Unbiased High-Precision Cloud Detection for Advanced Himawari Imager Using Automatic Machine Learning

Bochun Liu, Jinming Ge, Qingyu Mu, Chi Zhang, Xiaoyu Hu, Jiajing Du, Yanyan Wu, Bo Wang, Xiang Li, Jianping Huang, and Qiang Fu

***Abstract*—Clouds play a pivotal role in the global radiation budget and energy cycle, making accurate cloud detection via satellite remote sensing essential for understanding their evolutions. However, existing machine learning-based cloud detection algorithms have limitations, particularly in data preprocessing and feature engineering, and few are designed to cover the entire geostationary satellite observation region. In this study, we integrate three deviation elimination schemes for cloud cover differences and various surface type characterization modes to develop an unbiased, high-precision cloud detection algorithm for the advanced Himawari imager (AHI) onboard Himawari-8, leveraging automatic machine learning (AutoML) techniques. In addition, we provide a new perspective on model interpretability by incorporating concepts from cooperative game theory. Our results indicate that cloud detection algorithms constructed separately for different surfaces yield better results than those applied uniformly across the entire region by inputting surface feather. The optimal algorithm achieves unbiased cloud detection across multiple surfaces by adjusting the output threshold of the mapping function, allowing for cloud detection results with varying confidence levels as required. The true skill statistic of the optimal AutoML-based cloud detection algorithm is 87.32%, which is 13.72% higher than that of the AHI official cloud mask. The deviation rate of the optimal algorithm is only 0.79%, significantly lower than the 20.48% of the AHI official cloud mask. Long-term cloud frequency (CF) tests show that our optimal algorithm's CF distribution is more consistent with cloud-aerosol lidar with orthogonal polarization results compared with the AHI official products.**

***Index Terms*—Automatic machine learning (AutoML), geostationary satellite, marginal contribution, unbiased cloud detection.**

Received 25 July 2024; revised 7 October 2024 and 19 December 2024; accepted 25 January 2025. Date of publication 28 January 2025; date of current version 28 February 2025. This work was supported in part by the National Science Foundation of China under Grant 42275076 and Grant 41875028, in part by the Science and Technology Projects of Gansu Province under Grant 22JR5RA398, in part by the Excellent Graduate Student "Innovation Star" Project of Gansu Province under Grant 2023CXZX-107, and in part by the Fundamental Research Funds for the Central Universities under Grant lzujbky-2022-ct06. *(Corresponding author: Jinming Ge.)*

Bochun Liu, Jinming Ge, Qingyu Mu, Chi Zhang, Jiajing Du, Yanyan Wu, Bo Wang, Xiang Li, and Jianping Huang are with the Key Laboratory for Semi-Arid Climate Change, Ministry of Education, College of Atmospheric Sciences, Lanzhou University, Lanzhou 730000, China (e-mail: gejm@lzu.edu.cn).

Xiaoyu Hu is with the Laboratory for Intelligent Weather Consultation and Forecasting, Peking University Chongqing Big Data Research Institute, Chongqing 401333, China, and also with the School of Mathematical Sciences, Peking University, Beijing 100871, China.

Qiang Fu is with the Department of Atmospheric Sciences, University of Washington, Seattle, WA 98105 USA.

Digital Object Identifier 10.1109/JSTARS.2025.3535762

I. Introduction

CLOUDS cover two-third of the Earth's surface and play a crucial role in the radiative budget and hydrological cycle of the Earth-atmosphere system [1], [2]. Different cloud types exert distinct radiation effects: low clouds primarily reflect shortwave radiation back to space, contributing to surface cooling. However, high clouds predominantly absorb longwave radiation emitted by the surface and lower atmosphere, resulting in a warming effect [3], [4], [5], [6]. Given the intricacy of cloud radiative effects, even slight changes in cloud cover can lead to significant variations in global energy balance, thereby substantially impacting climate change. In the context of climate warming, continuous adjustment of the atmospheric circulation can induce the variation of cloud distribution. This alternation of cloud cover will further impact radiation budget, causing a feedback effect on the climate. General circulation model (GCM) simulations have found that different types of clouds tend to vary differently with global warming [7], [8]. However, these feedback changes cannot be accurately represented in GCMs, making cloud feedback a major source of uncertainty in estimating climate sensitivity [9], [10]. Hence, the accurate observation of cloud cover is essential for constraining and reducing these uncertainties in climate models.

Satellite observations are powerful platforms for observing cloud cover on a global scale [11], [12]. Spaceborne passive spectral imagers, such as advanced geosynchronous radiation imager, advanced Himawari imager (AHI), and moderate-resolution imaging spectroradiometer (MODIS), can monitor cloud evolutions on different spatiotemporal scales [13], [14]. Classical cloud detection algorithms for such instruments are developed based on the spectral feature differences that clouds typically exhibit a higher albedo and a lower emissivity compared with underlying surfaces, allowing for cloud detection by setting appropriate thresholds [15]. However, these methods face challenges due to varied surface conditions, cloud types, spectral bands, and complex threshold tests [16], [17]. For instance, when the radiative features of the cloud and the surface are similar, selecting the threshold becomes difficult. High albedo surfaces, such as desert and snow, are often misclassed as clouds. Meanwhile, the threshold applied to thick clouds can cause thin clouds being misidentified as clear skies. In addition, the differences in the number and center wavelength of spectral

channels among various imagers necessitate adjustments to the corresponding cloud detection algorithms, which also affects the detection outcomes across different instruments [18]. Moreover, the reliance on the albedo channel in classical algorithms makes nighttime cloud detection unfeasible, indirectly hampering a comprehensive understanding of cloud diurnal variations. In comparison, spaceborne active radar systems, such as cloud profile radar (CPR) and cloud-aerosol lidar with orthogonal polarization (CALIOP), can penetrate cloud layers to obtain cloud vertical structure due to their high sensitivity to cloud droplets [19], [20], enabling contemporary algorithms to better discriminate between cloud and noise signals and resulting in enhanced cloud detection [21], but their narrow beamwidth limits the spatial coverage and temporal continuity of observed clouds. Therefore, integrating active radar with a passive imager can leverage the strengths of both instruments, improving cloud cover observations from space and, consequently, enhancing the development of downstream cloud-related product.

Machine learning (ML) techniques excel at handling complex nonlinear relationships between multiple variables through data-driven approaches, providing solutions for improving the accuracy of cloud detection for spaceborne instruments [22], [23]. Numerous studies have leveraged the advantages of the ML methods to develop high-precision cloud detection algorithms. By combining both active and passive observations, these approaches can result in improved cloud detections compared with the manual threshold test results [24], [25], [26]. For instance, Wang et al. [27] utilized spectral observations from the visible infrared imaging radiometer and integrated with CALIOP feather classification data to train two random forest (RF) models for clouds and their phase detection. The evaluation results show a higher accuracy compared with official products. Likewise, Li et al. [28] developed two deep neural network models: one for the entire day and another for daytime only by using merged cloud products derived from CPR and CALIOP as reference labels. These models exhibited better performance in cloud detection compared with traditional methods.

The performance of ML-based cloud detection algorithms typically relies on the surface conditions, selected spectral channels, the quality of training data, etc. Previous ML studies commonly selected specific factors tailored to individual surface types for training independent models [27], [29], [30] or encoded varying numbers of surface features to develop models applicable to vary surface types [31]. These approaches have been demonstrated to enhance model accuracy. However, geostationary satellites, such as Himawari-8, observe wide range, experiencing significant variations in both surface type and cloud distribution within their domains. This often induces an imbalanced distribution of cloud and clear-sky samples, which will cause overfitting in ML models. Models may disproportionately favor categories with a larger sample size while underrepresenting those with smaller sample sizes, thereby inflating the accuracy of dominant categories at the expense of others. Although this issue is important, prior studies have not fully considered such imbalances. Consequently, addressing uneven cloud cover distribution, as well as reconciling the relationship between this imbalance and surface-type representation, are critical questions.

In addition to these challenges, observation parallax between different instruments is also important [32], [33]. The zenith angle of geostationary satellites varies significantly. Bieliński [34] demonstrates that this could lead to significant discrepancies in cloud top detections from different instruments over high-latitude regions, where far from the subsatellite point of geostationary satellite. This issue can lead to mislabeled pixels, especially at cloud edges, and introduce substantial noise that affects the stability and accuracy of the trained models. Moreover, the efficacy of an ML algorithm is contingent upon its alignment with specific problems and data [35]. Many previous studies manually select multiple ML models and iteratively identify the best hyperparameter combination based on the empirical values [22], [31]. This process, constrained by human experience, may result in suboptimal selection of models, leading to compromise in capturing the importance of each feature. The development of automatic machine learning (AutoML) techniques offers new possibilities for more objective model training and hyperparameter optimization.

In this study, an objective and efficient AutoML framework [36] is adopted to develop the cloud detection algorithm for the AHI onboard Himawari-8 [37], [38]. To elucidate the relationship between cloud cover variations and surface type mode selection, resampling techniques, model weights modification, and output threshold adjustment of the model mapping function are employed to ensure balanced cloud and clear-sky pixels' detections over various surfaces. In addition, cooperative game theory is applied to calculate the marginal contribution value of each detection factor for each sample, and these results are statistically analyzed to enhance the interpretability of the detection factors.

The rest of this article is organized as follows. Section II introduces the construction of the datasets and the cloud detection algorithms. In Section III, the influence of cloud cover on model outcomes is quantitatively evaluated. The results between the optimization model and AHI official cloud mask are compared by some cases and long-time data. In Section IV, the influence of parallax on algorithm training, the threshold selection of reference factors and vector angle, and different cloud cover balance schemes are discussed. Finally, Section V concludes this article.

II. Data and Methods

Fig. 1 illustrates the fundamental steps involved in constructing a cloud detection algorithm for AHI using AutoML. The process encompasses data matching, data preprocessing, model training, and outputs. This section provides a detailed account of the model's development, including the multisource satellite data and their matching methods, as well as the employed preprocessing techniques, such as the strong collinearity elimination method and surface type characterization modes. In addition, we outline the advantages of the AutoML framework and specify the parameter settings used during training. Finally, we describe the calculation of marginal contribution values for the detection factors and the evaluation metrics for the model.

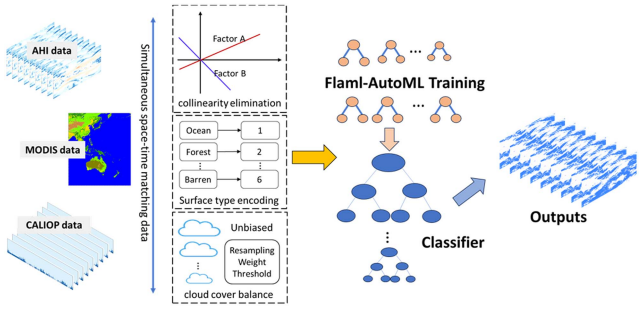

Fig. 1. Flowchart of AHI cloud detection algorithm construction.

