

Article

# Retrieval of Total Precipitable Water from Himawari-8 AHI Data: A Comparison of Random Forest, Extreme Gradient Boosting, and Deep Neural Network

Yeonjin Lee <sup>1</sup>, Daehyeon Han <sup>2</sup> , Myoung-Hwan Ahn <sup>1,\*</sup> , Junggho Im <sup>2</sup>  and Su Jeong Lee <sup>1</sup> 

<sup>1</sup> Department of Atmospheric Sciences and Engineering, Ewha Womans University, 52 Ewhayeodae-gil, Seodaemun-gu, Seoul 03760, Korea

<sup>2</sup> School of Urban and Environmental Engineering, Ulsan National Institute of Science and Technology, Ulsan 44919, Korea

\* Correspondence: terryaahn65@ewha.ac.kr; Tel.: +82-2-3277-4462

Received: 11 May 2019; Accepted: 19 July 2019; Published: 24 July 2019



**Abstract:** Total precipitable water (TPW), a column of water vapor content in the atmosphere, provides information on the spatial distribution of moisture. The high-resolution TPW, together with atmospheric stability indices such as convective available potential energy (CAPE), is an effective indicator of severe weather phenomena in the pre-convective atmospheric condition. With the advent of high performing imaging instrument onboard geostationary satellites such as Advanced Himawari Imager (AHI) onboard Himawari-8 of Japan and Advanced Meteorological Imager (AMI) onboard GeoKomsat-2A of Korea, it is expected that unprecedented spatiotemporal resolution data (e.g., AMI plans to provide 2 km resolution data at every 2 min over the northeast part of East Asia) will be provided. To derive TPW from such high-resolution data in a timely fashion, an efficient algorithm is highly required. Here, machine learning approaches—random forest (RF), extreme gradient boosting (XGB), and deep neural network (DNN)—are assessed for the TPW retrieved from AHI over the clear sky in Northeast Asia area. For the training dataset, the nine infrared brightness temperatures (BT) of AHI (BT8 to 16 centered at 6.2, 6.9, 7.3, 8.6, 9.6, 10.4, 11.2, 12.4, and 13.3  $\mu\text{m}$ , respectively), six dual channel differences and observation conditions such as time, latitude, longitude, and satellite zenith angle for two years (September 2016 to August 2018) are used. The corresponding TPW is prepared by integrating the water vapor profiles from Interim European Centre for Medium-Range Weather Forecasts Re-Analysis data (ERA-Interim). The algorithm performances are assessed using the ERA-Interim and radiosonde observations (RAOB) as the reference data. The results show that the DNN model performs better than RF and XGB with a correlation coefficient of 0.96, a mean bias of 0.90 mm, and a root mean square error (RMSE) of 4.65 mm when compared to the ERA-Interim. Similarly, DNN results in a correlation coefficient of 0.95, a mean bias of 1.25 mm, and an RMSE of 5.03 mm when compared to RAOB. Contributing variables to retrieve the TPW in each model and the spatial and temporal analysis of the retrieved TPW are carefully examined and discussed.

**Keywords:** total precipitable water; Himawari-8 AHI; machine learning; random forest; deep neural network; XGBoost

## 1. Introduction

Water vapor, one of the most influential constituents of the atmosphere, is responsible for determining the amount of precipitation that a region can receive [1]. Total precipitable water (TPW) is a meteorological factor that shows the amount of water vapor contained in the column of air per unit area of the atmosphere in terms of the depth of liquid [2]. Although TPW does not represent the vertical structure of moisture, it describes horizontal gradients of integrated water vapor content.

Furthermore, the amount of water vapor contained in the troposphere has significant implications for determining the strength and severity of a severe weather event [3]. Therefore, TPW is one of the critical variables used by forecasters when severe weather conditions are expected [4].

Himawari-8, Japan's geostationary (GEO) meteorological satellite launched in October 2014, includes a primary instrument called Advanced Himawari Imager (AHI). AHI has 16 channels—four visible (VIS), two near-infrared (NIR), and 10 infrared (IR) channels—with a temporal resolution of 10 min for the full disk (less than 10 min for limited areas) and a spatial resolution of 2 km at nadir for the infrared channels (less than 2 km for the visible channels) [5]. Unlike low earth orbit (LEO) satellites, which have limited temporal resolution, GEO satellites can produce data more timely and frequently. The retrieved high temporal resolution TPW from GEO satellite sensor data can be utilized to monitor pre-convective environments and predict heavy rainfall, convective storms, and clouds that may cause serious damage to human life and infrastructure [6–8]. For example, Lee et al. [8] showed that the 10-min interval measurements from the AHI sensor successfully provided information about the pre-landfall environment for typhoon Nangka that occurred in 2015.

In the remote sensing field, there are several approaches for the retrieval of TPW from IR channels of GEO satellite observations, including (1) a physical method using the one-dimensional variational system, (2) a split-window algorithm, and (3) machine learning algorithms. The physical modeling based on a nonlinear optimal estimation method has been traditionally used for vertical profiles of temperature and humidity (T-q profile) [9]. While the TPW derived from the T-q profile retrieved with a physical method usually has high accuracy, it does not fully use the original resolution information from satellite observations due to the high computing load. The split-window method is based on a different absorptive response to the water vapor content at two channels (near 11.0 and 12.0  $\mu\text{m}$ ) [10–12]. This method can be classified into the linear approach, look-up table approach, physical approach, and covariance-variance ratio approach [13–16]. These approaches have limitations which are easily affected by uncertainties in the surface emissivity, errors in first guess field, and the instrument noise [16,17]. On the other hand, machine learning approaches such as random forest (RF), extreme gradient boosting (XGB), and deep neural network (DNN) are capable of predicting the nonlinear relationship between the parameters [18] when compared to the other methods. Since these approaches have been often adopted as ways for the fast and reliable calculation of a target variable, various machine learning techniques have been used to derive TPW from satellite sensor data. For instance, Wang et al. [19] used a multi-layer perceptron neural network with NIR radiances from Moderate-resolution Imaging Spectro-radiometer (MODIS) as input to retrieve TPW. Zhang et al. [20] also employed a radial basis function neural network algorithm using infrared data from the hyperspectral sounder, Atmospheric Infrared Sounder, to retrieve TPW. Although these machine learning based models have shown good performance, there has been little exploration in comparison of multiple machine learning techniques especially for GEO satellite data-based TPW retrievals.

In this study, machine learning based approaches for the retrieval of TPW from GEO satellite data over clear sky pixels are proposed. The objectives of this study are to (1) develop the TPW retrieval algorithms from Himawari-8 AHI data using RF, XGB, and DNN over Northeast Asia, (2) quantify and examine relative variable importance and contribution by model, and (3) analyze the temporal and spatial distribution of the errors compared to ERA-Interim in Northeast Asia for 2 years. Section 2 describes the study area and data used for the algorithm development and Section 3 explains the retrieval methods (RF, XGB, and DNN) based on the machine learning approaches. Variable importance, validation results, and discussion are described in Section 4. The conclusions are presented in Section 5.

## 2. Data

Himawari-8 launched on October 7, 2014 carries out a meteorological mission on a GEO orbit (centered on 140.68°E). The considerably improved AHI sensor in temporal, spatial, and spectral resolutions over its predecessors is now better suited for nowcasting and has improved the accuracy of

numerical forecasts [21]. AHI has a total of 16 spectral channels (0.47–13.3  $\mu\text{m}$ ) consisting of VIS, NIR, and IR—short-wavelength infrared (SWIR), mid-wavelength infrared (MWIR) and thermal infrared (TIR) channels (Table 1). AHI collects data every 10 min from the full disk and every 2.5 min from the north-eastern and south-western areas of Japan. The brightness temperature (BT) data from Himawari-8 AHI infrared channels are used as input variables to develop machine learning-based models to retrieve TPW.

