remote sensing

MDPI

Article

A Machine Learning Model for FY-4A Cloud Detection Based on Physical Feature Fusion

Yanning Liang [1], Li Zhao [2], Yuan Sun [2], Zhihao Feng [1], Xiaogang Huang [1] and Wei Zhong [2,*]

[1] College of Meteorology and Oceanography, National University of Defense Technology, Changsha 410073, China
[2] College of Advanced Interdisciplinary Studies, National University of Defense Technology, Changsha 410073, China
* Correspondence: zhongwei@nudt.edu.cn

Highlights

What are the main findings?

- An MLP-TSAR model incorporating meteorological factors was developed.
- Achieved a 9.7% improvement compared with the FY-4A cloud detection product.
- Reduced the overall accuracy difference between day and night from 3% to 0.7%.

What is the implication of the main finding?

- Meteorological factors drove over 30% of MLP-TSAR predictions, day and night.

Abstract

Clouds critically influence Earth's radiation balance and climate, making accurate cloud detection essential for improving climate models. This study develops the TSAR model to improve the cloud detection accuracy of the FY-4A CLM product by incorporating physical features. The input features include FY-4A brightness temperature (BT) data from channels 8–14, geometric parameters (satellite zenith angle (SAZ), satellite azimuth angle (SAA), solar zenith angle (SOZ), solar azimuth angle (SOA), and latitude), and four ERA5 meteorological factors (2 m air temperature (T2m), skin temperature (SKT), air temperature profiles (ATP), and relative humidity profiles (RH)). Using the CALIPSO cloud detection product as labels, the model outputs cloud/clear-sky classification results. Additionally, four machine learning (ML) algorithms—RF, LightGBM, XGBoost, and MLP—achieved overall accuracies of 91.5%, 92.2%, 92.5%, and 92.8%, respectively, considerably outperforming the FY-4A L2 CLM product (83.1%). The results demonstrate that incorporating physical factors significantly improves cloud detection performance regardless of the algorithm employed. Incorporating meteorological factors notably improved nighttime and water–cloud detection, narrowing day–night accuracy gaps. Shapley additive explanation (SHAP) analysis indicated feature contributions of 15.8%, 50.8%, and 33.3% from geometric, BT, and meteorological variables, respectively, with stronger meteorological effects at mid- to high-latitudes. These findings demonstrate that integrating meteorological factors significantly improves FY-4A cloud detection accuracy and consistency, highlighting the MLP-TSAR model's effectiveness for reliable all-day operational applications.

Keywords: cloud detection; machine learning model; MLP; Shapley additive explanations (SHAP); Fengyun-4A (FY-4A)

Check for updates

Academic Editor: Yaoming Ma

Received: 5 December 2025
Revised: 22 January 2026
Accepted: 23 January 2026
Published: 7 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.

1. Introduction

According to the International Satellite Cloud Climatology Project (ISCCP), the global mean annual cloud cover is approximately 66% [1,2]. As an essential component of the Earth's atmosphere system, clouds influence global climate change by absorbing and scattering solar and infrared radiation and also affect the atmospheric environment through photochemical processes [3–6]. In satellite remote sensing, cloud coverage has a significant impact on the accuracy and reliability of parameter retrieval. Therefore, accurately retrieving the macro- and microphysical parameters of clouds from satellite observations is crucial for improving our understanding of cloud processes. Among these, cloud detection serves as a fundamental step that directly affects the quality of subsequent processes such as aerosol property retrieval, cloud type classification, cloud phase identification, and the retrieval of cloud optical thickness and effective particle radius [7–9].

In the last few decades, cloud detection techniques—such as ground-based remote sensing, in situ observations, and satellite remote sensing—have been extensively studied. Compared with the first two methods, satellite remote sensing offers the advantages of wide spatial coverage and high spatiotemporal resolution, enabling cloud observations on regional and global scales [10,11]. Passive satellite sensors, represented by the operational Moderate-resolution Imaging Spectroradiometer (MODIS), the Visible Infrared Imaging Radiometer Suite (VIIRS), the Advanced Very High-Resolution Radiometer (AVHRR), and the Advanced Himawari Imager (AHI), perform cloud detection by setting channel thresholds and utilizing differences in spectral information and spatial structures between clouds and underlying surfaces [12,13]. Generally, clouds exhibit higher reflectance and lower temperatures than the underlying surface. Although the traditional threshold-based method is simple and computationally efficient, its limitations have become increasingly evident [14]. On one hand, the threshold must be adjusted for different satellite payloads or even for the same payload under different observation conditions, which limits the applicability of the method. On the other hand, under complex surface types (e.g., snow- or ice-covered areas) and special atmospheric conditions (e.g., thin cirrus or nocturnal low-level stratus clouds), where the spectral characteristics of clouds and the surface are similar, the performance of traditional threshold-based methods decreases significantly [15–17].

With the improvement of computer technology and the optimization of machine learning (ML) algorithms, ML-based cloud detection methods have demonstrated significant advantages. As a typical binary classification problem, cloud detection essentially involves multivariate data analysis, which aligns well with the characteristics of ML methods [18–20]. Unlike traditional approaches that require manually defined thresholds or spectral pattern matching, ML can learn the hidden relationships between independent parameters and target variables from large training datasets, making the training process both efficient and flexible [21,22]. At present, ML models such as the Bayesian algorithm, Support Vector Machine (SVM), Convolutional Neural Network (CNN), Artificial Neural Network (ANN), and U-Net have achieved remarkable progress in cloud detection [23–28]. These models typically use observed radiance data as input features for training, combining both spectral and textural information [22]. Notably, some studies have introduced lidar and radar observations as reference labels, significantly improving the accuracy of pixel-level cloud identification [29,30]. In addition, atmospheric conditions, corresponding radiative transfer simulation data, and near-surface parameters have also been proven to be effective training datasets for improving the accuracy of ML-based cloud detection models [31,32]. Therefore, systematically quantifying the impact of atmospheric conditions and near-surface features on cloud detection algorithms and optimizing the accuracy of cloud cover products remain urgent scientific challenges.

The Fengyun-4A (FY-4A), launched in 2016, is China' s new-generation geostationary satellite. The Advanced Geostationary Radiation Imager (AGRI) onboard FY-4A provides a spatial resolution of up to 500 m and a 16-bit radiometric accuracy, with 14 spectral channels covering wavelengths from the visible to the thermal infrared bands. The National Satellite Meteorological Center (NSMC) provides a cloud mask product (FY-4A CLM) based on the traditional threshold method. Xu (2021) [33] evaluated the accuracy of FY-4A CLM over the Tibetan Plateau using MODIS CLM data and found that the accuracy for cloudy scenes was 93%, while that for clear-sky scenes was 73%. However, few studies have systematically validated this product using active radar observations.

Therefore, two critical limitations persist in current research. First, although ML approaches have been increasingly applied to cloud detection tasks, most existing studies rely predominantly on spectral features or single-level meteorological factors, lacking adequate integration of physical features that govern cloud formation processes. The vertical thermodynamic structure of the atmosphere, characterized by temperature and humidity profiles, fundamentally determines cloud development and maintenance, yet this three-dimensional physical information has rarely been incorporated into intelligent detection frameworks. Second, the validation of cloud detection products has traditionally depended on ground-based observations or other passive remote sensing datasets, which are inherently limited in their ability to capture the vertical distribution of cloud layers. The absence of high-precision active remote sensing data with three-dimensional vertical detection capabilities constrains the reliability of accuracy assessments.

In view of these limitations, this study proposes a physics-oriented ML framework for improving the cloud detection accuracy of the FY-4A Cloud Mask (CLM) product. This study employs the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) cloud detection data as the reference standard to evaluate the FY-4A CLM product, leveraging CALIPSO's capability for three-dimensional vertical detection of clouds and aerosols. The core innovation of this study lies in the development of a hierarchical feature engineering framework that systematically integrates physical features with spectral information. First, the BT model was constructed by combining the FY-4A brightness temperature (BT) data from infrared channels 8–14 with geometric parameters, including satellite zenith angle (SAZ), satellite azimuth angle (SAA), solar zenith angle (SOZ), solar azimuth angle (SOA), and latitude. These baseline features capture the fundamental radiative characteristics and observation geometry essential for cloud identification. Subsequently, the TSAR model was developed based on the BT model by incorporating meteorological factors derived from the ERA5 reanalysis dataset, including 2 m air temperature (T2m), land surface skin temperature (SKT), air temperature profiles (ATPs), and relative humidity (RH) profiles. Unlike previous research that relied mainly on threshold-based approaches or single-level meteorological variables, this work incorporated multi-dimensional atmospheric profiles to capture the vertical thermodynamic structure that governs cloud formation. Building upon the physics-oriented feature engineering framework, this study further compares the potential of four machine learning algorithms—Random Forest (RF), Light Gradient Boosting Machine (LightGBM), eXtreme Gradient Boosting (XGBoost), and Multilayer Perceptron (MLP)—in improving the cloud detection of the FY-4A CLM product. The optimal method was then selected to develop a day-night separated detection model based on the SOZ. To enhance the scientific credibility and operational reliability of the proposed framework, a Shapley Additive Explanations (SHAP)-based interpretability system was implemented. The SHAP analysis quantitatively links model outputs with atmospheric mechanisms, revealing the relative importance and interaction effects of individual features in the decision-making process, providing a

robust scientific foundation for the operational application of the improved FY-4A cloud detection products.

The remainder of this paper is organized as follows. Section 2 describes the data and methodology employed in this study. Section 3 presents the main results. Section 4 discusses the methodological advantages, physical mechanisms, and limitations. Section 5 is the conclusion.

2. Data and Methods

2.1. FY-4A AGRI Data

The AGRI, one of the main payloads onboard the FY-4A meteorological satellite, possesses highly efficient observation capabilities, requiring approximately 15 min to complete a full-disk scan. The instrument is equipped with 14 spectral bands, including three visible and near-infrared bands, three shortwave infrared bands, two mid-infrared bands, two water-vapor bands, and four longwave infrared bands [34]. In this study, Level-1 (L1) data from all spectral channels of FY-4A AGRI during 2019 and 2020 were selected as input features for model development. The detailed parameters of each spectral band are listed in Table 1. Considering the significant influence of observation geometry on data quality, key geometric parameters such as SAZ, SAA, SOZ, and SOA were also incorporated into the construction of the cloud detection model. In addition, the accuracy of the FY-4A Level-2 (L2) cloud detection product was evaluated and compared with that of the improved algorithm.

Table 1. Characteristics of AGRI (Advanced Geosynchronous Radiation Imager) channels [34].

Bands	Wavelength (μm)	Resolution (km)	Applications
1	0.47	1	Clouds, dust, and aerosols
2	0.65	0.5	Clouds, dust, and snow
3	0.825	1	Clouds, aerosols, vegetation, and ocean
4	1.375	2	Cirrus (ice crystal particles)
5	1.61	2	Low cloud, snow, water/ice cloud
6	2.25	2	Cirrus, aerosol
7	3.75 H	2	High albedo surface
8	3.75 L	4	Low albedo surface
9	6.25	4	Water vapor
10	7.1	4	Water vapor
11	8.5	4	Water vapor, cloud
12	10.7	4	Cloud, surface temperature
13	12.0	4	Water vapor, cloud, surface temperature
14	13.5	4	Water vapor, cloud

2.2. ERA5 Reanalysis Data

ERA5, developed by the European Centre for Medium-Range Weather Forecasts (ECMWF), is the fifth-generation global atmospheric reanalysis dataset with a spatial resolution of 0.25° × 0.25° and a temporal coverage extending from 1950 to the present. Based on an advanced four-dimensional variational assimilation system (4D-Var), ERA5 integrates satellite remote sensing data, ground-based observations, and numerical weather prediction model outputs. Compared with its predecessor ERA-Interim, improvements in the data assimilation scheme and physical parameterizations have significantly enhanced data accuracy [35]. In this study, ATP and RH at 1000, 925, 850, 700, 500, 300, 200, and 100 hPa, as well as hourly T2m and SKT, were selected for developing the cloud detection model. Cloud formation and evolution are closely related to atmospheric temperature:

when water vapor reaches saturation, temperature decreases lead to condensation and cloud formation. Therefore, significant differences exist in the vertical distributions of relative temperature and humidity between cloudy and clear-sky conditions. Clouds modify the environmental lapse rate through radiative–convective interactions, causing a sharp temperature gradient at cloud boundaries that deviates from the adiabatic lapse rate observed under clear-sky conditions (Figure 1a). The relative humidity within clouds typically approaches 100%, whereas it remains much lower in clear-sky regions (Figure 1b). Moreover, due to the radiative forcing effects of clouds, T2m and SKT exhibit noticeable differences between cloudy and clear-sky conditions, which can therefore help distinguish clouds from clear skies. Section 3 further evaluates the roles of these meteorological variables in cloud detection.