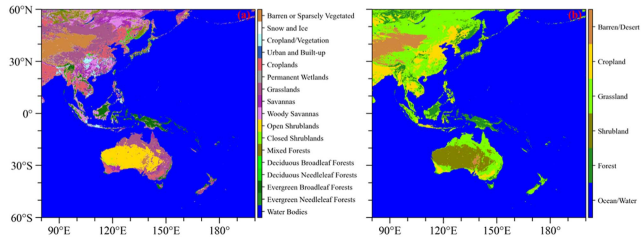

Fig. 2. LCs (a) before and (b) after simplification in the study area.

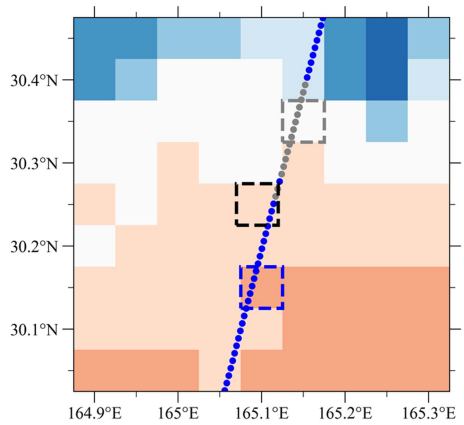

Fig. 3. Schematic diagram of CALIOP and AHI data matching process. The background color represents different AHI BT observation values. The scatters represent the CALIPSO nadir points, the blue scatters represent the cloudy profiles, and the gray scatters represent the clear-sky profiles. The dotted rectangular frame highlights the AHI observation pixels.

A. *Passive and Active Satellite Observations*

The AHI with 16 bands covers a broad spectrum of wavelengths ranging from 0.46 to 13.3 μm onboard Himawari-8, which is a geostationary meteorological satellite launched on 7 October 2014 and currently positioned at 140.7°E [39]. The full disk scan of AHI has a temporal resolution of 10 min and a spatial resolution ranging from 0.5 to 2 km, allowing for the detection and tracking of rapid weather phenomena [26], [40], [41]. In this study, the level 1 grid full disk data with a spatial resolution of 0.05° × 0.05° and a temporal resolution of 10 min from the Japan Aerospace Exploration Agency Himawari Monitor (P-Tree) system spanning over 2016–2017 is utilized. The official cloud mask data are from the level 2 cloud products, including cloud, probably cloud, clear sky, and probably clear sky [16], [17], that are employed to compare our algorithm's results. For the evaluation, these data are grouped into two scenarios: "official-1," where cloud and probably cloud are classified as cloud, while the remaining two are classified as clear sky, and "official-2," where only the confident cloud level is classified as cloud.

CALIOP is a dual-wavelength orthogonal polarization lidar onboard the cloud-aerosol lidar and infrared pathfinder satellite observations (CALIPSOs) [42]. It has high sensitivity to cloud droplet particles, enabling the accurate identification of cloud tops and a relatively complete depiction of the cloud's vertical structure, especially for thin clouds. Thus, CALIPSO data are selected to label the AHI observations and compare the results with AHI official products. The collected CALIOP level 2 cloud layer product during 2016–2017 is utilized, with a horizontal resolution of 1 km and the maximum vertical profile consists of ten layers to determine the presence of clouds or clear sky at a specific position. In detail, layers with at least one identified cloud are marked as cloudy, while those with no cloud layer are labeled as clear sky.

We incorporate MODIS annual International Geosphere–Biosphere Programmer (IGBP) global land cover (LC) type data, which implement the IGBP global vegetation classification scheme and encompass 16 distinct LC types and can differentiate between various surface conditions. However, more surface classification may lead to overly complex trained models. In addition, the absorption and emission characteristics of some surfaces are similar. For example, cropland and vegetation have almost the same reflection and emission features (see Figs. 2 and 3 in [27]). Therefore, to effectively represent surface conditions and streamline the model's complexity, the land type simplification method proposed by Wang et al. [27] is adopted to convert the original 16 LCs into five broader categories (refer to Fig. 2).

B. *Data Training, Validating, and Testing*

Simultaneous colocated data from active and passive satellite observations are selected to construct the training and validation datasets. The observing time difference between CALIOP and AHI must be less than 5 min. Since the resolution of the CALIOP (1 km) is higher than that of the AHI data (approximately 5 km), the CALIOP data are aligned with the AHI pixels, as shown in Fig. 3. When CALIPSO passes through the observation field of Himawari-8, the larger AHI pixel serves as a reference area (dashed rectangle). CALIOP nadir points indicating clear sky (blue dots) or clouds (gray dots) are selected as training or validation samples, while mixed observations shown in black rectangle are excluded. The AHI official product does not provide nighttime cloud top height (CTH) product. Thus, the CALIOP and AHI observations are matched first according to the way mentioned above, and the averaged CTH from CALIOP within these matched AHI pixels is used as a reference in the 9×9 grid area centered on the matched AHI pixels. Then, parallax correction on these pixels is performed to rematch the corrected

pixels [32], [33], [34]. Here, the range of 9×9 is obtained by multiple matching experiments, which can maintain the small change of the sample size before and after the parallax correction and maximize the regional representativeness of CALIOP CTH data. It should be noted that this parameter is a statistical result and does not have practical physical significance.

To enable our algorithm to effectively perform in both day and night, we select ten AHI infrared brightness temperature observation channels (BT07–BT16) and their differences (e.g., BTD0708 represents the difference between BT07 and BT08) as potential detection factors. Prior to training, collinearity analyses are conducted to understand the relationships among these potential actors. Specifically, all potential detection factors are standardized, and the six factors, which have the highest correlation with the training labels, are selected as the reference basis. Then, the standardized potential factors are treated as vectors and the vector angles between the remaining factors and the six reference factors are calculated. Any remaining factor with a vector angle greater than 160° or less than 20°, which is identified as optimal through the performance tests, is considered strongly collinear with the references and is eliminated. It is noted that each reference factor removes strong collinear factors in order of decreasing correlation. Factors eliminated by a highly correlated reference factor are not reconsidered in subsequent steps. Through this process, the remaining factors are used as the primary detection factors for training the models.

In addition, auxiliary detection factors, such as LC, satellite azimuth angle (SAA), solar zenith angle (SOZ), solar azimuth angle (SOA), time, latitude, and longitude, are included to account for the sun's position and Himawari-8s orientation. For LC, it can be directly input into the model as a predictor or used to train the model separately for each LC. Here, three coding methods are used for the former:

1) integer coding denoted as LC1, which maps each LC to a unique integer;
2) one-hot coding (LC2) [43], [44], which converts each LC into a binary vector with 1 at the corresponding position and 0 elsewhere;
3) frequency coding (LC3), which replaces each LC with its overall frequency in the dataset.

Training models separately for different LCs are denoted as LC4, with specific models, such as LC4-ocean for the ocean.

Through the above accurate data matching and reasonable feature selection processes, the samples after the 7th of each month in 2017 are used to form the dataset for model training, while those before the 7th are selected to validate performance. The training set includes 5 008 940 samples (3 542 711 cloudy), and the validation set includes 1 317 860 samples (943 531 cloudy). To further test the generalization of our optimal algorithms, we selected CALIOP and AHI data from January, April, July, and October 2016 to represent seasonal variations as the test dataset.

C. Three Processes for Cloud Cover Data Balance

The terrain in the field of view of Himawari-8 is complex, with significant spatial–temporal variations in water vapor transport and other meteorological conditions, leading to substantial differences in cloud occurrence. As a result, the ratio of clear sky to cloud samples size in the training set is imbalanced, approximately 1:3 in our study. To balance the sample size, the construction of AutoML cloud detection model is intervened in three processes separately:

1) data preprocessing (downsampling scheme and oversampling scheme);
2) model training (weight scheme);
3) model output (threshold scheme).

In data preprocessing, downsampling is used to randomly remove a portion of cloud samples, while oversampling is applied to randomly duplicate clear samples [45], [46]. It is important to note that the resampling process is entirely random. During model training, higher weights are assigned to clear-sky samples to increase model attention to this kind of sample [47], [48]. In the model output phase, the thresholds of the mapping function are adjusted to improve clear-sky detection and reduce false positives [49], [50].

D. AutoML Model Construction and Interpretability

To avoid suboptimal models from manual selection and empirical hyperparameters, we use Flaml, a lightweight Python AutoML library to optimize model performance through adaptive search strategies by considering learner selection, hyperparameters, training data size, resampling costs, and errors [39]. Compared with other AutoML frameworks, Flaml offers a lightweight design, low memory usage, and high accuracy [36], [38], showing competitiveness with human-tuned models in various tasks [38], [51]. The Flaml learner list includes three algorithms: RF, XGBoost, and a highly efficient gradient boosting decision tree (LightGBM). RF is an ensemble algorithm that uses decision trees as base learners, training multiple trees through random sampling and combining their results [52]. While both XGBoost and LightGBM are gradient boosting models that iteratively train decision trees, with each model correcting the errors of the previous ones to improve prediction accuracy [53], [54]. The optimal model can be achieved by setting a training time budget and using the early stopping option, which is set to 300 s in this study. In addition, the weighted scheme is implemented by specifying a class weight ratio for the model during training, which is based on the actual sample sizes of the two classes in the training set.

We also introduce a posteriori method—SHapley Additive exPlanations (SHAP) [55], [56]—to complement the feather importance score's role in actual classification. SHAP values are calculated using cooperative game theory to explain the impact of different detection factors on the final results. For instance, a sample is predicted as a cloud pixel by three detection factors (e.g., A, B, and C) in a trained model. The contribution of these factors can be translated into calculating the SHAP values, with the calculation process being as follows. First, the benchmark output probability of cloud pixels is calculated without any feature input based on training dataset. Each feature factor is then input separately, and the change in cloud pixel probability is determined compared with benchmark output probability, representing the SHAP value of each feature. A SHAP value

TABLE I
CONFUSION MATRIX

	AutoML&AHI (clear sky)	AutoML&AHI (cloudy sky)
CALIOP (clear sky)	TN	FP
CALIOP (cloudy sky)	FN	TP

of 0.4 for feature A indicates that a one-unit increase in A raises the probability of predicting cloud pixels by 0.4 units. From multiple cases of statistical analysis, it can be determined whether a detection factor's prediction result is more inclined toward clear-sky or cloud pixels at a specific threshold. This combined analysis with the feature importance scores is detailed, as described in Section III.

E. AutoML Model Performance Evaluation Metrics

The detection accuracy of both cloud pixels and clear-sky pixels is equally important for model performance. Ideally, the algorithm should predict cloud and clear-sky pixels with consistent accuracy, resulting in high overall accuracy. In other words, the algorithm should be unbiased. Therefore, we define some evaluation metrics based on the confusion matrix (see Table I) to evaluate the performance of each model.

