	95.64	96.49

Note: RF = Random Forest; LGBM = Light Gradient Boosting Machine; NF = Neighborhood Filtering; LGBM + NF = Light Gradient Boosting Machine with neighborhood filtering; CRR = Clutter Rejection Rate; MER = Meteorological Echo Retention Rate; CS = Comprehensive Score.

Fog echoes are typically weak (5 dBz) and concentrate near the surface, making them easily confused with near-surface turbulent clutter. The traditional threshold filtering is prone to the loss of fog echoes. In this research, we selected two typical fog cases for analysis, which occurred at Ankang station (No. 57245) in Shaanxi Province from 0800 UTC to 2000 UTC on 4 June 2024 (Figure 8a) and at Shapingba in Chongqing station at around 1200 UTC on 24 April 2024 (Figure 8b). The original echo at Ankang station showed that the fog echo at 1–2 km and the turbulent clutter below 1 km near the surface overlapped,

making fog echo coverage indistinguishable. The quality control of the LightGBM alone can filter part of the clutter, but the edge of the fog echo remains fuzzy. Due to the continuous distribution of near-surface fog echo, the joint method proposed in this research accurately removes the discrete turbulence clutter. As a result, the near-surface distribution of the fog echo at 1–2 km and the intensity pattern between −20 dBz and 5 dBz are clearer. The fog echo at Shapingba station in Chongqing between 1.2 km and 3 km suffered severe interference from the near-surface clutter below 1 km. Joint quality control algorithm completely eliminates the clutter at the edge of the fog echo and fully retains weak fog echo intensity without attenuation.

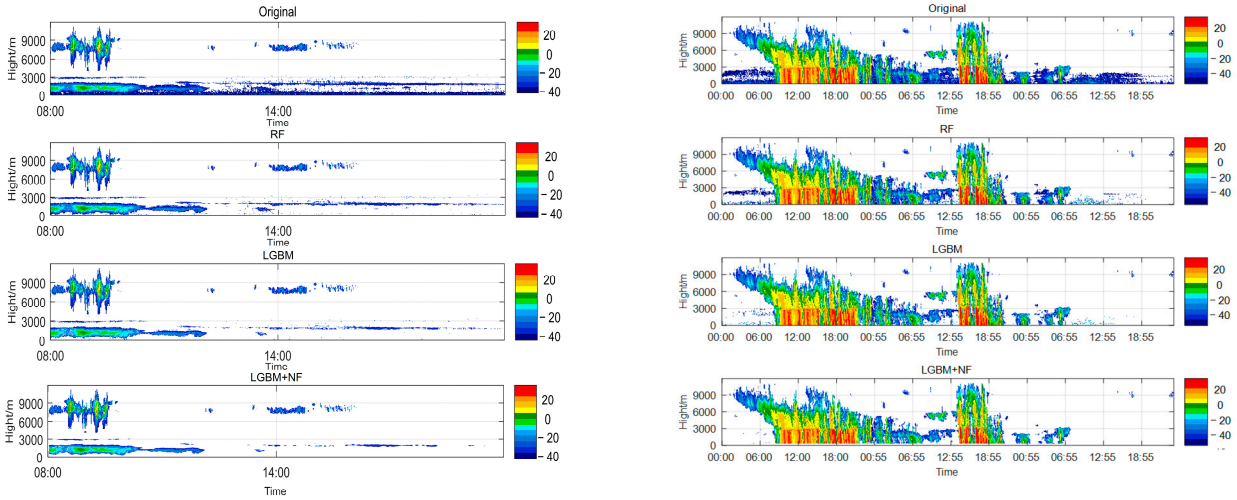


Figure 8. Same as Figure 5, but for fog cases (**a**) at Ankang station in Shaanxi Province from 0800 UTC to 2000 UTC on 4 June 2024; (**b**) Shapingba station in Chongqing at 1200 UTC on 24 April 2024.

Seen from Table 4, we find that in the fog scenarios, the average clutter rejection rate of the "LightGBM + neighborhood filtering" method is notably higher than that of the Random Forest (78.62%), exceeding the LightGBM algorithm by 2.73%. The clutter rejection rate increases from 98.30% to 99.30% at Ankang station (No. 57245) and rises from 99.22% to 100% at Zhengzhou station (No. 57083). Although the meteorological echo retention rate is slightly lower than that of the single algorithm, the comprehensive score 91.11 is still higher than that of the single algorithm 90.22. This indicates that the proposed method can effectively remove clutter in fog scenarios while maintaining fog-echo integrity, providing reliable data support for fog monitoring.