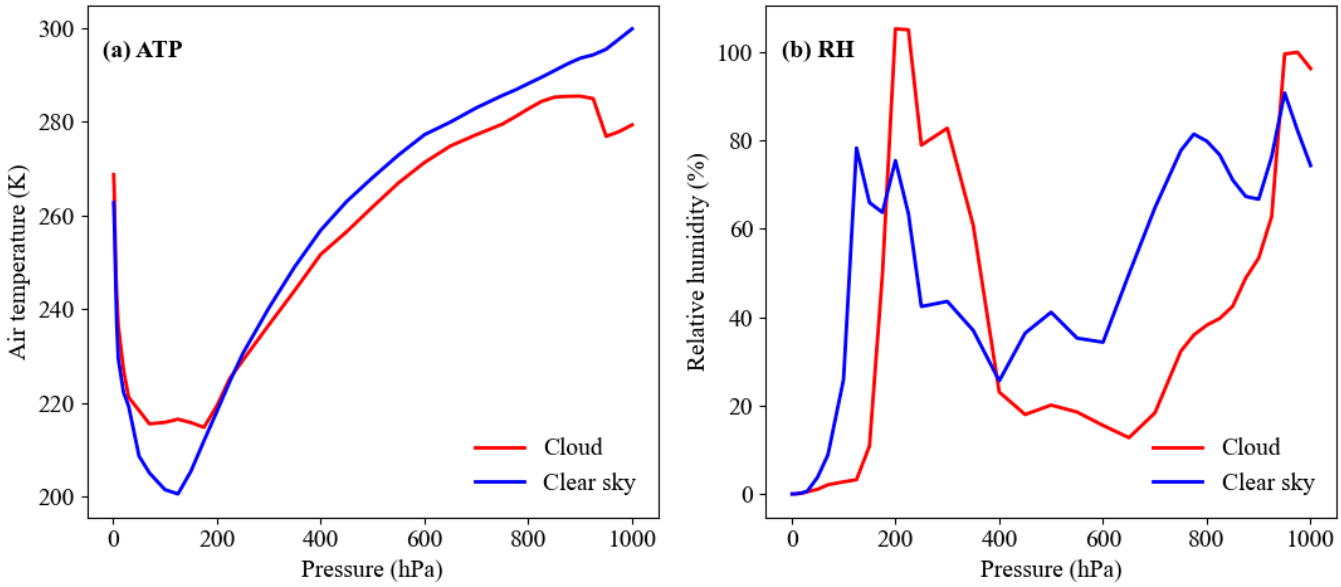


Figure 1. Individual (**a**) air temperature profile (ATP) and (**b**) relative humidity (RH) profile under cloudy and clear-sky conditions on 1 July 2020. The red and blue lines represent single-pixel profiles extracted from cloudy and clear-sky samples, respectively, illustrating the characteristic vertical structure differences between the two conditions.

2.3. CALIPSO Cloud Layer Product

The Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) mission was launched on 28 April 2006 and operated continuously until 1 August 2023. Its primary scientific objective was to accurately measure the vertical structure of the Earth's atmosphere and thereby investigate the roles of clouds and aerosols in the climate system and weather processes. The CALIPSO satellite is equipped with the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP), a dual-wavelength (532 and 1064 nm) polarization-sensitive lidar. This instrument provides high-vertical-resolution (30–60 m) atmospheric profiles and delivers highly accurate vertical distributions of clouds and aerosols [36]. In this study, Level-2 (L2) CALIPSO cloud layer data with a horizontal resolution of 1 km were used as ground-truth references for training and evaluating the cloud detection models.

2.4. Data Process

The FY-4A AGRI L1 full-disk data were preprocessed with quality inspection, geolocation, and radiometric calibration. The radiometric calibration converted raw digital count values into BTs for infrared channels using the calibration coefficients embedded in the data files. To ensure accurate spatiotemporal matching between FY-4A AGRI observations and CALIPSO cloud products, two collocation criteria were applied: (1) the time difference

between the two observations was required to be within 5 min, and (2) the spatial distance was limited to within 2 km. In addition, when five consecutive CALIPSO pixels yielded consistent classification results, the central pixel was retained to minimize the impact of temporal and spatial discrepancies between FY-4A and CALIPSO observations. Based on this collocation, the ERA5 reanalysis data were linearly interpolated in both time and space to match the corresponding pixel locations.

2.5. Machine Learning Methods

In ML-based cloud detection, the choice of algorithm directly affects the model's ability to represent complex spectral features and its generalization performance. To balance nonlinear feature extraction and model robustness, this study employed four algorithms—RF, LightGBM, XGBoost, and MLP—for cloud detection training. The first three possess efficient ensemble learning structures capable of handling high-dimensional and heterogeneous data, whereas MLP learns nonlinear mappings between spectral and spatial information through deep neural networks, thereby improving cloud recognition accuracy [20,25,37,38]. By comparing the performance of these four methods in terms of detection accuracy and feature adaptability, the optimal model incorporating meteorological variables for cloud detection was identified, providing methodological support for multi-source data-integrated cloud detection.

2.5.1. RF

RF solves classification and regression tasks by integrating multiple decision trees. It builds each tree on a bootstrap sample and selects split points from random feature subsets to enhance diversity. Final predictions are obtained by majority voting for classification or averaging for regression. Through this double randomness, RF reduces overfitting while maintaining high efficiency and interpretability, making it suitable for high-dimensional and nonlinear modeling [39,40].

2.5.2. LightGBM

LightGBM is an ensemble learning algorithm built on the gradient boosting decision tree (GBDT) framework. Its key optimizations include Gradient-based One-Side Sampling (GOSS), Exclusive Feature Bundling (EFB), the histogram-based algorithm, and a Leaf-wise growth strategy. Compared with traditional GBDT, LightGBM retains high-gradient samples via GOSS to maintain information gain accuracy, merges sparse features through EFB to reduce dimensionality, and speeds up split-point search using histogram discretization. By adopting a Leaf-wise rather than Level-wise growth strategy, it focuses on expanding nodes with the highest gain, thereby improving convergence speed, training efficiency, memory usage, and predictive accuracy on large-scale datasets [41].

2.5.3. XGBoost

XGBoost is also an ensemble learning algorithm based on GBDT. It builds multiple trees iteratively, combining their outputs to optimize model performance. By adding regularization terms (e.g., limits on leaf weights and node counts) to the objective function, XGBoost reduces complexity and mitigates overfitting. It uses a second-order Taylor expansion to approximate the loss function and adjusts sample weights through gradients and Hessians, allowing greater focus on samples with large errors. Additionally, XGBoost supports automatic feature selection, missing value handling, and parallel computation—enhancing training efficiency, generalization, and scalability for high-dimensional and imbalanced datasets [42].

2.5.4. MLP

MLP is a feedforward neural network that models complex nonlinear relationships by simulating biological neural processing. It consists of an input layer, multiple hidden layers, and an output layer, with full interconnections between adjacent layers through weight matrices. The model is trained via backpropagation: forward propagation computes predictions, gradients are obtained from a loss function (e.g., cross-entropy or mean squared error), and parameters are updated through gradient descent to minimize error. Nonlinear activation functions such as ReLU enable the network to capture higher-order feature interactions [43–45].

In this study, the four aforementioned methods are employed for cloud detection. A systematic comparison is conducted to evaluate their performance in terms of detection accuracy and feature adaptability, thereby providing methodological insights for multi-source data-driven cloud detection applications.

2.6. Model Training and Configuration

This study employed a temporal split strategy to partition the dataset, ensuring independent evaluation and preventing data leakage between the training and test sets. Specifically, data from the entire year of 2019 (January to December) were used as the training set, while data from the entire year of 2020 served as the independent test set (sample sizes are provided in Table S1). This approach evaluates the model's generalization ability across different time periods. Within the training set, an 80/20 random split was applied to create training and validation subsets, with a fixed random seed (random_state = 42) to ensure reproducibility. A grid search method was adopted for systematic hyperparameter tuning to optimize model performance. To robustly evaluate candidate configurations, five-fold stratified cross-validation was integrated during the training process (Figure S1). Model evaluation comprehensively considered accuracy, precision, recall, and F1-score, with the F1-score ultimately serving as the criterion for selecting the optimal hyperparameter configuration to balance precision and recall capabilities. To mitigate overfitting, all models implemented validation-based regularization mechanisms: the RF model utilized out-of-bag score estimation for internal validation, while LightGBM, XGBoost, and the MLP employed early stopping criteria based on validation set performance. The final hyperparameter configurations for the four models are summarized in Table S2.

All models were implemented in Python 3.8 using well-established ML libraries. RF was implemented using scikit-learn (version 1.0+). LightGBM and XGBoost were implemented using their respective official Python packages. The MLP model was constructed using TensorFlow/Keras (version 2.10.0).

2.7. McNemar's Test

McNemar's test is a non-parametric statistical test for paired categorical data, suitable for comparing the performance differences between two related classifiers. This test analyzes the classification results of two models on the same dataset, focusing on cases where the two models produce inconsistent results. Compared to directly comparing accuracies, McNemar's test can more effectively evaluate the statistical significance of differences between models, avoiding misjudgments caused by random fluctuations. In ML model comparisons, this method is widely used to verify whether a new model is statistically significantly superior to the baseline model [46].

2.8. Interpretation of Feature Contributions Based on SHAP

In this study, the SHapley Additive exPlanations (SHAP) method was employed to quantitatively evaluate the contribution of each physical feature to the model predictions,

i.e., feature importance. This approach is based on the Shapley value from cooperative game theory, which decomposes the model output into the sum of the marginal contributions of individual features, thereby quantifying each feature's impact on the prediction results [47,48]. Specifically, when a feature is introduced into the model, the system assigns a SHAP value to it according to the average marginal change in the model's predicted output. Considering two models f_S and $f_{S\cup\{i\}}$ trained with features in sets S and S∪{i}, respectively, SHAP values of the feature can be calculated as a weighted average of their differences:

$$\phi_i = \sum_{S\subseteq F\setminus\{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!}\left[f_{S\cup\{i\}}\left(x_{S\cup\{i\}}\right) - f_S(x_S)\right] \quad (1)$$

where F is the set of all features and S is all possible feature subsets ($S \subseteq F$). x_S and $x_{S\cup\{i\}}$ represent the values of the input features in sets S and $S \cup \{i\}$, respectively.

Then, the relative contribution of the feature is calculated by dividing its mean absolute SHAP values by the sum of mean absolute SHAP values of all considered features:

$$\phi_i = \frac{\overline{|\phi_i|}}{\sum_F \overline{|\phi_i|}} \quad (2)$$

3. Results

3.1. Overview of the Overall Accuracy

Figure 2 illustrates the clear-sky detection accuracy, as well as the cloud detection accuracy under water-cloud and ice-cloud conditions, for both the BT model (without meteorological factors) and the TSAR model (with four meteorological factors as additional inputs) by the RF, LightGBM, XGBoost, and MLP. The BT models of all four methods exhibit a pronounced day-night accuracy difference, particularly in clear-sky detection, where nighttime accuracy is 5–6% lower than that during the daytime (Figure 2a). This notable discrepancy is attributed to the inherent observational limitations of passive remote sensing techniques when detecting clear skies and clouds in the absence of solar radiation. Compared to the BT model, the inclusion of meteorological factors (Figure S2)—particularly the combination of all four—in the TSAR model significantly enhanced cloud detection accuracy at night. This improvement was most pronounced for water clouds, where nighttime accuracy increased by approximately 2%, even surpassing daytime performance (Figure 2e). Consequently, the TSAR model mitigated the substantial day–night discrepancy in accuracy observed with the BT model.

