

Article

Estimating Rainfall with Multi-resource Data over East Asia Based on Machine Learning

Yushan Zhang ¹, Kun Wu ^{1,2,*}, Jinglin Zhang ³, Feng Zhang ^{2,4}, Haixia Xiao ^{5,6}, Fuchang Wang ^{4,7}, Jianyin Zhou ⁸, Yi Song ⁹ and Liang Peng ⁹

¹ Department of Atmospheric Sciences, Nanjing University of Information Science & Technology, Nanjing 210044, China; 20191201103@nuist.edu.cn (Y.Z.)

² Shanghai Qi Zhi Institute, Shanghai 200232, China

³ Department of Computer and Software, Nanjing University of Information Science & Technology, Nanjing 210044, China; jinglin.zhang@nuist.edu.cn

⁴ Department of Atmospheric and Oceanic Sciences & Institute of Atmospheric Sciences, Fudan University, Shanghai 200438, China; fengzhang@fudan.edu.cn (F.Z.); fcwang@sjtu.edu.cn (F.W.)

⁵ Nanjing Joint Institute for Atmospheric Sciences, Nanjing 210008, China; xiaohx@cma.gov.cn

⁶ Key Laboratory of Transportation Meteorology, China Meteorological Administration, Nanjing 210008, China

⁷ Shanghai Typhoon Institute of China Meteorological Administration, Shanghai 200030, China

⁸ College of Meteorology and Oceanography, National University of Defense Technology, Changsha 410073, China; zhoujianyin17@nudt.edu.cn

⁹ Meteorological Remote Sensing Application Center, Beijing Piesat Information Technology Co., Ltd., Beijing 100195, China; songyi@piesat.cn (Y.S.); pengliang@piesat.cn (L.P.)

* Correspondence: wukun@nuist.edu.cn; Tel.: +86-151-9590-0108



Citation: Zhang, Y.; Wu, K.; Zhang, J.; Zhang, F.; Xiao, H. Wang, F.; Zhou, J.; Song, Y.; Peng, L. Estimating Rainfall with Multi-resource Data over East Asia Based on Machine Learning. *Remote Sens.* **2021**, *13*, 3332. <https://doi.org/10.3390/rs13163332>

Academic Editor: Trista Yang

Received: 19 July 2021

Accepted: 19 August 2021

Published: 23 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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 (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: The lack of accurate estimation of intense precipitation is a universal limitation in precipitation retrieval. Therefore, a new rainfall retrieval technique based on the Random Forest (RF) algorithm is presented using the Advanced Himawari Imager-8 (Himawari-8/AHI) infrared spectrum data and the NCEP operational Global Forecast System (GFS) forecast information. And the gauge-calibrated rainfall estimates from the Global Precipitation Measurement (GPM) product served as the ground truth to train the model. The two-step RF classification model was established for (1) rain area delineation and (2) precipitation grades' estimation to improve the accuracy of moderate rain and heavy rain. In view of the imbalance categories' distribution in the datasets, the resampling technique including the Random Under-sampling algorithm and Synthetic Minority Over-sampling Technique (SMOTE) was implemented throughout the whole training process to fully learn the characteristics among the samples. Among the features used, the contributions of meteorological variables to the trained models were generally greater than those of infrared information; in particular, the contribution of precipitable water was the largest, indicating the sufficient necessity of water vapor conditions in rainfall forecasting. The simulation results by the RF model were compared with the GPM product pixel-by-pixel. To prove the universality of the model, we used independent validation sets which are not used for training and two independent testing sets with different periods from the training set. In addition, the algorithm was validated against independent rain gauge data and compared with GFS model rainfall. Consequently, the RF model identified rainfall areas with a Probability Of Detection (POD) of around 0.77 and a False-Alarm Ratio (FAR) of around 0.23 for validation, as well as a POD of 0.60–0.70 and a FAR of around 0.30 for testing. To estimate precipitation grades, the value of classification was 0.70 in validation and in testing the accuracy was 0.60 despite a certain overestimation. In summary, the performance on the validation and test data indicated the great adaptability and superiority of the RF algorithm in rainfall retrieval in East Asia. To a certain extent, our study provides a meaningful range division and powerful guidance for quantitative precipitation estimation.

Keywords: Himawari-8 satellite; numerical product; rainfall retrieval; random forest; precipitation grades

1. Introduction

Precipitation is one of the most important indicators that reflects global and regional climate system changes. In particular, summer precipitation plays a vital role in atmospheric circulation, hydrological cycle, and thermal momentum exchange [1–3]. The interdecadal variation of precipitation is tightly linked to the circulation anomaly, water vapor budget, soil humidity, and wind speeds in East Asia [4,5]. High-resolution rainfall data have an extensive application in the fields of agriculture, forestry, transportation, and marine monitoring [6–9]. There are obvious weaknesses in traditional ground station measurement, due to the limited coverage of rain gauges, which is confined by the complex terrain and coastline, and the uncertainty of accuracy in the weather radar detection [10]. Contrastingly, satellites can make comprehensive observations and provide intuitive remote sensing image information; therefore, satellite rainfall products are superior and have good prospects to improve the ability of monitoring grid rainfall [11]. Satellite rainfall products are retrieved by microwave remote sensing, visible/infrared remote sensing, and multisensor rainfall estimation, with microwave sensors mainly carried by polar-orbiting satellites, such as NOAA, METOP, and FY-3. Although the detecting precision of polar-orbiting satellites is higher than geostationary satellites, the time sampling is low. Fortunately, since the new generation of geostationary meteorological satellites (such as GOES, MSG, Himawari-8, and FY-4) has been launched, visible/infrared remote sensing can provide high spatiotemporal resolution rainfall products. However, the rainfall accuracy still needs to be improved as a result of the indirect relationship between the infrared signal and precipitation [12–14].

Multi-channels retrieval overcomes many shortcomings of the past and has become the mainstream of high-quality rainfall products' retrieval, including the Tropical Rainfall Measuring Mission (TRMM), Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN), and the Climate Prediction Center Morphing Method (CMORPH), which have been used in the accuracy assessment of precipitation at different time scales [15–17]. Global Precipitation Measurement (GPM), as the successor of TRMM, has been equipped with more advanced passive microwave radiometer GMI and dual-frequency radar DPR sensors, enhancing the skill of observing solid and trace precipitation and providing a more accurate calibration reference for multisatellite precipitation estimation [18,19]. The Final-Run products have been corrected by the monthly scale of the ground station data, and the accuracy is closest to the real precipitation, which can be well studied and applied to climate change, drought identification, flood forecasting, and so on [20,21]. It is gratifying that the high-resolution infrared data can match the GPM products well in both time and space, which provides a favorable condition for rainfall retrieval systems [22,23].

Currently, many studies have established forecasting algorithms using real-time geostationary Infrared (IR) satellite data to retrieve high-resolution rainfall products based on the infrared–precipitation relationship. At an early stage, the single-channel infrared data were considered due to technical limitations. Arkin [24] found that there was a high correlation between the cold cloud coverage area and the 6 h precipitation accumulation in a large enough area where the cloud top temperature was lower than 225 K to 255 K in the GARP Atlantic Tropical Experiment (GATE). From this discovery, Arkin et al. [25,26] proposed a rainfall estimation algorithm, the GOES Precipitation Index (GPI); the application of the GPI requires a large enough spatial and temporal scale; however, with the improvement of resolution, the accuracy of the satellite-based estimation of small-scale precipitation is limited [27]. Todd et al. [28] combined the instantaneous value of passive remote sensing with frequent infrared images and introduced the MIRA algorithm to optimize the relationship between infrared and precipitation by using the probability matching method; they further reduced the spatial and temporal scale of precipitation estimation. Adler et al. [29] proposed the “inside to outside” Convective-Stratiform Technology (CST), which took the minimum value of the local brightness temperature as the convective precipitation center, allocated the precipitation rate and precipitation area to the surrounding points,

and used the brightness temperature threshold value of thunderstorm anvils to determine the stratified precipitation. Grecu et al. [30] combined lightning information into the CST to improve the algorithm. However, Ebert et al. [31] and Fruh et al. [32] found that CST algorithm was seriously inadequate for midlatitude tropical cyclone precipitation, limited especially to deep convective systems in the tropics.

With the improvement of the resolution of spectral data from geostationary satellites, Ba et al. [33] and Nauss et al. [34] extracted cloud radiation characteristics and physical parameters such as cloud optical thickness, cloud top height, and cloud phase to improve the algorithm. Ebert et al. [35] found that enhancing the classification of convective and stratified precipitation helps to increase the accuracy of satellite precipitation estimation, so some algorithms focus on the classification of convective-stratified regions. Torricella et al. [36] and Cattani et al. [37] pointed out that the height of the cloud top is representative of convective precipitation intensity, and the effective particle radius of ice clouds is positively correlated with precipitation intensity. Thies et al. [38] used Meteosat Second-Generation-Spinning Enhanced Visible and InfraRed Imager (MSG-SEVIRI) data to calculate the precipitation probability matrix at day and night to classify the convective-stratified precipitation area and precipitation intensity. And the combination of the brightness temperature and brightness temperature difference was introduced to represent cloud parameter information, so did as Feidas et al. [39]. These techniques generally rely on the parameter relationship between cloud attributes and the rainfall process. A few inputs have been used based on this theoretical background. However, the parameter algorithm has to develop to add more input features, which will gradually be unable to meet the assumptions of parameter testing and the conceptual model, let alone the nonlinear connections between remote sensing and precipitation in fact. Thus, more suitable approaches need to be introduced to overcome these deficiencies.

In recent years, the machine learning algorithm has become a powerful tool to link multichannel infrared data and rainfall. Many studies have integrated products from different sources, including satellite remote sensing data, numerical prediction products, ground observation station data, and satellite rainfall product data. Machine learning techniques provide new directions for the simulation of highly nonlinear relations to retrieval rainfall, such as neural networks [40–46], support vector machines [47], and gradient boosting decision trees [48]. The Random Forest (RF) algorithm also has been concerned in rainfall retrieval [49–55]. As an ensemble learning algorithm, random forest can generate reasonable predictions from a plurality of inputs by establishing multiple decision trees, and it is not easy to overfit compared to others and is usually superior to a single decision tree algorithm [56]. Kuhnlein et al. [49] used MSG-SEVIRI data and RF to differentiate convective-stratified rain and regressed the intensity in the day and night and at twilight, respectively. Min et al. [51] integrated Himawari-8 and Numerical Weather Product (NWP) data to separate precipitation and nonprecipitation pixels by the RF classification method, then successfully simulated rainfall intensity. Turini et al. [52] estimated high-resolution rainfall (i.e., 15 min and 3 km) from MSG1 and Integrated Multi-satellite Retrievals for GPM (IMERG) from the GPM based on RF. Counting the results of the validation set among these studies, the probability of detection and false-alarm rate can roughly reach about 0.6 and 0.5 when discriminating precipitation areas. However, when estimating precipitation intensity, the imbalance of samples often leads to the underestimation of higher-intensity rainfall, which is a universal problem that remains to be solved.

In summary, promising results were obtained in the previous study through the training dataset and validation set for the same period, while validation datasets for different periods have not been tested yet, resulting in an insufficient ability to convincingly prove the model. A significant issue was shown in that the intensity samples predicted by regression were concentrated below 0.5 mm/h, and the higher rate was underestimated by 3 mm/h in GPM [49,51,52,54,55]. To solve the above problems as much as possible, the test dataset was extracted from an independent period, which differed from both

the training and validation dataset with no overlap. For the purpose of testing, the data used in this paper consisted of real-time Himawari-8/AHI multispectral observations, the NWP from the Global Forecast System (GFS), and GPM IMERG data. The RF model was trained for East Asia throughout the whole day, and then, the grades of precipitation rather than the rainfall intensity were estimated to improve the accuracy of moderate rain and heavy rain prediction, expecting to further promote the application of rainfall retrieval in nowcasting systems.

The remainder of this paper is arranged as follows. The training, testing, and validation datasets used in the RF model are carefully recommended in Section 2. Section 3 describes the implementation of two-step RF model and the process of feature selection and tuning. The classification performance of the training, validation, and testing of RF are thoroughly displayed in Section 4. Section 5 provides the summary and discussion.

2. Materials and Methods

2.1. Materials

2.1.1. Satellite Observations and Numerical Products

The Himawari-8/AHI multispectral data, which have the high spatial and temporal resolution, were used in our study. The satellite “Himawari-8” was successfully launched by the Japan Meteorological Agency (JMA) on 7 October 2014 and commenced operation in 2015. The equipped AHI instrument has a full-disk observation over 16 spectral channels from 0.47 to 13.3 μm including 3 visible, 3 nearinfrared, and 10 infrared channels every 10 min; the satellite data can be obtained from <http://www.eorc.jaxa.jp/ptree/index.html> (accessed on 20 September, 2020). To decrease the impacts of sunlight, only the brightness temperature of water vapor and the longwave infrared channels were considered. The central wavelengths of the former are 6.2, 6.9, and 7.3 μm , and those of the latter are 8.6, 9.6, 10.4, 11.2, 12.4, and 13.3 μm . Moreover, the cloud properties’ information that allows a reliable identification of a deep convective area can be inferred by appropriate combinations of brightness temperature differences. Commonly, differences between the water vapor-IR channels ($\Delta T_{6.2-11.2}$ and $\Delta T_{7.3-12.4}$) are sensitive to different cloud top height, while the information of the cloud phase referring to ice or water cloud types of the upper parts of the cloud is acquired by differences between 8–11 μm and 11–12 μm ($\Delta T_{8.6-11.2}$ and $\Delta T_{11.2-12.4}$) [48,49]. It has been shown that the combination of these differences can supply useful information to determine convectively dominated precipitation areas [38]. The Satellite Zenith Angle (SAZ) was also included.