**Table 1.** Himawari-8 Advanced Himawari Imager (AHI) specifications.

Channel Number	Central Wavelength ( $\mu\text{m}$ )	Band Width ( $\mu\text{m}$ )	Spatial Resolution at Sub Satellite Point (km)
1	0.47	0.05	1
2	0.51	0.02	1
3	0.64	0.03	0.5
4	0.86	0.02	1
5	1.6	0.02	2
6	2.3	0.02	2
7	3.9	0.22	2
8	6.2	0.37	2
9	6.9	0.12	2
10	7.3	0.17	2
11	8.6	0.32	2
12	9.6	0.18	2
13	10.4	0.30	2
14	11.2	0.20	2
15	12.4	0.30	2
16	13.3	0.20	2

ERA-Interim, the reanalysis of the global atmospheric dataset, has been released by the European Centre for Medium-range Weather Forecasts (ECMWF). The data have been continuously updated and provided since 1979 with one month to two months delay. The data consist of analysis fields provided four times (00, 06, 12, and 18 UTC) a day and forecasts fields provided with 3, 6, 9, and 12 h steps at 00 and 12 UTC [22]. ERA-Interim TPW covering both ocean and land with a spatial resolution of  $0.125^\circ \times 0.125^\circ$  (Table 2) are selected as a target variable corresponding to the input variables [23,24]. TPW of ERA-Interim can be downloaded from the ECMWF website (<https://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/>).

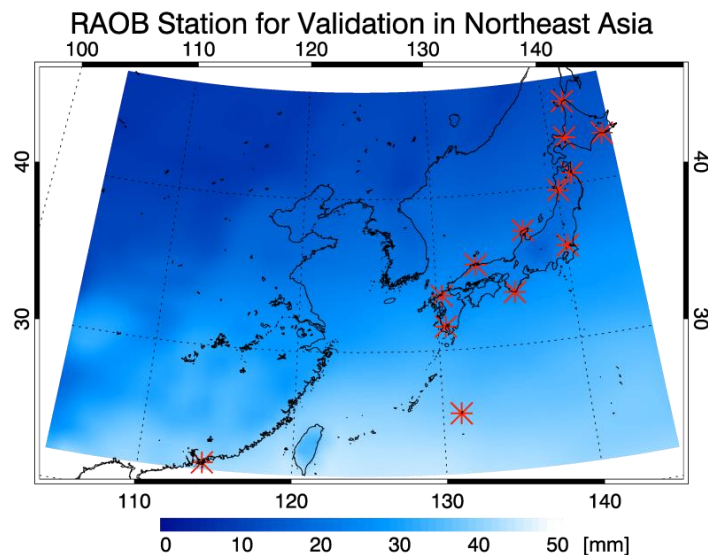
**Table 2.** Description of the reference data and usage.

Reference Data	Temporal and Spatial Resolution	Period and Usage
ERA-Interim	6 h/ $0.125^\circ \times 0.125^\circ$	September 2016–August 2018 Four days per month (5th, 10th, 15th, 20th) Training data (90%)/test data (10%) September 2016–August 2018 Two days per month (1st, 25th) Validation data
RAOB	12 h <sup>1</sup> / 13 sites over Northeast Asia	September 2016–August 2018 Two days per month (1st, 25th) Validation data

<sup>1</sup> The nominal time of a radiosonde launch is 0000 or 1200 UTC. Occasionally, this launches at 0600 and 1800 UTC.

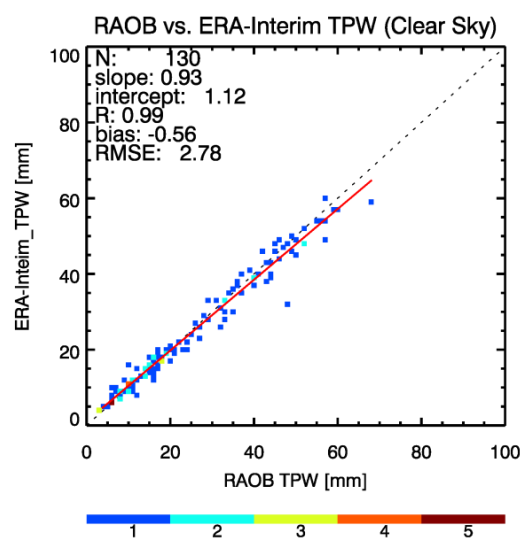
Radiosondes measure the state of the atmosphere at each altitude level and thus the radiosonde-derived water vapor content can be considered as true reference data. For the quantitative validation of the retrieved TPW, radiosonde observations (RAOB) from the University of Wyoming were used (Table 2). However, it should be noted that there might be uncertainty in RAOB, especially

humidity profiles, depending on the type of the sensor used. Most sensors used in the study are located in Northeast Asia and have high biases compared to the humidity profiles from ECMWF [25]. Thus, only the data from 13 observation stations (Figure 1) located in Northeast Asia (mainly in Japan) having the same sensor types (i.e., Vaisala RS92 and Meisei RS-11G) with generally good accuracy are used to assess the retrieved TPW. These radiosonde data are available on the website from the University of Wyoming (<http://weather.uwyo.edu/upperair/sounding.html>).



**Figure 1.** The red asterisks indicate 13 radiosonde observations (RAOB) sites with high accuracy for validation within the study area. The averaged ERA-Interim total precipitable water (TPW) from October 2016 to August 2018 was used as a background image.

For a comparison between two reference data (i.e., RAOB and ERA-interim TPW), ERA-Interim TPW was collocated at the grid-point near the 13 radiosonde stations in Northeast Asia during the validation period. The comparison result is shown in Figure 2. The error metrics were calculated from 130 collocated matches over clear sky regions. Specifically, the two reference data agree with each other with mean bias (ERA-Interim TPW—RAOB TPW) of  $-0.56$  mm and root mean square error (RMSE) of 2.78 mm.



**Figure 2.** Comparison between RAOB TPW and ERA-Interim TPW in Northeast Asia during the validation period (1st, 25th per month from September 2016 to August 2018). ERA-Interim TPW is collocated with the 13 RAOB stations over a clear sky region. Scatter plots are colored by density: the x-axis is RAOB TPW and the y-axis is ERA-Interim TPW. The red line represents a regression line.

### 3. Methods

#### 3.1. Preparation of Training Dataset

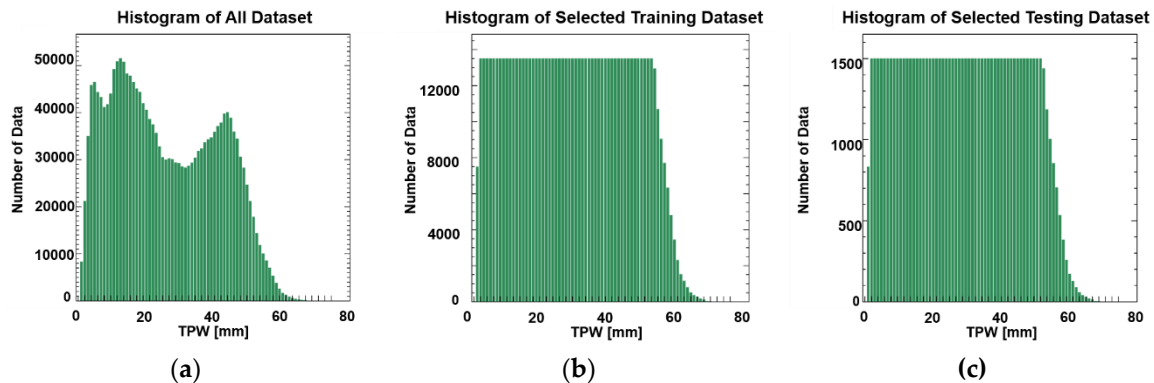
Himawari-8 AHI infrared BTs and ERA-Interim TPW were used as the input variables and the target variable of machine learning models, respectively. Table 3 shows the input variables used in model training. The channel centered at 3.9  $\mu\text{m}$  among the AHI IR channels was excluded from the input variables due to the contamination problem by solar radiation during the daytime [26]. The BT at each channel used as the training data contributes to the models at a different level. For example, the window channels (i.e., BT11, BT13, BT14, and BT15) are related to the temperature of land and sea surfaces, while the water vapor channels (i.e., BT8, BT9, and BT10) have characteristics related with the distribution of water vapor in three different vertical layers. BT12 and BT16 channels, which correspond to O<sub>3</sub> and CO<sub>2</sub> absorption bands respectively, are related to the information of atmospheric air mass [26]. In addition, dual channel differences (DCD), including the difference between the window channel and water vapor channel (i.e., DCD BT14–BT8, DCD BT14–BT9, DCD BT14–BT10), the difference between window channels (i.e., DCD BT14–BT11, DCD BT14–BT15), and the difference between water vapor channels (i.e., DCD BT10–BT8) carry information on the column water vapor amount. The DCDs tend to increase as TPW increases except for DCD BT14–BT15, which shows an opposite tendency.