RF performs relatively weakly under clear-sky conditions, with both the BT model and the TSAR model exhibiting lower detection accuracy compared to other models (Figure 2a,e). LightGBM and XGBoost demonstrate consistent performance across various meteorological factor combinations (Figure S2), though LightGBM performs slightly worse than XGBoost in both clear-sky and cloud detection tasks (Figure 2). In contrast to the other three models, MLP exhibits superior comprehensive performance in cloud detection. Regardless of the inclusion of meteorological factors, its overall detection accuracy is significantly higher than that of the other three models (Figures S1 and S2 and Table S3). With the inclusion of meteorological factors, particularly in cloud detection, the detection accuracy of MLP-TSAR exceeds 92.5%, reaching over 96% under ice-cloud conditions and outperforming the other three models (Figure 2b,e), because MLP can perform nonlinear fusion of continuous BT and ATP data in a high-dimensional space to capture the complex physical interactions that determine cloud. In contrast, tree-based models like RF, LightGBM, and XGBoost are constrained by their axis-aligned splitting rules, making it difficult for them to efficiently learn smooth and nonlinear decision boundaries among high-dimensional continuous meteorological variables [49].

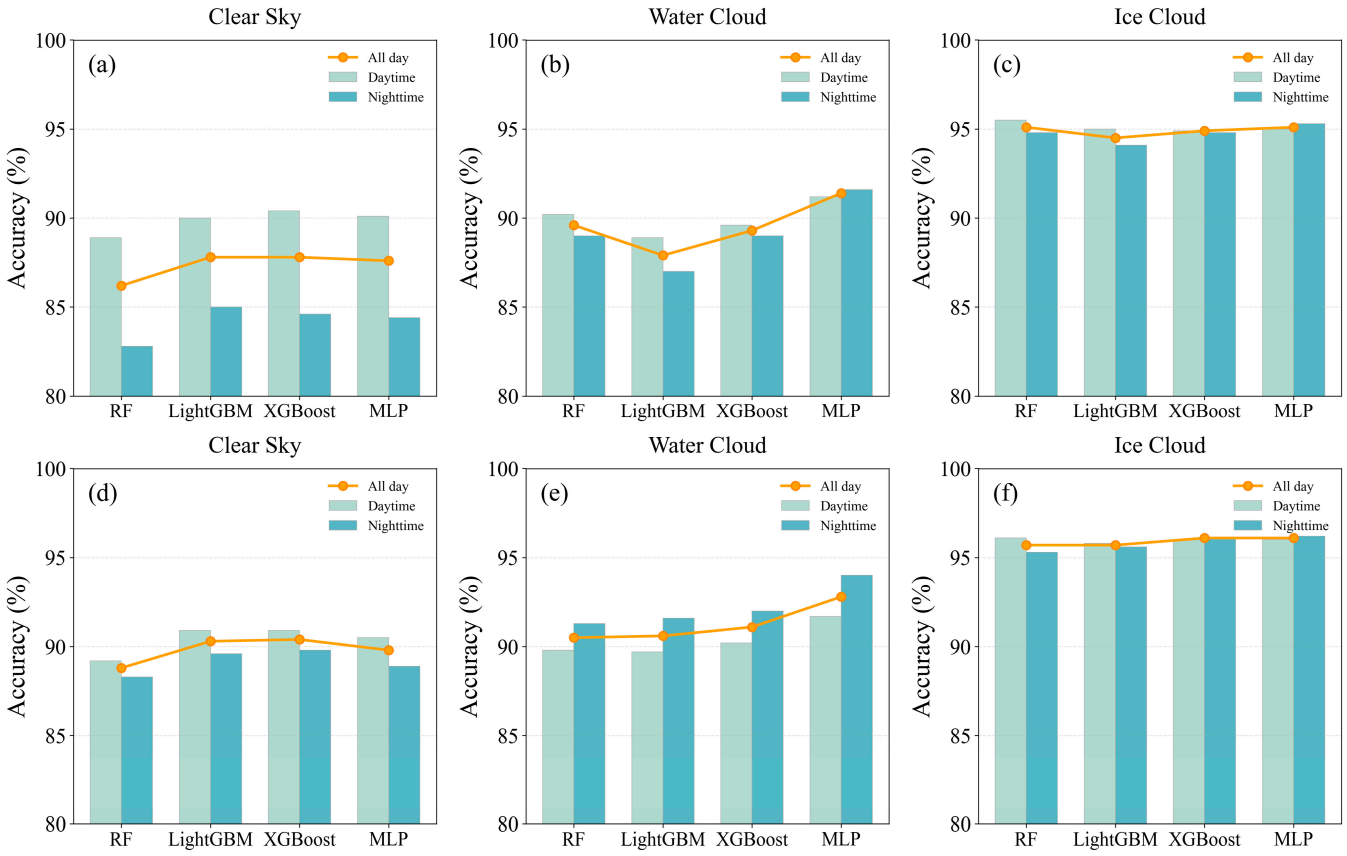


Figure 2. (**a**–**c**) Detection accuracy distributions of clear sky, water clouds, and ice clouds by the BT model, and (**d**–**f**) detection accuracy distributions by the TSAR model. The models include RF, LightGBM, XGBoost, and MLP under all-day, daytime, and nighttime conditions.

In summary, the incorporation of meteorological factors improves the performance of all four methods, particularly enhancing nighttime accuracy and reducing the day-night accuracy difference. Among the four methods, RF, LightGBM, and XGBoost perform slightly worse than MLP in cloud detection. Moreover, MLP maintains high detection accuracy regardless of time period or conditions, demonstrating strong robustness.

3.2. Comparison to FY-4A L2 Cloud Mask Products

The overall accuracy of the FY-4A L2 CLM product exceeds 80% (Table 2). During the daytime, the FY-4A L2 CLM product performs remarkably well in cloud detection, achieving an accuracy of over 95%, although its nighttime cloud detection performance decreases noticeably. In contrast, the accuracy of clear-sky detection reaches only about 70%. To further assess the limitations of the FY-4A L2 CLM product, this study illustrates the spatial distribution of detection accuracy (Figure 3). The clear-sky detection accuracy is lower near the disk edges, particularly in mid- to high-latitude regions, regardless of day or night (Figure 3a,d). In comparison, the daytime discrimination of water and ice clouds performs better across the entire disk (Figure 3b,c). However, the nighttime accuracy declines significantly, with weakened performance observed over land areas and low-latitude ocean regions (Figure 3e,f). The spatial error patterns at the disk edge are primarily driven by increased SAZ, which extends the atmospheric path length and enhances absorption and scattering effects, thereby reducing the BT contrast between clouds and clear-sky backgrounds. Additionally, the prevalence of oceanic and ice/snow-covered surfaces at the disk edge introduces spectral confusion that further compromises detection accuracy. The overall accuracies of the RF, LightGBM, XGBoost, and MLP (91.5%,

92.2%, 92.5%, and 92.8%) are all significantly higher than that of the FY-4A L2 CLM product (83.1%) (Table 2).

Table 2. Accuracy of FY-4A L2 cloud mask products and different models.

Models	Time	Accuracy (%)			
		Clear Sky	Water Clouds	Ice Clouds	Overall
FY-4A L2	Daytime	68.3	95.0	97.3	84.6
	Nighttime	72.2	83.5	89.1	81.6
	All day	70.0	89.3	92.9	83.1
RF	Daytime	89.2	89.8	96.1	91.4
	Nighttime	88.3	91.3	95.3	91.7
	All day	88.8	90.5	95.7	91.5
LightGBM	Daytime	90.9	89.7	95.8	92.0
	Nighttime	89.6	91.6	95.6	92.4
	All day	90.3	90.6	95.7	92.2
XGBoost	Daytime	90.9	90.2	96.0	92.3
	Nighttime	89.8	92.0	96.1	92.7
	All day	90.4	91.1	96.1	92.5
MLP	Daytime	90.5	91.7	96.0	92.4
	Nighttime	88.9	94.0	96.2	93.1
	All day	89.8	92.8	96.1	92.8

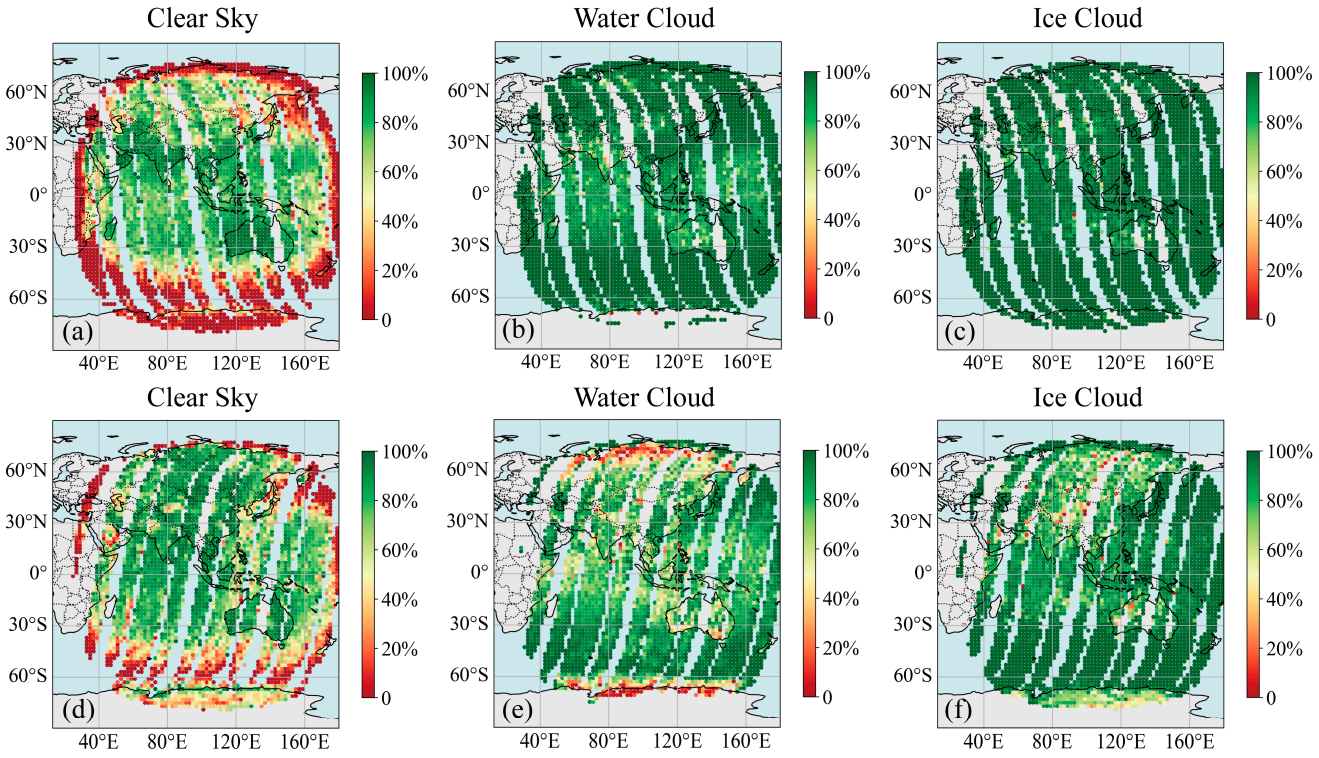


Figure 3. Detection accuracy distributions of clear sky (**a**,**d**), water clouds (**b**,**e**), and ice clouds (**c**,**f**) for FY-4A L2 cloud mask products in the daytime (**top**) and nighttime (**bottom**).

A significant improvement is seen in clear-sky identification accuracy during both daytime and nighttime observations at mid to high latitudes (>30°N/S), along with enhanced detection performance at the disk edge region (Figure 4a,d). Additionally, cloud detection performance improved during nighttime over both mid- to high-latitude land areas (>30°N/S) and low-latitude oceans (<30°N/S). Notably, a substantial increase in water

cloud identification accuracy was observed at night over the high-latitude ocean areas in the Southern Hemisphere (>60°S). However, a slight performance degradation in daytime cloud detection was observed in certain regions, such as Africa, the Arctic, and Antarctica (Figure 4b,c). This can be attributed to two primary factors: (1) From a physical perspective, the complex terrain of Africa makes it difficult to distinguish surface thermal signals from thin clouds, and daytime solar heating further exacerbates this confusion [20]. In the Arctic, the cold snow–ice surface exhibits BTs similar to those of cloud tops, thereby reducing thermal contrast and increasing the difficulty of detection [50,51]. (2) From the perspective of the FY-4A sensor's observation geometry and radiative characteristics, observations near the edge of the satellite disk or at high latitudes have larger viewing angles relative to the sub-satellite point. The increased solar zenith angle (SAZ) extends the atmospheric path length, enhancing absorption and scattering effects and thus reducing the radiative contrast between the cloud tops and the surface [52,53]. In summary, the TSAR model effectively compensates for the shortcomings of the FY-4A L2 CLM product, significantly enhancing the accuracy of clear-sky detection and nighttime cloud detection. Similar conclusions were observed for the other models (Figures S3 and S4).