Beyond that, taking into account the auxiliary effect of atmospheric environmental variables on rainfall retrieval, the meteorological variables in the NCEP operational GFS forecast products were used to enhance the performance of the algorithm, which can be obtained daily at 0000, 0600, 1200, and 1800 UTC with time steps of 3 h at a horizontal resolution of $0.25^\circ \times 0.25^\circ$. To conduct the real-time rainfall retrieval, we used the real-time GFS forecasting products. For real-time access, we set the forecast timeliness as 48 h, referring to the initial forecast time as 0000 UTC two days beforehand. The historical product can be downloaded at the National Center for Atmospheric Research, Computational and Information Systems Laboratory (<https://rda.ucar.edu/datasets/ds084.1/index.html>) (accessed on 15 October, 2020). The meteorological factors in terms of the temperature index, thermodynamic stability parameters, moisture indices, momentum were involved. The parameters revealed various aspects of the atmospheric characteristics. The Land Surface Temperature (LST) indicated the thermal state, Convective Inhibition (CIN), Convection Available Potential Energy (CAPE), and Best 4-layer Lifted Index (LI) described the instability of the atmosphere, Precipitable Water (PW) and Relative Humidity (RH) were empirical parameters for moisture, U-component of the planetary boundary layer wind (U) and V-component of the planetary boundary layer wind (V) were dynamic parameters for rainfall, these variables were bound up with the cloud and precipitation physical mechanism [1,3–5]. The study region (Figure 1) was East Asia (80°E – 135°E and 15°N – 55°N)

from June to August in 2018 and 2019. The predictors applied in the two-step model after feature selection were listed in the Table 1.

Table 1. The predictors used in the RF model for rain area determination.

Products	Variable	Unit
Himawari-8	SAZ (Satellite Zenith Angle)	°
	$\Delta T_{6.2-11.2}$	K
	$\Delta T_{7.3-12.4}$	K
	$\Delta T_{8.6-11.2}$	K
	$\Delta T_{11.2-12.4}$	K
	$T_{6.2}$	K
	$T_{6.9}$	K
	$T_{7.3}$	K
	$T_{9.6}$	K
	LST (Land Surface Temperature)	K
GFS	CIN (Convective Inhibition)	J/kg
	CAPE (Convection Available Potential Energy)	J/kg
	LI (Best 4-layer Lifted Index)	K
	PW (Precipitable Water)	kg/m ²
	RH (Relative Humidity)	%
	U (U-component of the planetary boundary layer)	m/s
	V (V-component of the planetary boundary layer)	m/s

2.1.2. Satellite-Gauge Product

The quantitative precipitation data from the Final-Run of the integrated multisatellite retrievals for the Level-3 gridded product GPM IMERG (Version V06B) were introduced, which is a satellite-gauge product that is published about 3.5 months after observation, provided by NASA (<https://disc.gsfc.nasa.gov/datasets>) (accessed on 17 October, 2020). Rainfall estimates with $0.1^\circ \times 0.1^\circ$ resolution and a 30 min interval were produced from a variety of precipitation-related satellite Passive Microwave (PMW) sensors, IR estimates, rain gauge observations, and so on. The final merged data from multiple fields provided the calibrated precipitation variable: precipitationCal (unit: mm/h), defined as the multisatellite precipitation estimate under gauge calibration. Then we considered the precipitationCal values as the ground truth when training the machine learning model. For the purpose of rainfall retrieval, four grades of precipitation were divided based on rainfall intensity according to meteorological standards [57], and the list of grades used in our study is presented in Table 2.

Table 2. Classification of precipitation grades based on meteorological standards.

Class	Grades	Range
1	No rain	<0.1 mm/h
2	Light rain	0.1 mm/h–1.5 mm/h
3	Moderate rain	1.6 mm/h–6.9 mm/h
4	Heavy rain	≥ 7.0 mm/h

2.1.3. Rain gauge observations

Ground-based data used for the comparison and validation were acquired from the China Integrated Meteorological Information Sharing System (CIMISS). The rain-gauge data were collected from observations at approximately 2,500 national weather stations every hour. A quality control program was developed which take into account the historical extremes and differences between adjacent sites. The observation datasets of August 2018 and summer 2019 in study area were applied. The distribution of these stations in the study region was demonstrated in the Figure 1.

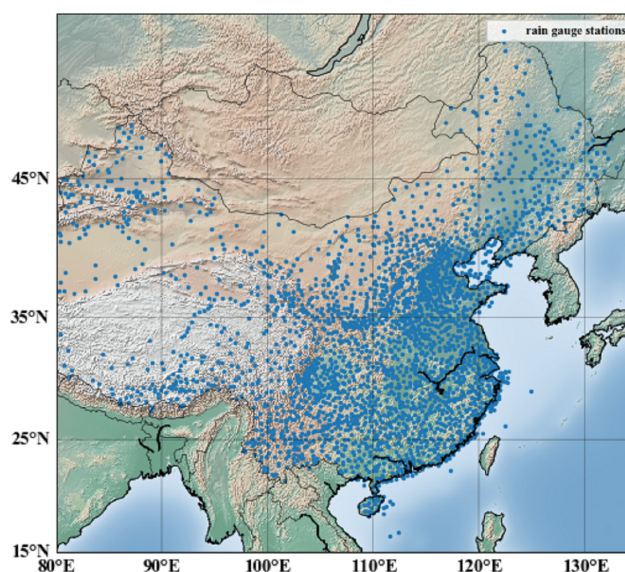


Figure 1. Distribution of the rain gauge stations in the study area.

2.1.4. Data Processing

Different characteristics of spatial–temporal-resolution satellite observations and numerical products match the satellite-gauge products. The temporal matching between the instantaneous state image and the precipitation rates needs to consider the relevant factors about the physical mechanism of cloud precipitation, such as the development and evolution of cloud, the phase change of water matter, the drop and growth of droplets, and the influence of turbulent dynamic process. To cover the rain process, we applied the median for the instantaneous image of the geostationary satellite over half an hour, corresponding to the calibrated precipitation value of multiple satellites within the same period. For example, the instantaneous values of Himawari-8 at 0120 UTC and 0150 UTC are parallel to the precipitation rates of GPM IMERG from 0100 UTC to 0130 UTC and from 0130 UTC to 0200 UTC, respectively. Meanwhile, we integrated the GFS data with 6-h temporal-resolution into 30 minute intervals by adopting the cubic-spline-interpolation method and the spatial matching technique was used on GFS variables and GPM data, keeping the same spatial-resolution as Himawari-8. After the preprocessing, a series of datasets ($0.05^\circ \times 0.05^\circ$, 30min) was established to build the RF classification model.

2.2. Methods

2.2.1. Random Forests

The Random Forests technique is an ensemble algorithm initially proposed by Breiman [56] and has been developed for numerous nonlinear classifications and regression estimations. Based on the bootstrap sampling method and random feature selection, a vast unrelated forest composed of a great number of decision trees was established, with the final output the category with the most votes in the classification or the average prediction among all decision trees in the regression model. Notably, it enables great speed for highly parallel computing on large samples or multidimensional vector training with a high prediction accuracy. As a result of the random sampling, the generalization ability of the training model is enhanced and stable, showing that the adaptability is high in other circumstances with similar characteristics [56].

A remarkable advantage is that it can be evaluated internally due to the bootstrap sampling technique, eventually giving the importance score of each input feature to the model. An unbiased estimation of the error in the process of training with the Out-Of-Bag (OOB) score using out-of-bag samples was calculated to evaluate the model performance. For the classification algorithm, with the judgment of all the trees, the category with the highest votes was identified as the predicted category of this sample [56,58].

In this study, the mature scikit-learn toolkit (<http://scikit-learn.org/stable/>) (accessed on 1 April, 2021) in the ML Python module was used to implement the RF algorithm.

2.2.2. Resampling Technique for Imbalanced Data

Resampling technique is one of the most common methods to adjust the distribution of the training samples with unbalanced data. In the field of machine learning, the number of training samples in each category has varied greatly in some situations such as disease detection and fraud monitoring [59]. There are some problems that researchers have focused on, such as minority classes, where the classifier tends to predict the majority classes under this circumstance even if a high accuracy is acquired. The same phenomenon is also described in precipitation classification, which suffers from category imbalance. Random undersampling and oversampling are found to modify the distribution; however, the former randomly discards the partial data in the majority samples, but has the risk of removing the potentially useful information, while the latter duplicates samples from the minority and tends to overfit [60,61]. An improved scheme based on random oversampling, named the Synthetic Minority Oversampling Technique (SMOTE), has been proposed to synthesize new minority samples on the basis of the original sample to reduce the impact of excessive fitting [60]. The key steps of this method are as follows: (i) for each sample x_i in the minority class, the Euclidean distance is taken as the criterion to compute the distance from it through all samples to obtain its k -nearest neighbor; (ii) determine the multiplier Y based on the sampling ratio, and the nearest neighbor sample \hat{x}_i is randomly selected around x_i ; (iii) a point \bar{x}_i on the line between x_i and \hat{x}_i is randomly selected as the new synthesized minority samples, expressed as:

$$\bar{x}_i = x_i + |x_i - \hat{x}_i| * \text{rand}(0, 1) \quad (1)$$

In the above equation, the function of $\text{rand}(0,1)$ is to generate random numbers from 0 to 1. However, there are still imperfections that need to be explained. If the selected minority samples are also encompassed by minority samples, the newly synthesized samples cannot offer more useful characteristic information. Conversely, when the selected minority samples are encompassed by majority samples, this may produce noise, and then, the synthesized samples will overlap most of the surrounding majority samples, which makes the classification difficult [62].

For the implementation of the resampling techniques in the paper, we took advantage of the Python toolbox Imbalanced-learn [63]. More detailed information about the above project can be obtained at the scikit learn website at <https://imbalanced-learn.org/stable> (accessed on 10 April, 2021).

3. Implementation Details

The flow chart of the two-step RF-based rainfall retrieval technique is presented in Figure 2. First, we made the spatiotemporal resolution of the whole dataset containing the Himawari-8 scan, NWP data, and GPM IMERG uniform. Then, the first module was established to delineate the rainfall area. We sampled a small dataset to optimize the RF model by feature selection and parameter tuning. Then, we applied the trained model to the independent dataset to determine whether there was precipitation or no precipitation for the given pixels in the satellite images at a specific time. In the following step, we removed the clear sky pixels from the dataset used in the first module, analogously, we developed the second module to estimate the grades of precipitation, and the model was used to confirm if the pixels were light rain, moderate rain, or heavy rain on the geostationary satellite remote sensing images. Finally, we respectively interpreted and evaluated the accuracy of the two-step model and compared the result with the calibrated precipitation of GPM IMERG. The training of the two-step model in each step was carried out independently.

After the step of data preprocessing, considering that the number of pixels of June and July in 2018 was too large to be able to directly train the classification model, we randomly

sampled 1% of the pixels every day in June and July of 2018 as the sample dataset covering 25,822,032 samples to capture the distribution characteristics of the summer rainfall, then the sample dataset was split into the training (20%) and validation (80%) datasets to make a considerably fair assessment of the module. To test the universality of the model, we set the whole dataset in August 2018 and summer 2019 as an independent testing dataset separately. Because a lot of pixels in the IMERG product were marked as rainy, whereas flagged as clear sky in the H08/AHI cloud mask, we did not make use of the cloud mask product to avoid the process error when delineating the rainfall area. In addition, we combined random undersampling and the SMOTE technique in the training for the first classification model, and the SMOTE technique was used independently in the second RF model to improve the performance of the two-step retrieval algorithm.

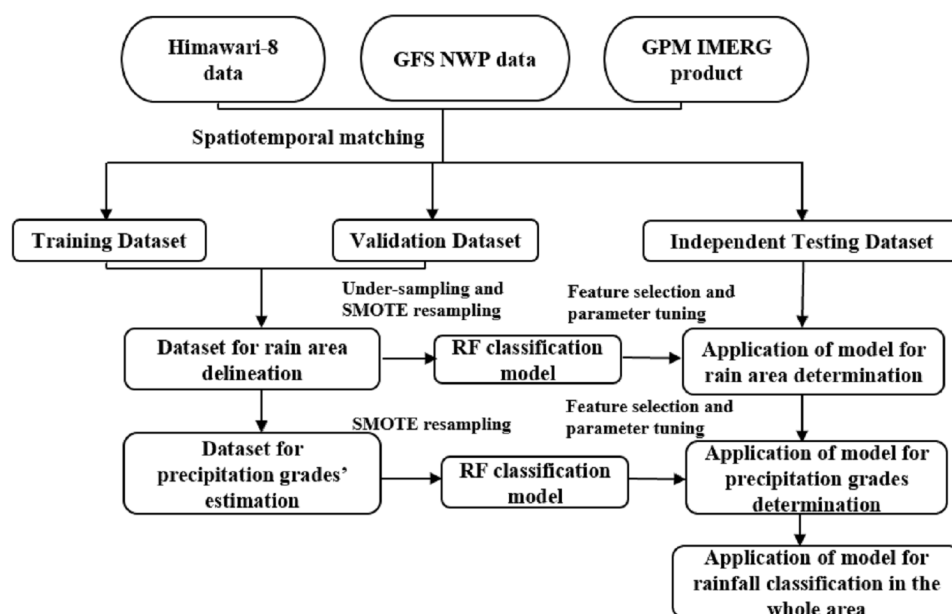


Figure 2. Framework of the two-step RF-based precipitation retrieval model.

3.1. Model Tuning and Training

For the tuning of the two-step RF model, first, the sample balance technique was adopted to choose the most suitable sampling method to process the training dataset. Then, we screened the relatively important features from the initially chosen variables and later optimized the hyperparameter of the modules. Ultimately, the model was trained and optimized. To decrease the optimization time and enhance the overall performance, 10,000 samples were randomly sampled from the prepared sample dataset for optimization.