**Table 3.** Summary of the physical characteristics of input variables in the retrieval models of TPW. Here, each brightness temperature (BT) of channels 8 to 16 are named BT8 to BT16, respectively.

Input Variable	Physical Characteristics
Cyclic_day	Temporal characteristics
Latitude	Spatial characteristics
Longitude	Spatial characteristics
Satellite zenith angle	Optical depth
BT8 (IR 6.2 $\mu\text{m}$ )	Upper tropospheric water vapor
BT9 (IR 6.9 $\mu\text{m}$ )	Mid and upper tropospheric water vapor
BT10 (IR 7.3 $\mu\text{m}$ )	Mid tropospheric water vapor
BT11 (IR 8.6 $\mu\text{m}$ )	Total water for stability, dust, SO <sub>2</sub> , rainfall
BT12 (IR 9.63 $\mu\text{m}$ )	Total ozone
BT13 (IR 10.4 $\mu\text{m}$ )	Surface
BT14 (IR 11.2 $\mu\text{m}$ )	Sea surface temperature and rainfall
BT15 (IR 12.4 $\mu\text{m}$ )	Total water and SST
BT16 (IR 13.3 $\mu\text{m}$ )	Air temperature
DCD BT14–BT8	Upper tropospheric moisture
DCD BT14–BT9	Mid and upper tropospheric moisture
DCD BT14–BT10	Mid tropospheric moisture
DCD BT14–BT11	Amount of water vapor
DCD BT14–BT15	Split-window channels (amount of water vapor)
DCD BT10–BT8	Difference between water vapor channels

Training data were collected in Northeast Asia region, four days a month (5th, 10th, 15th, and 20th of each month) and four times a day (00, 06, 12, and 18 UTC for each day) from September 2016 to August 2018. The study area over Northeast Asia centered at 37.588 N, 124.044 E (4300 km in E-W direction and 2900 km in the N-S direction) is covering South Korea, North Korea, Japan, Taiwan, and parts of China and Russia (Figure 1). To prevent cloud contamination, cloudy pixels were removed based on the empirical cloud mask algorithm proposed by Lee et al. (2019) [27]. When 5 × 5 neighboring pixels of a grid of target data (ERA-Interim TPW) are all clear, the AHI data are averaged to consider different spatial scales. The construction of the representative training data is crucial to develop successful retrieval models using machine learning [28–30]. Since the raw data have a skewed distribution toward low TPW (Figure 3a), it is necessary to adjust them to have a balanced distribution to avoid biased training [30]. Thus, the original dataset was randomly divided into the training dataset

(80%, 698,033 samples) and the testing dataset (20%, 174,510 samples) with the same number of data for each bin (i.e., 1 mm in TPW) as shown in Figure 3b,c. Here, each subset has the same distributions of the original dataset for the balanced samples. For the validation, data that are not used for the training are selected from two days a month (1st and 25th of each month), four times a day (00, 06, 12, and 18 UTC for each day) from September 2016 to August 2018.

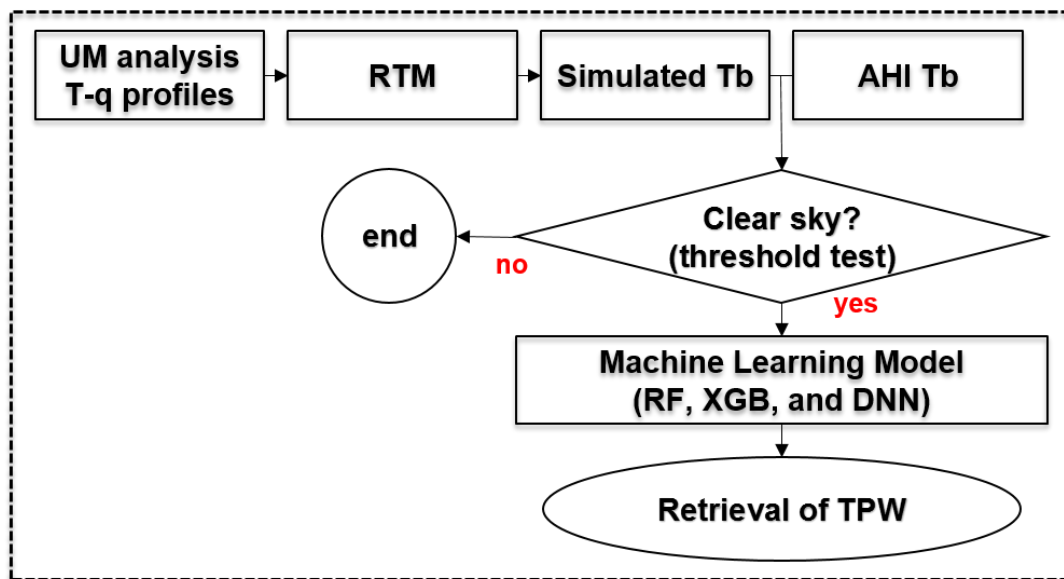


**Figure 3.** Histograms of training data for machine learning in Northeast Asia for two years (i.e., September 2016 to August 2018). The x-axis is the TPW range (0–80 mm) and the y-axis is histogram density. The bin size is 1 mm. (a) Histogram of all TPW during the period, (b) histogram of the selected TPW (13,500 datasets per bin size) for training, and (c) histogram of the selected TPW (1500 datasets per bin size) for testing.

### 3.2. Machine Learning Approaches

Machine learning-based approaches have been widely applied in the field of remote sensing for both classification and regression [4,6,19,20,23,24,31–47]. Here, three machine learning models (i.e., RF, XGB, and DNN) for TPW retrieval using AHI data are assessed. The process flow of the TPW retrieval based on machine learning algorithms is illustrated in Figure 4. For the cloud screening, the radiative transfer model (RTM) BT simulation was conducted and here, the Radiative Transfer for TIROS Operational Vertical Sounder (RTTOV) version 11.2 [48] was used to simulate BT from the unified model analysis data. Through an empirical thresholding test [27] using the simulated BT and the observed AHI BT, the clear-sky data were selected. If data were cloud free, TPW was retrieved by the machine learning models.

RF is an ensemble of rule-based algorithm based on classification and regression trees (CART) [49]. RF has been widely applied to various fields such as remote sensing and geographic information science [31–38]. Each independent tree in RF is created by the randomly selected subsets of training samples and input variables. The results from a multitude of independent trees are averaged to produce the final output of RF. For a fast implementation of RF in the R environment, a method of “ranger” is used to take advantage of the parallel processing of RF especially suited for the large and high dimensional dataset [50]. The RF has two basic model parameters to tune: the number of trees (num.trees) and the number of variables sampled in random when splitting at each node (mtry). The other parameters except for the two most representative parameters are used by applying default values. To find two optimum parameters, mtry is tested with 2, 4, 8, 10, and 19 and num.trees is tested with 100, 250, 500, and 1000. The parameter optimization is conducted based on the mean square error (MSE). The mtry and num.trees were determined at the 10 and 1000, respectively.