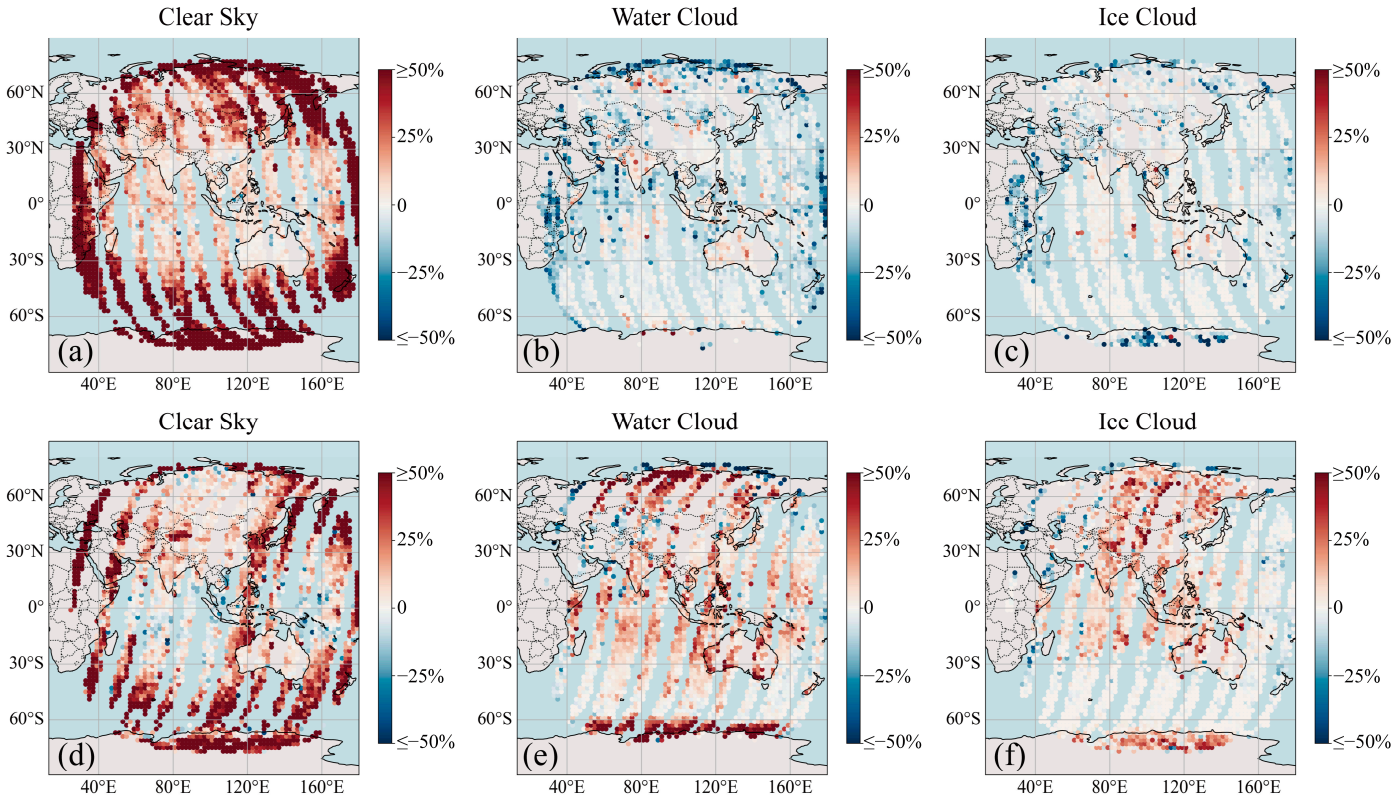


Figure 4. Accuracy differences in detection of clear sky (**a**,**d**), water clouds (**b**,**e**) and ice clouds (**c**,**f**) between MLP-TSAR model and FY-4A L2 cloud mask products during daytime (**top**) and nighttime (**bottom**).

3.3. Effects of Meteorological Factors on Model Performance

3.3.1. Clear Sky Detection

The MLP-BT model demonstrates a significantly higher accuracy under clear-sky conditions during the daytime (90.1%) than at night (84.4%). And the MLP-BT model exhibits a notable decline in clear-sky detection accuracy over mid- to high-latitude oceans (30–60°N/S), with a more pronounced reduction at night (Figure 5a,b). This is likely attributable to the prevalent presence of water clouds over these oceanic regions, where the weak thermal infrared signal contrast between water clouds and the sea surface leads to an

increased false alarm rate for clear-sky conditions. Consequently, relying solely on spectral band information proves inadequate for accurate clear-sky identification. Compared to the MLP-BT model, the MLP-TSAR model demonstrates superior performance over mid- to high-latitude oceans (30–60°N/S), improving accuracy by 5.5% during the day and by 13.2% at night (Figure 5c,d). Notably, the MLP-TSAR model achieves an accuracy improvement of approximately 32% over high-latitude land areas (beyond 70°N/S) at night. Similar performance advantages are observed in models utilizing single meteorological factors as input, such as T2m, SKT, ATP, and RH (Figures S5 and S6).

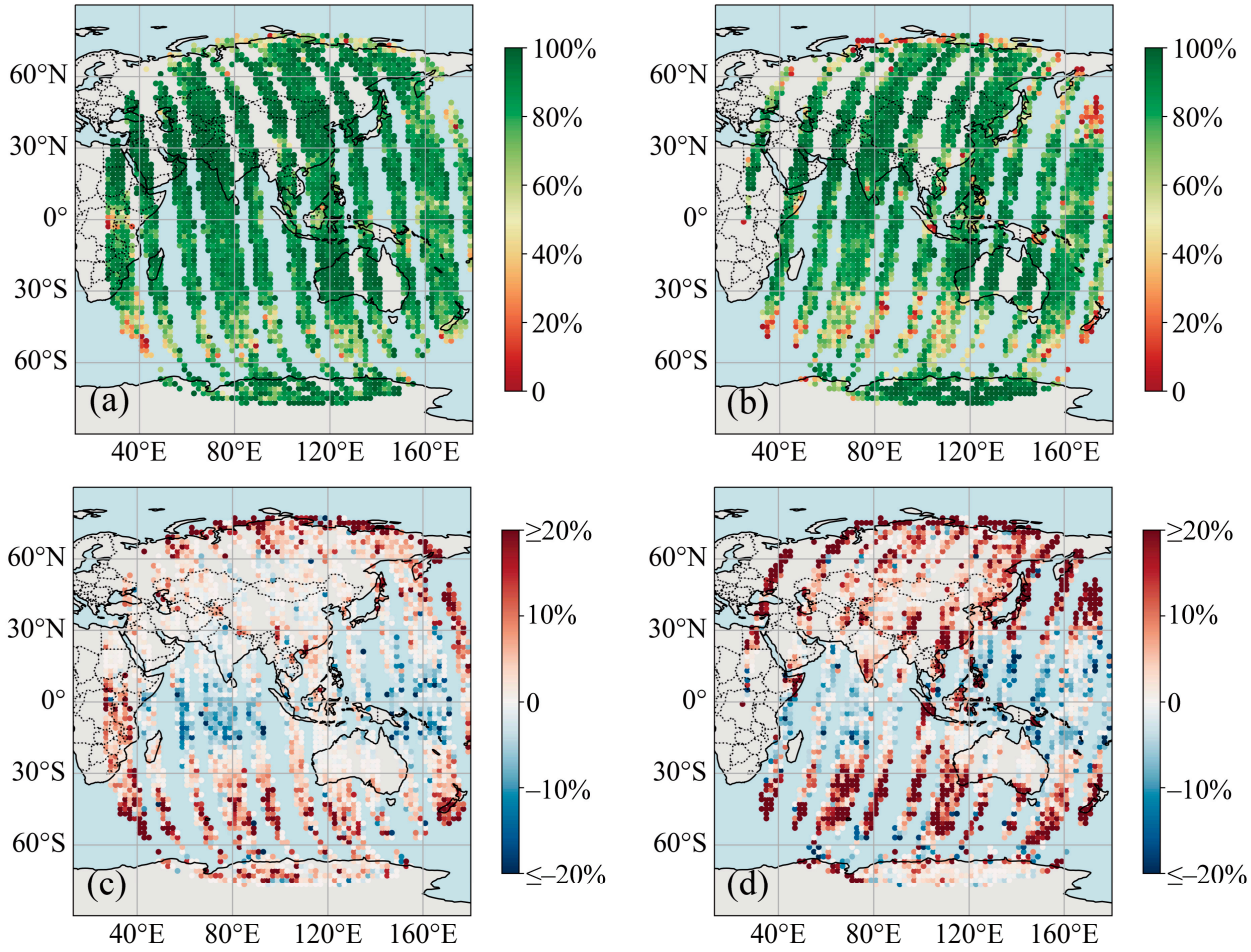


Figure 5. Distribution of (**a**) daytime and (**b**) nighttime clear sky detection accuracy for the MLP-BT model, and the accuracy differences between the TSAR and BT models during (**c**) daytime and (**d**) nighttime.

Furthermore, to investigate the impact of meteorological factors on model performance, this study compares the performance differences between various meteorological factor models and the BT model during daytime and nighttime, analyzing the latitudinal variation in clear-sky detection accuracy across different models. During daytime, all models enhance clear-sky detection accuracy in mid-to-high- latitude regions (>20°N/S). At night, with the exception of the SKT model, which shows reduced accuracy in the Southern Hemisphere, all other models improve accuracy in mid- to high-latitude regions (>20°N/S). Overall, the TSAR model consistently maintains a leading advantage in clear-sky detection, demonstrating superior performance at night compared to other meteorological factor models (Figure S11a,b).

3.3.2. Cloud Detection

Water Cloud Conditions

It should be noted that the primary objective of this study is to enhance and compare the capability of models to distinguish between clear-sky and cloudy conditions, rather than to retrieve cloud phase information. Accordingly, water cloud and ice cloud samples were merged into a single "cloudy" category for binary classification against the "clear-sky" category during model training. However, to provide a more comprehensive evaluation, this study separately examined the influence of meteorological factors on cloud detection accuracy under water cloud and ice cloud conditions.

The MLP-BT model achieves detection accuracies of 91.2% and 91.6% for daytime and nighttime under water cloud conditions, respectively. It exhibits lower accuracy in low-latitude regions (0–30°N/S) during the daytime and in mid- to high- latitude land areas of the Northern Hemisphere (20–60°N) during the nighttime (Figure 6a,b). Compared to the MLP-BT model, during the daytime, the MLP-TSAR model primarily enhances detection accuracy over low-latitude oceanic areas, albeit with a decrease observed in the mid-to-high- latitude land regions of the Northern Hemisphere. At night, the MLP-TSAR model significantly improves detection accuracy over oceanic areas (30°N–60°S) (Figure 6c,d). Models utilizing single meteorological factors as input (T2m, SKT, ATP, and RH) also demonstrate similar conclusions (Figures S7 and S8).