In regard to the first classification for rain area delineation, nonprecipitation and precipitation pixels were divided to effectively identify the precipitation area. During the process of model training, due to the significant imbalance of the two classes in the dataset that lowered the number of rainy pixels, directly randomly selecting the training dataset led to the underestimation of precipitation. To solve this problem, the sampling technique has been introduced by several researchers [51–53,55]. Notably, reducing the number of nonprecipitation pixels and then increasing the amount of precipitation pixels can alleviate the effects of imbalance. Afterward the SMOTE sampling was performed after downsampling to balance the number of pixels for the two classes. We selected four scenarios and controlled the ratio of clear sky pixels to rainy pixels as 4:1, 3:1, 2:1, and 1.25:1 to test the optimal ratio of undersampling. Then, 2000 samples were randomly selected from the whole sample dataset as the training dataset to pick the optimal model, and 8000

samples were used as the validation dataset. For the purpose of the second classification for the precipitation grades' estimation, precipitation pixels were extracted and classified into different grades of precipitation. Similarly, direct training can overestimate the amount of light rain and underestimate the quantity of moderate rain and heavy rain, owing to the extreme imbalance. The training dataset can be balanced by dramatically increasing the amount of moderate rain and heavy rain. Later, 2000 precipitation samples were selected to acquire the optimum parameter collocation for classification.

We eliminated relatively unimportant features to reduce the number of input variables according to the variable importance score to increase the accuracy and efficiency of the RF model learning and training. The classifier gave each input feature a certain weight so that the feature importance score could be calculated from the sum of Gini impurity reduction of all nodes split on the feature across all decision trees. We finally determined the first 17 important predictors from the initial variables. The parameter tuned in the RF model was the number of decision trees in the entire forest ($n_{estimators}$), which played a vital role in affecting the sensitivity of model performance, meaning the maximum number of iterations of the weak learner in the algorithm. We determined the parameters to be as large as possible while comprehensively taking the objective hardware conditions into consideration. During the tuning process, the values of the OOB score were regarded as the evaluation index for parameter adjustment. A great quantity of RF models were applied to train with the number of trees being 10, 50, 100, 200, 300, 500, 800, and 1000.

After building the optimal model, the training dataset was shuffled out of order to improve the generalization ability of the simulation, including 5,164,407 pixels in the first classification sampled from the sample dataset, and 840,650 pixels were subsequently employed in the second classification. Then, the two datasets were applied for training in the rain area delineation and precipitation grades' estimation separately.

3.2. Validation and Testing

For the assessment of the presented algorithm above, we independently introduced the validation dataset including 20,657,625 pixels in the module for the rain area delineation and 3,362,604 pixels in the module for the precipitation grades' estimation. Moreover, we synchronously tested the data in August 2018 and the whole summer of 2019 to further analyze the applicability of the model.

Different metrics commonly used in meteorology were calculated pixel-by-pixel separately in the two-class and multiclass modules. In the case of the two-class classification delineating the rainfall area, we introduced the Probability Of Detection (POD), False-Alarm Ratio (FAR), Critical Success Index (CSI), Heike Skill Score (HSS), Equitable Threat Score (ETS), and Bias. The POD represents the reliability of correctly classifying precipitation pixels when rain is observed. The FAR shows the error rate to identify precipitation pixels when rain was simulated. The CSI reveals the fraction of correctly classified precipitation pixels while removing the correctly classified nonprecipitation pixels. The HSS measures the percentage of correctly classified pixels, excluding those simulations that were correct purely due to random chance. The ETS shows the fraction of correctly classified pixels after accounting for accidental correctness and is relatively fairer. The Bias describes the frequency of correctly classified precipitation pixels compared to that of observed rain and the tendency of underforecasting (Bias < 1) or overforecasting (Bias > 1). They can be calculated for four situations: a reveals that both the IMERG product and the estimated pixels by the RF model as rain; c indicates that the RF classification identified the pixels as rain, but the IMERG product did not; b shows that the pixels in the IMERG product are rain, but in the RF model are not; d displays that pixels both in the IMERG product and the RF classification are considered as nonprecipitation. The detailed information for these metrics is shown in Table 3.

Table 3. Names, calculation formula, range, and optimum values of the various metrics for delineating the rainfall area.

Name	Formula	Range	Optimum
Probability of detection	$POD = \frac{a}{a+c}$	[0, 1]	1
False-alarm ratio	$FAR = \frac{b}{a+b}$	[0, 1]	0
Critical success index	$CSI = \frac{a}{a+b+c}$	[0, 1]	1
Heike skill score	$HSS = \frac{2(ad-bc)}{(a+c)(c+d)+(a+b)(b+d)}$	$[-\infty, 1]$	1
Equitable threat score	$ETS = \frac{a-dr}{a+b+c-dr}$ where $dr = \frac{(a+c)(a+b)}{a+b+c+d}$	$[-1/3, 1]$	1
Bias	$Bias = \frac{a+b}{a+c}$	$[0, +\infty]$	1

With reference to the multiclass classification estimating the precipitation grades, the Accuracy (ACC) was used to illustrate the ratio of pixels classified as the correct grades. ACC can be computed by dividing the correctly classified pixels by the total number of pixels and easily shows a preference for the majority in comparison to the minority; it can be expressed as:

$$ACC = \frac{1}{N} \sum_{i=1}^N I(PR_{Simi} = PR_{Obsi}) \quad (2)$$

where PR_{Simi} is the category of the i -th simulation of the RF model, PR_{Obsi} is the result of the corresponding i -th GPM IMERG products, N is the total number of simulation items, and I is the indicator function.

4. Results

4.1. Rain Area Delineation

This section summarizes the results of the tuning and training for the two-class precipitation to identify the rain area, and then, the metrics are applied in the validation dataset and testing dataset to assess the model overall performance. The ratio of nonprecipitation pixels to precipitation pixels in the sample set was close to 5:1, manifesting a large quantity difference. The 2000 prepared samples for the training dataset for parameter tuning were used to establish four scenarios to obtain the optimal ratio of random undersampling. Table 4 shows the validation scores for the 8000 validation samples with different proportions.

Table 4. Validation scores under four scenarios with different proportions of clear sky pixels to rainy pixels.

Model Name	Clear Sky/Rainy	POD	FAR	CSI	HSS	ETS	Bias
Scenario_0	4:1	0.70	0.16	0.61	0.56	0.39	0.83
Scenario_1	3:1	0.72	0.17	0.63	0.57	0.39	0.87
Scenario_2	2:1	0.76	0.19	0.64	0.58	0.39	0.94
Scenario_3	1.25:1	0.80	0.21	0.66	0.59	0.39	1.02

With the decrease of the ratios in the training dataset, the performance of all verification scores improved except the FAR. The POD had the most significant alteration from 0.70 to 0.80. The Bias varied from underestimating to overestimating a little. The CSI, HSS, and ETS displayed a slight change. Therefore, despite the FAR becoming worse due to the reduction of nonprecipitation pixels, Scenario₃ was finally comprehensively considered.

After the implementation of the sampling technology, the next step was feature selection based on the feature importance score. The single channels of IR 8.6, IR 10.4, IR 11.2, IR 12.4, and IR 13.3 showed lower importance compared to the other input variables. The Water Vapor channels (i.e., $T_{6.2}$) and IR channel (i.e., $T_{9.6}$) were retained, along with the differences between the water vapor-IR channels and longwave IR channels (i.e., $\Delta T_{6.2-11.2}$, which represents the differences between $T_{6.2}$ and $T_{11.2}$), and the physically

related meteorological variables were used. Table 1 shows the 17 input variables for nonprecipitation and precipitation pixels' classification in detail.

During the model tuning process, we tuned the number of decision trees to determine the optimum by examining the variance of the OOB scores obtained from the out-of-bag samples that did not participate in the training. Figure 3 displays how the values of the OOB score changed when iteratively adjusting the parameters. The OOB score increased significantly until the number of trees reached 200, showing that there was no longer a decided change in scores, but a large consumption in internal storage and processing time. Finally, we took 200 decision trees as the optimal RF classification model.

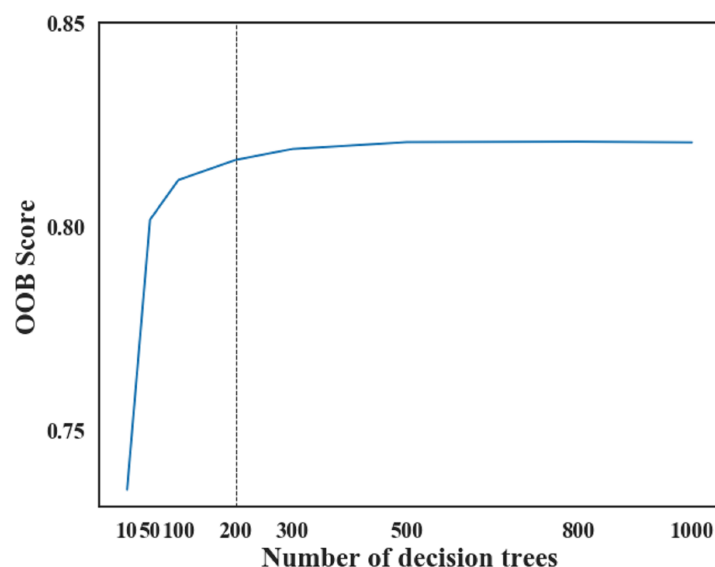


Figure 3. The parameter tuning in the RF model for rain area determination.

Next, the validation process of the RF model for the identification of precipitation regions was carried out after the training procedure. Meanwhile, the metrics for the RF model forecasting on each pixel were calculated and compared with the corresponding pixels in the IMERG product. The scores of the validation dataset for rain area delineation were $POD = 0.77$, $FAR = 0.23$, $CSI = 0.63$, $HSS = 0.54$, $ETS = 0.37$, and $Bias = 1.01$. Then, the contributions of each feature were obtained. Figure 4 shows the ranking of these predictors for the nonprecipitation and precipitation classification. Among all the variables used, PW occupied the most important position, indicating that abundant water vapor is indeed an indispensable element for the occurrence and development of convective precipitation. The SAZ was the second-most important predictor. The convective instability parameter comprising the LI, CAPE, and CIN also showed extremely high rankings, demonstrating the favorable conditions for triggering convective precipitation. Moreover, wind speed also showed a high importance, because the low-level jet in the planetary boundary layer may be a crucial dynamic mechanism of convective precipitation. In addition, the brightness temperature differences between the infrared channels, which represent the cloud attribute information, showed a certain importance. Meanwhile, cloud top temperature belonging to $T_{9.6}$ and the single water vapor channel made a modest contribution.

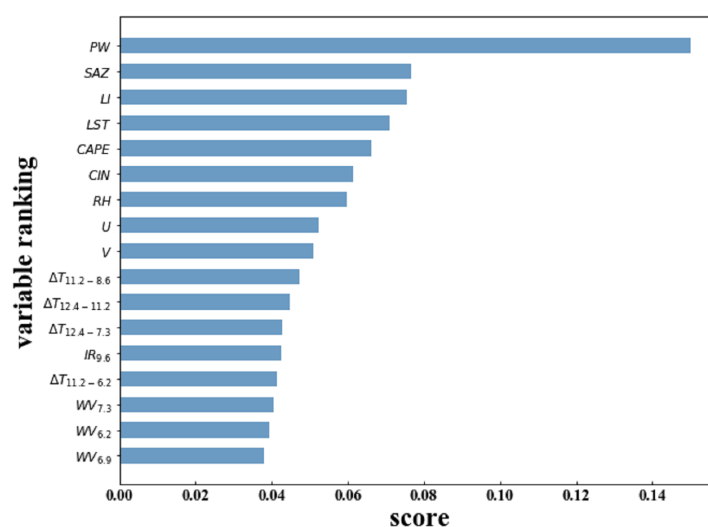


Figure 4. The ranking of the predictor variable in the RF model for rain area determination.

After the RF algorithm showed a better result on the validation dataset, we primarily utilized the August data in 2018 to test the model for rain area determination. To evaluate the continuous performance of the model over a course of the day, a boxplot of the validation metrics (Figure 5) was used to express the simulation results of the model over half-hour period. As can be seen from Figure 5, in the afternoon (0830 UTC to 1330 UTC), the model presented a higher POD of 0.75, while the distribution of the other period was basically around 0.68. In the same period, the FAR and Bias scores were also slightly higher than the others, revealing that the model resulted in some overestimation. This may be due to that the convective clouds often formed in the afternoon and gradually developed into the deep convective precipitation in the mesoscale convective system, making the model better at identifying precipitation, especially convective precipitation. The FAR scores were lower than 0.30. There were no noticeable temporal variations of the CSI, HSS, and ETS scores, and they were around 0.50, 0.40, and 0.30, respectively. It is worth noting that the testing results were superior to the previous studies: The precipitation region estimated by Kuhnlein et al. [49] showed a POD of 0.5, a FAR of 0.5, and a CSI of 0.3. Min et al. [51] discriminated the raining pixels with a POD of 0.58, a FAR of 0.33, and a CSI of 0.45. Ma et al. [48] obtained a POD of 0.6 and a FAR of 0.47 in rain area classification. Turini et al. [52] detected precipitation areas from cloudy pixels with a POD of 0.8 and a FAR of 0.3. In summary, the RF model can accurately classify the nonprecipitation and precipitation pixels.

Figure 6 shows an example of identifying the rain area at 0700 UTC 21 August 2018. Here, the rain area outline predicted by RF was largely consistent with that of the GPM, despite some overestimation, which can also be reflected in the performance of the POD and FAR. The most likely reason is that the undersampling for the clear sky pixels led to the bias of the predicted precipitation samples and partly because the PW contributed most and is large-scale, continuous, and zonal, while rainfall is locally small-scale, which led to the mismatch between them. Perhaps adding local physical parameters such as the convergence index could improve this, and we will try to achieve this in future work. Therefore, to some extent, the sample balance technology will sacrifice the accuracy of the nonprecipitation area.