Among the metrics, true positives (TPs) are the pixels classified as clouds by both AutoML/AHI and CALIOP, and true negatives (TNs) are the pixels classified as clear by both. False negatives (FNs) occur when AutoML/AHI classifies a pixel as clear but CALIOP detects it as cloudy, while false positives (FPs) occur when AutoML/AHI classifies a pixel as cloudy but CALIOP classifies it as clear. The true positive rate (TPR) measures the proportion of correctly detected cloudy pixels, and the true negative rate (TNR) measures the proportion of correctly detected clear pixels. These two metrics directly characterize the prediction accuracy of cloud and clear-sky pixels. Conversely, the false positive rate (FPR) and false negative rate (FNR), which can be calculated by $1-\mathrm{TNR}$ or $1-\mathrm{TPR}$, respectively, measure the proportions of clear pixels misidentified as cloudy and cloudy pixels misidentified as clear. The deviation rate (DR) is the absolute difference between TPR and TNR, indicating the detection bias of the models regarding cloud and clear-sky pixels. A lower DR means less bias. The true skill statistic (TSS) ranges from -1 to 1, with higher values indicating better overall detection performance

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{1}$$

$$\mathrm{FNR} = 1 - \mathrm{TPR} \tag{2}$$

$$\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}} \tag{3}$$

$$\mathrm{FPR} = 1 - \mathrm{TNR} \tag{4}$$

$$\mathrm{DR} = \mathrm{ABS}\left(\mathrm{TPR} - \mathrm{TNR}\right) \tag{5}$$

$$\mathrm{TSS} = \mathrm{TPR} - \mathrm{FPR}. \tag{6}$$

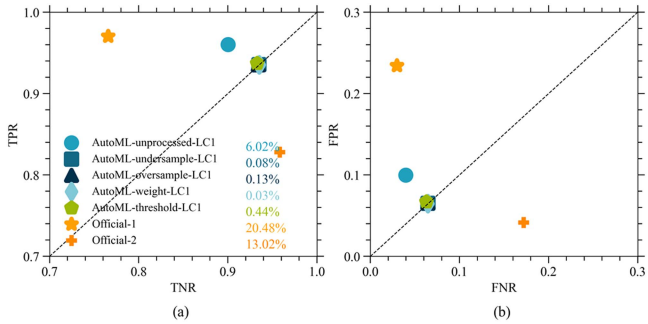

Fig. 4. TPR, TNR, FPR, and FNR are calculated to compare the detection results of AutoML without sample balance processing with those of AutoML enhanced by resampling, weight, and threshold methods, as well as the detection results of official products.

III. RESULTS

In this section, the performance of the models constructed through the above processes is systematically evaluated by comparing the cloud detection results with CALIOP and AHI official cloud mask products.

A. Evaluation of Model Detection Bias on Cloud and Clear-Sky Pixels

The unbiasedness of model cloud detection results is crucial for the accurate cloud cover quantification. To demonstrate our improved model performance, we evaluate the results using both the AHI official cloud mask and the unprocessed AutoML outcomes (with no cloud cover balance treatment) cloud mask is conducted to explicit improved model performance. The LC1 scheme is used as an example to illustrate the distinctions and improvements in the model. This comparison can show the effectiveness of the AutoML adjustments in enhancing cloud detection accuracy and reducing biases in cloud cover quantification.

Fig. 4 displays the TPR, TNR, FPR, FNR, and DR for AutoML-unprocessed-LC1, AutoML-downsample-LC1, AutoML-oversample-LC1, AutoML-weight-LC1, AutoML-threshold-LC1, official-1, and official-2 cloud mask on the validation set. The proximity of the mark to the black dotted line indicates a smaller deviation in detecting cloud and clear-sky pixels. Compared with CALIOP, the TPR of AutoML-unprocessed-LC1 is greater than TNR, and the FPR is greater than FNR, indicating that the cloud pixels are overestimated at the expense of clear-sky pixels. As previously mentioned, this result is due to the difference in the number of cloud and clear-sky samples in the training set because, in the Himawari-8 observation area, most of the areas are oceans, and the cloud cover over the ocean is large that results in an imbalance in the actual class sample size. The overestimation observed in official-1 cloud mask is more pronounced. Compared with CALIOP, the TPR of official-1 exceeds 97%, meaning that almost all cloud pixels can be detected. However, the TNR is small, and the FPR is more than 20%, indicating that the identified cloud pixels contain a considerable number of misdetected clear-sky pixels, and therefore are the farthest from the black dotted line. Official-2 is almost the opposite scenario, with a TPR of only

TABLE II
QUANTITATIVE ANALYSIS OF CLOUD DETECTION RESULTS USING DIFFERENT CLOUD COVER BALANCE METHODS ON VALIDATION SET

	TSS	TPR	TNR	DR
unprocessed	86.04%	96.03%	90.01%	6.02%
undersample	87.04%	93.56%	93.48%	0.08%
oversample	86.95%	93.54%	93.41%	0.13%
weight	87.07%	93.55%	93.52%	0.03%
threshold	87.01%	93.73%	93.28%	0.44%
Official-1	73.60%	97.04%	76.56%	20.48%
Official-2	78.63%	82.81%	95.83%	13.02%

82%, while the TNR is more than 95%, but with a value of the TSS larger than that of official-1. This implies that the probably cloudy level of the official product contains fewer real cloud pixels than real clear-sky pixels inside.

In Fig. 4(a), all the cloud detection algorithms based on AutoML have lower DR. Quantitative values are provided in Table II. AutoML-downsample-LC1, AutoML-oversample-LC1, AutoML-weight-LC1, and AutoML-threshold-LC1 show lower DR than AutoML-unprocessed-LC1. It can be seen that this low DR is at the expense of reducing TPR to increase TNR. The results of the equilibrium process reduce the FPR but also reduce the TPR. However, the TSS of the model with enhanced detection bias increases by about 1%, indicating that these synchronous reductions improve the performance of the algorithms. In addition, we can see that the cloud detection models constructed by the three different cloud cover equalization schemes have smaller DR and larger TSS, which effectively solves the interference of the uneven category in the training samples on the results (in the supporting information section, we also provide the evaluation results of LC2, LC3, and LC4 schemes, which lead to similar conclusions), but in fact, some of them have potential problems. In Section IV, we further discuss the influence of these three schemes on our algorithm in detail. In the later part, in order to simplify the analysis process, we choose the threshold scheme as the equilibrium scheme of cloud cover difference to further evaluate the performance of the model. The specific reason is given in Section IV.

B. Cloud Detection Under Different Surface Types

We evaluate the performance of the AutoML models based on the threshold scheme and two surface-type characterization methods in the whole field of view of the Himawari-8 and the local surface through quantitative and representative case analysis in this part. The outcomes from the models, including AutoML-threshold-LC1, AutoML-threshold-LC2, AutoML-threshold-LC3, and AutoML-threshold-LC4, are compared first. In the evaluation of the whole region of Himawari-8, LC1, LC2, LC3, and LC4 modes exhibit similar values for TPR, TNR, and TSS, treating cloud and clear-sky pixels equally. The specific quantitative results are shown in Table III. Numerically, the TSS

TABLE III
EVALUATION OF CLOUD DETECTION RESULTS ON DIFFERENT LCS USING VALIDATION SETS

	TSS	TPR	TNR	DR
All	87.01%	93.73%	93.28%	0.44%
	87.24%	*93.96%*	*93.28%*	*0.68%*
	(87.03%)	(93.85%)	(93.18%)	(0.67%)
	(87.32%)	*(94.06%)*	*(93.27%)*	*(0.79%)*
	73.60%	**97.04%**	**76.56%**	**20.48%**
	78.63%	***82.81%***	***95.83%***	***13.02%***
Ocean	85.94%	94.14%	91.79%	2.35%
	86.15%	*94.37%*	*91.78%*	*2.60%*
	(85.96%)	(94.28%)	(91.68%)	(2.60%)
	(86.48%)	*(94.07%)*	*(92.41%)*	*(1.65%)*
	69.89%	**98.40%**	**71.49%**	**26.91%**
	80.67%	***85.60%***	***95.07%***	***9.47%***
Forest	87.00%	94.77%	92.23%	2.55%
	87.40%	*94.92%*	*92.48%*	*2.44%*
	(86.92%)	(94.57%)	(92.35%)	(2.22%)
	(87.52%)	*(94.53%)*	*(93.00%)*	*(1.53%)*
	67.43%	**91.42%**	**76.01%**	**15.41%**
	67.17%	***72.55%***	***94.62%***	***22.07%***
Shrub-land	87.95%	89.20%	98.75%	9.54%
	88.94%	*90.41%*	*98.53%*	*8.12%*
	(87.63%)	(89.08%)	(98.55%)	(9.48%)
	(90.98%)	*(93.31%)*	*(97.67%)*	*(4.36%)*
	89.96%	**92.41%**	**97.55%**	**5.14%**
	59.23%	***59.74%***	***99.50%***	***39.76%***
Grass-land	86.44%	91.53%	94.91%	3.38%
	86.61%	*91.73%*	*94.88%*	*3.15%*
	(86.59%)	(91.81%)	(94.78%)	(2.98%)
	(87.51%)	*(93.97%)*	*(93.53%)*	*(0.44%)*
	73.45%	**89.58%**	**83.87%**	**5.71%**
	68.14%	***71.69%***	***96.46%***	***24.77%***
Crop-land	88.11%	91.86%	96.25%	4.39%
	88.81%	*92.12%*	*96.70%*	*4.58%*
	(87.76%)	(91.41%)	(96.35%)	4.94%)
	(89.58%)	*(94.61%)*	*(94.97%)*	*0.36%)*
	82.43%	**93.49%**	**88.94%**	**4.54%**
	72.72%	***73.89%***	***98.83%***	***24.94%***
Barren	83.76%	87.76%	96.00%	8.24%
	83.86%	*87.86%*	*96.00%*	*8.14%*
	(84.47%)	(88.53%)	(95.94%)	(7.41%)
	(86.81%)	*(92.25%)*	*(94.55%)*	*(2.30%)*
	78.37%	**92.53%**	**85.84%**	**6.68%**
	58.15%	***60.50%***	***97.65%***	***37.16%***

Statistics are denoted as follows: Normal (AutoML-threshold-LC1), italicized (AutoML-threshold-LC2), bracketed normal (AutoML-threshold-LC3), bracketed italicized (AutoML-threshold-LC4), bolded (official-1), and bolded italicized (official-2).

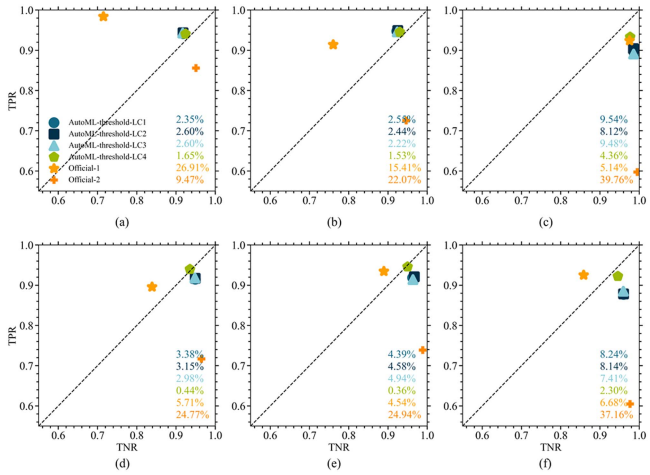

Fig. 5. TPR, FPR, and DR of different AutoML algorithms on different LCs in the validation set. The subgraphs (a)–(f) correspond to LCs of ocean, forest, shrubland, grassland, farmland, and barren, respectively.