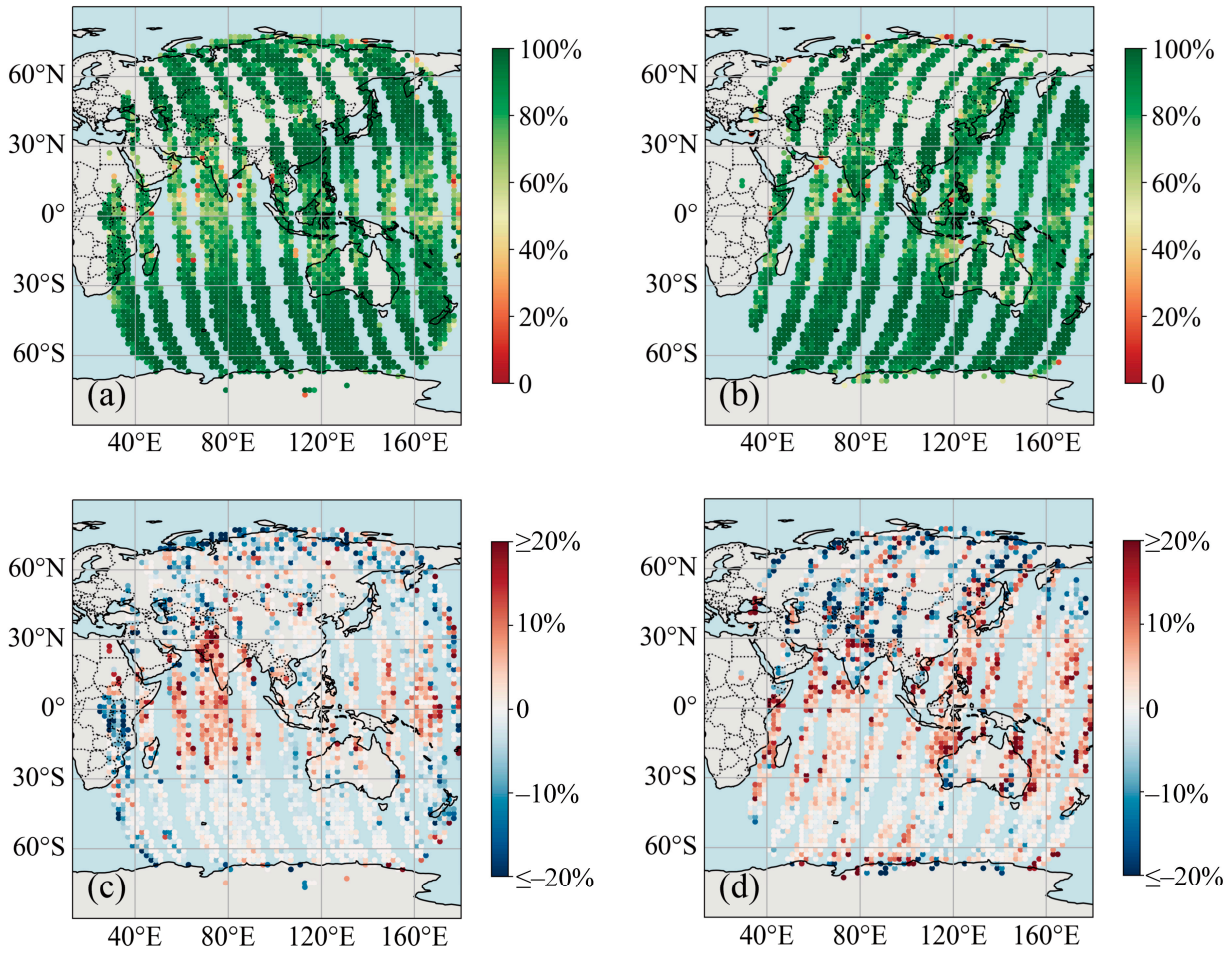


Figure 6. Distribution of (**a**) daytime and (**b**) nighttime cloud detection accuracy in water cloud conditions for the MLP-BT model, and the accuracy differences between the TSAR and BT models during (**c**) daytime and (**d**) nighttime.

This study further compares the performance differences between various meteorological factor models and the BT model during daytime and nighttime, analyzing the latitudinal variations in water cloud detection accuracy across the models. During the daytime, the performance differences among the T2m, SKT, and RH models relative to the BT model are not significant. However, the ATP and TSAR models exhibit varying degrees of accuracy degradation with increasing latitude. At night, the ATP model reduces water cloud detection accuracy across almost all latitudes, whereas the other four meteorological factor models improve accuracy to different extents in the low- to mid-latitude regions (40°S–30°N), with a particularly notable enhancement over Antarctica (Figure S11c,d).

Ice Cloud Conditions

The MLP-BT model demonstrates high accuracy levels under ice cloud conditions during both periods (95.0% and 95.3%). However, regions of relatively lower accuracy are observed over mid- to high-latitude land areas (30–80°N) (Figure 7a,b). Compared to the MLP-BT model, the results indicate that meteorological factors contribute to improved ice cloud detection accuracy in mid-latitude regions and over low-latitude oceans (Figure 7c,d). Specifically, the MLP-TSAR model shows a 2.0% accuracy improvement over mid-latitude land areas in the Northern Hemisphere during daytime, and a 1.2% improvement over low-latitude oceanic regions during nighttime. Similar findings are confirmed in the T2m, SKT, ATP, and RH models (Figures S9 and S10).

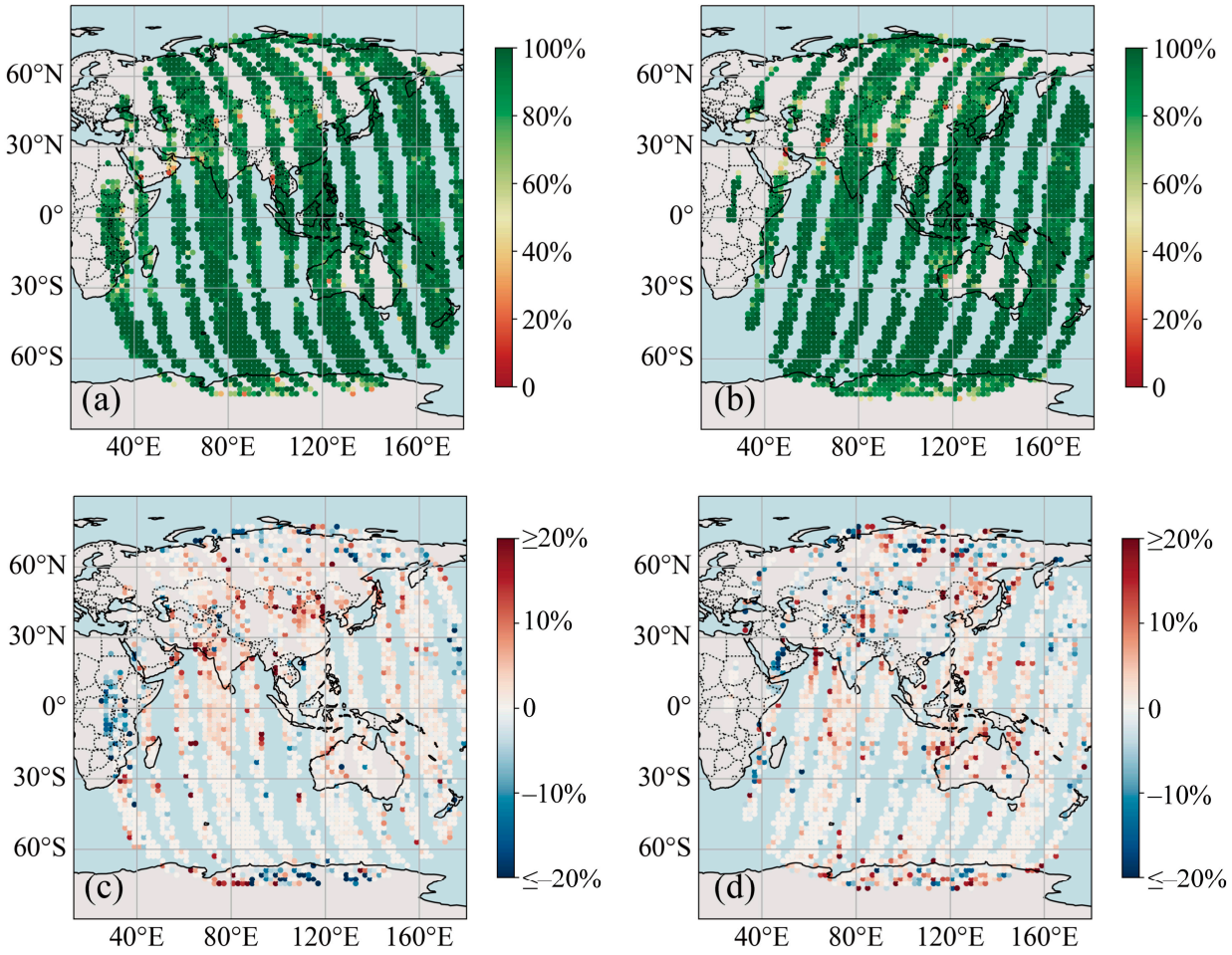


Figure 7. Distribution of (**a**) daytime and (**b**) nighttime cloud detection accuracy in ice cloud conditions for the MLP-BT model, and the accuracy differences between TSAR and BT during (**c**) daytime and (**d**) nighttime.

This study further compares the performance differences between various meteorological factor models and the BT model during daytime and nighttime, analyzing the latitudinal variations in ice cloud detection accuracy across different models. Overall, the capability of meteorological factors to enhance ice cloud detection accuracy by the MLP is limited during both daytime and nighttime. The detection accuracy of various meteorological factor models exhibits fluctuating patterns across different latitude zones, with no substantial overall differences observed (Figure S11e,f).

3.4. Evaluation of Feature Contributions

3.4.1. Geometry Parameters

Figure 8a,b show the latitudinal variations in individual geometric parameter contributions and the total contribution for MLP-TSAR during daytime (solid lines) and nighttime (dashed lines), respectively. The SAZ emerges as the primary contributing factor (11.2%), as it alters the radiative transfer path, thereby enhancing the contrast between cloud and clear-sky observation signals. The contributions of other parameters are relatively minor (<5%), though they still play important roles at specific latitudes. Furthermore, the diurnal differences in the contributions of geometric parameters are small in low-latitude regions but become significantly more pronounced at high latitudes in the Northern Hemisphere. This is primarily because the low solar elevation angle at high latitudes extends the thermal infrared path and amplifies geometric effects, thus enhancing the SAZ's contribution to daytime cloud detection. This geometric amplification diminishes at night with a negative solar angle, reducing its influence.

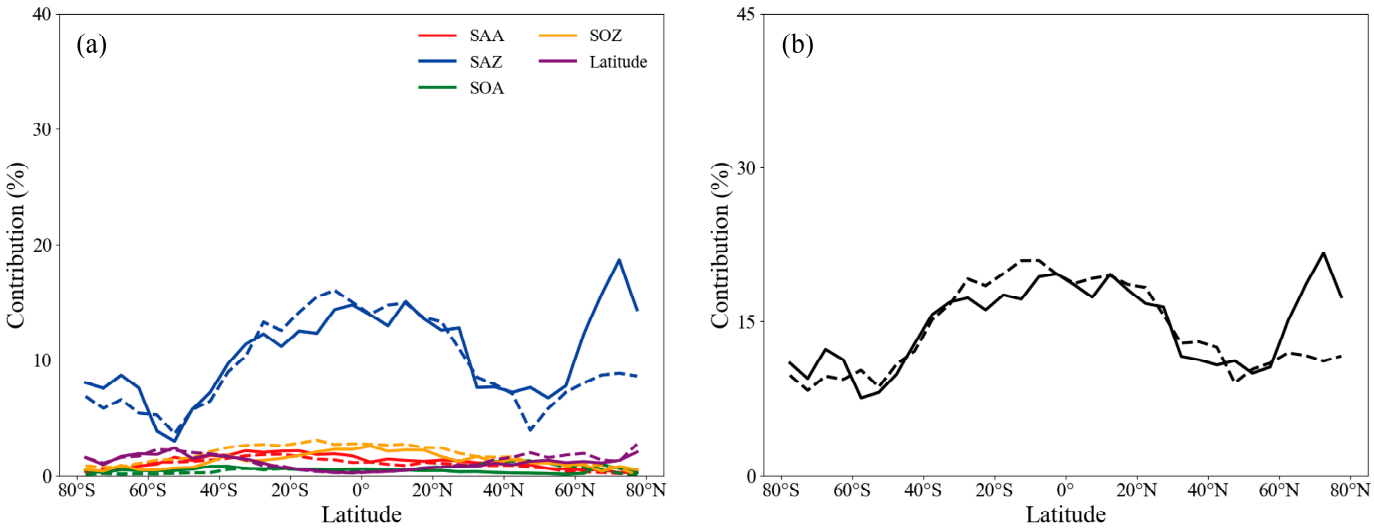


Figure 8. Latitudinal variation in contributions to the MLP-TSAR model from geometry parameters (**a**) and combined geometry parameters (**b**). Curves represent latitudinal averages. Solid and dashed lines represent daytime and nighttime, respectively (SHAP analysis performed on the model using the 2020 dataset with a sample size of 10,000).