Then, the model for identifying rain area was further tested with the following whole summer in 2019. We used the boxplot (Figure 7) to describe the testing results of the RF model at half-hour intervals. The result revealed that the evaluation metrics in each time period of the whole day were relatively stable. However, the overall performance was worse than that of August 2018.

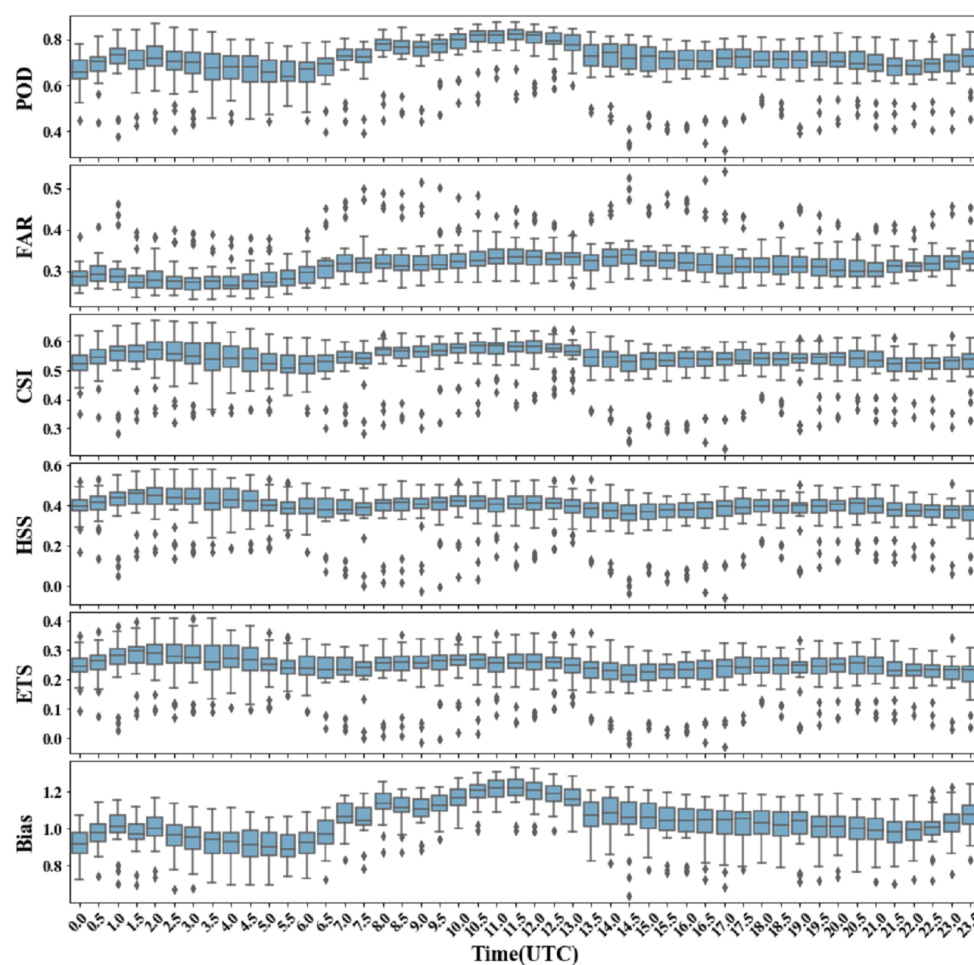


Figure 5. The diurnal variation of the evaluation scores for rain area determination on the testing dataset during August 2018; the distributions of the statistical values by half-an-hour are interpreted in box plots. Box diagrams indicate the 25th, 50th, and 75th percentiles, whereas the periphery of the box extends to 1.5-times the quartile deviation (25th–75th percentile). Outliers are indicated by black dots.

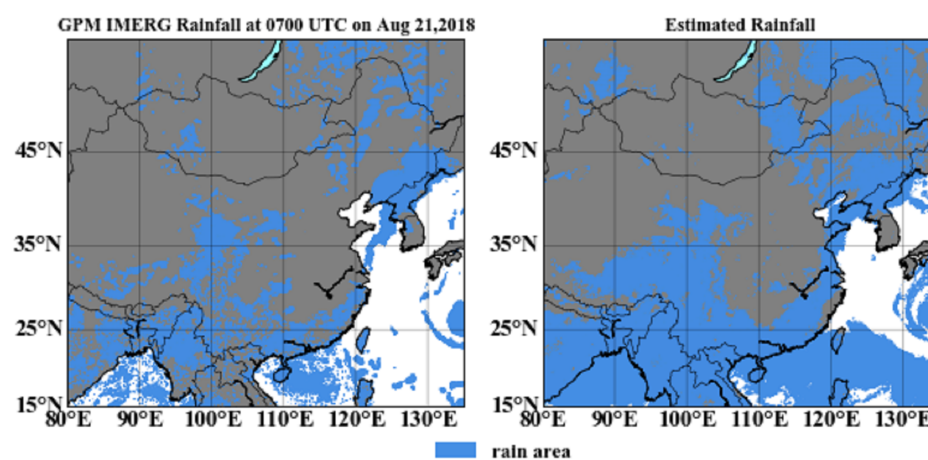


Figure 6. Comparisons of rain area delineation between the GPM (left) and RF model (right) at 0700 UTC 21 August 2018.

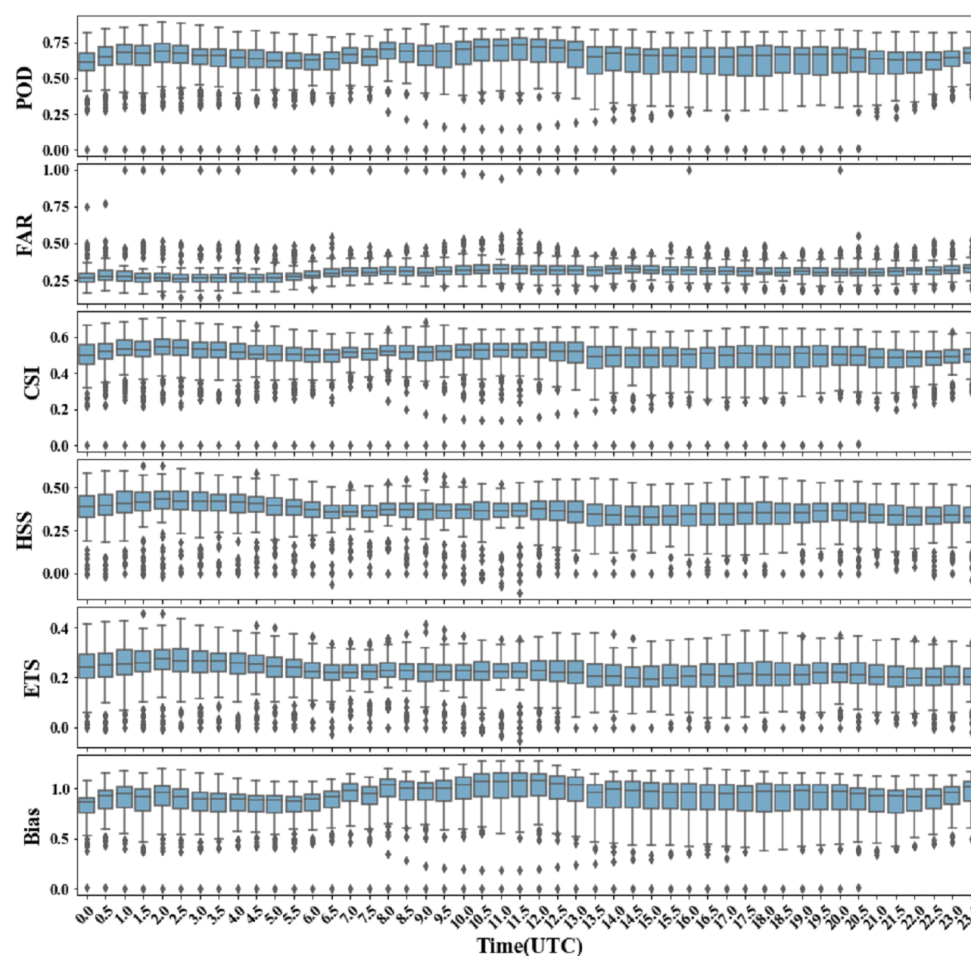


Figure 7. The diurnal variation of evaluation scores for rain area determination in the testing dataset in summer 2019; the distributions of the statistical values by half-an-hour are interpreted in box plots. Box diagrams indicate the 25th, 50th, and 75th percentiles, whereas the periphery of the box extends to 1.5-times the quartile deviation (25th–75th percentile). Outliers are indicated by black dots.

Figure 8 shows an example for identifying rain area at 0630 UTC 1 July 2019. The rain area profile predicted by RF was still similar to that of the GPM precipitation with a certain overestimation of the rain area, which was also caused by the characteristics of the undersampling algorithm and the possible influence of the PW, which had the main contribution.

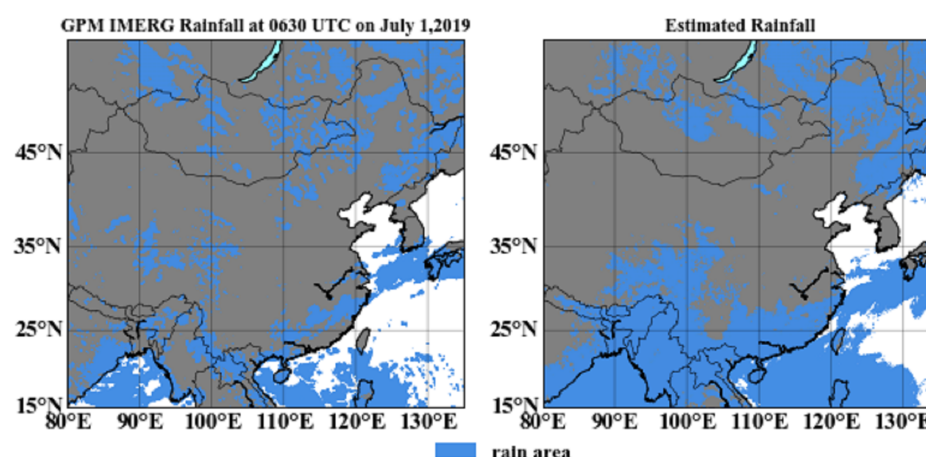


Figure 8. Comparisons of rain area delineation between the GPM (left) and RF model (right) at 0630 UTC 1 July 2019.

4.2. Precipitation Grades' Estimation

This section analyzes the results of the tuning and training of the three-class classification to estimate precipitation grades after fixing the rain area, and the accuracy was applied to evaluate the performance. There was also a large difference with respect to the ratio of light rain, moderate rain, and heavy rain in the sample dataset, close to 20:6:1. Therefore, the training model could not detect heavy rain. To solve this, the SMOTE sampling was used to keep the number of samples of the three classes consistent. Twenty-thousand samples were randomly selected from the sample dataset after removing nonprecipitation pixels and then used for feature selection and tuning. The 17 variables ultimately chosen were the same as those in the RF model for rain area determination, and then, the number of decision trees was also optimized. Figure 9 depicts how the values of the OOB score changed when iteratively adjusting the parameters. The results showed that the model was subtle for the number of decision trees, and there was also no noticeable variation in the scores over 200 iterations, so the 200 iterations were considered as optimal.

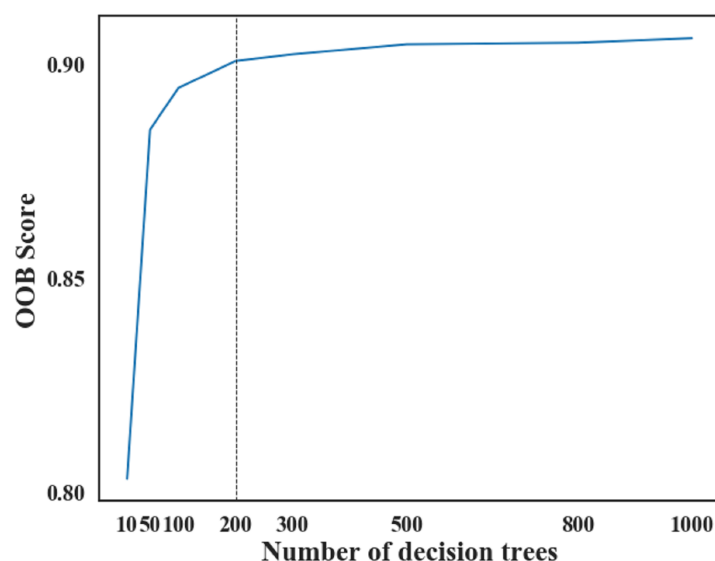


Figure 9. The parameter tuning in the RF model for the precipitation grades' estimation.

The trained RF model through the tuning process was applied to the validation datasets. Moreover, the metrics were calculated on each pixel, then the results were compared with the IMERG products. The ACC of the validation dataset was 0.70. Figure

10 shows the ranking of the variable importance scores in the RF model for measuring precipitation grades. The PW had the most contribution to the classification model due to the effect of sufficient moisture on precipitation. The importance of the CAPE, which represents the convective instability condition, increased. SAZ also showed plenty of contribution. Next were the other meteorological variables of the environmental background field in the NWP model and the Himawari-8 longwave spectral channels, including the brightness temperature and their difference, which reflects the cloud property information.