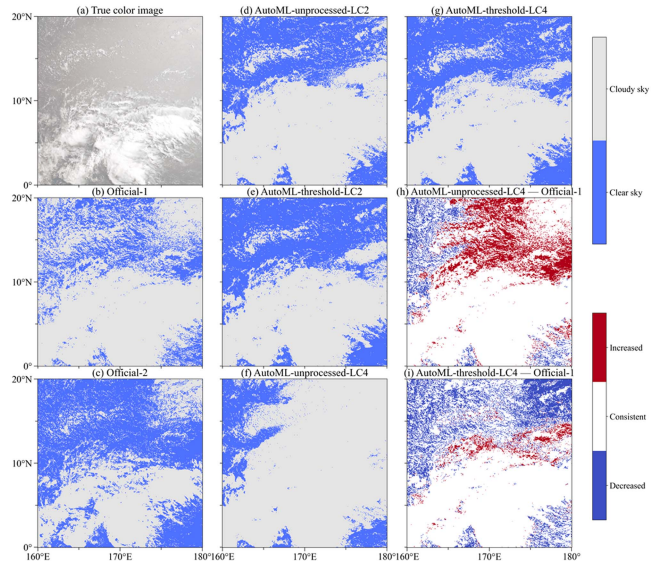

Fig. 6. Comparison of AHI operational cloud mask products and our algorithm results at 2050 UTC on 20 June 2017. (a) AHI true-color image. (b) Official-1 cloud mask. (c) Official-2 cloud mask. (d) AutoML-unprocessed-LC2 cloud mask. (e) AutoML-threshold-LC2 cloud mask. (f) AutoML-unprocessed-LC4 cloud mask. (g) AutoML-threshold-LC4 cloud mask. (h) AutoML-threshold-LC4 cloud mask minus official-1 cloud mask. (i) AutoML-threshold-LC4 cloud mask minus official-2 cloud mask. The terms "increased," "decreased," and "consistent" are used to describe the differences in cloud and clear-sky pixels detection: "Increased" represents cloud pixels identified by the former algorithm but missed by the latter, "decreased" indicates cloud pixels detected by the latter algorithm but not by the former, and "consistent" denotes cases where both algorithms produce identical results.

of the models exceed 87%, with individual accuracies for cloud and clear-sky predictions surpassing 93%, and DR consistently below 1%. Compared with official-1's TSS (73.60%) and DR (20.48%), the overall TSS of the models has improved by nearly 15%, while DR has decreased by approximately 20%.

Among AutoML-threshold-LC1, AutoML-threshold-LC2, and AutoML-threshold-LC3, AutoML-threshold-LC2 exhibits better overall performance (higher TSS), indicating that one-hot encoding better captures differences between categorical features. However, the models demonstrate different detection accuracies for different LCs. In Fig. 5(a) and (b), TSS and DR values for LC1, LC2, LC3, and LC4 modes are relatively consistent. LC4 is slightly better because it takes into account the changes in different LCs and selects different detection factors. However, official-1 results show that cloud pixels over oceans and forests are significantly overestimated, resulting in smaller TSS values and larger DR values. Official-2 shows the opposite result, where clear-sky pixels are significantly overestimated, but the TSS values are about 10% larger and the DR is better compared with Official-1. This means that, over the ocean, the cloud detection results of official-2 are much more practical. Over other LCs, each scheme shows results different from the overall assessment. In Fig. 5(c)–(f), it can be seen that LC1, LC2, and LC3 significantly deviate from the black dashed line than LC4, showing overestimations of clear-sky pixels. Even over shrubland, the TSS of the former three is lower than the official-1 results, mainly because using LC types as detection factors in the model does not allow modifying the output mapping function for different LCs. Due to the large area of the ocean and forest, with higher cloud cover overhead, the overall number of cloud pixels exceeds clear-sky pixels. Modifying the mapping function does not significantly impact these regions but leads to an overestimation of clear-sky pixels in other regions with lower cloud cover. Hence, the observed deviations are shown in Fig. 5(c)–(f). LC4 allows for the modification of output threshold values based on the actual cloud cover over different LC types. Thus, in Fig. 5, LC4 aligns most closely with the black dashed line. In addition, Table III also lists TSS values for LC4 that surpass both official results and other AutoML algorithms, with DR consistently lower than official results and some AutoML models. If we just focus on the official results, we can see that, on land, official-2 has a lower TSS and a larger DR compared with official-1. This means that over land (except for the forest), official-1 has better utility for cloud detection results.

To gain a detailed understanding of the various cloud detection models, we selected three examples for qualitative analysis, representing an all-ocean area, a land–sea intersection, and a barren region, respectively. Fig. 6 presents a case from 20 June 2017, at UTC 2050, situated entirely over the ocean. Fig. 6(a) shows a true-color image synthesized from the AHI visible channel with enhanced contrast and brightness, which reveals the true distribution of clouds and clear skies. Fig. 6(b) and (c) displays the official-1 and official-2 cloud mask. The cloud mask using AutoML-unprocessed-LC2, AutoML-threshold-LC2, AutoML-unprocessed-LC4, and AutoML-threshold-LC4 is shown in Fig. 6(d)–(g). To highlight differences in detail, Fig. 6(h)–(i) depicts the subtraction of the AutoML model cloud mask (LC4 scheme) from the official-1 cloud mask. Overall, compared with Fig. 6(c), the AutoML results exhibit a lower missed detection. The false detection [refer to the "Decreased" part in Fig. 6(i)] of AutoML is lower compared with Fig. 6(b). The results after balancing processing are also more consistent with the actual distribution in the true-color image, as evident from the changes in Fig. 6(f) and (g). In Fig. 6(f), considering only ocean, the

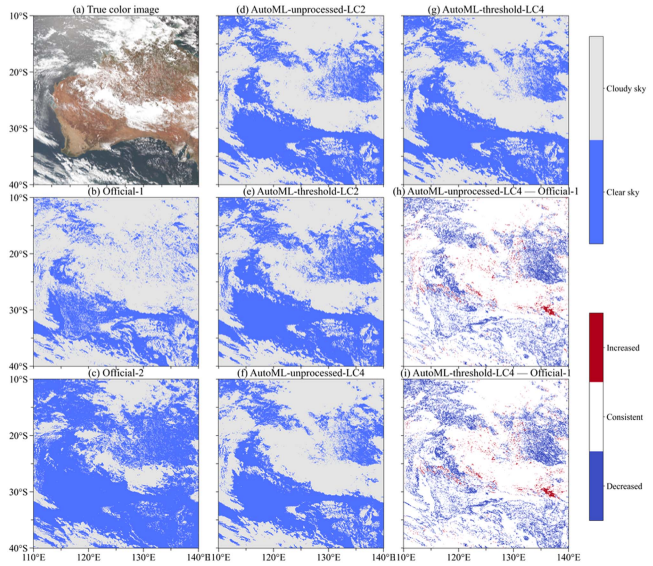

Fig. 7. Comparison of AHI operational cloud mask products and our algorithm results at 05:50 UTC on 20 January 2017. (a) AHI true-color image. (b) Official-1 cloud mask. (c) Official-2 cloud mask. (d) AutoML-unprocessed-LC2 cloud mask. (e) AutoML-threshold-LC2 cloud mask. (f) AutoML-unprocessed-LC4 cloud mask. (g) AutoML-threshold-LC4 cloud mask. (h) AutoML-threshold-LC4 cloud mask minus official-1 cloud mask. (i) AutoML-threshold-LC4 cloud mask minus official-2 cloud mask. The terms "increased," "decreased," and "consistent" are used to describe the differences in cloud and clear-sky pixels detection: "Increased" represents cloud pixels identified by the former algorithm but missed by the latter, "decreased" indicates cloud pixels detected by the latter algorithm but not by the former, and "consistent" denotes cases where both algorithms produce identical results.

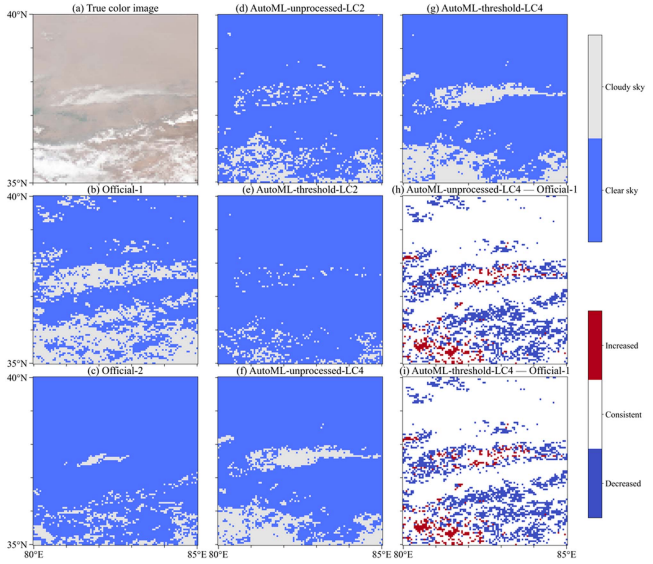

Fig. 8. Comparison of AHI operational cloud mask products and our algorithm results at 05:00 UTC on 28 April 2017. (a) AHI true-color image. (b) Official-1 cloud mask. (c) Official-2 cloud mask. (d) AutoML-unprocessed-LC2 cloud mask. (e) AutoML-threshold-LC2 cloud mask. (f) AutoML-unprocessed-LC4 cloud mask. (g) AutoML-threshold-LC4 cloud mask. (h) AutoML-threshold-LC4 cloud mask minus official-1 cloud mask. (i) AutoML-threshold-LC4 cloud mask minus official-2 cloud mask. The terms "increased," "decreased," and "consistent" are used to describe the differences in cloud and clear-sky pixels detection: "Increased" represents cloud pixels identified by the former algorithm but missed by the latter, "decreased" indicates cloud pixels detected by the latter algorithm but not by the former, and "consistent" denotes cases where both algorithms produce identical results.

overestimation of cloud pixels is more severe than in Fig. 6(d) because the proportion of cloud samples to clear-sky samples exceeds three times. This also reflects the importance of balance processing. Therefore, the algorithms with modified output threshold significantly reduce the proportion of clear-sky mistakenly detected as clouds, with only a partial misclassification of clouds as clear sky, which is acceptable.

Fig. 7 presents a case from 20 January 2017, at UTC 05:50, with dominant surface types of shrubland, barren, grassland, and surrounding ocean. Taking Fig. 7(f) as the reference, Fig. 7(b) tends to overestimate clouds, notably misidentifying the land edge as clouds in the southwest of Australia [refer to the "decreased" part in Fig. 7(h)]. Fig. 7(c) shows a clear underestimation of cloud in the western ocean, southern ocean, and central part of the continent, compared with the true-color image. If Fig. 7(g) is considered as the reference, the overestimation [refer to the "decreased" part in Fig. 7(i)] in Fig. 7(b) and underestimation in Fig. 7(c) become more pronounced over the Australian continent and surrounding ocean. The results for oceanic regions in Fig. 7(d)–(g) are relatively consistent, but within the continent, Fig. 7(d) and (e) exhibits an underestimation of cloud pixels. This is consistent with the previous evaluation results. Overall, Fig. 7(g) avoids overestimating clouds over the ocean while retaining a higher similarity to the true-color image in the presence of genuine cloud pixels over land.