**Figure 4.** Process flow of the proposed machine learning (random forest (RF), extreme gradient boosting (XGB), and deep neural network (DNN))-based TPW retrieval models from AHI observations under clear-sky conditions.

Extreme gradient boosting (XGB) is a kind of tree-based boosting method, which is a sequential ensemble learning model of a decision tree. XGB weighs data with unexpected error from training to make the model better predict such data iteratively. The XGB based algorithm has been widely used in the latest studies [38,51,52]. The algorithm is highly effective in reducing the computing time and provides optimal use of memory resources because of parallel and distributed computing [51]. XGB has produced higher classification accuracy and faster execution time when compared to other models in several studies [41–43]. The XGB model is developed in the Python environment using XGBoost packages, here. For parameter optimization of the XGB model, Bayesian optimization is used as an effective tool for optimizing parameters in machine learning models [52]. Four parameters—number of trees used (`n_trees`), maximum depth of a tree (`max_depth`), minimum loss reduction required to make a further partition on a leaf node of the tree (`gamma`), and the subsample ratio of columns when constructing each tree (`colsample_bytree`)—were empirically tuned based on root mean square error (RMSE). The optimum `n_trees`, `gamma`, `colsample_bytree`, and `max_depth` were 4553, 0.7, 1.0, and 10, respectively.

Artificial neural network (ANN) is a biologically inspired machine learning approach that consists of neurons showing particularly good performance at modeling the nonlinear relationships between input and target data using a backpropagation [53]. Backpropagation adjusts the connection strength (weight) between neurons by minimizing the prediction error iteratively. ANN has been used for various purposes in remote sensing fields [42,43]. DNN is a subset of ANN with multiple hidden layers. Here, the DNN is developed with “Keras” which is a high-level Python library for deep learning. The DNN model has parameters including the number of hidden layers and neurons, optimizer, and activation function. The DNN parameters were adjusted to achieve high performance for the prepared training dataset. The activation functions, i.e., linear, sigmoid, tanh, and rectified linear unit (ReLU), hidden layers with 1 to 4 and neurons with 5 to 30 in five intervals in each hidden layer were tested. Besides, several widely used optimizers (i.e., stochastic gradient descent, rmsprop, and Adam) that update the weights in the direction that error decreases were tested by comparing the calculated results with the target value per iteration. The other parameters of the DNN are set as follows: `batch_size` = 256, `dropout_rate` = 0.5, `stop_steps` = 100 (if there is no improvement within `n` steps, training will be terminated), and learning rate = 0.001. Through the empirical testing, ReLU was set as the activation

function with the number of the hidden layers = 4, and the number of the hidden neuron = 60 using Adam optimizer.

### 3.3. Accuracy Metrics

For the selection of model configuration, K-fold cross-validation is used [54]. This method has been widely used to estimate overall performance. When a specific value for k is chosen (here, k is 10), datasets are randomly and equally distributed into k groups. One group is the test fold and the (k−1) groups are the training folds. In total k-times validation, performance is calculated by using the different test folds for each validation. Finally, the averaged validation results are used to tune the hyperparameters of each model. Typical accuracy metrics including correlation coefficient (R), mean bias, and RMSE are used (Equations (1) to (3), respectively). These statistical error metrics are calculated between the target TPW value (or reference data) and averaged TPW using the retrieved value based on the collocation criteria (Table 4).

$$R = \frac{\sum_{i=1}^n (A_i - \bar{A})(B_i - \bar{B})}{\sqrt{\sum_{i=1}^n (A_i - \bar{A})^2 \sum_{i=1}^n (B_i - \bar{B})^2}}, \quad (1)$$

$$\text{bias} = \frac{1}{n} \sum_{i=1}^n (A_i - B_i)^2, \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (A_i - B_i)^2}{n}}. \quad (3)$$

**Table 4.** Description of reference data and collocation criteria for the validation of the proposed models.

Reference Data	Temporal Resolution	Spatial Resolution	Collocation Criteria
ERA-Interim	6 h	0.125° × 0.125°	Averaging one or more retrieved TPW within a 0.1-degree radius
RAOB	12 h <sup>1</sup>	13 sites over Northeast Asia	Averaging one or more retrieved TPW within a 0.1-degree radius

<sup>1</sup> The nominal time of a radiosonde launch is 0000 or 1200 UTC. Occasionally, this launches at 0600 and 1800 UTC.

Additionally, to determine the contribution of each input variable to the performance of the three models, relative variable importance indicating how much a given model uses that variable to make accurate predictions is analyzed. To measure variable importance in a tree-based model for regression, the impurity reductions are summed over all split nodes in the tree [43]. In contrast to tree models, there is no clear way to assess the variable importance in the DNN model. For the same criterion for relative importance, the RMSE difference between the results obtained using all variables and the results calculated through the ‘leave-one-variable-out’ method is utilized. This method starts with all variables and removes each variable iteratively to determine the performance and robustness of the model [55].

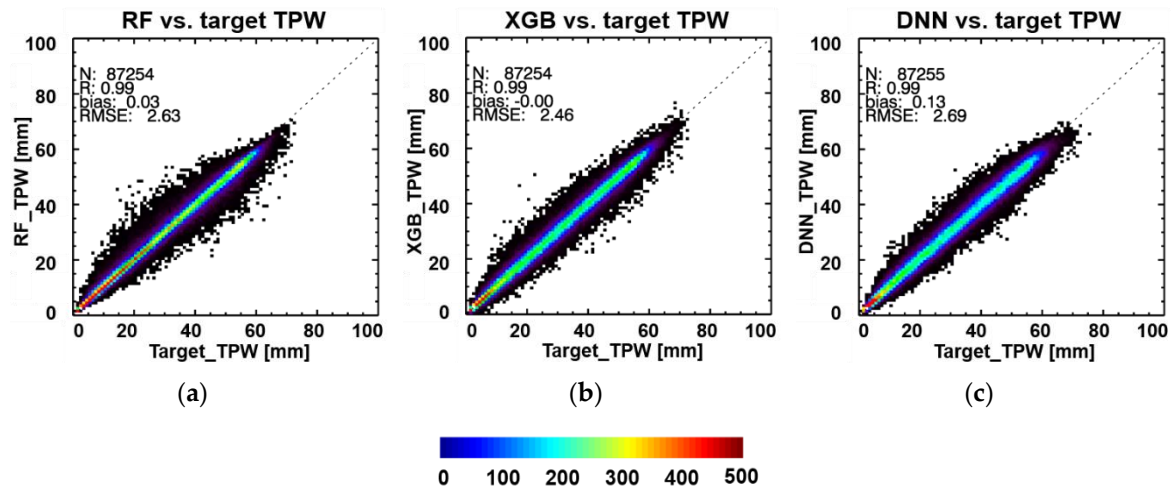
## 4. Results and Discussion

### 4.1. Model Performance

Figure 5 illustrates the averaged model performance using 10-fold cross-validation with the accuracy metrics (Table 2). The XGB model shows the highest accuracy with the RMSE of 2.46 mm (Figure 5b), followed by RF with 2.63 mm (Figure 5a) while the DNN model results in relatively less accurate performance with the RMSE of 2.69 mm (Figure 5c). The mean biases are under the absolute value of 0.15 mm for all models (0.03 mm for RF, 0.00 mm for XGB, and 0.13 mm for DNN) and the



correlation coefficients are 0.99 in all models. In all statistical results, the RF and XGB models are very similar and show higher performance than the DNN model. These model performances using the test dataset imply that the rule-based algorithms (i.e., RF and XGB) are more suitable for the retrieval of TPW. Overall, the XGB model outperforms the RF and DNN models. This might be because the XGB models highly weigh the weak learner.