We continued to test the RF model with nonprecipitation pixels removed in August. To evaluate the effect of the RF model throughout the whole day, a boxplot (Figure 11) of ACC is used to present the performance at half-hour intervals. ACC fluctuated markedly throughout all the day with an average of 0.60, exhibiting a promising result for forecasting the strength of precipitation regarding the accuracy of heavy rain prediction. The inhomogeneity may be caused by the large proportion of light rain samples at some time, e.g. 0730 UTC. For unbalanced samples, the accuracy of the model will be higher when the number of samples in majority classes is large.

To visualize the classification results for light rain, moderate rain, and heavy rain in space, Figure 12 gives an example for identifying rainfall intensity at 0700 UTC 21 August 2018. In this case, the distribution of light rain areas was consistent, and heavy rain areas could be detected. A direct cause was that the oversampling algorithm led to the prediction results being biased towards majority classes for heavy rain samples. There was a some overestimation for heavy rain areas with a concentrated distribution, while an underestimation for heavy rain areas with a scattered distribution.

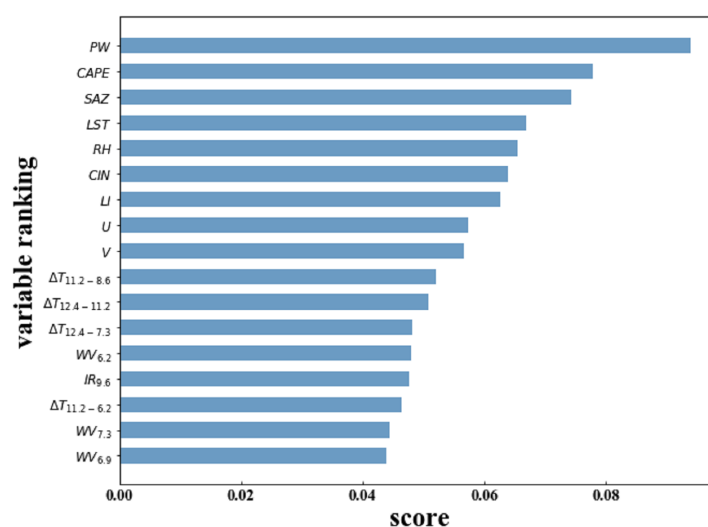


Figure 10. The ranking of the predictor variable in the RF model for the precipitation grades' estimation.

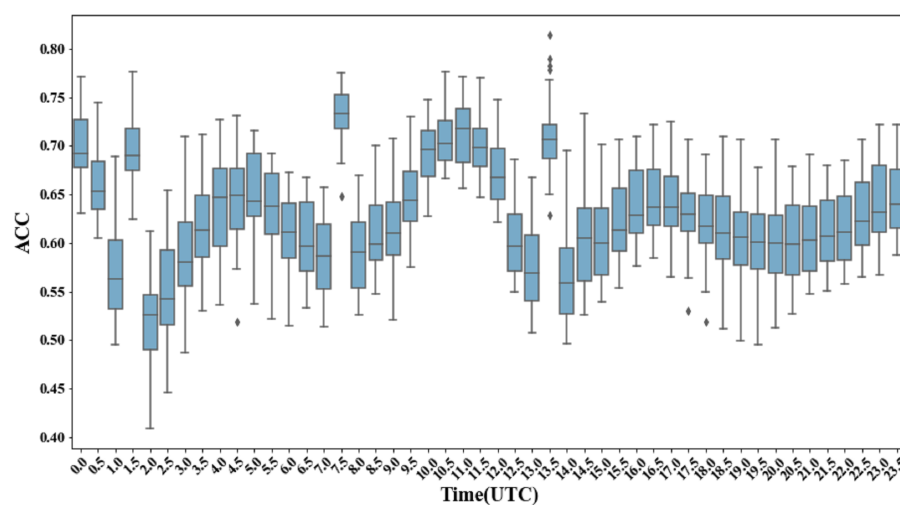


Figure 11. The diurnal variation of ACC for the precipitation grades' estimation in testing dataset during August 2018; the distributions of the statistical values by half-an-hour are interpreted in box plots. Box diagrams indicate the 25th, 50th, and 75th percentiles, whereas the periphery of the box extends to 1.5-times the quartile deviation (25th–75th percentile). Outliers are indicated by black dots.

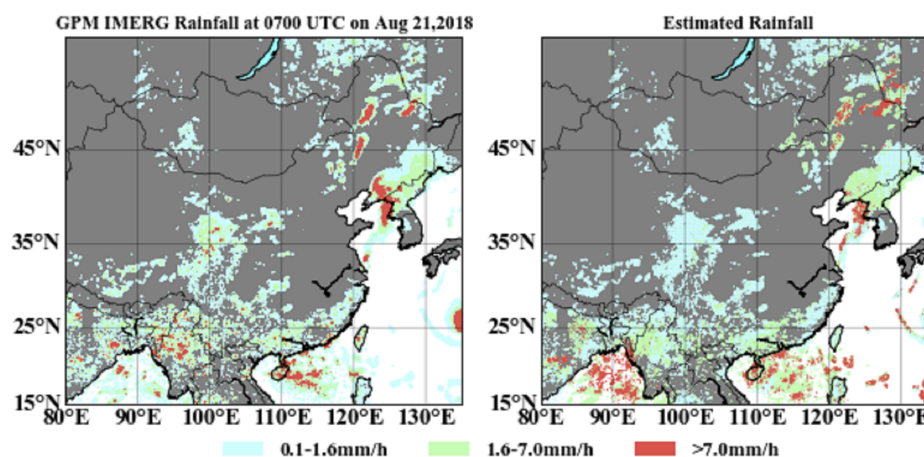


Figure 12. Comparisons of the precipitation grades' estimation between the GPM (left) and RF model (right) on 0700 UTC 21 August 2018.

We then removed nonprecipitation pixels in summer 2019 and tested the RF model for precipitation grades' estimation. To examine the simulation result of the RF model for the whole day, the boxplot (Figure 13) of ACC was used to present the performance at half-hour intervals. ACC maintained the prominent fluctuation trend throughout all the day, which was consistent with August 2018, and the average ACC was approximately 0.60. The inhomogeneity at certain moments (e.g. 0730 UTC) may also be due to the large number of light rain samples.

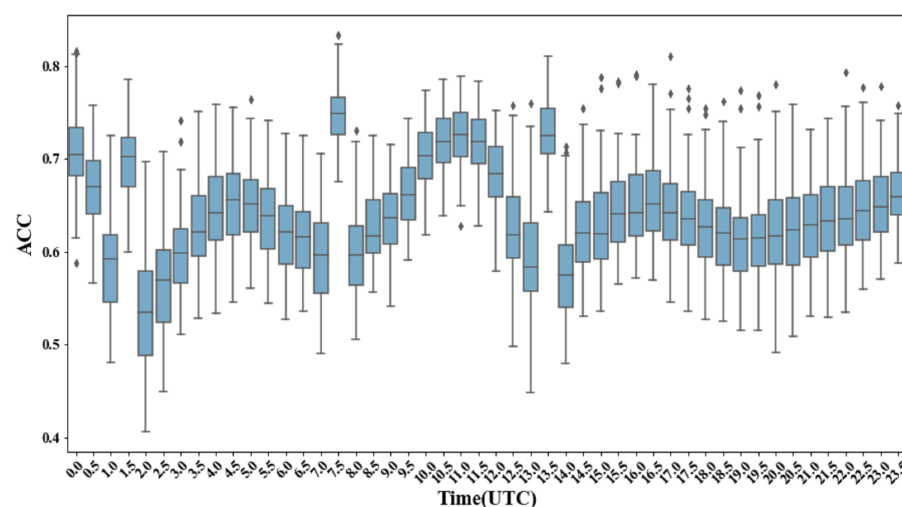


Figure 13. The diurnal variation of ACC for the precipitation grades' estimation in testing dataset in summer 2019; the distributions of the statistical values by half-an-hour are interpreted in box plots. Box diagrams indicate the 25th, 50th, and 75th percentiles, whereas the periphery of the box extends to 1.5-times the quartile deviation (25th–75th percentile). Outliers are indicated by black dots.

To visualize the classification results for light rain, moderate rain, and heavy rain in space, Figure 14 draws the example for identifying precipitation grades at 0630 UTC 1 July 2019. In this situation, the prediction accuracy of light rain areas was still high, and heavy rain areas could basically be estimated. This was still caused by the feature of the oversampling algorithm. The spatial overestimation was in high-density heavy rain areas, and underestimation was in thin heavy rain areas.

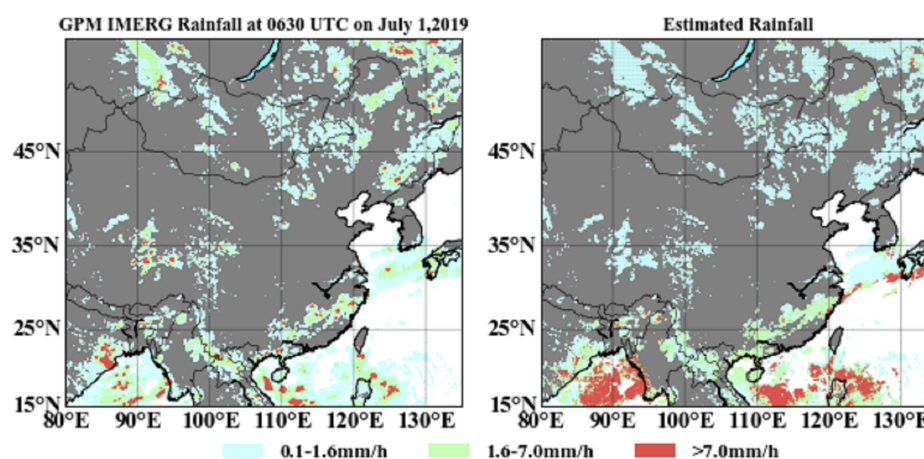


Figure 14. Comparisons of the precipitation grades' estimation between the GPM (left) and RF model (right) at 0630 UTC 1 July 2019.

4.3. The rainfall Retrieval Integrated Model

To assess the overall performance of trained rainfall retrieval model, RF classification model was used for rain area determination and precipitation grades' estimation to assess the final fusion product. The boxplot (Figure 15) of ACC is used to demonstrate the fusion testing results at half-hour intervals during August 2018. The average ACC was about 0.65, and the ACC from 0230 UTC to 0530 UTC was about 0.70, which was higher than other times. It may be that the convective precipitation in this period was weak, and the samples tended to be nonprecipitation and light rain, which made the forecasting accuracy higher.

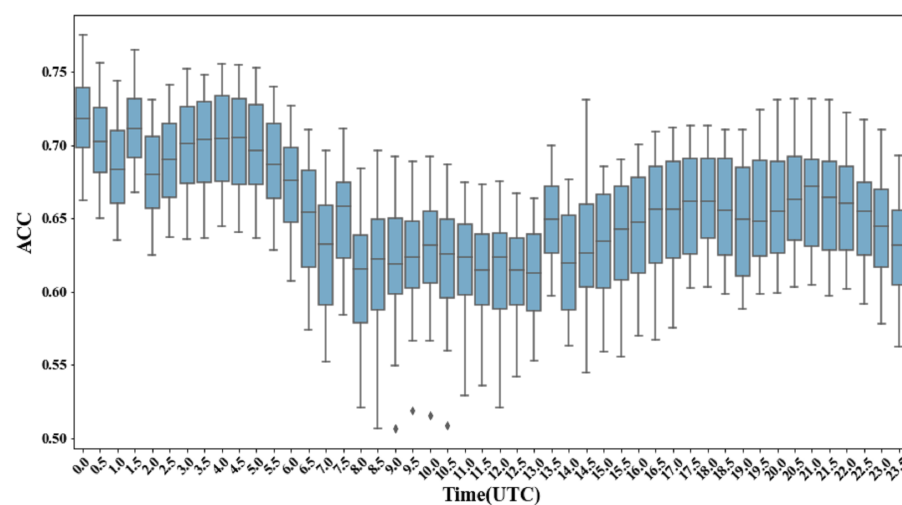


Figure 15. The diurnal variation of the accuracy for rainfall retrieval in the testing dataset during August 2018; the distributions of the statistical values by half-an-hour are interpreted in box plots. Box diagrams indicate the 25th, 50th, and 75th percentiles, whereas the periphery of the box extends to 1.5-times the quartile deviation (25th–75th percentile). Outliers are indicated by black dots.

To visualize the classification results for the whole rain area in space, Figure 16 describes the example for identifying rainfall intensity at 0700 UTC 21 August. In this scenario, the spatial distribution distinctly revealed that the errors of the previous two RF classification models were accumulated in the integrated model. What is more remarkable is that the strength of rainfall predicted by the RF model was overestimated, especially for the moderate rain and heavy rain in the densely distributed areas of heavy rain; however, the model underestimated rainfall in areas with a sparse distribution of heavy rain. The results were mainly affected by the imbalance of the rainfall data, which can be improved by using sampling method to some extent. A more suitable sampling method may be introduced to deal with this situation and improve the algorithm in the future. For instance, subsequent study can take into account the EasyEnsemble algorithm, which conducts multiple downsampling to generate multiple training sets and finally fuses the trained model together, can make up for the possible loss of important information in the random downsampling technique. Beyond that, adding more important input variables may also result in better performance. Besides the local small-scale climate variables mentioned above, the topographic types and texture features can also be introduced into the model because of their effects on precipitation distribution, like the mountain can block and guide the movement of water vapor.