3.4.2. BT

Figure 9a,b show the latitudinal variations in the contributions from different band albedos and the total contribution for MLP-TSAR during daytime (solid lines) and nighttime (dashed lines). The pronounced latitudinal dependence of band contributions is closely related to complex surface environments (e.g., land-sea distribution and type) and satellite viewing geometry. The primary contributions come from the 8.5 μm, 10.8 μm, and 12.0 μm bands, accounting for 15.3%, 10.8%, and 9.2%, respectively. The contributions of the 3.72 μm and 7.1 μm bands are relatively small at 5.9% and 4.8%, respectively. The 6.25 μm and

13.5 μm bands make minimal contributions, approximately 2% each. The 8.5 μm band contributes significantly at high latitudes and exhibits notable diurnal variation in the Northern Hemisphere. The 10.8 μm band also shows significant diurnal variation across all latitudes, with both bands demonstrating higher contributions during daytime than at night. Two mechanisms collectively enhance daytime cloud detection: (1) At 8.5 μm, solar radiation interaction with ice crystals elevates the cloud-top brightness temperature relative to the nighttime emission state, increasing sensitivity to thin cirrus; (2) Daytime surface heating maximizes the thermal contrast with cloud tops at 10.8 μm. These effects amplify the daytime clear-sky/cloud brightness temperature difference, leading to higher model contributions for both bands than at night. Furthermore, the total BT contribution remains relatively stable in low-latitude regions (20°S–20°N) but decreases sharply in mid-to-high latitude regions (>60°S, >30°N).

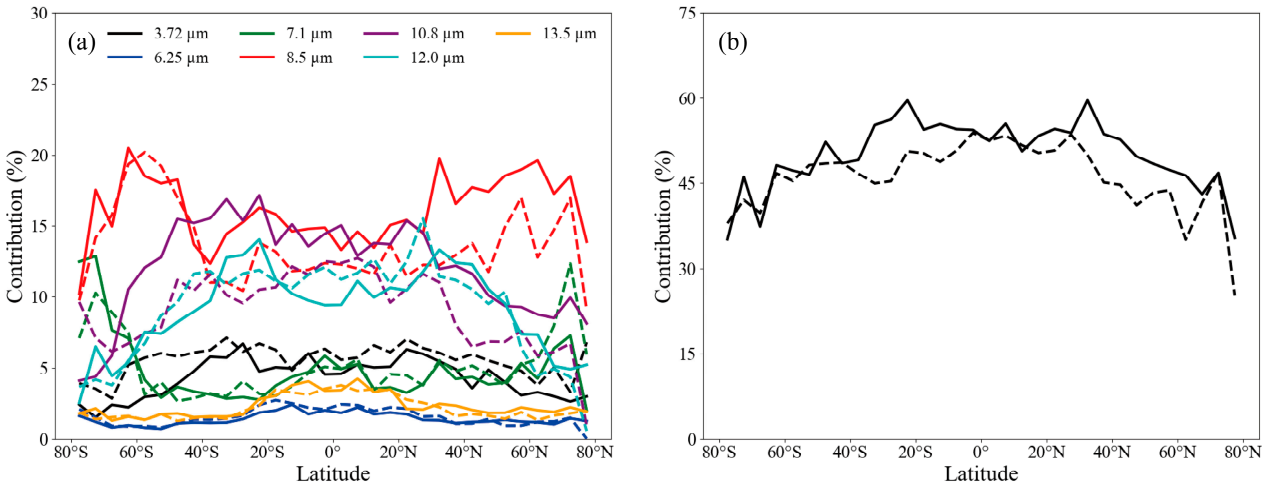


Figure 9. Latitudinal variation in contributions to the MLP-TSAR model from BTs (**a**) and combined BTs (**b**). Curves represent latitudinal averages. Solid and dashed lines represent daytime and nighttime, respectively (SHAP analysis performed on the model using the 2020 dataset with a sample size of 10,000).

3.4.3. Meteorological Factors

To minimize potential interactions among meteorological factors, we initially conducted an independent evaluation of single-factor models. All meteorological factors made significant contributions (>12%) to the model predictions (Table 3). Within the MLP, the influence of each factor was generally more pronounced during nighttime than during daytime. Among these single-factor models, ATP demonstrated particular prominence, achieving a total contribution of 18.8%.

Table 3. Contributions of different single meteorological factors to the MLP model.

Meteorological Fator	Contributions (%)		
	Daytime	Nighttime	All Day
T2m	12.5	12.6	12.6
SKT	14.1	14.3	14.2
ATP	18.4	19.1	18.8
RH	12.0	13.9	12.9

To further investigate the impact of meteorological factor combinations on model performance, the contributions of these factors to MLP-TSAR were calculated, as shown in Table 4. ATP emerged as the primary contributing factor with a contribution rate of 12.7%, substantially higher than other meteorological factors—a finding consistent

with the observations in Table 3. Compared to single-factor models (Table 3), the TSAR model leveraged synergistic effects between multiple meteorological factors (combined contribution > 30%), achieving significantly enhanced detection capability over single-factor models (Table 3) during both day and night.

Table 4. Contributions of meteorological factors to the MLP-TSAR model.

Meteorological Factors	Contributions (%)		
	Daytime	Nighttime	All Day
T2m	6.1	7.6	6.8
SKT	5.1	7.2	6.1
ATP	12.6	12.8	12.7
RH	7.7	7.8	7.7
T2m + SKT + ATP + RH	31.5	35.3	33.3

Figure 10a,b show the latitudinal variations in contributions from individual meteorological factors and the total contribution for MLP-TSAR during daytime (solid lines) and nighttime (dashed lines). The contribution of ATP is substantially higher than other factors, particularly (20%) in the Southern Hemisphere's mid to high latitudes (40–60°S). This prominence stems from the region's strong baroclinicity and frequent cyclonic activity, which generate complex, stratified temperature structures. These structures provide robust signals for cloud distribution discrimination, and due to its considerably higher information entropy, ATP is assigned a dominant weight in the model. The contributions of other meteorological factors exhibit latitudinal fluctuations, though the differences are not significant in low-latitude regions (20°S–20°N). However, SKT and T2m show pronounced diurnal variations in their contributions across the Northern Hemisphere (nocturnal > diurnal), particularly within 20–80°N. The mechanism involves contrasting nocturnal radiative processes: clear skies promote strong surface cooling via atmospheric window emission, while clouds provide an insulating effect by emitting downward longwave radiation. This creates a pronounced nighttime warm anomaly under clouds that is clearly reflected in both SKT and T2m. The combined contribution of meteorological factors is generally substantial (>20%), with higher values observed at mid to high latitudes, particularly at night, where it can reach 50%.

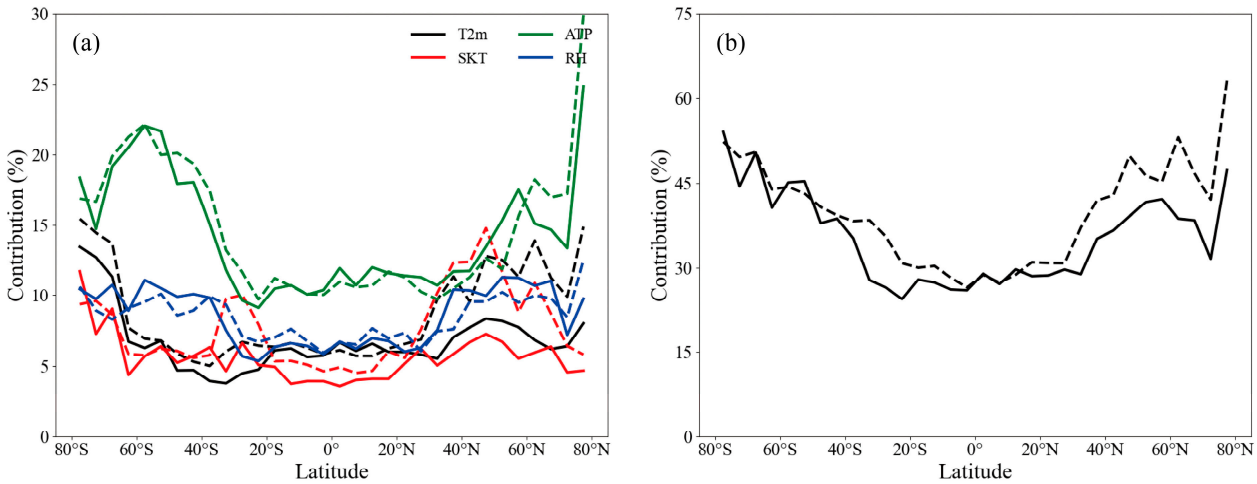


Figure 10. Latitudinal variation in contributions to the MLP-TSAR model from individual meteorological factors (**a**) and combined meteorological factors (**b**). Curves represent latitudinal averages. Solid and dashed lines represent daytime and nighttime, respectively (SHAP analysis performed on the model using the 2020 dataset with a sample size of 10,000).

We also examined the longitudinal variations in the contributions of the aforementioned geometric parameters, BTs, and meteorological factors (Figure S12). The dominant influencing factors along longitude showed no significant differences from those along latitude. Furthermore, although the contributions of individual factors exhibited fluctuations across different longitudes, these variations were relatively gradual with no apparent longitudinal dependence.

4. Discussion

4.1. Methodological Advantages and Physical Mechanisms

The FY-4A L2 CLM product employs a threshold-based method grounded in physical principles, primarily relying on empirical analysis of the contrast between cloudy and clear-sky pixels to determine classification thresholds [35]. However, threshold-based methods have inherent limitations in adapting to complex and variable surface and atmospheric conditions. They are prone to misclassification when distinguishing clouds from other high-reflectance targets (such as snow and desert) and frequently fail to detect optically thin clouds such as thin cirrus [14,54]. To overcome these limitations, this study implements multi-source data fusion through an active-passive sensor synergy strategy, integrating multi-dimensional features encompassing satellite spectral data, geometric parameters, and meteorological factors. In particular, the incorporation of ATP and RH provides atmospheric state constraints that facilitate the identification of spurious cloud signals induced by variations in atmospheric water vapor [55,56]. The model tends to predict clear skies at higher temperatures and lower humidity and the presence of clouds at lower temperatures and higher humidity (Figure S13). This comprehensive utilization of BTs and meteorological factors enables more accurate discrimination between clear-sky and thin cloud scenarios. Compared with the FY-4A L2 CLM product, overall accuracy improved by approximately 9%, with particularly pronounced advantages in clear-sky detection (approximately 20% improvement). Furthermore, the results demonstrate that incorporating physical features significantly improves cloud detection performance regardless of the ML algorithm employed, with MLP achieving the best performance. Compared with threshold-based methods, the ML models are capable of automatically learning complex nonlinear mappings between input features and cloud labels, adaptively determining optimal decision boundaries from training data, and demonstrating enhanced robustness under diverse surface and atmospheric conditions [14,19]. Moreover, compared with ML-based cloud detection studies utilizing geostationary satellites such as GOES-16 and Himawari-8 (reported accuracies of 88–93%) [57–59], the methods proposed in this study achieve comparable performance.

Furthermore, the SHAP feature importance analysis in this study is highly consistent with the physical mechanisms of cloud radiative transfer. SAZ influences cloud detection by modulating optical path length, cloud geometric projection, and anisotropic effects [60]. The model's reliance on this parameter indicates that it has successfully captured the effects of viewing geometry. The 8.5 μm channel plays a critical role in thin cirrus detection and cloud thermodynamic phase discrimination due to its unique optical sensitivity to ice crystals; a lower BT indicates a colder radiation source, which is typically identified as cloud (Figure S13). The infrared window channels (10.8 μm and 12 μm) provide cloud-top temperature characterization and split-window information, serving as essential parameters for distinguishing clouds from clear-sky conditions and determining cloud thermodynamic phase [61,62]. The atmospheric profile data from ERA5 reanalysis effectively enhances low cloud identification and detection consistency across varying atmospheric conditions by providing temperature profile constraints and water vapor correction information [54]. The physical interpretability of these features demonstrates that machine learning models are

not merely "black-box" statistical fitting tools but rather implicitly capture the essential physical processes underlying cloud radiative transfer. Since the model exclusively utilizes thermal infrared channels (FY-4A bands 8–14), cloud detection relies entirely on thermal emission rather than solar reflectance. Consequently, the diurnal variations in feature importance can be attributed to differences in thermal infrared radiative transfer between daytime and nighttime. During the day, solar heating elevates land surface temperatures, creating strong thermal contrast between low-level clouds and the underlying surface. At night, radiative cooling reduces surface temperatures and weakens this thermal contrast, making low cloud detection considerably more challenging. This effect is particularly pronounced in the Northern Hemisphere, where the larger continental land mass exhibits a greater diurnal temperature range compared to oceanic regions. The pronounced day–night temperature variations directly affect the discriminative capability of the infrared window channels (10.8 μm and 12.0 μm) for cloud–surface separation [15]. Furthermore, nocturnal surface-based temperature inversions, which commonly develop under clear-sky conditions due to radiative cooling, can cause the BT of low clouds to approach or even exceed that of the cooled ground surface [63]. This phenomenon contradicts the typical thermal contrast assumption and poses significant challenges for infrared cloud detection. The enhanced importance of meteorological factors during nighttime reflects the model's adaptive dependence on these variables. These physical mechanisms collectively explain the latitudinal and diurnal variation patterns observed in the SHAP analysis, demonstrating that the MLP-TSAR model has successfully captured the temporal variations and regional characteristics inherent to thermal infrared cloud detection.