Table 4. Comparison of the scores of quality control results in the fog scenario.

Station/No.	Time	RF			LGBM			LGBM + NF		
		CRR (%)	MER (%)	CS	CRR (%)	MER (%)	CS	CRR (%)	MER (%)	CS
Ankang/57245	4 June 2024	87.60	82.08	84.84	98.30	79.10	88.70	99.30	78.03	88.67
Shapingba/57516	23 April 2024	69.02	97.37	83.19	81.37	90.74	86.05	89.64	90.07	89.85
Zhengzhou/57083	4 February 2024	76.96	88.11	82.53	99.22	91.20	95.21	100.00	90.60	95.30
Dali/56046	25 March 2024	80.90	77.99	79.44	97.73	84.09	90.91	98.61	82.61	90.61
Average Score		78.62	86.39	82.50	94.16	86.28	90.22	96.89	85.33	91.11

Note: RF = Random Forest; LGBM = Light Gradient Boosting Machine; NF = Neighborhood Filtering; LGBM + NF = Light Gradient Boosting Machine with neighborhood filtering; CRR = Clutter Rejection Rate; MER = Meteorological Echo Retention Rate; CS = Comprehensive Score.

For the dust scenario, two representative cases from Aksu (Xinjiang) and Ejin Banner (Inner Mongolia) are selected based on ground-based observations (Figure 9). Seen from Table 5, the results show that the "LightGBM + neighborhood filtering" method outperforms the comparative models in both recognition rate (96.47%) and comprehensive score (95.79). Since the number of dust-weather cases is limited, and the selected cases feature low clutter interference and high data quality, these conclusions require further verification in more complex weather conditions. Future work will involve the systematic collection of additional rare weather cases to improve model robustness under imbalanced sample distributions.

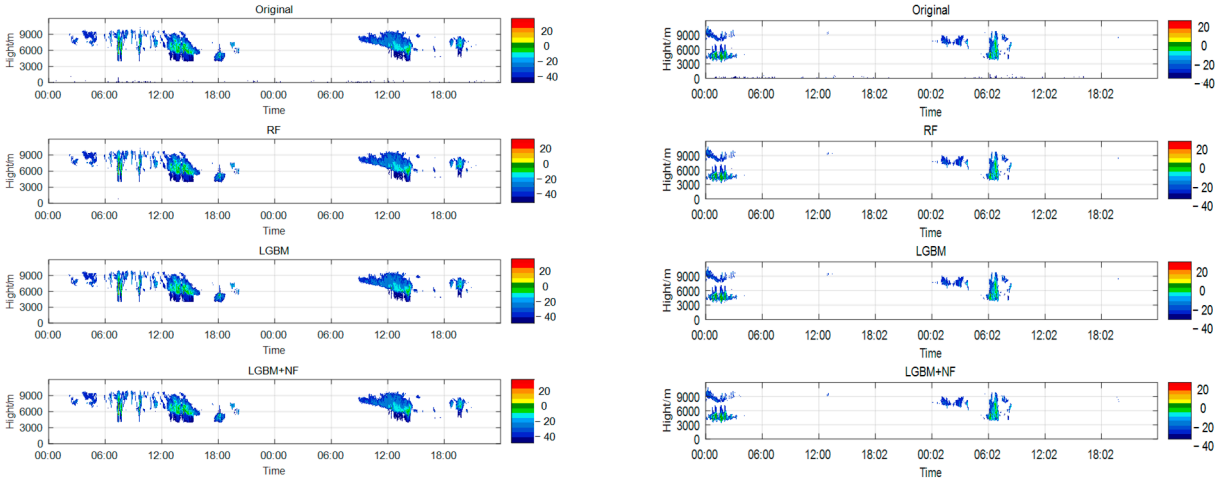


Figure 9. Same as Figure 5, but for dust (**a**) at Akesu station in Xinjiang Province from 0000 UTC on 22 December 2023 to 2348 UTC on 23 December 2023; (**b**) Ejin Banner station in Inner Mongolia from 0000 UTC to 1800 UTC on 18 May 2024.

Table 5. Comparison of the average scores of quality control performance at Akesu and Ejin Banner stations in the dust scenario.