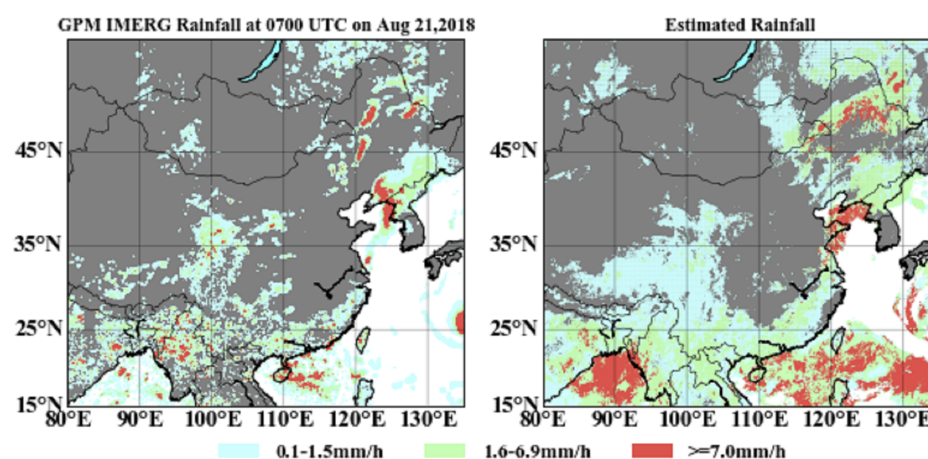


Figure 16. Comparisons of the rainfall retrieval integrated model between the GPM (left) and RF model (right) at 0700 UTC 21 August 2018.

We then tested the final merged product on the basis of the rainfall retrieval integrated RF models in summer 2019. The boxplot (Figure 17) of ACC is used to demonstrate the fusion testing results at half-hour intervals. The average ACC of the testing dataset was about 0.65, and the tendency of the ACC to vary with time became smooth and steady compared to that of August 2018, the possible causes being that the model's adaptability was lower in the following summer, as well as the number of samples being more than before.

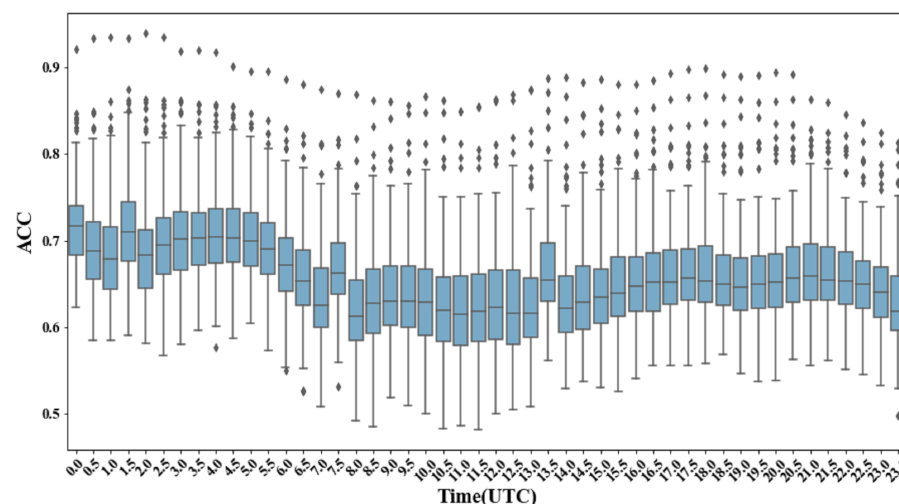


Figure 17. The diurnal variation of the accuracy for rainfall retrieval in the testing dataset in summer 2019; the distributions of the statistical values by half-an-hour are interpreted in box plots. Box diagrams indicate the 25th, 50th, and 75th percentiles, whereas the periphery of the box extends to 1.5-times the quartile deviation (25th–75th percentile). Outliers are indicated by black dots.

To visualize the classification results for the whole rain area, Figure 18 shows the circumstance for identifying precipitation grades at 0630 UTC 1 July 2019. In summary, the accuracy of the rainfall retrieval still needs to be improved, due to the overestimation of rain areas on the whole area. However, when compared with previous research, great progress has been made in the identification of moderate rain areas and heavy rain areas. Therefore, the two-step model can provide an important reference for further quantitative precipitation estimation and forecasting of the key weather systems and meteorological elements.

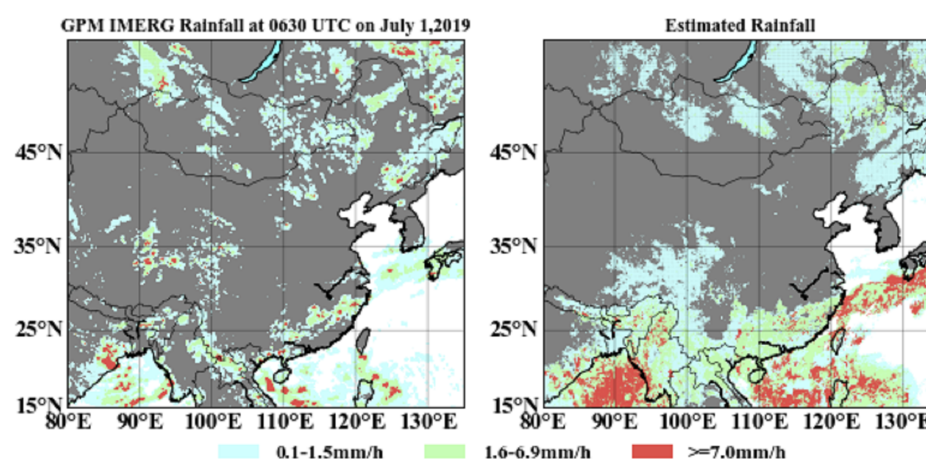


Figure 18. Comparisons of the rainfall retrieval integrated model between the GPM (left) and RF model (right) at 0630 UTC 1 July 2019.

4.4. Comparison with rain gauge stations

To further validate the performance of the RF model on independent datasets, the ground-based data from August 2018 and summer 2019 (Figure 1) were applied as a reference. The near-neighbor algorithm was implemented to interpolate the pixels of the testing datasets to the corresponding station, and we compared the model results of the second 30 minutes with the precipitation grades of the station in one hour. To assess the continuous performance of the RF model in August 2018 against rain gauge data over a course of the day, the boxplot of validation metrics (Figure 19) was used to express the simulation results of the model over one hour period. The results display the same characteristics as before. The values of POD scores from 0600 UTC to 1200 UTC are higher than that of other times and generally distribute above 0.6. The FAR and Bias in the same period are also a little higher, however, the accuracy becomes lower in this period, which displays that the overestimated precipitation samples cause great deviation to the final results of the model. The CSI, HSS, and ETS don't change significantly over time, they are around 0.5, 0.38, and 0.2. The overall performance indicates the RF model has shown expecting behavior in rainfall retrieval.

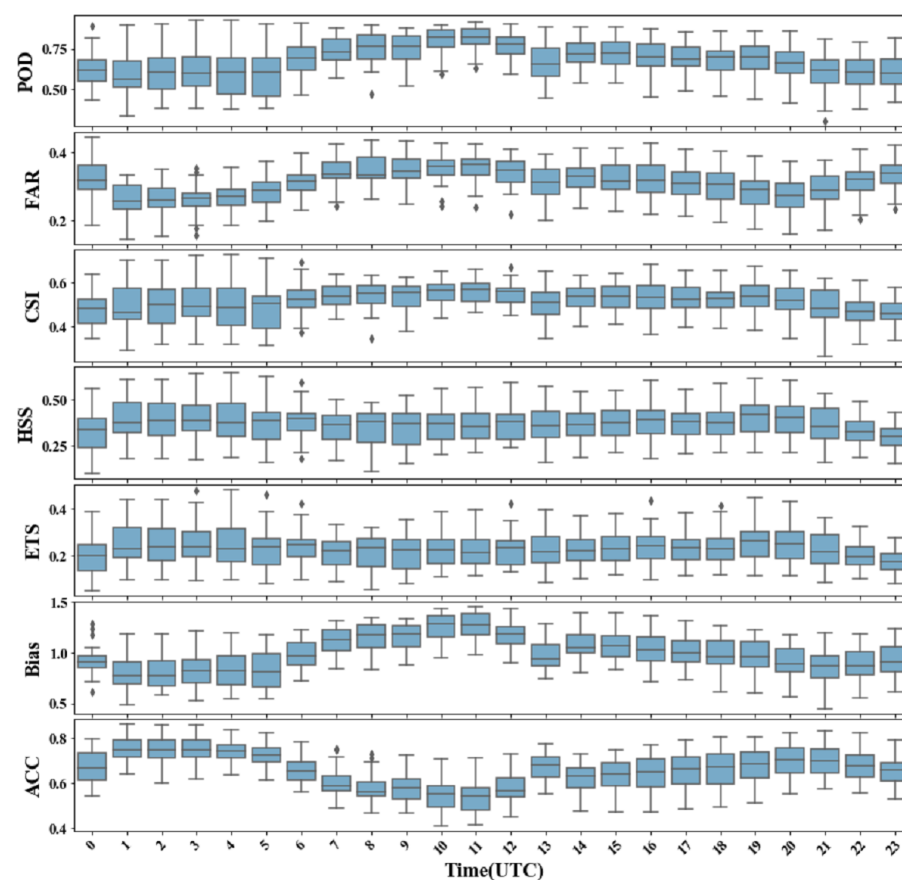


Figure 19. The diurnal variation of evaluation scores for rain area determination and precipitation grades estimation on RF model against gauge stations during August 2018, distributions of statistical values at one hour intervals are interpreted in box plots. Box diagrams indicate the 25th, 50th, and 75th percentiles, whereas the periphery of the box extends to 1.5 times of the quartile deviation (25th–75th percentile). Outliers are indicated by black dots.

We used the boxplot (Figure 20) to describe the testing results of RF model in summer 2019 against gauge data at one hour intervals. The results shows that the evaluation scores are relatively stable throughout the whole day and the performance is worse than that of August 2018.

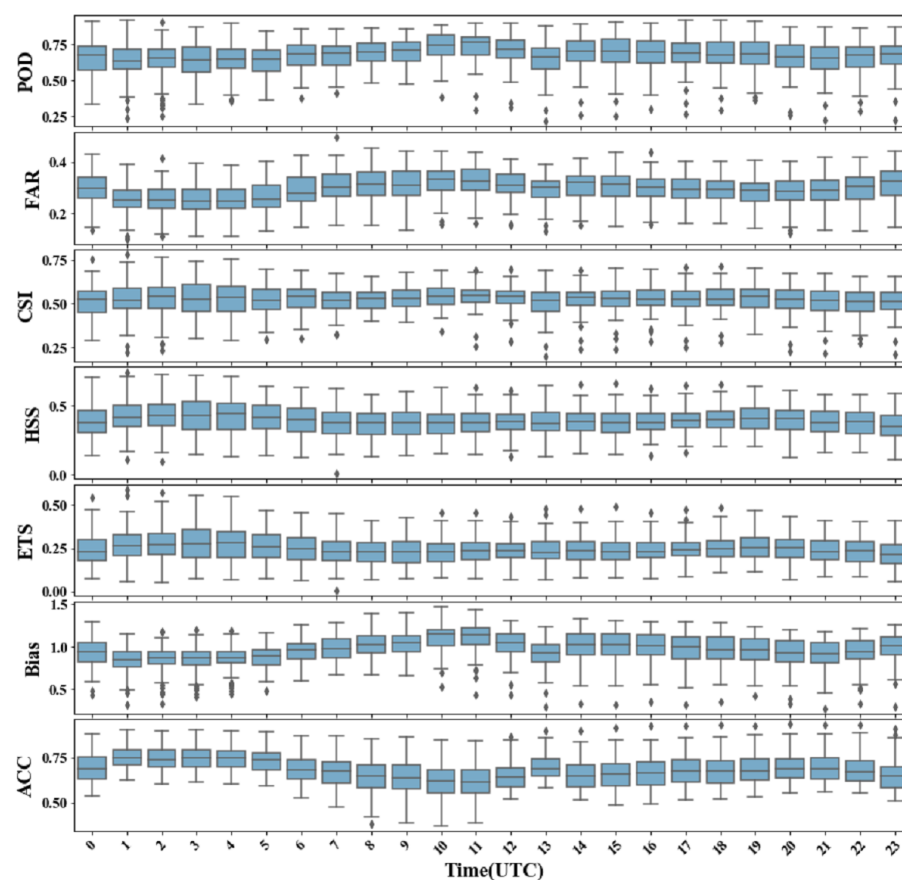


Figure 20. The diurnal variation of evaluation scores for rain area determination and precipitation grades estimation on RF model against gauge stations during summer 2019, distributions of statistical values at one hour intervals are interpreted in box plots. Box diagrams indicate the 25th, 50th, and 75th percentiles whereas the periphery of the box extends to 1.5 times of the quartile deviation (25th–75th percentile). Outliers are indicated by black dots.

4.5. Comparison with GFS model

The ranking of predictor variables in the RF model both reveals that the GFS meteorological variables are more important than the Himawari imager infrared brightness temperatures. To prove the contribution of infrared satellite data, we compared the 6-hour precipitation predicted in the GFS model with RF model. According to meteorological standards [57], the values less than 0.1 mm are considered as no rain. The pixels between 0.1 mm and 3.9 mm are regarded as light rain, between 4.0 mm and 12.9 mm are defined as moderate rain, more than 13 mm are considered as heavy rain. Due to the temporal resolution mismatch, we accumulated the 30-minute GPM IMERG product to 6 hours, then calculated the evaluation scores between the GFS model and GPM IMERG product, finally compared them with the 6-hour average value of the model testing dataset. The boxplot (Figure 21) to compare the GFS model rainfall with testing results of RF model in August 2018 at six hour intervals was showed. Compared with the GFS model, the evaluation scores of the RF model display better performance in general, especially the accuracy, which shows that the infrared satellite data can sufficiently reveal the information of precipitation intensity. However, the POD scores from 0000 UTC to 0600 UTC are lower than GFS model, the possible reason is that the GFS forecast field overestimates the precipitation area to a greater degree than the RF model in this period, so that the FAR is also higher in this case.