4.2. Limitations and Prospects

The proposed method in this study still has several limitations: (1) The CALIPSO satellite observations have inherent spatiotemporal sampling constraints, making it difficult to fully characterize the diurnal variation in clouds. Its narrow-swath observation geometry results in insufficient sampling for specific regions and cloud types, thereby limiting the model's generalization capability under all-sky conditions. (2) Sensor and observation geometry errors may still affect model performance. For instance, the observation accuracy of FY-4A near the edge of the disk is influenced by the increased SAZ, leading to higher radiometric calibration and geometric positioning errors, which introduce regional systematic biases at the model input level. (3) The ERA5 reanalysis data exhibit higher uncertainty in observation-sparse regions (e.g., Africa and high-latitude polar areas), particularly for cloud, aerosol, and radiative variables [64,65]. These reanalysis errors may affect the accuracy of model inputs. Moreover, the temporal (hourly) and spatial (0.25°) resolutions of ERA5 are coarser than the high-frequency, high-resolution observations from FY-4A, resulting in scale mismatches that may distort sub-grid processes and rapidly evolving cloud systems, thus introducing error propagation [66]. In addition, although ERA5 provides long-term historical coverage, its latency in updating near-real-time or rapidly changing weather events limits its applicability for operational use. (4) The model's generalization and transferability remain to be validated. Since the current model was trained and tested primarily on full-disk data, its adaptability to different climatic zones, surface types, and seasonal variations may be limited.

For operational applications, the update latency of ERA5 data remains the main bottleneck for real-time model operation. To enhance the model's practicality and physical consistency, future research will further improve the current cloud detection framework from multiple perspectives. On one hand, multi-model ensemble and adaptive frameworks can be adopted, or near-real-time meteorological data and numerical weather prediction (NWP) products can be introduced as alternative inputs to improve the timeliness and

stability of the model under varying meteorological conditions. On the other hand, future work will incorporate the physical processes associated with the formation and evolution of different cloud types, integrate additional physical variables (e.g., atmospheric stability indices), and combine feature selection strategies with interpretability analyses (e.g., SHAP) to identify key features and develop a more comprehensive, physically consistent, and intelligent cloud detection model. Despite these challenges, improved cloud detection accuracy holds the potential to enhance the quality of downstream products, including cloud-top height, cloud optical thickness, and cloud thermodynamic phase. Furthermore, high-quality cloud products can contribute to data assimilation for numerical weather prediction and to climate monitoring, thereby providing a robust data foundation for climate change research.

5. Conclusions

This study incorporates physical features to improve all-day cloud detection accuracy of the FY-4A CLM product. The models integrated BT data from FY-4A L1 full-disk channels 8–14, geometric parameters (SAZ, SAA, SOZ, SOA, and latitude), and four meteorological factors (T2m, SKT, ATP, and RH), using CALIPSO cloud products as labels. A key finding of this study is that the incorporation of atmospheric temperature and humidity profiles (ATP and RH) provides fundamental advantages over satellite-only approaches by imposing physical constraints derived from the vertical thermodynamic structure of the atmosphere. These profiles characterize the environmental conditions governing cloud formation and maintenance, enabling the models to better distinguish clouds from clear-sky backgrounds, particularly under ambiguous radiative signatures. The integration of multi-source data, particularly the three-dimensional atmospheric profiles, effectively reduced the day–night accuracy discrepancy, ensuring spatiotemporal consistency in cloud detection. Based on the integration of physical features, four ML models—RF, LightGBM, XGBoost, and MLP—were compared to the potential of the cloud detection. The results demonstrate that incorporating physical factors significantly improves cloud detection performance regardless of the ML algorithm employed. Specifically, the detection accuracies of RF, LightGBM, XGBoost, and MLP reach 91.5%, 92.2%, 92.5%, and 92.8%, respectively—substantially outperforming the FY-4A L2 product (83.1%). All four models markedly improved clear-sky detection accuracy over mid- to high-latitude oceans and enhanced nighttime cloud identification over mid- to high-latitude land areas. Comprehensive evaluation across various temporal and conditional scenarios confirms that the MLP-TSAR model delivers the most stable and reliable performance, demonstrating strong robustness.

A further investigation into the contributions of physical features to predictions was conducted using the SHAP method based on the MLP-TSAR model. Among the geometric parameters, the SAZ significantly influences the observation signals, making it the primary contributing geometric parameter. The total contribution of BT reached 50.8%, establishing it as the dominant factor in the model. In the MLP-TSAR model, the combined contribution of the four meteorological factors exceeded 30%, with this value increasing significantly to 45% in mid- to high-latitude regions. Notably, within the 40–80°N, the contribution of meteorological factors was substantially higher during nighttime than during daytime.

In summary, the incorporation of physical features into the cloud detection framework significantly enhances both accuracy and spatiotemporal consistency. Among the models evaluated, the MLP-TSAR model exhibits superior robustness and reliability under varying illumination and surface conditions, rendering it particularly well-suited for operational all-weather cloud detection applications utilizing FY-4A satellite data. Nevertheless, several limitations warrant consideration. First, the physical features employed in this study are derived from the ERA5 reanalysis dataset, which may introduce inherent uncertainties

associated with data assimilation processes and temporal interpolation schemes. Future research will explore the integration of numerical weather prediction (NWP) model outputs to potentially enhance the timeliness and accuracy of auxiliary meteorological information. Second, as the number of physical features increases, the development of a robust feature selection framework becomes essential to identify the optimal subset of predictive variables, thereby ensuring model parsimony, mitigating overfitting risks, and maintaining algorithmic stability across diverse atmospheric conditions.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/rs18040536/s1, Refs. [43–47] are cited in the Supplementary Materials.

Author Contributions: Conceptualization, L.Z. and W.Z.; methodology, L.Z. and Y.L.; software, Y.L.; validation, Y.L.; formal analysis, Y.L.; investigation, Y.L. and L.Z.; resources, W.Z.; data curation, Y.L. and L.Z.; writing—original draft preparation, Y.L.; writing—review and editing, Y.S., X.H. and W.Z.; visualization, Y.L. and Z.F.; supervision, W.Z.; project administration, W.Z.; funding acquisition, Y.S. and W.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was sponsored by the National Natural Science Foundation of China (Grants 42075035 and 42075011) and the China Postdoctoral Science Foundation (Grant 2025M784460).

Data Availability Statement: The Fengyun-4A data are available from the CMA National Satellite Meteorological Center (NSMC) at http://satellite.nsmc.org.cn/portalsite/default.aspx (accessed on 1 December 2025). The ERA5 reanalysis data are available from the Copernicus Climate Data Store (CDS) of the European Centre for Medium-Range Weather Forecasts (ECMWF) at https://cds.climate.copernicus.eu/ (accessed on 1 December 2025). The CALIPSO (Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations) data are available from NASA's Atmospheric Science Data Center (ASDC) at https://asdc.larc.nasa.gov/ (accessed on 1 December 2025).

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

Abbreviations

The following abbreviations are used in this manuscript:

ML	Machine learning
RF	Random Forest
LightGBM	Light Gradient Boosting Machine
XGBoost	eXtreme Gradient Boosting
MLP	Multilayer Perceptron
SAZ	satellite zenith angle
SAA	satellite azimuth angle
SOZ	solar zenith angle
SOA	solar azimuth angle
T2m	2 m air temperature
SKT	land surface skin temperature
ATP	air temperature profiles
RH	relative humidity profiles

References

1. Zhang, Y.-C.; Rossow, W.B.; Lacis, A.A.; Oinas, V.; Mishchenko, M.I. Calculation of radiative fluxes from the surface to top of atmosphere based on ISCCP and other global data sets: Refinements of the radiative transfer model and the input data. *J. Geophys. Res.* **2004**, *109*, D19105. [CrossRef]
2. Zhang, H.; Huang, Q.; Zhai, H.; Zhang, L. Multi-temporal cloud detection based on robust PCA for optical remote sensing imagery. *Comput. Electron. Agric.* **2021**, *188*, 106342. [CrossRef]
3. Slingo, A.; Slingo, J.M. The response of a general circulation model to cloud longwave radiative forcing. I: Introduction and initial experiments. *Q. J. R. Meteorol. Soc.* **1988**, *114*, 1027–1062. [CrossRef]