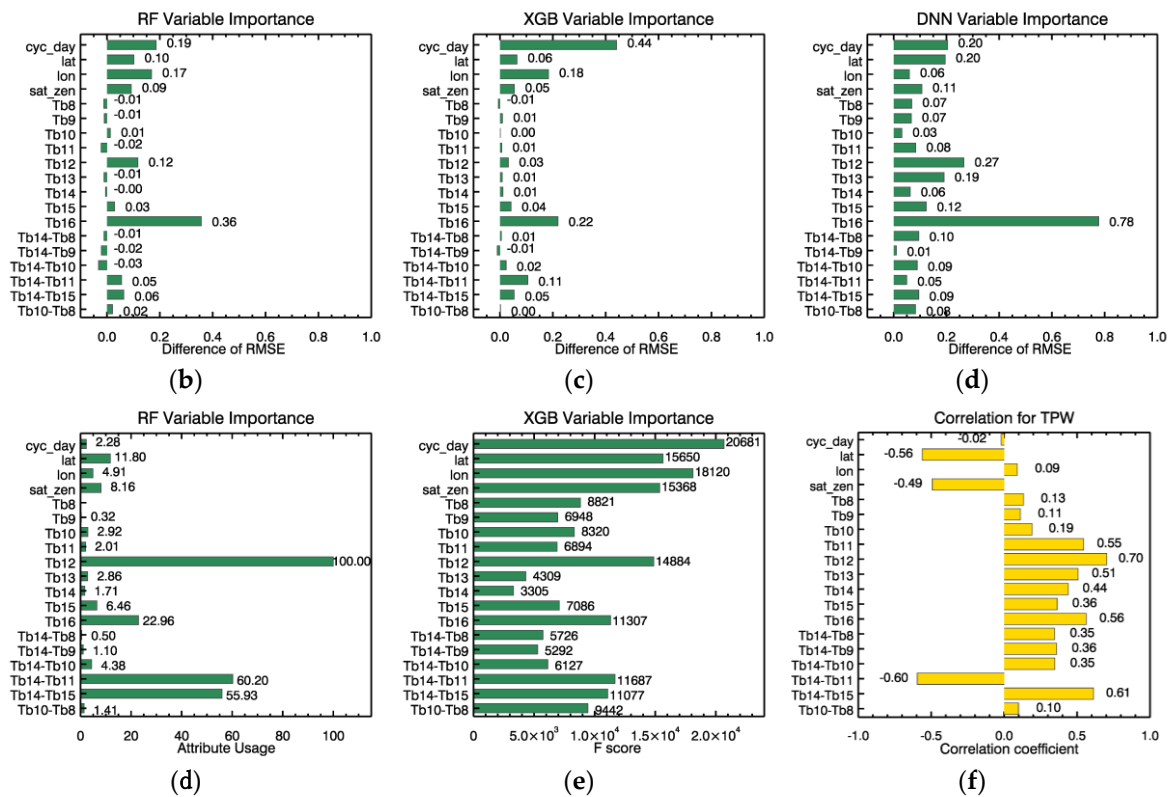
**Figure 5.** Model performance of machine learning models (RF, XGB, and DNN) using the test dataset. Scatter plots are colored by density: the x-axis is the target TPW (ERA-Interim TPW) and the y-axis is the retrieved TPW using each machine learning model (a) RF, (b) XGB, and (c) DNN.

#### 4.2. Variable Importance

Figure 6a–c shows the summary of calculated variable importance results of the RF, XGB, and DNN. BT16 is identified as the most contributing variable to the retrieval of TPW in all models while the cyc\_day and longitude are considered next significant in the RF and XGB and BT12, cyc\_day, and latitude are considered next significant in the DNN. The BT16, which is a sensitive channel to carbon dioxide, is used for the estimation of mean tropospheric air temperature. The cyc\_day reflects natural variations of TPW considering seasonal characteristics [56]. The longitude and latitude represent the spatial distribution of TPW. The BT12 is a channel sensitive to water vapor as well as ozone. Unlike the variable importance results identified in the DNN, some variables in the RF and XGB show improved accuracy when they are excluded (i.e., negative RMSE difference in Figure 6a,b). This implies that the variables might have redundant information and not be necessary for the RF and XGB to produce the best performance [38].

The ensemble tree models (RF and XGB) provide variable importance during model training, while DNN does not provide such information. Figure 6d,e shows the variable importance identified by using the provided library from RF and XGB. While the BT12 and the difference of window channels (i.e., BT14–BT11 and BT14–BT15) are identified to be very significant for the RF model, the cyc\_day and longitude are used as the most important variables in the XGB model. The BT14–BT11 and BT14–BT15, differences between two window channels, imply that the amount of atmospheric water vapor calculated by the difference in absorption coefficients is directly related to TPW. While the provided XGB variable importance (Figure 6e) has a similar pattern to the variable importance in terms of RMSE difference (Figure 6b), in RF, the rank of variable importance shows different patterns among the variables. This might be because the optimized RF model learns using randomly collected variables (here, mtry is 10) at each split time unlike XGB (1.0 for colsample\_bytree) and DNN (use all variables). To identify the linearity of the input variables to TPW, the correlation coefficients between TPW and each input variable (Figure 6f) are examined. If a high correlation variable has a high value in the variable importance identified by the model, it might indicate that the linearity between the input variables and TPW is relatively strong. In contrast, the nonlinearity is strong in the opposite

case. Interestingly, despite the lowest correlation of *cyc\_day* and longitude with TPW, they have the highest variable importance in the XGB model. These results imply that the XGB model well utilizes the nonlinear characteristics of the variables when compared to the other two models.



**Figure 6.** Relative variable importance identified by (a) RF, (b) XGB, and (c) DNN. Bar plots using input variables of the training dataset: the x-axis is the difference of root mean square error (RMSE) and the y-axis is the corresponding name of input variables. The difference in RMSE (the original RMSE – new RMSE) is calculated using the leave-one-variable-out method. The original RMSE of RF, XGB, and DNN are 2.63, 2.46, 2.69 mm, respectively. The variable importance is calculated after the construction of ensemble trees for the rule-based models (d) RF and (e) XGB. The longer the length of the bar is, the more important the variable is. (f) summarizes correlation coefficients between the input variables and TPW.

### 4.3. Validation Results

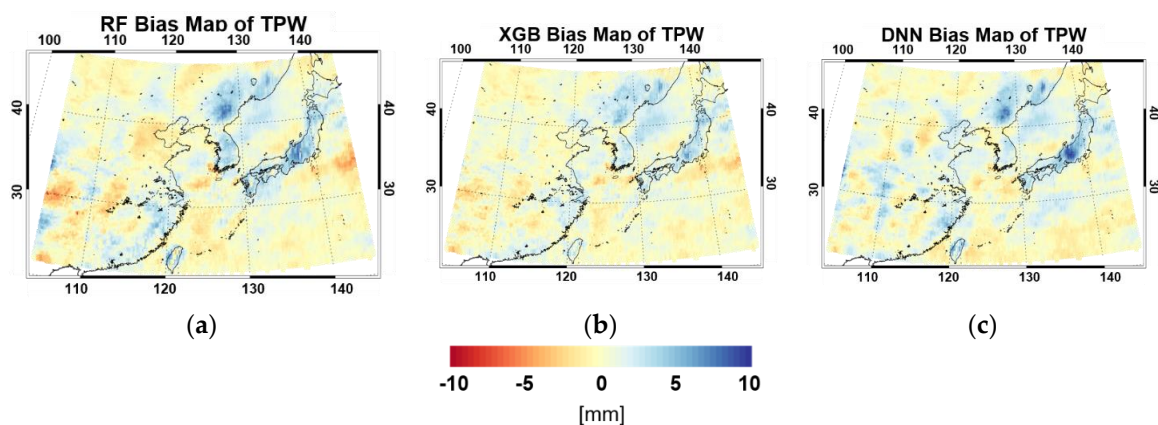
To validate the developed RF, XGB, and DNN models, the observed Himawari-8 AHI data that were not used for both the training and testing were utilized (validation dataset in Section 3.1). Table 5 shows the quantitative validation results between ERA-Interim TPW and the retrieved TPW from each model over Northeast Asia region during the validation period (Table 2). Approximately 500,000 collocated data are used for each model over the ‘all’, ‘land’, ‘sea’, and ‘coast’ regions as categorized in the table. In all areas, the DNN model yields the highest performance in terms of RMSE, followed by the XGB and RF models. The algorithms tend to overestimate TPW. To analyze the spatial distribution of the averaged errors (i.e., bias and RMSE), the retrieved TPW and the ERA-Interim TPW for two years (September 2016 to August 2018) are mapped over Northeast Asia (Figure 7). The mean biases of the RF, XGB, and DNN are about 1.60, 1.30, and 1.47 mm, respectively. The mean RMSE value is 5.02 mm for RF, 4.79 mm for XGB, and 4.56 mm for DNN. The spatial distributions of the averaged bias and RMSE show the characteristics of the discontinuity between land and sea. The lowest performance along the coastal regions (Table 5) can be attributed to these characteristics in the ERA-Interim grid. Additionally, the averaged errors of the RF, XGB, and DNN models appear relatively high in the black circled regions as shown in Figure 7g. These regions have relatively lower surface pressures when