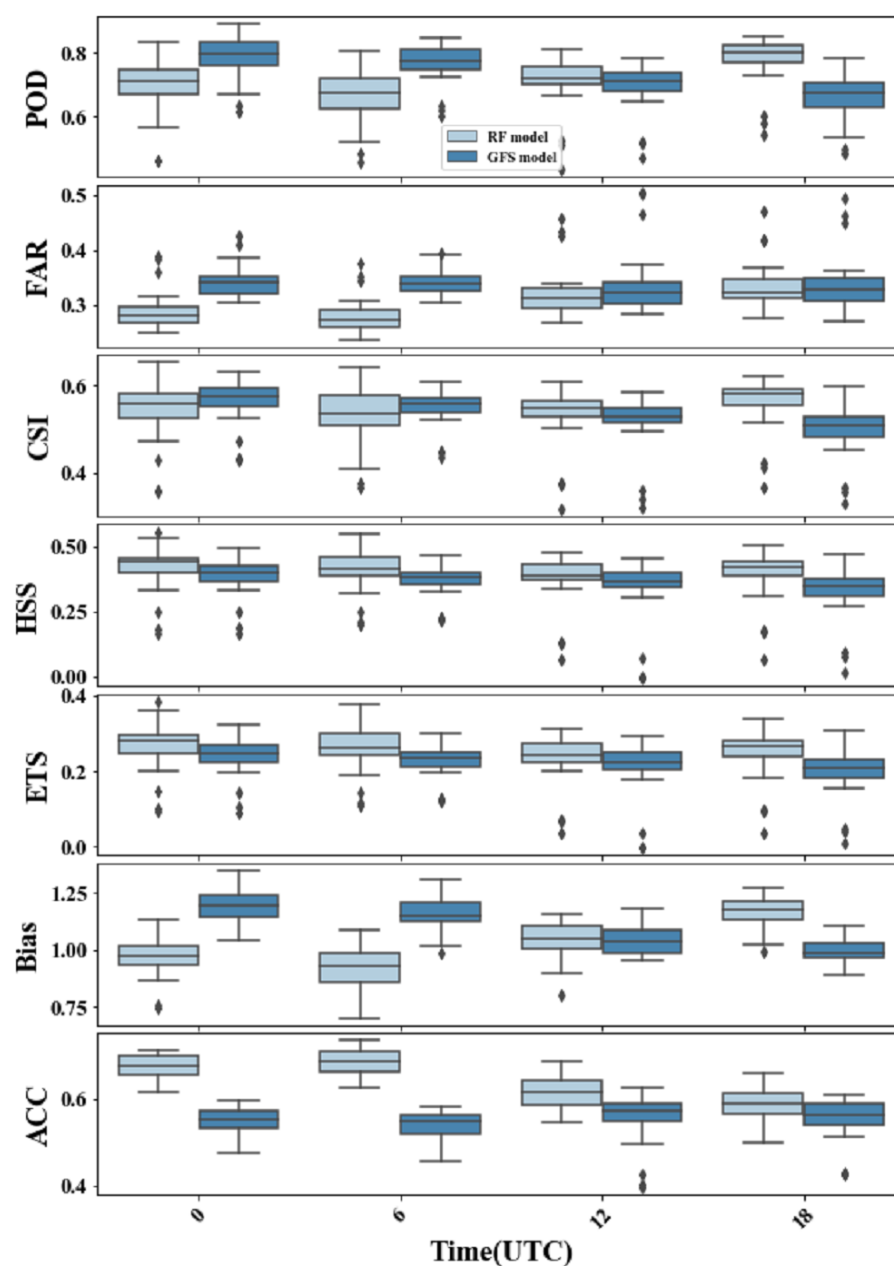


Figure 21. The diurnal variation of evaluation scores for rain area determination and precipitation grades estimation on RF model and GFS model against GPM data during August 2018, distributions of statistical values at six hour intervals are interpreted in box plots. Box diagrams indicate the 25th, 50th, and 75th percentiles whereas the periphery of the box extends to 1.5 times of the quartile deviation (25th–75th percentile). Outliers are indicated by black dots.

The boxplot (Figure 22) was showed to compare the GFS model rainfall with RF model results in summer 2019 at six hour intervals. The results still shows that the RF model is superior as a result of the addition of infrared information. And the model performance is worse than that of August 2018.

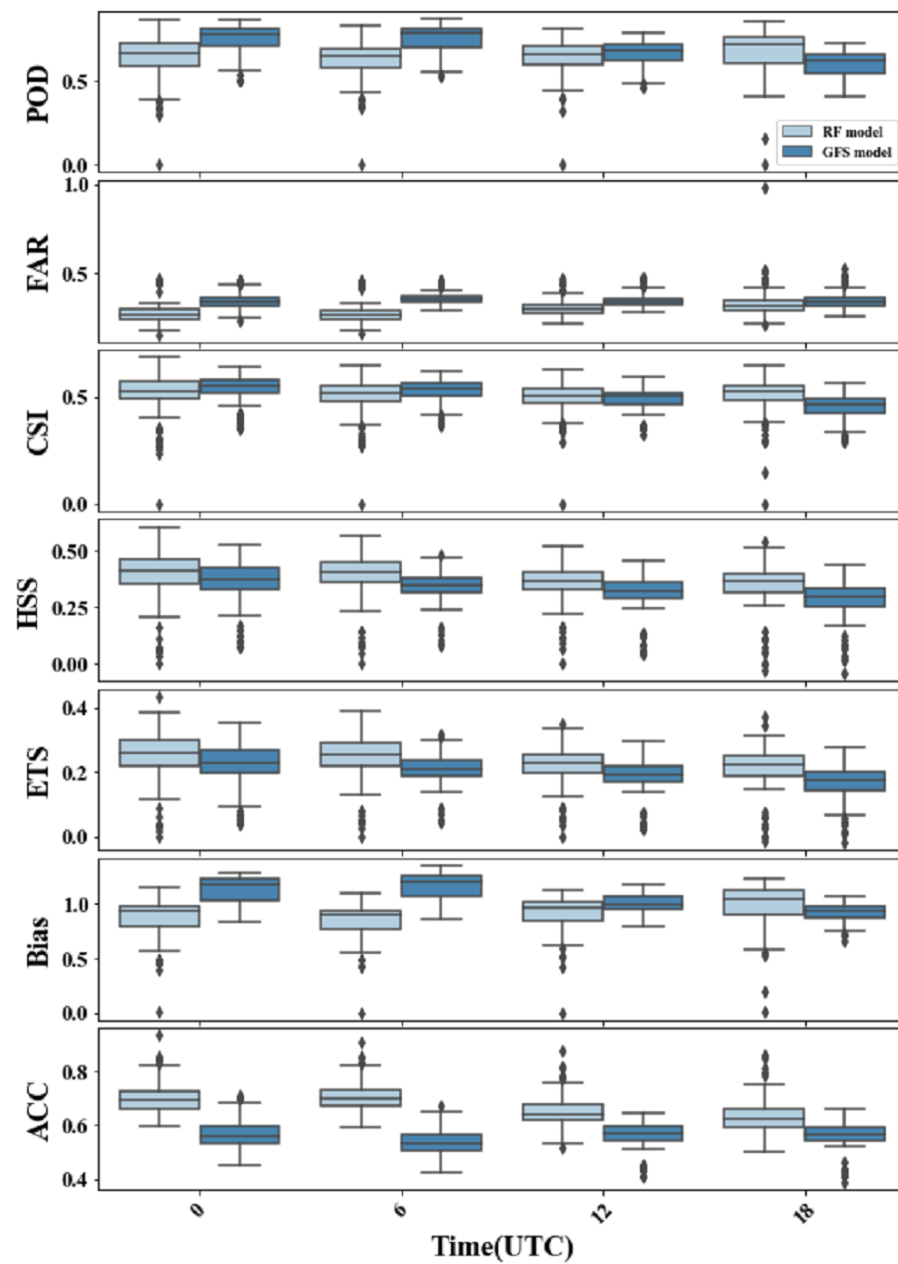


Figure 22. The diurnal variation of evaluation scores for rain area determination and precipitation grades estimation on RF model and GFS model against GPM data during summer 2019, distributions of statistical values at six hour intervals are interpreted in box plots. Box diagrams indicate the 25th, 50th, and 75th percentiles whereas the periphery of the box extends to 1.5 times of the quartile deviation (25th-75th percentile). Outliers are indicated by black dots.

5. Discussion

In our study, a new rainfall retrieval technique based on the RF model was proposed, which using multichannel Himawari-8/AHI satellite observations and GFS meteorological data. The calibrated precipitation information from GPM IMERG was regarded as the ground truth during the training process. We built a two-step rainfall retrieval integrated model to handle high-dimensional physically associated features and precipitation information to improve the retrieval precision of the rainfall area and precipitation grades.

The relevant atmospheric factors in the GFS product were introduced into the geostationary IR-based technique as additional variables. Their extremely high contribution to the RF classifications demonstrated that the environmental background field of atmospheric transportation surely has an evident effect on the precipitation forecasting. In response to the widespread underestimation of moderate and heavy rain areas in the quantitative precipitation estimation, the classification model was established to qualitatively estimate the precipitation grades after determining the rain area first. In addition, the resampling techniques were implemented on the sample dataset to increase the accuracy of the whole classification model. We examined the which test had the best random undersampling ratio in terms of the frequently used evaluation metrics counted pixel-by-pixel compared to the IMERG data. Then, more important predictors were screened and the model parameters optimized. Large validation and testing datasets were applied to train the RF model to test the general adaptability with respect to the independent datasets. Additionally, the RF model was validated against independent gauge data, then compared with the forecasting cumulative rainfall of the GFS model.

Overall, using the developed approach, the evaluation metrics obtained in the rain area detection module were excellent in comparison with previous satellite-based [47–52,54,55]. Regarding the effectiveness of the model optimization, the results directly proved the practicability and feasibility of the rain area delineation algorithm in forecasting summer precipitation the next month and the next year. At the moment when convective precipitation was developing vigorously, the model performed slightly better. In the next step of the grades' estimation, the accuracy of moderate and heavy rain prediction improved to some extent, giving the forecasting of rainfall intensity a certain guideline and instruction. This pixel-by-pixel model can also be rebuilt for different regions and periods after the application of the retuning and resampling method.

The proposed technique has made progress in the field of IR precipitation retrieval. The RF, as a popular machine learning algorithm in present research, achieved gratifying results with samples that required only a small memory footprint in the Interdisciplinary Course of Meteorological Research, even in the case of unbalanced data. In the future, we hope to realize high-resolution rainfall retrieval for other seasons with new geostationary satellite infrared information. Moreover, the presented algorithm will provide the premise for quantitative rainfall intensity estimations.

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

Funding: This research work was supported by the National Key Research and Development Program of China (2018YFC1507002) and National Natural Science Foundation of China (42075125).

Acknowledgments: We gratefully thank JMA for freely offering the Himawari-8 satellite data and NOAA for providing the NCEP GFS historical Forecast data archive. Furthermore, the IMERG V06B Final-Run products used in our study were kindly provided by NASA. The authors would like to sincerely thank the Python scikit-learn and Imbalanced-learn groups for offering the powerful analysis computational tools.

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

Abbreviations

The following abbreviations are used in this manuscript:

ACC	Accuracy
CAPE	Convection Available Potential Energy
CIMISS	China Integrated Meteorological Information Sharing System
CIN	Convective Inhibition
CMORPH	Climate Prediction Center Morphing Method
CSI	Critical Success Index
CST	Convective-Stratiform Technology
DPR	Dual-frequency Precipitation Radar
ETS	Equitable Threat Score
FAR	False-Alarm Ratio
GATE	GARP Atlantic Tropical Experiment
GFS	Global Forest System
GMI	GPM Microwave Imager
GOES	Geostationary Operational Environmental Satellite
GPI	GOES Precipitation Index
GPM	Global Precipitation Measurement
Himawari-8/AHI	Advanced Himawari Imager-8
HSS	Heike skill score
IMERG	Integrated Multi-satellite Retrievals for GPM
IR	Infrared
JMA	Japan Meteorological Agency
LI	Best four-layer Lifted Index
LST	Land Surface Temperature
MSG-SEVIRI	Meteosat Second Generation-Spinning Enhanced Visible and InfraRed Imager
NWP	Numerical Weather Product
OOB score	Out-Of-Bag score
PERSIANN	Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks
PMW	Passive Microwave
POD	Probability Of Detection
PW	Precipitable Water
RH	Relative Humidity
RF	Random Forest
SAZ	Satellite Zenith Angle
SMOTE	Synthetic Minority Over-sampling Technique
TRMM	Tropical Rainfall Measuring Mission
U	U component of planetary boundary layer wind
V	V component of planetary boundary layer wind

References

1. Cheng, H.; Wu, T.; Dong, W. Thermal contrast between the middle-latitude Asian continent and adjacent ocean and its connection to the East Asian summer precipitation. *J. Clim.* **2008**, *21*, 4992–5007.
2. Yao, C.; Yang, S.; Qian, W.; Lin, Z.; Wen, M. Regional summer precipitation events in Asia and their changes in the past decades. *J. Geophys. Res. Atmos.* **2008**, *113*, D17.
3. Li, S.; Hou, W.; Feng, G. Atmospheric circulation patterns over East Asia and their connection with summer precipitation and surface air temperature in Eastern China during 1961–2013. *J. Meteorol. Res.* **2018**, *32*, 203–218.
4. Huang, R.; Liu, Y.; Feng, T. Interdecadal change of summer precipitation over Eastern China around the late-1990s and associated circulation anomalies, internal dynamical causes. *Chin. Sci. Bull.* **2013**, *58*, 1339–1349.
5. Hu, P.; Wang, M.; Yang, L.; Wang, X.; Feng, G. Water vapor transport related to the interdecadal shift of summer precipitation over northern East Asia in the late 1990s. *J. Meteorol. Res.* **2018**, *32*, 781–793.
6. Neeck, S.P.; Kakar, R.K.; Azarbarzin, A.A.; Hou, A.Y. Global precipitation measurement (gpm) implementation. In *Sensors, Systems, and Next-Generation Satellites XIV. International Society for Optics and Photonics. SPIE: Toulouse, France*, **2010**.
7. Du, L.; Tian, Q.; Yu, T.; Meng, Q.; Jancso, T.; Udvardy, P.; Huang, Y. A comprehensive drought monitoring method integrating MODIS and TRMM data. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *23*, 245–253.