4. Fueglistaler, S.; Dessler, A.E.; Dunkerton, T.J.; Folkins, I.; Fu, Q.; Mote, P.W. Tropical tropopause layer. *Rev. Geophys.* **2009**, *47*, RG1004. [CrossRef]
5. Haynes, J.M.; Vonder Haar, T.H.; L' Ecuyer, T.; Henderson, D. Radiative heating characteristics of Earth' s cloudy atmosphere from vertically resolved active sensors. *Geophys. Res. Lett.* **2013**, *40*, 624–630. [CrossRef]
6. Li, J.D.; Wang, W.C.; Dong, X.Q.; Mao, J.Y. Cloud-radiation-precipitation associations over the Asian monsoon region: An observational analysis. *Clim. Dyn.* **2017**, *49*, 3237–3255. [CrossRef]
7. King, M.D.; Tsay, S.-C.; Platnick, S.E.; Wang, M.; Liou, K.-N. Cloud retrieval algorithms for MODIS: Optical thickness, effective particle radius, and thermodynamic phase. In *MODIS Algorithm Theoretical Basis Document No. ATBD-MOD-05*; NASA Goddard Space Flight Center: Greenbelt, MD, USA, 1997.
8. Mei, L.; Rozanov, V.; Vountas, M.; Burrows, J.P.; Levy, R.C.; Lotz, W. A cloud masking algorithm for the XBAER aerosol retrieval using MERIS data. *Remote Sens. Environ.* **2017**, *197*, 141–160. [CrossRef]
9. Jafariserajehlou, S.; Mei, L.; Vountas, M.; Rozanov, V.; Burrows, J.P.; Hollmann, R. A cloud identification algorithm over the Arctic for use with AATSR-SLSTR measurements. *Atmos. Meas. Tech.* **2019**, *12*, 1059–1076. [CrossRef]
10. Zhao, Z.; Zhang, F.; Wu, Q.; Li, Z.; Tong, X.; Li, J.; Han, W. Cloud identification and properties retrieval of the Fengyun-4A satellite using a ResUnet model. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4102318. [CrossRef]
11. Du, M.; Luo, S.; Shi, J.; Guo, W.; Zhang, J.; Gu, H.; Gu, W. Operational application of Fengyun geostationary meteorological satellites to cloud observation products. *Sci. Rep.* **2024**, *14*, 17780. [CrossRef]
12. Huang, X.; Ali, S.; Wang, C.; Ning, Z.; Purushotham, S.; Wang, J.; Zhang, Z. Deep domain adaptation based cloud type detection using active and passive satellite data. In Proceedings of the 2020 IEEE International Conference on Big Data (Big Data 2020), Atlanta, GA, USA, 10–13 December 2020; IEEE: New York, NY, USA, 2020; pp. 1330–1337.
13. Huang, C.; Wang, Z.; Li, Q.; Feng, L.; Zhang, M.; Qin, W.; Wang, L.; Tong, M.; Wang, Y. Cloud mask detection by combining active and passive remote sensing data. *Remote Sens.* **2025**, *17*, 3315.
14. Mahajan, S.; Fataniya, B. Cloud detection methodologies: Variants and development—A review. *Complex Intell. Syst.* **2020**, *6*, 251–261. [CrossRef]
15. Jedlovec, G.J.; Haines, S.L.; LaFontaine, F.J. Spatial and temporal varying thresholds for cloud detection in satellite imagery. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1705–1717. [CrossRef]
16. Xiong, Q.; Wang, Y.; Liu, D.; Ye, S.; Du, Z.; Liu, W.; Huang, J.; Su, W.; Zhu, D.; Yao, X.; et al. A Cloud Detection Approach Based on Hybrid Multispectral Features with Dynamic Thresholds for GF-1 Remote Sensing Images. *Remote Sens.* **2020**, *12*, 450. [CrossRef]
17. Shang, H.; Nagao, T.M.; Letu, H.; Wei, L.; He, J.; Chen, L.; Nakajima, T.Y.; Shao, J.; Xu, R.; Wu, L.; et al. A hybrid cloud detection and cloud phase classification algorithm using classic threshold-based tests and extra randomized tree model. *Remote Sens. Environ.* **2023**, *302*, 113957. [CrossRef]
18. Mateo-Garcia, G.; Laparra, V.; López-Puigdollers, D.; Gómez-Chova, L. Transferring deep learning models for cloud detection between Landsat-8 and Proba-V. *ISPRS J. Photogramm. Remote Sens.* **2020**, *160*, 1–17. [CrossRef]
19. Ma, N.; Sun, L.; Zhou, C.; He, Y. Cloud Detection Algorithm for Multi-Satellite Remote Sensing Imagery Based on a Spectral Library and 1D Convolutional Neural Network. *Remote Sens.* **2021**, *13*, 3319. [CrossRef]
20. Shi, X.; Fan, Y.; Sun, L.; Liu, X.; Liu, C.; Pang, S. Cloud detection sample generation algorithm for nighttime satellite imagery based on daytime data and machine learning application. *Sci. Rep.* **2024**, *14*, 27917. [CrossRef]
21. Zhang, Q.; Yu, Y.; Zhang, W.; Luo, T.; Wang, X. Cloud Detection from FY-4A's Geostationary Interferometric Infrared Sounder Using Machine Learning Approaches. *Remote Sens.* **2019**, *11*, 3035. [CrossRef]
22. Liu, C.; Yang, S.; Di, D.; Yang, Y.; Di, D.; Zhou, C.; Hu, X.; Sohn, B. A Machine Learning-based Cloud Detection Algorithm for the Himawari-8 Spectral Image. *Adv. Atmos. Sci.* **2022**, *39*, 1994–2007. [CrossRef]
23. Xu, L.; Wong, A.; Clausi, D.A. A novel Bayesian spatial-temporal random field model applied to cloud detection from remotely sensed imagery. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 4913–4924. [CrossRef]
24. Miroszewski, A.; Mielczarek, J.; Czelusta, G.; Spychała, J. Detecting clouds in multispectral satellite images using quantum-kernel support vector machines. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 7601–7613. [CrossRef]
25. Huang, W.; Li, Z.; Sun, L.; Zhu, X.; Yuan, Q.; Liu, L.; Cribb, M. Cloud detection for Landsat imagery by combining the random forest and superpixels extracted via energy-driven sampling segmentation approaches. *Remote Sens. Environ.* **2020**, *248*, 112005.
26. Shao, Z.; Pan, Y.; Diao, C.; Cai, J. Cloud detection in remote sensing images based on multiscale features-convolutional neural network. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 4062–4076. [CrossRef]
27. Kalesse-Los, H.; Schimmel, W.; Luke, E.; Seifert, P. Evaluating cloud liquid detection against Cloudnet using cloud radar Doppler spectra in a pre-trained artificial neural network. *Atmos. Meas. Tech.* **2022**, *15*, 279–295. [CrossRef]
28. Cao, X.; Liu, B.; Gao, M.; Guo, Y. Cloud Detection for Satellite Imagery Using Attention-Based U-Net Convolutional Neural Network. *Symmetry* **2020**, *12*, 1056.
29. Wang, M.; Balmes, K.A.; Thorsen, T.J.; Willick, D.; Fu, Q. An investigation of the ice cloud detection sensitivity of cloud radars using the Raman lidar at the ARM SGP site. *Remote Sens.* **2022**, *14*, 3466. [CrossRef]

30. Zhao, W.; Wang, Y.; Huo, J.; Liu, B.; Lyu, D.; Li, J. Unveiling cloud vertical structures over the interior Tibetan Plateau through anomaly detection in synergetic lidar and radar observations. *Adv. Atmos. Sci.* **2024**, *41*, 2381–2398. [CrossRef]
31. Deng, M.; Xu, X.; Ma, Y.; Gong, W.; Jin, S.; Hu, R. Multi-layer perceptron combined with radiative transfer model for complex land surface cloud detection. *Acta Electron. Sin.* **2022**, *50*, 932–942.
32. Jiao, Y.; Zhang, M.; Wang, L.; Qin, W. A new cloud and haze mask algorithm from radiative transfer simulations coupled with machine learning. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4101216. [CrossRef]
33. Xu, W. Evaluation of cloud mask and cloud top height from Fengyun-4A with MODIS cloud retrievals over the Tibetan Plateau. *Remote Sens.* **2021**, *13*, 1418. [CrossRef]
34. Yang, J.; Zhang, Z.; Wei, C.; Lu, F.; Guo, Q. Introducing the New Generation of Chinese Geostationary Weather Satellites, Fengyun-4. *Bull. Am. Meteorol. Soc.* **2017**, *98*, 1637–1658. [CrossRef]
35. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **2020**, *146*, 1999–2049. [CrossRef]
36. Winker, D.M.; Pelon, J.; Coakley, J.A., Jr.; Ackerman, S.A.; Charlson, R.J.; Colarco, P.R.; Flamant, P.; Fu, Q.; Hoff, R.M.; Kittaka, C.; et al. The CALIPSO Mission: A Global 3D View of Aerosols and Clouds. *Bull. Am. Meteorol. Soc.* **2010**, *91*, 1211–1229. [CrossRef]
37. Singh, R.; Biswas, M.; Pal, M. Cloud detection using sentinel 2 imageries: A comparison of XGBoost, RF, SVM, and CNN algorithms. *Geocarto Int.* **2022**, *38*, 1–32. [CrossRef]
38. Zhang, C.; Pan, X.; Li, H.; Gardiner, A.; Sargent, I.; Hare, J.; Atkinson, P.M. A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification. *ISPRS J. Photogramm. Remote Sens.* **2018**, *140*, 133–144. [CrossRef]
39. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
40. Belgiu, M.; Drăguţ, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]
41. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems 30 (NIPS 2017)*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; pp. 3146–3154.
42. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'16), San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
43. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [CrossRef]
44. Nair, V.; Hinton, G.E. Rectified linear units improve restricted Boltzmann machines. In Proceedings of the 27th International Conference on Machine Learning (ICML-10), Haifa, Israel, 21–24 June 2010; pp. 807–814.
45. Du, K.-L.; Leung, C.-S.; Mow, W.H.; Swamy, M.N.S. Perceptron: Learning, Generalization, Model Selection, Fault Tolerance, and Role in the Deep Learning Era. *Mathematics* **2022**, *10*, 4730. [CrossRef]
46. Dietterich, T.G. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Comput.* **1998**, *10*, 1895–1923. [CrossRef] [PubMed]
47. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. In Proceedings of the Advances in Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; pp. 4765–4777.
48. Li, M.; Sun, H.; Huang, Y.; Chen, H. Shapley value: From cooperative game to explainable artificial intelligence. *Auton. Intell. Syst.* **2024**, *4*, 2. [CrossRef]
49. Grinsztajn, L.; Oyallon, E.; Varoquaux, G. Why do tree-based models still outperform deep learning on typical tabular data? *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 507–520.
50. Zhou, Y.; Yang, Y.; Gao, M.; Zhai, P. Cloud detection over snow and ice with oxygen A- and B-band observations from the Earth Polychromatic Imaging Camera (EPIC). *Atmos. Meas. Tech.* **2020**, *13*, 1575–1591. [CrossRef]
51. Paul, S.; Huntemann, M. Improved machine-learning-based open-water-sea-ice-cloud discrimination over wintertime Antarctic sea ice using MODIS thermal-infrared imagery. *Cryosphere* **2021**, *15*, 1551–1565. [CrossRef]
52. Payez, A. Dual view on clear-sky top-of-atmosphere albedos from Meteosat Second Generation satellites. *Remote Sens.* **2021**, *13*, 1655. [CrossRef]
53. Zhong, B.; Ma, Y.; Yang, A. Radiometric performance evaluation of FY-4A/AGRI based on Aqua/MODIS. *Sensors* **2021**, *21*, 1859. [CrossRef]
54. Zhu, Z.; Wang, S.; Woodcock, C.E. Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4-7, 8, and Sentinel-2 images. *Remote Sens. Environ.* **2015**, *159*, 269–277. [CrossRef]
55. Wang, X.; Iwabuchi, H.; Yamashita, T. Cloud identification and property retrieval from Himawari-8 infrared measurements via a deep neural network. *Remote Sens. Environ.* **2022**, *275*, 113026. [CrossRef]

56. Guo, B.; Zhang, F.; Zhao, Z.; Guo, L.; Li, W. Retrieval of cloud macro-physical properties using the FY-4A Advanced Geostationary Radiation Imager (AGRI) and the Geostationary Interferometric Infrared Sounder (GIIRS). *Geophys. Res. Lett.* **2024**, *51*, e2024GL109772. [CrossRef]
57. Lee, Y.; Kummerow, C.D.; Ebert-Uphoff, I. Applying machine learning methods to detect convection using Geostationary Operational Environmental Satellite-16 (GOES-16) Advanced Baseline Imager (ABI) data. *Atmos. Meas. Tech.* **2021**, *14*, 2699–2716. [CrossRef]
58. Chai, D.; Huang, J.; Wu, M.; Yang, X.; Wang, R. Remote sensing image cloud detection using a shallow convolutional neural network. *ISPRS J. Photogramm. Remote Sens.* **2024**, *209*, 66–84. [CrossRef]
59. Yang, Y.; Sun, W.; Chi, Y.; Yan, X.; Fan, H.; Yang, X.; Ma, Z.; Wang, Q.; Zhao, C. Machine learning-based retrieval of day and night cloud macrophysical parameters over East Asia using Himawari-8 data. *Remote Sens. Environ.* **2022**, *273*, 112971. [CrossRef]
60. Roy, D.P.; Zhang, H.K.; Ju, J.; Gomez-Dans, J.L.; Lewis, P.E.; Schaaf, C.B.; Sun, Q.; Li, J.; Huang, H.; Kovalskyy, V. A general method to normalize Landsat reflectance data to nadir BRDF adjusted reflectance. *Remote Sens. Environ.* **2016**, *176*, 255–271. [CrossRef]
61. Strabala, K.I.; Ackerman, S.A.; Menzel, W.P. Cloud properties inferred from 8–12-μm data. *J. Appl. Meteorol.* **1994**, *33*, 212–229. [CrossRef]
62. Garnier, A.; Pelon, J.; Pascal, N.; Mark, A.; Philippe, D.; Yang, P.; Mitchell, D. Version 4 CALIPSO Imaging Infrared Radiometer ice and liquid water cloud microphysical properties—Part II: Results over oceans. *Atmos. Meas. Tech.* **2021**, *14*, 3277–3299. [CrossRef]
63. Ellrod, G.P.; Gultepe, I. Inferring low cloud base heights at night for aviation using satellite infrared and surface temperature data. *Pure Appl. Geophys.* **2007**, *164*, 1193–1205. [CrossRef]
64. Hersbach, H.; Bell, B.; Berrisford, P.; Dragani, R.; Fuentes, M.; Janicki, P.; Simmons, A.; Soci, C.; Thépaut, J.-N.; Trémolet, Y. The ERA5 global reanalysis. In *ECMWF Technical Memorandum No. 859*; European Centre for Medium-Range Weather Forecasts: Reading, UK, 2020.
65. Nygård, T.; Tjernström, M.; Naakka, T. Winter thermodynamic vertical structure in the Arctic atmosphere linked to large-scale circulation. *Weather. Clim. Dyn.* **2021**, *2*, 1263–1282. [CrossRef]
66. Yao, B.; Teng, S.; Lai, R.; Xu, X.; Yin, Y.; Shi, C.; Liu, C. Can atmospheric reanalyses (CRA and ERA5) represent cloud spatiotemporal characteristics? *Atmos. Res.* **2020**, *244*, 105091. [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.