Station/No.	Time	RF			LGBM			LGBM + NF		
		CRR (%)	MER (%)	CS	CRR (%)	MER (%)	CS	CRR (%)	MER (%)	CS
Akesu/51628	4 June 2024	93.82	99.99	96.91	100.00	96.98	98.49	96.06	99.13	97.59
Ejin Banner/52267	23 April 2024	85.08	99.91	92.50	85.43	99.72	92.58	96.88	91.10	93.99
Average Score		89.45	99.95	94.70	92.71	98.35	95.53	96.47	95.12	95.79

Note: RF = Random Forest; LGBM = Light Gradient Boosting Machine; NF = Neighborhood Filtering; LGBM + NF = Light Gradient Boosting Machine with neighborhood filtering; CRR = Clutter Rejection Rate; MER = Meteorological Echo Retention Rate; CS = Comprehensive Score.

To comprehensively evaluate the overall performance of the proposed method, Table 6 and Figure 10 summarize the test results from independent cases across different weather scenarios, with a total duration of 853 h. Performance scores are calculated via weighted averaging based on observation durations, providing a more objective representation of data contribution in different scenarios. The scenario-specific results indicate that the "LightGBM + neighborhood filtering" method consistently achieves higher clutter rejection rates than the LightGBM-only approach across all weather conditions. For example, in the snowfall scenario, the clutter rejection rate increases from 97.76% to 99.08%, and in the dust scenario, from 92.72% to 96.47%. Although the meteorological echo retention rate decreases slightly, the comprehensive score exhibits a stable improvement, with an increase ranging

from 0.02 to 0.25. The most pronounced enhancement in the meteorological echo retention rate is observed in the dust scenario (from 95.54% to 95.79%).

Table 6. Comparison of average quality-control effect scores in multiple scenarios.

Weather Scenario	Total Duration (h)	Weight	LGBM			LGBM + NF			Improvement
			CRR(%)	MER(%)	CS	CRR(%)	MER(%)	CS	
Snow	255	37.50%	97.76	83.66	90.71	99.08	82.37	90.73	0.02
Rain	208	30.50%	94.01	92.67	93.34	96.46	92.01	94.23	0.89
Low Cloud	106	15.60%	95.93	96.79	96.36	97.35	95.64	96.49	0.13
Fog	248	36.40%	96.51	85.2	90.86	98.42	83.79	91.10	0.24
Dust	36	5.30%	92.72	98.35	95.54	96.47	95.12	95.79	0.25
Weighted Average	853	100%	96.58	86.93	91.76	98.47	85.79	92.13	0.37

Note: LGBM = Light Gradient Boosting Machine; NF = Neighborhood Filtering; LGBM + NF = Light Gradient Boosting Machine with neighborhood filtering; CRR = Clutter Rejection Rate; MER = Meteorological Echo Retention Rate; CS = Comprehensive Score.

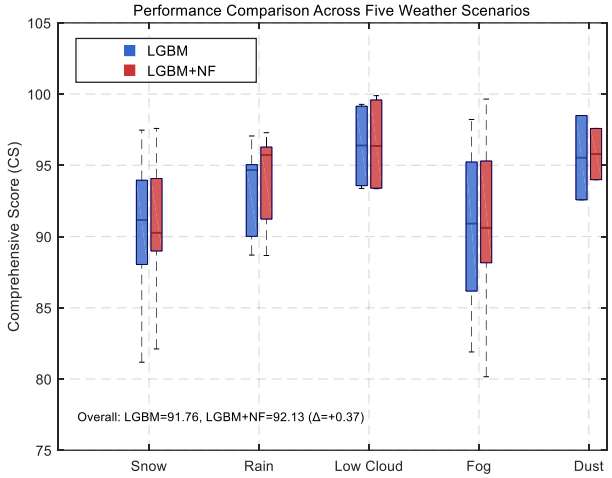


Figure 10. Performance distribution of quality-control methods evaluated.

The weighted-average results further confirm the superiority of the "LightGBM + neighborhood filtering" method. The weighted-average clutter rejection rate reaches 98.47%, representing an improvement of 1.91% over the LightGBM-only method. In addition, the weighted-average comprehensive score increases to 92.13%, corresponding to an improvement of 0.37%. These results demonstrate that the "LightGBM + neighborhood filtering" approach offers enhanced quality control stability and cross-scenario adaptability.