compared to the surrounding areas and the models tend to overestimate TPW over these regions with low surface pressure. It is clear that the relatively lower surface pressure is related to the relatively higher terrain elevation and this is the main cause for the overestimated TPW in the region. Thus, further study will need to consider including elevation as one of the predictors. In the meantime, the XGB model has a lower bias in the region with lower surface pressure compared to the RF and DNN model since the XGB model learns repetitively to generate the weighted mean of weak learners [44].

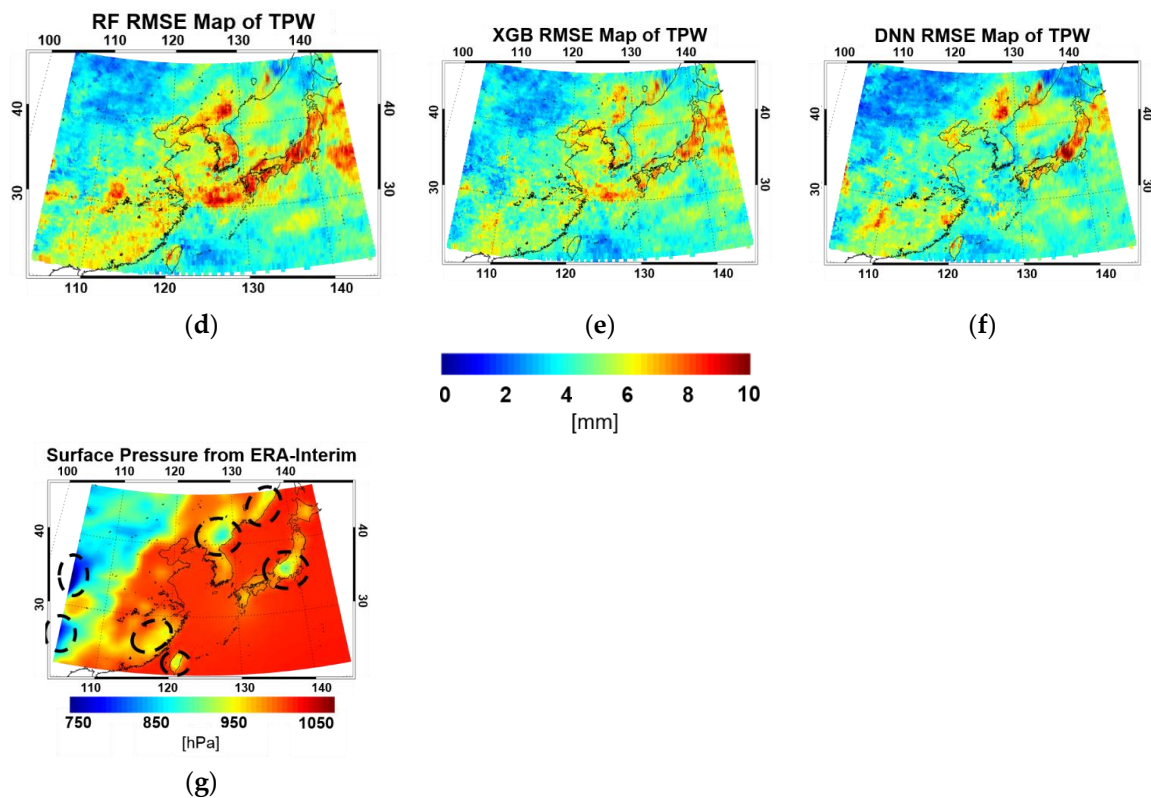
**Table 5.** Accuracy assessment based on ERA-Interim TPW for the RF, XGB, and DNN in Northeast Asia (1st, 25th per month from September 2016 to August 2018). Validation metrics (i.e., bias and RMSE) are calculated over the ‘all’, ‘land’, ‘sea’, and ‘coast’ regions, respectively.

	All		Land		Sea		Coast	
	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE
<b>RF</b>	0.62	5.09	0.66	4.94	0.55	5.29	0.88	5.98
<b>XGB</b>	0.70	4.85	0.67	4.63	0.72	5.14	1.04	5.65
<b>DNN</b>	0.90	4.65	0.94	4.50	0.83	4.86	1.20	5.22

In the validation results (Table 5), the overall accuracies of three models decrease compared to the results of the model performance as shown in Section 4.1. For example, the retrieved TPW using test datasets from all models agrees well with the ERA-Interim TPW overall both on land and sea within 0.15 mm in terms of bias while they have positive bias ranging from 0.62 to 0.90 in the validation results. The datasets randomly selected for the evaluation of the model performance can be different from those used for the validation of retrieval accuracy, and this discrepancy might be caused by overfitting since the optimal model is not chosen based on the final validation result. The model performance is used as an index to internally validate each model and to optimally tune the model that is sufficiently learned. The model accuracy needs to be evaluated with a dataset that is not used for the training or testing.