Fig. 8 provides a cloud mask case entirely over barren on 28 April 2017, at 05:00 UTC. Fig. 8(b) and (c) shows an overestimation and underestimation of clouds [refer to Fig. 8(a)], respectively. The algorithm under the LC2 scheme exhibits considerable underestimation both before and after processing, compared with the true-color image. The underestimation in Fig. 8(d) is mainly due to the less adaptability of the detection factors selected for the entire region to the barren, resulting in lower detection accuracy. Fig. 8(e) is due to further underestimation of cloud detection results in Fig. 8(d) caused by modifying the output threshold for the entire region, and the cloud cover in this area is less than the result in Fig. 8(c). In comparison, the algorithm under the LC4 scheme fits more closely to the true-color image.

C. *Interpretability of AutoML Models*

Here, we further explain the cloud detection model of LC2 and LC4 modes. The top-ranking importance scores for AutoML-threshold-LC2 and AutoML-threshold-LC4 are shown in Fig. 9. In the AutoML-threshold-LC2 model, BTD1516 is identified as the most important detection factor. The importance score of LC (ocean in Fig. 9) is also ranked in the top five, indicating that surface properties play a significant indicative role. In AutoML-threshold-LC4-ocean, BTD1516 continues to be the most important feature, accounting for an even higher proportion. This suggests that when surface properties are consistent and the surface area is large, it greatly influences the detection factor in the entire regional model. To explain the significance of the

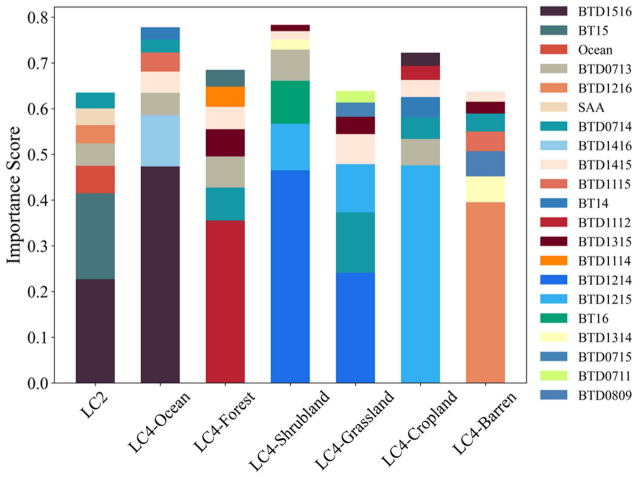

Fig. 9. Top seven detection factors with importance scores and their importance values in the AutoML algorithms based on the LC2 and LC4 modes.

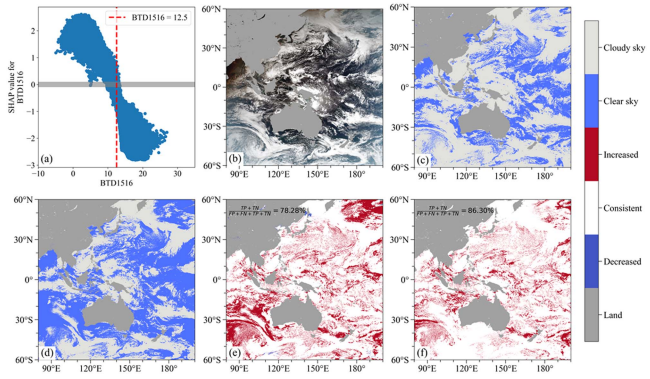

Fig. 10. Interpretation of the AutoML-threshold-LC4-ocean algorithm based on the SHAP values using an example at 01:10 UTC on 11 January 2017. (a) SHAP values of BTD1516 in the training set and AHI observations of BTD1516. (b) AHI true-color map. (c) AutoML-threshold-LC4-ocean algorithm detection results. (d) Threshold-based BTD1516 detection results. (e) Detection results of AutoML-threshold-LC4-ocean algorithm subtract the detection results of BTD1516 based on the threshold. (f) Detection results of AutoML-threshold-LC4-ocean algorithm subtract the detection results of BTD1516 and BTD1416 based on threshold. The terms "increased," "decreased," and "consistent" indicate differences in cloud and clear-sky pixel detection and have the same definitions as those provided in Fig. 6. In (e) and (f), the TP, FP, TN, and FN represent the evaluation results of BTD1516 and BTD1416, calculated based on a specific threshold, using the detection results in (c) as the ground truth.

detection factors, the reliability of detection factor contribution is validated using the SHAP framework, with a focus on the LC4-ocean model. SHAP values are calculated for the top two features (BTD1516 and BTD1416) in here to demonstrate their importance.

Fig. 10(a) shows the distribution of BTD1516 SHAP values. All data within the range of 0.1 are extracted when the BTD1516 SHAP value is equal to 0. The value of BTD1516 observed values with a cumulative probability equal to 0.5 is then determined by the cumulative probability density, which is approximately 12.5, as shown by the red dashed line in Fig. 10(a). The AHI observation from 11 January 2017 at 01:10 is selected to demonstrate the function of this threshold. Fig. 10(b) is the true-color image, with the gray area representing the masked land regions. Fig. 10(c) illustrates the cloud mask based on

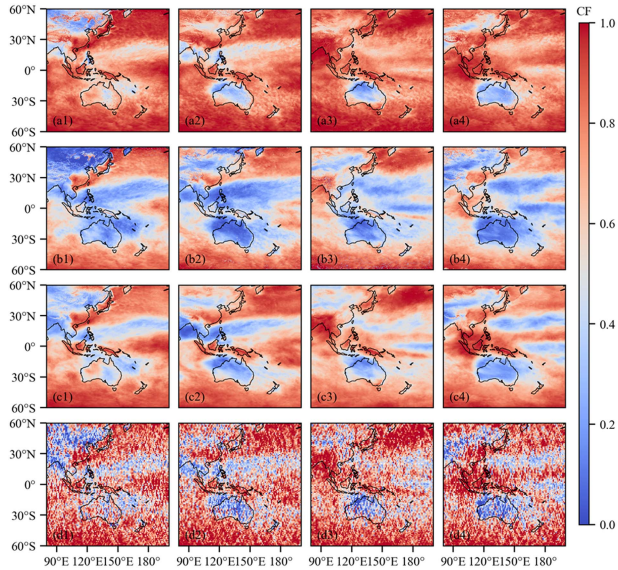

Fig. 11. Distribution of monthly averaged CF in January, April, July, and October 2016. (a) Average CF distribution calculated under the official-1 scenario that only includes daytime cloud detection results. (b) Average CF distribution calculated under the official-2 scenario that only includes daytime cloud detection results. (c) Average CF distribution calculated by AutoML-threshold-LD4 day–night cloud detection results. (d) Average CF distribution calculated after the CALIOP cloud layer products after interpolating to the $1° \times 1°$ grid points.

the AutoML-threshold-LC4-ocean. Here, for comparison convenience, this result is assumed to be the truth. Fig. 10(d) provides the observation based on BTD1516, with a cloud mask obtained using the threshold of 12.5. To visually depict the accuracy of this threshold, Fig. 10(e) shows the results of subtracting the real scene from the threshold detection, showing a large number of red areas [refer to the "increased" part in Fig. 10(e)] indicating a significant number of missed detections of cloud pixels, especially at high latitudes. Utilizing the confusion matrix, the detection accuracy in this region is roughly calculated to be 78.28%. When incorporating the observation of BTD1416 in a similar manner, further suppression of missed detections [refer to the "increased" part in Fig. 10(f)] and a reduction in false detections [refer to the "decreased" part in Fig. 10(f)] are observed, resulting in an improved accuracy of 86.30%. Similar approaches can be applied to include different factors for further enhanced detection accuracy. It should be noted that the internal construction process of the algorithm can be relatively complex, and here, a straightforward illustration of why certain factors exhibit higher importance scores is provided. Other algorithms can similarly showcase the role of factors visually, and further details are not shown.

D. Evaluation of Long-Time-Series Cloud Detection Results

As the previous analysis results, the LC4 mode based on the threshold scheme shows good performance in both local quantization and case details. In order to further explore the robustness of the cloud detection results of the model in the whole region of Himawari-8, Fig. 11(a)–(d) shows the monthly average cloud frequency (CF) distribution calculated by official-1, official-2, AutoML-threshold-LC4, and CALIOP cloud detection results in

TABLE IV
EVALUATION OF CLOUD DETECTION RESULTS ON DIFFERENT LCS USING TEST SETS

	TSS	TPR	TNR	DR
January	*(85.86%)*	*(93.06%)*	*(92.81%)*	*(0.25%)*
	75.03%	**95.94%**	**79.09%**	**16.85%**
	77.19%	***81.23%***	***95.96%***	***14.73%***
April	*(86.32%)*	*(93.63%)*	*(92.69%)*	*(0.94%)*
	72.22%	**96.63%**	**75.29%**	**21.65%**
	76.67%	***81.06%***	***95.62%***	***14.56%***
July	*(84.77%)*	*(93.30%)*	*(91.46%)*	*(1.84%)*
	69.80%	**98.26%**	**71.55%**	**26.71%**
	78.65%	***84.23%***	***94.42%***	***10.19%***
October	*(87.51%)*	*(94.16%)*	*(93.35%)*	*(0.81%)*
	72.70%	**97.41%**	**75.29%**	**22.12%**
	79.27%	***83.30%***	***95.97%***	***12.67%***
Total	*(86.12%)*	*(93.53%)*	*(92.58%)*	*(0.95%)*
	72.50%	**97.19%**	**75.31%**	**21.88%**
	78.00%	***82.51%***	***95.49%***	***12.99%***

Statistics are denoted as follows: Bracketed italicized (AutoML-threshold-LC4), bolded (official-1), and bolded italicized (official-2).

test dataset, respectively. Here, CF is the number of cloud pixels or cloud profiles divided by the total number of pixels or total profiles. The CALIOP data are interpolated on a $1^\circ \times 1^\circ$ grid to facilitate the comparison. In general, official-1, official-2, and AutoML-threshold-LC4 are roughly consistent with CALIOP in the regional distribution of CF. However, there are still some misclassifications in the details. For example, compared with the monthly average CF obtained by CALIOP, the results obtained by the official-1 data are overvalued on the CF as a whole, while the results obtained by the official-2 data are undervalued on the CF as a whole, which indicates that some clear sky is misclassified as cloudy in the probable cloud category. Compared with the results obtained by CALIOP, the results obtained by AutoML-threshold-LC4 are in good agreement with the results obtained by CALIOP, both in the ocean and above the land.

To further evaluate the generalization ability of the AutoML-threshold-LC4 cloud detection results, we compared the performance of each algorithm on the test set against the CALIOP cloud layer product. Table IV presents the detailed evaluation metrics. The TSS values for AutoML-threshold-LC4 remain consistently around 86% across different months, whereas official-1 and official-2 are both below 80%. The DR for AutoML-threshold-LC4 is stable near 1%, while official-1 and official-2 exceed 10%. The TPR and TNR results are consistent with the analysis in Section III-B, with both stabilizing around 93% for AutoML-threshold-LC4. In contrast, TPR is higher and TNR is lower for official-1, while the opposite is true for official-2.