4. Discussion

This study aims to solve the key challenges of limited non-meteorological clutter identification accuracy and spatial mixing between meteorological and non-meteorological echoes in cloud radar observations under complex weather scenarios. Combined with ground-based meteorological observations, Ka-band cloud radar base data for the one-year period of 2023–2024 from 49 stations (covering 24 provinces) in China are selected to construct a training dataset comprising over 20 million sample sets using a dual-mechanism approach of "manual labeling + automatic quality control verification". Based on this, we propose a two-step quality control method integrating the LightGBM model and neighborhood filtering technology and have completed the method construction for cloud-radar data quality control, performance verification, and cross-scenario adaptability assessment under various weather scenarios.

The proposed method achieves an effective balance between clutter rejection and meteorological echo retention, outperforming the single machine-learning algorithms. Its weighted average clutter rejection rate reaches 98.47%, which is higher than that of the single LightGBM method (96.58%). Its weighted average meteorological echo retention rate is 85.79%, and the weighted average comprehensive score of the proposed method is 92.13, higher than that of the single LightGBM method (91.76). In addition, the "LightGBM + neighborhood filtering" method exhibits stable performance across various complex meteorological scenarios. In the four scenarios with substantial sample support—rain, snow, low clouds, and fog—the comprehensive scores are 94.23, 90.73, 96.49, and 91.10, respectively. Notably, even in the dust scenario, where the sample size is relatively small, the proposed method achieves a clutter rejection rate of 96.47% and a comprehensive score of 95.79, maintaining stable performance compared with the LightGBM-only method, which further confirms its robust cross-scenario applicability. The proposed method establishes a data foundation for cloud-radar-based algorithms such as precipitation particle type retrieval and precipitation estimation. Nonetheless, this method still requires further improvements. For instance, the meteorological echo retention rate needs to be further enhanced in snow scenarios, and the recognition accuracy between weak echoes at the edge of low clouds and clutter remains insufficient.

Future research will focus on further optimizing the proposed approach. First, efforts will be made to reduce reliance on manual labeling and develop semi-supervised or unsupervised learning frameworks that incorporate multi-source observations, such as temperature, humidity, cloud fraction and satellite images, to enhance the objectivity of distinguishing low clouds from near-surface clutter. Second, the weather adaptivity of neighborhood filtering thresholds should be quantitatively assessed, and dynamic threshold adjustment mechanisms tailored to different weather types should be established to further enhance the retention and identification accuracy of weak meteorological echoes near cloud boundaries. Moreover, the fusion framework proposed in this study demonstrates potential for extension to emerging observation platforms, including unmanned aerial vehicles and satellite-borne radars. Future research will explore the deployment of this method on such emerging platforms to support data quality control for integrated air–space–ground meteorological observation networks.

5. Conclusions

This study proposes a novel technical pathway for cloud-radar data quality control by integrating machine learning with physical filtering. The innovative two-step "machine learning + neighborhood filtering" framework leverages the fitting advantage of LightGBM for high-dimensional features and employs the K-neighborhood frequency method to eliminate residual clutter, thereby effectively resolving the contradiction between clutter rejection and valid echo retention. Supported by a standardized, multi-regional, long-term, and multi-model dataset, the method overcomes the generalization limitations of traditional single-station, short-term data. The main findings and quantitative results are summarized as follows:

- Enhanced clutter rejection capability: The proposed method achieves a weighted average clutter rejection rate of 98.47%, which is 1.89 percentage points higher (improvement) than that of the standalone LightGBM algorithm (96.58%) and substantially outperforms the Random Forest algorithm.
- Balanced echo retention and overall performance: The weighted average meteorological echo retention rate is 85.79%, and the weighted average comprehensive score reaches 92.13 (verification statistic), representing a 0.37-point improvement over the

standalone LightGBM (91.76). This demonstrates that a slight trade-off in echo retention yields a significant gain in overall quality control performance.

- Robust cross-scenario adaptability: Under five complex weather scenarios—snow, rain, low clouds, fog, and dust—the method achieves comprehensive scores of 90.73, 94.23, 96.49, 91.10, and 95.79, respectively. These scores are consistently higher than those of both the Random Forest and standalone LightGBM, confirming the method's stability, adaptability, and operational robustness across diverse conditions.
- Reusable paradigm and extensibility: The proposed algorithm can be extended to the quality control of X-band and C-band radar data, thereby providing a more reliable data foundation for refined meteorological observation and product generation.