8. Panegrossi, G.; Casella, D.; Dietrich, S.; Marra, A.C.; Sano, P.; Mugnai, A.; Baldini, L.; Roberto, N.; Adirosi, E.; Cremonini, R.; et al. Use of the GPM constellation for monitoring heavy precipitation events over the Mediterranean region. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 2733–2753.
9. Yu, Y.; Wang, J.; Cheng, F.; Deng, H.; Chen, S. Drought monitoring in Yunnan Province based on a TRMM precipitation product. *Nat. Hazards* **2020**, *104*, 2369–2387.
10. Anagnostou, E.N.; Maggioni, V.; Nikolopoulos, E.I.; Meskele, T.; Hossain, F.; Papadopoulos, A. Benchmarking high-resolution global satellite rainfall products to radar and rain-gauge rainfall estimates. *IEEE Trans. Geosci. Remote Sens.* **2009**, *48*, 1667–1683.
11. Stampoulis, D.; Anagnostou, E.N. Evaluation of global satellite rainfall products over continental Europe. *J. Hydrometeorol.* **2012**, *13*, 588–603.
12. Dinku, T.; Chidzambwa, S.; Ceccato, P.; Connor, S.; Ropelewski, C. Validation of high-resolution satellite rainfall products over complex terrain. *Int. J. Remote Sens.* **2008**, *29*, 4097–4110.
13. Bitew, M.M.; Gebremichael, M. Evaluation of satellite rainfall products through hydrologic simulation in a fully distributed hydrologic model. *Water Resour. Res.* **2011**, *47*, doi:10.1029/2010WR009917.
14. Satgé, F.; Xavier, A.; Pillco Zolá, R.; Hussain, Y.; Timouk, F.; Garnier, J.; Bonnet, M.P. Comparative assessments of the latest GPM mission's spatially enhanced satellite rainfall products over the main Bolivian watersheds. *Remote Sens.* **2017**, *9*, 369.
15. Sorooshian, S.; Hsu, K.L.; Gao, X.; Gupta, H.V.; Imam, B.; Braithwaite, D. Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bull. Am. Meteorol. Soc.* **2000**, *81*, 2035–2046.
16. Dinku, T.; Connor, S.J.; Ceccato, P. Comparison of CMORPH and TRMM-3B42 over mountainous regions of Africa and South America. In *Satellite Rainfall Applications for Surface Hydrology*. Springer: Berlin/Heidelberg, Germany, **2010**, 193–204.
17. Chen, F.; Li, X. Evaluation of IMERG and TRMM 3B43 monthly precipitation products over mainland China. *Remote Sens.* **2016**, *8*, 472.
18. Skofronick-Jackson, G.; Petersen, W.A.; Berg, W.; Kidd, C.; Stocker, E.F.; Kirschbaum, D.B.; Kakar, R.; Braun, S.A.; Huffman, G.J.; Iguchi, T.; et al. The Global Precipitation Measurement (GPM) mission for science and society. *Bull. Am. Meteorol. Soc.* **2017**, *98*, 1679–1695.
19. Wang, Z.; Zhong, R.; Lai, C.; Chen, J. Evaluation of the GPM IMERG satellite-based precipitation products and the hydrological utility. *Atmos. Res.* **2017**, *196*, 151–163.
20. Foelsche, U.; Kirchengast, G.; Fuchsberger, J.; Tan, J.; Petersen, W.A.; others. Evaluation of GPM IMERG Early, Late, and Final rainfall estimates using WegenerNet gauge data in southeastern Austria. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 6559–6572.
21. Hosseini-Moghari, S.M.; Tang, Q. Validation of gpm imerg v05 and v06 precipitation products over iran. *J. Hydrometeorol.* **2020**, *21*, 1011–1037.
22. Yu, L.; Leng, G.; Python, A.; Peng, J. A Comprehensive Evaluation of Latest GPM IMERG V06 Early, Late and Final Precipitation Products across China. *Remote Sens.* **2021**, *13*, 1208.
23. Li, X.; Sungmin, O.; Wang, N.; Huang, Y.; et al. Evaluation of the GPM IMERG V06 products for light rain over Mainland China. *Atmos. Res.* **2021**, *253*, 105510.
24. Arkin, P.A. The relationship between fractional coverage of high cloud and rainfall accumulations during GATE over the B-scale array. *Mon. Weather Rev.* **1979**, *107*, 1382–1387.
25. Arkin, P.A.; Joyce, R.; Janowiak, J.E. The estimation of global monthly mean rainfall using infrared satellite data: The GOES Precipitation Index (GPI). *Remote Sens. Rev.* **1994**, *11*, 107–124.
26. Arkin, P.A.; Meisner, B.N. The relationship between large-scale convective rainfall and cold cloud over the western hemisphere during 1982–84. *Mon. Weather Rev.* **1987**, *115*, 51–74.
27. Mishra, A.K.; Gairola, R.; Varma, A.; Agarwal, V.K. Improved rainfall estimation over the Indian region using satellite infrared technique. *Adv. Space Res.* **2011**, *48*, 49–55.
28. Todd, M.C.; Kidd, C.; Kniveton, D.; Bellerby, T.J. A combined satellite infrared and passive microwave technique for estimation of small-scale rainfall. *J. Atmos. Ocean. Technol.* **2001**, *18*, 742–755.
29. Adler, R.F.; Negri, A.J. A satellite infrared technique to estimate tropical convective and stratiform rainfall. *J. Appl. Meteorol. Climatol.* **1988**, *27*, 30–51.
30. Grecu, M.; Anagnostou, E.N.; Adler, R.F. Assessment of the use of lightning information in satellite infrared rainfall estimation. *J. Hydrometeorol.* **2000**, *1*, 211–221.
31. Ebert, E.E.; Janowiak, J.E.; Kidd, C. Comparison of near-real-time precipitation estimates from satellite observations and numerical models. *Bull. Am. Meteorol. Soc.* **2007**, *88*, 47–64.
32. Früh, B.; Bendix, J.; Nauss, T.; Paulat, M.; Pfeiffer, A.; Schipper, J.W.; Thies, B.; Wernli, H. Verification of precipitation from regional climate simulations and remote-sensing observations with respect to ground-based observations in the upper Danube catchment. *Meteorol. Z.* **2007**, *16*, 275–293.
33. Ba, M.B.; Gruber, A. GOES multispectral rainfall algorithm (GMSRA). *J. Appl. Meteorol.* **2001**, *40*, 1500–1514.
34. Nauss, T.; Kokhanovsky, A.A. Discriminating raining from non-raining clouds at mid-latitudes using multispectral satellite data. *Atmos. Chem. Phys.* **2006**, *6*, 5031–5036.
35. Ebert, E.E.; Manton, M.J. Performance of satellite rainfall estimation algorithms during TOGA COARE. *J. Atmos. Sci.* **1998**, *55*, 1537–1557.

36. Torricella, F.; Cattani, E.; Levizzani, V. Rain area delineation by means of multispectral cloud characterization from satellite. *Adv. Geosci.* **2008**, *17*, 43–47.
37. Cattani, E.; Torricella, F.; Laviola, S.; Levizzani, V. On the statistical relationship between cloud optical and microphysical characteristics and rainfall intensity for convective storms over the Mediterranean. *Nat. Hazards Earth Syst. Sci.* **2009**, *9*, 2135–2142.
38. Thies, B.; Nauß, T.; Bendix, J. Precipitation process and rainfall intensity differentiation using Meteosat second generation spinning enhanced visible and infrared imager data. *J. Geophys. Res. Atmos.* **2008**, *113*, D23.
39. Feidas, H.; Giannakos, A. Classifying convective and stratiform rain using multispectral infrared Meteosat Second Generation satellite data. *Theor. Appl. Climatol.* **2012**, *108*, 613–630.
40. Bellerby, T.; Todd, M.; Kniveton, D.; Kidd, C. Rainfall estimation from a combination of TRMM precipitation radar and GOES multispectral satellite imagery through the use of an artificial neural network. *J. Appl. Meteorol.* **2000**, *39*, 2115–2128.
41. Hong, Y.; Hsu, K.L.; Sorooshian, S.; Gao, X. Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system. *J. Appl. Meteorol.* **2004**, *43*, 1834–1853.
42. Meyer, H.; Kühnlein, M.; Appelhans, T.; Nauss, T. Comparison of four machine learning algorithms for their applicability in satellite-based optical rainfall retrievals. *Atmos. Res.* **2016**, *169*, 424–433.
43. Chen, H.; Chandrasekar, V.; Cifelli, R.; Xie, P. A machine learning system for precipitation estimation using satellite and ground radar network observations. *IEEE Transactions on Geoscience and Remote Sensing* **2019**, *58*, 982–994.
44. Ramanujam, S.; Radhakrishnan, C.; Subramani, D.; Chakravarthy, B. On the effect of non-raining parameters in retrieval of surface rain rate using TRMM PR and TMI measurements. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2012**, *5*, 735–743.
45. Chen, H.; Chandrasekar, V.; Tan, H.; Cifelli, R. Rainfall estimation from ground radar and TRMM precipitation radar using hybrid deep neural networks. *Geophysical Research Letters* **2019**, *46*, 10669–10678.
46. Balaji, C.; Krishnamoorthy, C.; Chandrasekar, R. On the possibility of retrieving near-surface rain rate from the microwave sounder SAPHIR of the Megha-Tropiques mission. *Current Science* **2014**, pp. 587–593.
47. Hamidi, O.; Poorolajal, J.; Sadeghifar, M.; Abbasi, H.; Maryanaji, Z.; Faridi, H.R.; Tapak, L. A comparative study of support vector machines and artificial neural networks for predicting precipitation in Iran. *Theoretical and applied climatology* **2015**, *119*, 723–731.
48. Ma, L.; Zhang, G.; Lu, E. Using the gradient boosting decision tree to improve the delineation of hourly rain areas during the summer from advanced Himawari imager data. *Journal of Hydrometeorology* **2018**, *19*, 761–776.
49. Kühnlein, M.; Appelhans, T.; Thies, B.; Nauß, T. Precipitation estimates from MSG SEVIRI daytime, nighttime, and twilight data with random forests. *Journal of Applied Meteorology and Climatology* **2014**, *53*, 2457–2480.
50. Das, S.; Chakraborty, R.; Maitra, A. A random forest algorithm for nowcasting of intense precipitation events. *Advances in Space Research* **2017**, *60*, 1271–1282.
51. Min, M.; Bai, C.; Guo, J.; Sun, F.; Liu, C.; Wang, F.; Xu, H.; Tang, S.; Li, B.; Di, D.; others. Estimating summertime precipitation from Himawari-8 and global forecast system based on machine learning. *IEEE Transactions on Geoscience and Remote Sensing* **2018**, *57*, 2557–2570.
52. Turini, N.; Thies, B.; Bendix, J. Estimating high spatio-temporal resolution rainfall from MSG1 and GPM IMERG based on machine learning: case study of Iran. *Remote Sensing* **2019**, *11*, 2307.
53. Kolbe, C.; Thies, B.; Egli, S.; Lehnert, L.; Schulz, H.M.; Bendix, J. Precipitation Retrieval over the Tibetan Plateau from the Geostationary Orbit—Part 1: Precipitation Area Delineation with Elektro-L2 and Insat-3D. *Remote Sensing* **2019**, *11*, 2302.
54. Hirose, H.; Shige, S.; Yamamoto, M.K.; Higuchi, A. High temporal rainfall estimations from Himawari-8 multiband observations using the random-forest machine-learning method. *Journal of the Meteorological Society of Japan. Ser. II* **2019**.
55. Turini, N.; Thies, B.; Horna, N.; Bendix, J. Random forest-based rainfall retrieval for Ecuador using GOES-16 and IMERG-V06 data. *European Journal of Remote Sensing* **2021**, *54*, 117–139.
56. Breiman, L. Random forests. *Machine learning* **2001**, *45*, 5–32.
57. Qiao, L.; Li, Y.; Fu, J.; Tian, C.; Bi, B.; Zhou, Q.; Committee, C.N.S.M.; others. Grade of Precipitation, GB/T 28592–2012, **2012**.
58. Belgiu, M.; Drăguț, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS journal of photogrammetry and remote sensing* **2016**, *114*, 24–31.
59. Hoens, T.R.; Chawla, N.V. Imbalanced datasets: from sampling to classifiers. In *Imbalanced learning: Foundations, algorithms, and applications*. Wiley Online Library, **2013**, pp. 43–59.
60. Yap, B.W.; Abd Rani, K.; Abd Rahman, H.A.; Fong, S.; Khairudin, Z.; Abdullah, N.N. An application of oversampling, undersampling, bagging and boosting in handling imbalanced datasets. *Proceedings of the first international conference on advanced data and information engineering (DaEng-2013)*. Springer, Singapore, **2014**, 13–22.
61. Sáez, J.A.; Krawczyk, B.; Woźniak, M. Analyzing the oversampling of different classes and types of examples in multi-class imbalanced datasets. *Pattern Recognition* **2016**, *57*, 164–178.
62. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research* **2002**, *16*, 321–357.
63. Lemaître, G.; Nogueira, F.; Aridas, C.K. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *The Journal of Machine Learning Research* **2017**, *18*, 559–563.