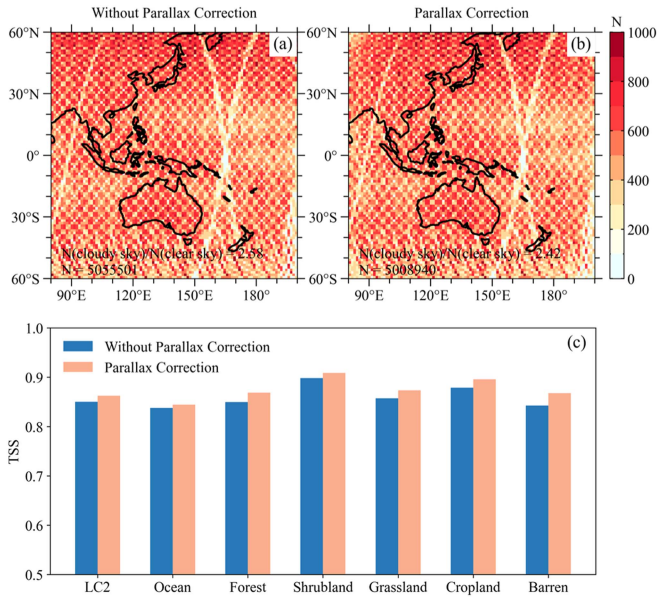

Fig. 12. Comparison of results before and after parallax correction. Sample size (N) distribution of the training set (a) without parallax correction and (b) with parallax correction. (c) TSS of models based on LC2 and LC4 modes before and after parallax correction.

IV. DISCUSSION

In this part, we discuss the influence of the parallax caused by the stationary satellite's fixed subsatellite point and high altitude on the model training results, the threshold selection of reference factor and vector angle, and further analyze the potential impact of the three cloud balance schemes.

A. Influence of AHI Observation Parallax

Here, we compare the changes of training set sample size (N) and model's TSS before and after parallax correction using the method mentioned in Section II. Fig. 12 shows the comparison results based on the LC2 and LC4 modes. In terms of sample size distribution, there is little difference between Fig. 12(a) and (b), but the ratio of cloud sample size to clear-sky sample size decreases after the revision, which indicates that some of the real clear-sky samples in the matching results before parallax correction are mistaken as cloud samples. This further shows that if the steps of parallax correction are omitted, there is a risk that the cloud pixels in the cloud detection results will be overestimated. In the geostationary satellite observation field with a large ocean area, this will lead to a further increase in the detection deviation of the untreated model. Meanwhile, the TSS of the whole region and different surfaces before and after parallax correction are compared using the LC2 and LC4 modes. It can be seen from Fig. 12(c) that the TSS slightly increases after parallax correction, which indirectly reflects the improvement by correcting the observation parallax.

B. Thresholds of Reference Factors and Vector Angle

The six selected detection factors as reference factors to eliminate strong collinearity are selected as a compromise method.

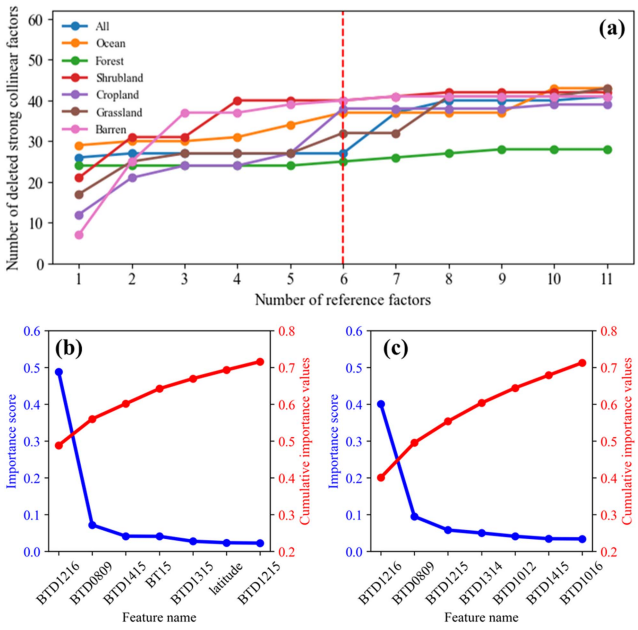

Fig. 13. Selection basis of the number of reference factors. Where (a) compares the number of strong collinear factors removed under different numbers of reference factors. Different color broken lines represent different datasets, and the red dotted line marks the number of strong collinear factors eliminated in different datasets when the number of reference factors is 6. Where (b) and (c) compare the feature importance scores before (b) and after (c) treatment of strong collinear factors based on LC4-barren.

These factors are selected based on their correlation with labels in the training dataset. As shown in Fig. 13(a), the number of strong collinear factors eliminated by only one reference factor is limited. In other words, fewer reference factors cannot completely eliminate strong collinear factors. As the reference factor increases, the number of strong collinear factors removed increases and tends to remain unchanged. The six reference factors are relatively eclectic choices. Fewer than six reference factors yield limited removals of collinear factors, while factors more than six do not significantly improve the results. In addition, selecting 6 reference factors can retain about 30 detection factors (the total detection factors of each dataset are 62, except for 63 in the whole surface), which can significantly reduce the model complexity when all factors are trained together and enhance model's interpretability. Fig. 13(b) and (c) compares feature importance scores before and after removing strong collinear factors in the barren dataset. Before the strong collinear factors are removed [see Fig. 13(b)], BTD1216 plays a very important role in the cloud detection, while the importance of subsequent features drops sharply. The cumulative importance score of the top seven features exceeds 0.7, which means that the total contribution of the remaining 55 features is less than 0.3. It shows that the model may have a serious overfitting phenomenon. After eliminating the strong collinear factor [see Fig. 13(c)], the importance score of BTD1216 in the model decreased, while the importance score of the other factors increased. The cumulative importance score is roughly the same as that of the previous case without eliminating the strong collinear factor, but the remaining 20 feature contributions after elimination are close to 0.3.

TABLE V
ACCURACY OF THE MODEL AND THE NUMBER OF DETECTION FACTORS OF DIFFERENT DATASETS UNDER DIFFERENT VECTOR ANGLE THRESHOLD INTERVALS

	30,150	20,160	10,170
All	(93.78%, 23)	(94.17%, 40)	(94.16%, 49)
Ocean	(93.91%, 20)	(94.14%, 30)	(94.26%, 49)
Forest	(95.00%, 22)	(95.20%, 42)	(95.17%, 52)
Shrubland	(96.32%, 20)	(96.29%, 27)	(96.30%, 50)
Cropland	(93.75%, 22)	(93.83%, 29)	(93.96%, 54)
Grassland	(94.51%, 22)	(94.77%, 35)	(94.92%, 49)
Barren	(93.57%, 19)	(94.02%, 27)	(94.03%, 54)

Value without brackets represents the interval value of the vector angle threshold test. In the bracket, the number with a percentile is the accuracy, and the number without is the number of detection factors.

Obviously, this simplification enhances model interpretability and highlights the importance of addressing strong collinearity. In addition, the first seven factors changed significantly before and after elimination. This reflects the importance of eliminating strong collinear factors from the side.

Regarding the threshold of vector angle, we chose several sets of parameters and tested them in combination with the model complexity and model accuracy. We select three threshold intervals: (10, 170), (20, 160), and (30, 150). We calculated the model accuracy after removing strong collinear factors using these thresholds, as shown in Table V. A broad range of threshold allows for more recognition factors but does not guarantee higher accuracy across different datasets. On the contrary, a stricter threshold results in fewer recognition factors, with most datasets showing lower accuracy than those with the other two intervals. Consequently, the (20, 160) interval is the optimal balance between the number of detection factors and the model accuracy.

C. *Differences in the Three Processes of Cloud Cover Data Balance*

Although the bias of cloud detection can be improved from different perspectives, some of these schemes have potential adverse effects. It should be noted that the downsampling and oversampling schemes may change the distribution of the original samples. As shown in Fig. 14(a1) and (a2), after the training set samples are processed by oversampling schemes, the median and average values of BT07 to BT16 are increased, and the SAA, SOZ, and SOA have also changed compared with the original dataset. Only the Lat and Lon information is consistently distributed before and after resampling. The downsampling scheme demonstrated similar results [see Fig. 14(b1) and (b2)]. This means that the models deviation improvement of this perspective may lead to the lack of some useful information, which, in turn, affects the subsequent learning process, especially in the relatively complex model construction. In addition, previous studies have found that the intervention of the model training process by the weight scheme will increase the complexity of the training, resulting in poor stability [57], [58], [59]. While

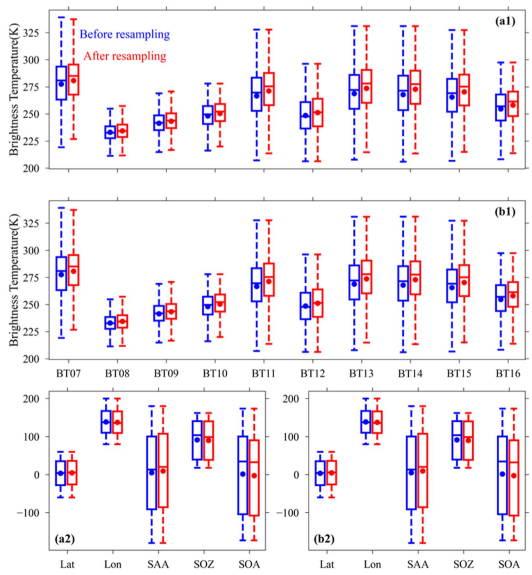

Fig. 14. Comparison of the data distribution before and after resampling in the training set. Where (a) represents oversampling and (b) represents downsampling. Blue indicates before resampling and red indicates after resampling. Dots denote the mean, the center line of the box denotes the median, the top and bottom edges of the box denote the 25th and 75th quartiles, and beyond the edges of the box denotes the remaining extreme value distribution.

due to the small difference between the number of cloud and clear-sky samples, this does not appear in our algorithm, and the DR of the models based on this method is low. In comparison, the threshold scheme modifies the output mapping function without destroying the distribution of the original data and the structure of the model itself after completing the model training. Meanwhile, the threshold can be adjusted according to the actual needs. For example, if a higher TPR is required, a lower output threshold can be selected, and if a larger TNR is required, a larger output threshold can be selected. In addition, this method can also be used to obtain cloud detection results with different confidence levels similar to the traditional algorithms, which only need to adjust the threshold to different levels.

V. Conclusion

In this study, we develop an unbiased, high-precision cloud detection algorithm for AHI using AutoML by integrating parallax correction, a vector angle-based collinearity factor selection, and three model deviation processing schemes. The algorithm is refined through a comparative analysis of cloud cover differences and surface characterization modes. Our evaluation results indicate that parallax correction can prevent the misidentification of clear-sky pixels as cloud pixels. The vector angle-based colinear factor selection effectively eliminates strong collinearity, enhancing model interpretability. By incorporating the marginal contribution value of the detection factor, we provide a novel model interpretation approach from the traditional threshold perspective. In addition, quantitative evaluations reveal that these deviation processing schemes effectively address class imbalance caused by cloud cover differences in the Himawari-8 field. The processed AutoML models exhibit a 5% reduction in DR and a 1% increase in TSS compared with the untreated models. However, we noted that the resampling scheme can unintentionally alter the original data distribution due to its inherent randomness, potentially affecting the accuracy of cloud detection results. The weight scheme, applied during model training, may disrupt the internal structure of the original algorithm when there is a large class sample size ratio. Conversely, the threshold scheme, as an a posteriori method, does not impact the original data distribution or the construction of the trained model. This approach allows for improved cloud detection results through threshold adjustments according to varying requirements.