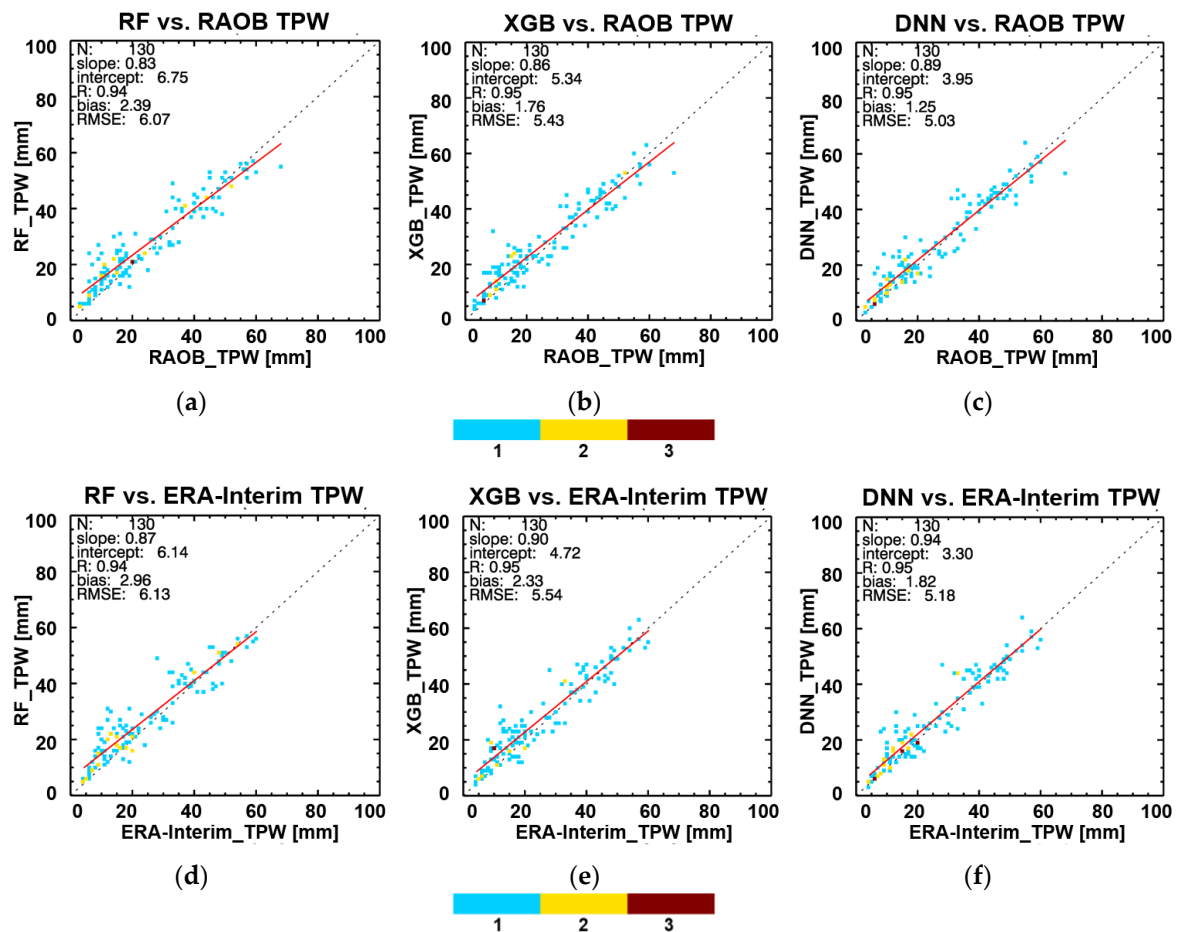


**Figure 7.** Cont.



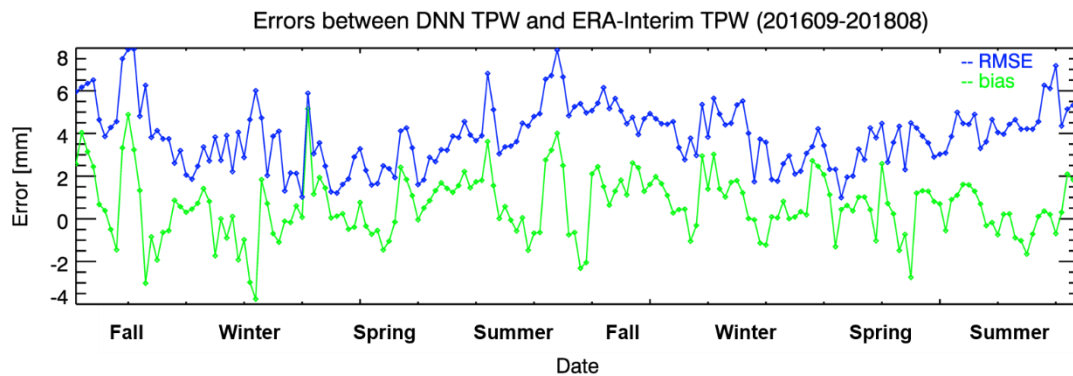
**Figure 7.** Mean error maps of TPW in Northeast Asia for two years (September 2016 to August 2018). The upper figures represent mean bias (ERA-Interim TPW – retrieved TPW) maps (a–c) and the lower figures represent mean RMSE maps (d–f). The left, middle, and right figures are the retrieved results from the RF, XGB, and DNN, respectively. (g) is the averaged surface pressures using ERA-Interim data from October 2016 to August 2018 in the study region. The black dotted circles indicate relatively lower surface pressures when compared to the surrounding areas.

To verify the performance of the models, a quantitative validation was carried out using RAOB from the University of Wyoming with an untrained dataset (Table 2). In addition to the ERA-interim data, RAOB data were also utilized for quantitative validation of model performance using an untrained dataset (Figure 8). Since only RAOB stations with high accuracy sensors were used, the number of in situ measurements used for the validation is relatively small (130 collocation data over the clear sky). The comparisons of TPW from the DNN, XGB, and RF models with RAOB and ERA-Interim show good performance in the order of RMSE during the validation period. The results from all models yield positive bias (2.39, 1.76, and 1.25 mm, respectively). This implies that all models tend to overestimate the TPW values concerning RAOB, which is similar to the validation results using ERA-Interim. The biases (retrieved TPW – reference TPW) using RAOB are smaller than the bias using ERA-Interim. The averaged difference value (about 0.56 mm) coincides with the difference between the two reference data (Figure 2). Additionally, the RMSE over the coastal region is larger compared to other regions (Table 5) since RAOB is collected mainly over land and coastal locations and also over islands as shown in Figure 1. The DNN is identified as the most optimal model for the retrieval of TPW through both validation results using ERA-Interim and RAOB. The RMSE value is comparable to or even better than the performance of TPW retrieval based on optical sensor data. For example, the neural network model developed from the MODIS near-infrared data over Western Europe and western Africa shows a validation RMSE of 6.4 mm when compared to RAOB [19].



**Figure 8.** Accuracy assessment results of the retrieved TPW from the DNN, RF, and XGB using RAOB (a–c) and ERA-Interim (d–f) in Northeast Asia (1st, 25th per month from September 2016 to August 2018). The retrieved TPW from each model and ERA-Interim TPW is collocated with the 13 RAOB stations over clear sky regions. Scatter plots are colored by density: the x-axis is the reference data (RAOB and ERA-Interim) and the y-axis is the retrieved TPW from each model. The red lines are the regression lines.

Figure 9 displays the time series of the averaged mean bias and RMSE between the retrieved TPW from the DNN model and ERA-Interim TPW per scene between October 2016 and August 2018 in the Northeast Asia area. The variabilities of the mean bias and RMSE are relatively high during the humid summer season, followed by fall and spring. On the other hand, both errors are relatively low and consistently stable during the dry winter season. This is because humidity errors are larger under moist conditions [57]. As described earlier, the study utilizes the criteria based on the single channel or dual-channel differences to detect clouds and they may not effectively discriminate snow from clouds, causing a low retrieval rate in the Northern Hemisphere winter [27]. The decreased retrieval rate can lead to relatively poor performance of the retrieval algorithm in winter.



**Figure 9.** Time series (1st, 25th per month from September 2016 to August 2018) of bias and RMSE of the retrieved TPW from DNN for ERA-Interim TPW. The x-axis shows time and the y-axis shows the value of errors. The blue line is RMSE and the green line is bias.

#### 4.4. Novelty and Limitations

The machine learning approaches (RF, XGB, and DNN) for the TPW retrieval from Himawari-8 AHI data in Northeast Asia were compared and analyzed in terms of their spatial and temporal characteristics of performances. The DNN model shows an overall good agreement with both types of reference data (ERA-Interim and RAOB data) and the RF model showed the lowest performance. The variable importance of each variable was calculated using the ‘leave-one-variable-out’ method. The retrieved TPW, provided every 10 min with 2 km resolution at nadir, together with atmospheric stability indices such as the Lifted index or CAPE, plays a good predictor of severe weather phenomena in the pre-convective atmospheric condition [4]. In addition to 10 min of AHI observations, the rapid scanning mode (about 2 min) data are also available, which can be used to monitor details of the temporal changes of the atmosphere. This near-real-time monitoring can give a promising result for the severe weather forecast for a more smooth transition of the atmosphere and information between cloudy scenes [8].