Author Contributions: Conceptualization, C.H. and F.T.; methodology, W.L.; investigation, C.H. and S.H.; writing—original draft preparation, C.H. and W.L.; writing—review and editing, F.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China (Grant No. 2022YFC3004101) and the Joint Fund Project for Collaborative Innovation in Meteorological Science and Technology of Northeast China (Grant No. 2026MS003).

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

H	Height
Z	Reflectivity
LDR	Linear Depolarization Ratio
V	Radial Velocity
W	Spectrum Width
SNR	Signal-to-noise Ratio
LGBM	Light Gradient Boosting Machine
NF	Neighborhood Filtering
LGBM + NF	Light Gradient Boosting Machine with neighborhood filtering
CRR	Clutter Rejection Rate
MER	Meteorological Echo Retention Rate
CS	Comprehensive Score
ROC	Receiver Operating Characteristic

References

1. Arakawa, A.; Schubert, W.H. Interaction of a cumulus cloud ensemble with the large-scale environment, Part I. *J. Atmos. Sci.* **1974**, *31*, 674–701. [CrossRef]
2. Liu, L.; Xie, L.; Cui, Z. Examination and application of Doppler spectral density data in drop size distribution retrieval in weak precipitation by cloud radar. *Chin. J. Atmos. Sci.* **2014**, *38*, 223–236.
3. Yang, L.; Wang, Y.; Wang, Z.K.; Yang, Q.; Fan, X.G.; Tao, F.; Zhen, X.Q.; Yang, Z.P. Automatic identification of clear-air echoes based on millimeter-wave cloud radar measurements. *Adv. Atmos. Sci.* **2020**, *37*, 912–924. [CrossRef]
4. Shi, Z.; Huang, L.; Wu, F.; Lei, Y.; Wang, H.; Tang, Z. An Improved Multi-Threshold Clutter Filtering Algorithm for W-Band Cloud Radar Based on K-Means Clustering. *Remote Sens.* **2024**, *16*, 4640. [CrossRef]
5. Kou, L.; Lin, Z.; Gao, H.; Liao, S.; Ding, P. Simulation and sensitivity analysis for cloud and precipitation measurements via spaceborne millimeter-wave radar. *Atmos. Meas. Tech.* **2023**, *16*, 1723–1744. [CrossRef]
6. Yao, Z.; An, L.; Zhang, P.; Gao, L.; Jia, S.; Wang, W. Characteristics of Cloud Water Resource and Precipitation Efficiency of Hydrometeors over Northwest China. *J. Meteorol. Res.* **2023**, *37*, 353–369. [CrossRef]
7. Hu, K.; Zhang, D.; Xia, M. CDUNet: Cloud Detection UNet for Remote Sensing Imagery. *Remote Sens.* **2021**, *13*, 4533. [CrossRef]