Building on these findings, we applied the LC1, LC2, LC3, and LC4 processing modes for the six LCs. Among these, the LC4 mode demonstrates superior performance under all surface conditions, as evidenced by statistical quantification and representative region case analysis. Consequently, the AutoML-threshold-LC4 algorithm achieves a much better TSS of 87.32% compared with 73.60% of the official-1 cloud mask, showing a significant improvement of 13.72%. The DR of AutoML-threshold-LC4 is 0.79%, considerably lower than the 20.48% of the official-1 cloud mask. In long-term series tests, the CF distribution obtained by AutoML-threshold-LC4 is more consistent with CALIOP than that of the AHI official product. This demonstrates that our method not only has good stability in detecting various clouds but also surpasses the accuracy of the official cloud mask product. Furthermore, the adjustable threshold in AutoML-threshold-LC4 allows for broader application scenarios, enhancing its versatility.

Our algorithm effectively improves the accuracy of Himawari-8 cloud detection and ensures the unbiasedness of detection. In addition, this algorithm construction processing is versatile and can be applied to a wide range of scenarios, irrespective of sensor features on different satellites, such as the number of channels or center wavelengths. By integrating a sample balance scheme, we aim to further identify and monitor strong convective systems characterized by heavy precipitation, significant harm, and substantial occurrence differences between land and sea.

Acknowledgment

The AHI data are obtained from JAXA's P-Tree system (https://www.eorc.jaxa.jp/ptree/index.html) and the CALIOP and MODIS data are obtained from the NASA (https://search.earthdata.nasa.gov/search).

References

[1] A. Battaglia et al., "Spaceborne cloud and precipitation radars: Status, challenges, and ways forward," *Rev. Geophys.*, vol. 58, no. 3, 2020, Art. no. e2019RG000686, doi: 10.1029/2019RG000686.

[2] J. Ge et al., "A novel liquid water content retrieval method based on mass absorption for single-wavelength cloud radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, May 2023, Art. no. 4102815, doi: 10.1109/TGRS.2023.3278735.

[3] J. Ge, C. Zheng, H. Xie, Y. Xin, J. Huang, and Q. Fu, "Midlatitude cirrus clouds at the SACOL site: Macrophysical properties and large-scale atmospheric states," *J. Geophysical Res., Atmos.*, vol. 123, no. 4, pp. 2256–2271, 2018, doi: 10.1002/2017JD027724.

[4] X. Hu, J. Ge, W. Li, J. Du, Q. Li, and Q. Mu, "Vertical structure of tropical deep convective systems at different life stages from CloudSat observations," *J. Geophysical Res., Atmos.*, vol. 126, no. 21, 2021, Art. no. e2021JD035115, doi: 10.1029/2021JD035115.

[5] C. J. Stubenrauch, G. Caria, S. E. Protopapadaki, and F. Hemmer, "3D radiative heating of tropical upper tropospheric cloud systems derived from synergistic A-train observations and machine learning," *Atmos. Chem. Phys.*, vol. 21, no. 2, pp. 1015–1034, Jan. 2021, doi: 10.5194/acp-21-1015-2021.

[6] G. Tana et al., "Retrieval of cloud microphysical properties from Himawari-8/AHI infrared channels and its application in surface shortwave downward radiation estimation in the sun glint region," *Remote Sens. Environ.*, vol. 290, May 2023, Art. no. 113548, doi: 10.1016/j.rse.2023.113548.

[7] X. Jiang et al., "Muted extratropical low cloud seasonal cycle is closely linked to underestimated climate sensitivity in models," *Nature Commun.*, vol. 14, no. 1, Sep. 2023, Art. no. 5586, doi: 10.1038/s41467-023-41360-0.

[8] M. D. Zelinka and D. L. Hartmann, "Why is longwave cloud feedback positive?," *J. Geophysical Res., Atmos.*, vol. 115, no. D16, 2010, Art. no. D16117, doi: 10.1029/2010JD013817.

[9] J. Bjordal, T. Storelvmo, K. Alterskjær, and T. Carlsen, "Equilibrium climate sensitivity above 5 °C plausible due to state-dependent cloud feedback," *Nature Geosci.*, vol. 13, no. 11, pp. 718–721, Nov. 2020, doi: 10.1038/s41561-020-00649-1.

[10] S. C. Sherwood, S. Bony, and J.-L. Dufresne, "Spread in model climate sensitivity traced to atmospheric convective mixing," *Nature*, vol. 505, no. 7481, pp. 37–42, Jan. 2014, doi: 10.1038/nature12829.

[11] Z. Feng et al., "A global high-resolution mesoscale convective system database using satellite-derived cloud tops, surface precipitation, and tracking," *J. Geophysical Res., Atmos.*, vol. 126, no. 8, 2021, Art. no. e2020JD034202, doi: 10.1029/2020JD034202.

[12] S. Jiang and L. Tao, "Classification and estimation of typhoon intensity from geostationary meteorological satellite images based on deep learning," *Atmosphere*, vol. 13, no. 7, Jul. 2022, Art. no. 1113, doi: 10.3390/atmos13071113.

[13] X. Wang, M. Min, F. Wang, J. Guo, B. Li, and S. Tang, "Intercomparisons of cloud mask products among Fengyun-4A, Himawari-8, and MODIS," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8827–8839, Nov. 2019, doi: 10.1109/TGRS.2019.2923247.

[14] M. Zhang et al., "A new method for tropospheric tomography using GNSS and Fengyun-4A data," *Atmos. Res.*, vol. 280, Dec. 2022, Art. no. 106460, doi: 10.1016/j.atmosres.2022.106460.

[15] S. A. Ackerman, K. I. Strabala, W. P. Menzel, R. A. Frey, C. C. Moeller, and L. E. Gumley, "Discriminating clear sky from clouds with MODIS," *J. Geophysical Res., Atmos.*, vol. 103, no. D24, pp. 32141–32157, 1998, doi: 10.1029/1998JD200032.

[16] H. Ishida and T. Y. Nakajima, "Development of an unbiased cloud detection algorithm for a spaceborne multispectral imager," *J. Geophysical Res., Atmos.*, vol. 114, no. D7, 2009, Art. no. D07206, doi: 10.1029/2008JD010710.

[17] I. Takahito and Y. Ryo, "Algorithm theoretical basis for Himawari-8 cloud mask product," *Meteorological Satell. Center Tech. Note*, 2016.

[18] R. Lai et al., "Comparison of cloud properties from Himawari-8 and FengYun-4A geostationary satellite radiometers with MODIS cloud retrievals," *Remote Sens.*, vol. 11, no. 14, Jan. 2019, Art. no. 1703, doi: 10.3390/rs11141703.

[19] C.-Y. Liu, C.-H. Chiu, P.-H. Lin, and M. Min, "Comparison of cloud-top property retrievals from advanced Himawari imager, MODIS, CloudSat/CPR, CALIPSO/CALIOP, and radiosonde," *J. Geophysical Res., Atmos.*, vol. 125, no. 15, 2020, Art. no. e2020JD032683, doi: 10.1029/2020JD032683.

[20] H. Shang, L. Chen, H. Letu, M. Zhao, S. Li, and S. Bao, "Development of a daytime cloud and haze detection algorithm for Himawari-8 satellite measurements over central and eastern China," *J. Geophysical Res., Atmos.*, vol. 122, no. 6, pp. 3528–3543, 2017, doi: 10.1002/2016JD025659.

[21] X. Hu, J. Ge, Y. Li, R. Marchand, J. Huang, and Q. Fu, "Improved hydrometeor detection method: An application to CloudSat," *Earth Space Sci.*, vol. 7, no. 2, 2020, Art. no. e2019EA000900, doi: 10.1029/2019EA000900.

[22] M. Min, J. Li, F. Wang, Z. Liu, and W. P. Menzel, "Retrieval of cloud top properties from advanced geostationary satellite imager measurements based on machine learning algorithms," *Remote Sens. Environ.*, vol. 239, Mar. 2020, Art. no. 111616, doi: 10.1016/j.rse.2019.111616.

[23] P. Xia et al., "Accurate nowcasting of cloud cover at solar photovoltaic plants using geostationary satellite images," *Nature Commun.*, vol. 15, no. 1, Jan. 2024, Art. no. 510, doi: 10.1038/s41467-023-44666-1.

[24] B. Guo, F. Zhang, W. Li, and Z. Zhao, "Cloud classification by machine learning for geostationary radiation imager," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, Jan. 2024, Art. no. 4102814, doi: 10.1109/TGRS.2024.3353373.

[25] J. H. Jeppesen, R. H. Jacobsen, F. Inceoglu, and T. S. Toftegaard, "A cloud detection algorithm for satellite imagery based on deep learning," *Remote Sens. Environ.*, vol. 229, pp. 247–259, Aug. 2019, doi: 10.1016/j.rse.2019.03.039.

[26] Z. Tan et al., "Detecting multilayer clouds from the geostationary advanced Himawari imager using machine learning techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4103112, doi: 10.1109/TGRS.2021.3087714.

[27] C. Wang, S. Platnick, K. Meyer, Z. Zhang, and Y. Zhou, "A machine-learning-based cloud detection and thermodynamic-phase classification algorithm using passive spectral observations," *Atmos. Meas. Techn.*, vol. 13, no. 5, pp. 2257–2277, May 2020, doi: 10.5194/amt-13-2257-2020.

[28] W. Li, F. Zhang, H. Lin, X. Chen, J. Li, and W. Han, "Cloud detection and classification algorithms for Himawari-8 imager measurements based on deep learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Feb. 2022, Art. no. 4107117, doi: 10.1109/TGRS.2022.3153129.

[29] A. K. Heidinger, A. T. Evan, M. J. Foster, and A. Walther, "A naive Bayesian cloud-detection scheme derived from CALIPSO and applied within PATMOS-x," *J. Appl. Meteorol. Climatol.*, vol. 51, no. 6, pp. 1129–1144, Jun. 2012, doi: 10.1175/JAMC-D-11-02.1.

[30] Y. Yang, C. Zhao, Y. Sun, Y. Chi, and H. Fan, "Convective cloud detection and tracking using the new-generation geostationary satellite over South China," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Jul. 2023, Art. no. 4103912, doi: 10.1109/TGRS.2023.3298976.

[31] C. Liu et al., "A machine learning-based cloud detection algorithm for the Himawari-8 spectral image," *Adv. Atmos. Sci.*, vol. 39, no. 12, pp. 1994–2007, Dec. 2022, doi: 10.1007/s00376-021-0366-x.

[32] C. Wang, Z. J. Luo, and X. Huang, "Parallax correction in collocating CloudSat and moderate resolution imaging spectroradiometer (MODIS) observations: Method and application to convection study," *J. Geophysical Res., Atmos.*, vol. 116, no. D17, 2011, Art. no. D17201, doi: 10.1029/2011JD016097.

[33] G. A. Vicente, J. C. Davenport, and R. A. Scofield, "The role of orographic and parallax corrections on real time high resolution satellite rainfall rate distribution," *Int. J. Remote Sens.*, vol. 23, no. 2, pp. 221–230, Jan. 2002, doi: 10.1080/01431160010006935.

[34] T. Bieliński, "A parallax shift effect correction based on cloud height for geostationary satellites and radar observations," *Remote Sens.*, vol. 12, no. 3, Jan. 2020, Art. no. 365, doi: 10.3390/rs12030365.

[35] D. H. Wolpert, "The lack of A priori distinctions between learning algorithms," *Neural Comput.*, vol. 8, no. 7, pp. 1341–1390, Oct. 1996, doi: 10.1162/neco.1996.8.7.1341.

[36] C. Wang, Q. Wu, M. Weimer, and E. Zhu, "FLAML: A fast and lightweight AutoML library," in *Proc. Mach. Learn. Syst.*, vol. 3, pp. 434–447, 2021.

[37] M. D. Adams, F. Massey, K. Chastko, and C. Cupini, "Spatial modelling of particulate matter air pollution sensor measurements collected by community scientists while cycling, land use regression with spatial cross-validation, and applications of machine learning for data correction," *Atmos. Environ.*, vol. 230, Jun. 2020, Art. no. 117479, doi: 10.1016/j.atmosenv.2020.117479.

[38] Z. Zheng et al., "Automated machine learning to evaluate the information content of tropospheric trace gas columns for fine particle estimates over India: A modeling testbed," *J. Adv. Model. Earth Syst.*, vol. 15, no. 3, 2023, Art. no. e2022MS003099, doi: 10.1029/2022MS003099.

[39] K. Bessho et al., "An introduction to Himawari-8/9— Japan's new-generation geostationary meteorological satellites," *J. Meteorol. Soc. Jpn., II*, vol. 94, no. 2, pp. 151–183, 2016, doi: 10.2151/jmsj.2016-009.

[40] T. Horinouchi, U. Shimada, and A. Wada, "Convective bursts with gravity waves in tropical cyclones: Case study with the Himawari-8 satellite and idealized numerical study," *Geophysical Res. Lett.*, vol. 47, no. 3, 2020, Art. no. e2019GL086295, doi: 10.1029/2019GL086295.

[41] M. C. M. Tierra and G. Bagtasa, "Identifying the rapid intensification of tropical cyclones using the Himawari-8 satellite and their impacts in the Philippines," *Int. J. Climatol.*, vol. 43, no. 1, pp. 1–16, 2023, doi: 10.1002/joc.7696.

[42] B. Gasparini, A. Meyer, D. Neubauer, S. Münch, and U. Lohmann, "Cirrus cloud properties as seen by the CALIPSO satellite and ECHAM-HAM global climate model," *J. Climate*, vol. 31, no. 5, pp. 1983–2003, Mar. 2018, doi: 10.1175/JCLI-D-16-0608.1.

[43] S. Okada, M. Ohzeki, and S. Taguchi, "Efficient partition of integer optimization problems with one-hot encoding," *Sci. Rep.*, vol. 9, no. 1, Sep. 2019, Art. no. 13036, doi: 10.1038/s41598-019-49539-6.
[44] P. Rodríguez, M. A. Bautista, J. Gonzàlez, and S. Escalera, "Beyond one-hot encoding: Lower dimensional target embedding," *Image Vis. Comput.*, vol. 75, pp. 21–31, Jul. 2018, doi: 10.1016/j.imavis.2018.04.004.
[45] A. Moreo, A. Esuli, and F. Sebastiani, "Distributional random oversampling for imbalanced text classification," in *Proc. 39th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2016, pp. 805–808, doi: 10.1145/2911451.2914722.
[46] P. Song et al., "Accelerated singular value-based ultrasound blood flow clutter filtering with randomized singular value decomposition and randomized spatial downsampling," *IEEE Trans. Ultrasonics, Ferroelectr., Freq. Control*, vol. 64, no. 4, pp. 706–716, Apr. 2017, doi: 10.1109/TUFFC.2017.2665342.
[47] A. Anand, G. Pugalenthi, G. B. Fogel, and P. N. Suganthan, "An approach for classification of highly imbalanced data using weighting and undersampling," *Amino Acids*, vol. 39, no. 5, pp. 1385–1391, Nov. 2010, doi: 10.1007/s00726-010-0595-2.
[48] W. Zong, G.-B. Huang, and Y. Chen, "Weighted extreme learning machine for imbalance learning," *Neurocomputing*, vol. 101, pp. 229–242, Feb. 2013, doi: 10.1016/j.neucom.2012.08.010.
[49] J. M. Johnson and T. M. Khoshgoftaar, "Deep learning and thresholding with class-imbalanced big data," in *Proc. 18th IEEE Int. Conf. Mach. Learn. Appl.*, 2019, pp. 755–762, doi: 10.1109/ICMLA.2019.00134.
[50] Q. Zou, S. Xie, Z. Lin, M. Wu, and Y. Ju, "Finding the best classification threshold in imbalanced classification," *Big Data Res.*, vol. 5, pp. 2–8, Sep. 2016, doi: 10.1016/j.bdr.2015.12.001.
[51] X. Xia et al., "Retrieving precipitable water vapor over land from satellite passive microwave radiometer measurements using automated machine learning," *Geophysical Res. Lett.*, vol. 50, no. 22, 2023, Art. no. e2023GL105197, doi: 10.1029/2023GL105197.
[52] G. Biau and E. Scornet, "A random forest guided tour," *TEST*, vol. 25, no. 2, pp. 197–227, Jun. 2016, doi: 10.1007/s11749-016-0481-7.
[53] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.
[54] G. Ke et al., "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, Accessed: Jul. 5, 2023. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html
[55] A. B. Parsa, A. Movahedi, H. Taghipour, S. Derrible, and A. (Kouros) Mohammadian, "Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis," *Accident Anal. Prevention*, vol. 136, Mar. 2020, Art. no. 105405, doi: 10.1016/j.aap.2019.105405.
[56] D. Slack, S. Hilgard, E. Jia, S. Singh, and H. Lakkaraju, "Fooling LIME and SHAP: Adversarial attacks on post hoc explanation methods," in *Proc. AAAI/ACM Conf. AI, Ethics, Soc.*, 2020, pp. 180–186, doi: 10.1145/3375627.3375830.
[57] M. Zhu et al., "Class weights random forest algorithm for processing class imbalanced medical data," *IEEE Access*, vol. 6, pp. 4641–4652, 2018, doi: 10.1109/ACCESS.2018.2789428.
[58] H. Zhang et al., "Combining a class-weighted algorithm and machine learning models in landslide susceptibility mapping: A case study of Wanzhou section of the Three Gorges Reservoir, China," *Comput. Geosci.*, vol. 158, Jan. 2022, Art. no. 104966, doi: 10.1016/j.cageo.2021.104966.
[59] M. Belgiu and O. Csillik, "Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis," *Remote Sens. Environ.*, vol. 204, pp. 509–523, Jan. 2018, doi: 10.1016/j.rse.2017.10.005.

Bochun Liu received the B.S. degree in atmospheric science in 2022 from Lanzhou University, Lanzhou, China, where he is currently working toward the Ph.D. degree in atmospheric science.

His research interests include cloud detection, convective cloud detection and tracking, and convective cloud characteristics research.

Jinming Ge received the B.S. and Ph.D. degrees in atmospheric science from Lanzhou University, Lanzhou, China, in 2005 and 2010, respectively.

He is currently a Professor with the Key Laboratory for Semi-Arid Climate Change, Ministry of Education, College of Atmospheric Sciences, Lanzhou University. His research has focused on atmosphere aerosols, cloud physical properties retrieval, and their radiative effects on climate.

Qingyu Mu received the B.S. degree in atmospheric science in 2020 from the College of Atmospheric Sciences, Lanzhou University, Lanzhou, China, where he is currently working toward the Ph.D. degree in atmospheric science, with a focus on contrail cirrus, cloud–aerosol interactions, and atmospheric infrared radiative transfer processes.

Chi Zhang received the B.S. degree in atmospheric science from the College of Atmospheric Sciences, Chengdu University of Information Technology, Chengdu, China, in 2021. He is currently working toward the Ph.D. degree in atmospheric science with Lanzhou University, Lanzhou, China, with a focus on diurnal cycle of clouds and its radiative effects, as well as research on millimeter-wave radar algorithms.

Xiaoyu Hu received the Ph.D. degree in atmospheric science from Lanzhou University, Lanzhou, China, in 2022.

He is currently a Postdoctoral Researcher with the Chongqing Research Institute of Big Data, Peking University, Beijing, China. His research interests include cloud remote sensing and artificial intelligence application.

Jiajing Du received the B.S. degree in atmospheric science in 2019 from Lanzhou University, Lanzhou, China, where she is currently working toward the Ph.D. degree in atmospheric science.

Her research interests include radar data processing and cloud physical properties retrieval algorithm.

Yanyan Wu received the B.S. degree in atmospheric science from the College of Atmospheric Sciences, Chengdu University of Information Technology, Chengdu, China, in 2022. She is currently working toward the M.S. degree in atmospheric science with Lanzhou University, Lanzhou, China.

Her research interests include the relationship between dust aerosols and sea temperature, near-surface wind speed, and atmospheric circulation.

Bo Wang received the B.S. degree in atmospheric science in 2022 from the College of Atmospheric Sciences, Lanzhou University, Lanzhou, China, where he is currently working toward the M.S. degree in atmospheric science, with a focus on convective cloud characteristics, WRF model simulation, and cloud–aerosol interactions.

Xiang Li received the B.S. degree in atmospheric science in 2023 from the College of Atmospheric Sciences, Lanzhou University, Lanzhou, China, where he is currently working toward the M.S. degree in atmospheric science with Lanzhou University, Lanzhou, China, with a focus on the relationship between meteorological field and cloud, and the integration of artificial intelligence and atmospheric science.

Jianping Huang received the Ph.D. degree in weather dynamics from Lanzhou University, Lanzhou, China, in 1988.

He performed postdoctoral work with Peking University, Beijing, China. He is an Academician of the Chinese Academy of Sciences, Beijing, China, and a Professor with the Key Laboratory for Semi-Arid Climate Change, Ministry of Education, College of Atmospheric Sciences, Lanzhou University. He is the main author of the first working group of the Working Group I of the Fifth Assessment Report of the Intergovernmental Panel on Climate Change and is one of the academic leaders in the field of atmospheric science in China who has combined systematic observation and theoretical research and made major breakthroughs. He plays an important role in semiarid climate research in China. He created and developed the first semiarid climate comprehensive observation station with international standards in western China. His research focuses on the global arid and semiarid regions and the interaction of dust aerosols with clouds and precipitation.

Qiang Fu received the B.E. and M.S. degrees in atmospheric physics from Peking University, Beijing, China, in 1983 and 1986, respectively, and the Ph.D. degree in atmospheric science from The University of Utah, Salt Lake City, UT, USA, in 1991.

He is currently a Professor of atmospheric sciences with the University of Washington, Seattle, WA, USA. His research interests include light scattering and radiative transfer, atmospheric radiation and clouds, atmospheric aerosols, remote sensing, climate, and climate change.