8. Liu, L.; Zheng, J. Algorithms for doppler spectral density data quality control and merging for the Ka-band solid-state transmitter cloud radar. *Remote Sens.* **2019**, *11*, 209. [CrossRef]
9. Zeng, Z.M.; Zheng, J.F.; Yang, H.; Zheng, M.; Zeng, Y. Quality control and evaluation on non-cloud echo of Ka-band cloud radar. *J. Appl. Meteorol. Sci.* **2021**, *32*, 347–357.
10. Zheng, J.F.; Liu, L.P.; Zeng, Z.M.; Xie, X.L.; Wu, J.Y.; Feng, K. Ka-band millimeter wave cloud radar data quality control. *J. Infrared Millim. Waves* **2016**, *35*, 748–757.
11. Habib, A.; Kersting, A.P.; Bang, K.I.; Lee, D.-C. Alternative Methodologies for the Internal Quality Control of Parallel LiDAR Strips. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 221–236. [CrossRef]
12. Frisch, A.S.; Fairall, C.W.; Snider, J.B. Measurement of stratus cloud and drizzle parameters in ASTEX with a Kα-band Doppler radar and a microwave radiometer. *J. Atmos. Sci.* **1995**, *52*, 2788–2799. [CrossRef]
13. Taconet, O.; Benallegue, M.; Vidal-Madjar, D.; Prevot, L.; Dechambre, M.; Normand, M. Estimation of soil and crop parameters for wheat from airborne radar backscattering data in C and X bands. *Remote Sens. Environ.* **1994**, *50*, 287–294. [CrossRef]
14. Wang, W.M.; Xu, B.L.; Lei, Y.; Zhang, Y.; Chen, L. Analysis of precipitation echoes from Ka-band millimeter wave cloud radar on the southeast margin of the Tibetan Plateau. *Meteorol. Mon.* **2024**, *50*, 291–302.
15. Petitdidier, M.; Sy, A.; Garrouste, A.; Delcourt, J. Statistical characteristics of the noise power spectral density in UHF and VHF wind profilers. *Radio Sci.* **1997**, *32*, 1229–1247. [CrossRef]
16. Yu, C.; Hu, X.; Lu, Y.; Wu, W.; Liu, D. A Machine-Learning-Based Cloud Detection and Cloud-Top Thermodynamic Phase Algorithm over the Arctic Using FY3D/MERSI-II. *Remote Sens.* **2025**, *17*, 3128. [CrossRef]
17. Pincus, R.; Platnick, S.; Ackerman, S.A.; Hemler, R.S.; Patrick Hofmann, R.J. Reconciling Simulated and Observed Views of Clouds: MODIS, ISCCP, and the Limits of Instrument Simulators. *J. Clim.* **2012**, *25*, 4699–4720. [CrossRef]
18. Wu, X.; Shi, Z.; Zou, Z. A geographic information-driven method and a new large scale dataset for remote sensing cloud/snow detection. *ISPRS J. Photogramm. Remote Sens.* **2021**, *174*, 87–104. [CrossRef]
19. Ghasemian, N.; Akhoondzadeh, M. Introducing two Random Forest based methods for cloud detection in remote sensing images. *Adv. Space Res.* **2018**, *62*, 288–303. [CrossRef]
20. Li, J.; Wu, Z.; Hu, Z.; Zhang, J.; Li, M.; Mo, L.; Molinier, M. Thin cloud removal in optical remote sensing images based on generative adversarial networks and physical model of cloud distortion. *ISPRS J. Photogramm. Remote Sens.* **2020**, *166*, 373–389. [CrossRef]
21. Zhang, Z.J.; Zhang, J.; Wu, G.S. A method for identifying abnormal wind direction at automatic weather stations based on machine learning. *Meteorol. Mon.* **2025**, *51*, 460–472.
22. Li, Q.; Qi, Y.; Zhang, Z.; Yang, Y.; Zhu, Z.; Wang, N.; Hu, Q. Nonmeteorological Echoes Identification Method Based on Bayesian Classifier and Echo Physical Characteristics Using C-Band Radar and Its Performance. *Chin. J. Atmos. Sci.* **2024**, *48*, 823–836.
23. Zhao, M.; Song, H.; Chen, Y.; Wang, L.; Liu, T. Semi-Supervised Classification of PolSAR Images Based on Co-Training of CNN and SVM with Limited Labeled Samples. *Sensors* **2023**, *23*, 2109. [CrossRef]
24. Liu, X.; Zhao, Z. SVM Based Land/Sea Clutter Classification Algorithm. *Appl. Mech. Mater.* **2012**, *236–237*, 1156–1160. [CrossRef]
25. An, Q.; Pan, Z.; Wang, L.; You, H. Target Detection Network for SAR Images Based on Semi-Supervised Learning and Attention Mechanism. *Remote Sens.* **2021**, *13*, 2686–2707.
26. Besson, O.; Pascal, F. Learning Strategies for Radar Clutter Classification. *IEEE Trans. Signal Process.* **2021**, *69*, 1070–1082. [CrossRef]
27. He, Z.; Wang, L.; Chen, X. Semantic Feature-Enhanced Graph ATtention Network for Radar Target Recognition in Heterogeneous Radar Network. *IEEE Trans. Aerosp. Electron. Syst.* **2023**, *59*, 1748–1760.
28. Saim, A.A.; Aly, M.H. Fusion-Based Approaches and Machine Learning Algorithms for Forest Monitoring: A Systematic Review. *Wild* **2025**, *2*, 7. [CrossRef]
29. Lei, W.; Wang, Y.; Zhong, J.; Xu, Q.; Jiang, Y.; Li, C. A Review of Clutter Suppression Techniques in Ground Penetrating Radar: Mechanisms, Methods, and Challenges. *J. Electron. Inf. Technol.* **2025**, *47*, 4079–4095.
30. Yang, Q.; Feng, Y.; Guan, L.; Wu, W.; Wang, S.; Li, Q. X-Band Radar Attenuation Correction Method Based on LightGBM Algorithm. *Remote Sens.* **2023**, *15*, 864. [CrossRef]
31. Ye, Y.; Liu, F.; Zhao, S.; Liang, Z. Ensemble Learning Based on GBDT and CNN for Adoptability Prediction. *Comput. Mater. Contin.* **2020**, *65*, 1361–1372. [CrossRef]
32. Kim, H.J.; Yoon, S.S.; Gu, Y.H.; Kim, S.H.; Choi, Y.-D.; Kwon, H.-H. A Hybrid Approach to Physical and Deep Learning Models for Radar-Based Precipitation Nowcasting. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 1–16. [CrossRef]
33. Cieślik, K.; Milczarek, W. Application of Machine Learning in Forecasting the Impact of Mining Deformation: A Case Study of Underground Copper Mines in Poland. *Remote Sens.* **2022**, *14*, 4755. [CrossRef]
34. Li, Z.; Lu, Z.; Li, Y.; Liu, X. SwinNowcast: A Swin Transformer-Based Model for Radar-Based Precipitation Nowcasting. *Remote Sens.* **2025**, *17*, 1550. [CrossRef]

35. Li, J.; Zhou, Z.; Feng, R.; Wang, L.; Zhang, H. Transformer-Based Cloud Detection Method for High-Resolution Remote Sensing Imagery. *J. Appl. Remote Sens.* **2024**, *18*, 661. [CrossRef]
36. Zhou, C.; Wei, C.; Yang, F.; Wei, J. A Quality Control Method for High Frequency Radar Data Based on Machine Learning Neural Networks. *Appl. Sci.* **2023**, *13*, 11826. [CrossRef]
37. Williams, P.D.L. Limitations of radar techniques for the detection of small surface targets in clutter. *Radio Electron. Eng.* **2010**, *45*, 379–389. [CrossRef]
38. Peters, J.F.; Suraj, Z.; Shan, S.; Ramanna, S.; Pedrycz, W.; Pizzi, N. Classification of meteorological volumetric radar data using rough set methods. *Pattern Recognit. Lett.* **2003**, *24*, 911–920. [CrossRef]
39. Harrison, D.L.; Driscoll, S.J.; Kitchen, M. Improving precipitation estimates from weather radar using quality control and correction techniques. *Meteorol. Appl.* **2010**, *7*, 135–144. [CrossRef]
40. Zhang, L.; You, W.; Wu, Q.M.J.; Qi, S.; Ji, Y. Deep Learning-Based Automatic Clutter/Interference Detection for HFSWR. *Remote Sens.* **2018**, *10*, 1517. [CrossRef]
41. Chen, Y.; Jin, Y.; Ding, D.; Bi, K.; Ma, X.; Li, Y. Preliminary Analysis on the Application of Millimeter Wave Cloud Radar in Snow Observation. *Chin. J. Atmos. Sci.* **2018**, *42*, 134–149.
42. Chen, P.; Song, J.; Bai, Y.; Wang, J.; Du, Y.; Tian, L. Range-Velocity Measurement Accuracy Improvement Based on Joint Spatiotemporal Characteristics of Multi-Input Multi-Output Radar. *Remote Sens.* **2024**, *16*, 2648. [CrossRef]
43. Zou, W.; Yin, Z.; Zhang, J.; Dai, Y.; Chen, Y.; Bu, Z.; Yun, H.D.; Lu, T.; Wei, Y.; Wang, X. Ka-band Cloud Radar Meteorological Echo Dataset with Complex Weather Coverage: A Baseline Model for Deep Learning Applications. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 1–16. [CrossRef]
44. Chen, Y.; Liu, X.; Bi, K.; Zhao, D. Hydrometeor Classification of Winter Precipitation in Northern China Based on Multi-Platform Radar Observation System. *Remote Sens.* **2021**, *13*, 5070. [CrossRef]
45. Grazioli, J.; Tuia, D.; Berne, A. Hydrometeor classification from polarimetric radar measurements: A clustering approach. *Atmos. Meas. Tech.* **2015**, *8*, 149–170. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.