Aside from the novelties of this study, there are several limitations. One of the main limitations is the relatively low accuracy of the cloud mask used. This problem can cause uncertainty in the retrieval of TPW and validation results. As a result, there are little data during the wintertime especially over land, leading to relatively high RMSE of the retrieved TPW. Another limitation is the robustness of the model depending on the dataset. This is typically caused by overfitting [58] and makes the model difficult to directly apply to other cases. One possible solution to mitigate the difference between the accuracy of model tuning and validation is ‘online learning’ keeping the up-to-date dataset by constantly updating new data to the model [59]. This can help the model generalization since the difference between the training, testing, and validation dataset is decreased. Another solution is to combine different models. This is called ‘ensemble method’ to create a new model with a various model combination. This method has the advantage to complement the weaknesses of each other. It is important to choose the model considering the given problem.

## 5. Conclusions

In this study, TPW retrieval models based on the machine learning approaches (RF, XGB, and DNN) were developed for Himawari-8 AHI data over Northeast Asia. Nine AHI BTs (BT8 to BT16 centered at 6.2, 6.9, 7.3, 8.6, 9.6, 10.4, 11.2, 12.4, and 13.3  $\mu\text{m}$ ), six kinds of dual channel differences, time, location information (latitude and longitude), and satellite zenith angles were used as the input variables while the TPW calculated from the atmospheric temperature and humidity profiles from ERA-Interim were used as a target variable for the models. The parameters of each model were optimized through 10-fold cross-validation with the testing dataset (10% of the training dataset). The BT16, temperature sounding channel, is identified as the most contributing variable to the TPW retrievals in all models. The ERA-Interim and in-situ data (i.e., RAOB) are used for the model validation

and characterization of each model. The DNN model yields the highest accuracy metrics (R, mean bias, and RMSE) regardless of the reference data used (i.e., R is 0.96, mean bias is 0.90 mm, and RMSE is 4.65 mm when using ERA-Interim TPW and R is 0.95 mean bias is 1.25 mm, and RMSE is 5.03 mm with respect to RAOB) and the RF model showed the lowest performance. The distribution of spatially averaged errors of each model reveals that the TPW is overestimated particularly in the regions with relatively lower surface pressures, mainly due to the relatively high elevations of the regions. This suggests the importance of considering the elevation as a predictor in a TPW retrieval study using machine learning techniques. The validation results also show a decreased accuracy compared to the accuracy obtained during the model training, implying a tendency for the overfitting. Nevertheless, the retrieved TPW with finer spatiotemporal resolution (2-min intervals with about 2 km spatial resolution) together with atmospheric instability is expected to provide quite useful information for analyzing and predicting possible severe weather phenomena such as convective storms and heavy rainfall in pre-convective environments. In the future, the proposed models are going to be used for the TPW retrievals from Advanced Meteorological Imager loaded in the Geostationary Korea Multi-Purpose Satellite (GeoKompasat)-2A, a South Korea's second geostationary meteorological satellite, which has similar specifications to Himawari-8 AHI.

**Author Contributions:** Y.L. led manuscript writing and contributed to data analysis and research design. D.H. contributed to data analysis and model design. M.-H.A. supervised this study, contributed to the research design and manuscript writing, and served as the corresponding author. J.I. contributed to research design and manuscript writing. S.J.L. contributed to the discussion of the results.

**Funding:** This work was supported by the Development of Geostationary Meteorological Satellite Ground Segment (NMSC-2014-01) program, funded by the National Meteorological Satellite Centre (NMSC) of the Korea Meteorological Administration (KMA). It was also partially supported by a grant (2019-MOIS32-015) of Disaster-Safety Industry Promotion Program funded by Ministry of Interior and Safety (MOIS), Korea, and by the Ministry of Science and ICT (MSIT), Korea, under the Information Technology Research Center (ITRC) support program (IITP-2019-2018-0-01424) supervised by the Institute for Information & communications Technology Promotion (IITP).

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

## References

1. Trenberth, K.E.; Dai, A.; Rasmussen, R.M.; Parsons, D.B. The changing character of precipitation. *Bull. Am. Meteorol. Soc.* **2003**, *84*, 1205–1218. [[CrossRef](#)]
2. Viswanadham, Y. The relationship between total precipitable water and surface dew point. *J. Appl. Meteorol.* **1981**, *20*, 3–8. [[CrossRef](#)]
3. Manning, T.; Zhang, K.; Rohm, W.; Choy, S.; Hurter, F. Detecting severe weather using GPS tomography: An Australian case study. *J. Glob. Position. Syst.* **2012**, *11*, 58–70. [[CrossRef](#)]
4. Lee, S.J.; Ahn, M.H.; Lee, Y. Application of an artificial neural network for a direct estimation of atmospheric instability from a next-generation imager. *Adv. Atmos. Sci.* **2016**, *33*, 221–232. [[CrossRef](#)]
5. Bessho, K.; Date, K.; Hayashi, M.; Ikeda, A.; Imai, T.; Inoue, H.; Inoue, H.; Kumagai, Y.; Miyakawa, T.; Murata, H.; et al. An introduction to Himawari-8/9—Japan's new-generation geostationary meteorological satellites. *J. Meteorol. Soc. Jpn.* **2016**, *94*, 151–183. [[CrossRef](#)]
6. Martinez, M.A.; Velazquez, M.; Manso, M.; Mas, I. Application of LPW and SAI SAFNWC/MSG satellite products in pre-convective environments. *Atmos. Res.* **2007**, *83*, 366–379. [[CrossRef](#)]
7. Liu, Z.; Min, M.; Li, J.; Sun, F.; Di, D.; Ai, Y.; Li, Z.; Qin, D.; Li, G.; Lin, Y.; et al. Local Severe Storm Tracking and Warning in Pre-Convection Stage from the New Generation Geostationary Weather Satellite Measurements. *Remote Sens.* **2019**, *11*, 383. [[CrossRef](#)]
8. Lee, Y.K.; Li, J.; Li, Z.; Schmit, T. Atmospheric temporal variations in the pre-landfall environment of typhoon Nangka (2015) observed by the Himawari-8 AHI. *Asia-Pac. J. Atmos. Sci.* **2017**, *53*, 431–443. [[CrossRef](#)]
9. Lee, S.J.; Ahn, M.-H.; Chung, S.-R. Atmospheric Profile Retrieval Algorithm for Next Generation Geostationary Satellite of Korea and Its Application to the Advanced Himawari Imager. *Remote Sens.* **2017**, *9*, 1294. [[CrossRef](#)]
10. Wan, Z.; Dozier, J.A. Generalized split-window algorithm for retrieving land-surface temperature from space. *IEEE Trans. Geosci. Remote Sens.* **1996**, *34*, 892–905. [[CrossRef](#)]

11. Tang, B.; Bi, Y.; Li, Z.L.; Xia, J. Generalized split-window algorithm for estimate of land surface temperature from Chinese geostationary FengYun meteorological satellite (FY-2C) data. *Sensors* **2008**, *8*, 933–951. [[CrossRef](#)] [[PubMed](#)]
12. Chesters, D.; Robinson, W.D.; Uccellini, L.W. Optimized retrievals of precipitable water from the VAS “Split Window”. *J. Appl. Meteorol. Climatol.* **1987**, *26*, 1059–1066. [[CrossRef](#)]
13. Dalu, G. Satellite remote sensing of atmospheric water vapour. *Int. J. Remote Sens.* **1986**, *7*, 1089–1097. [[CrossRef](#)]
14. Sobrino, J.A.; Jimenez, J.C.; Raissouni, N.; Soria, G. A simplified method for estimating the total water vapor content over sea surfaces using NOAA-AVHRR channels 4 and 5. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 357–361. [[CrossRef](#)]
15. Schroedter-Homscheidt, M.; Drews, A.; Heise, S. Total water vapor column retrieval from MSG-SEVIRI split window measurements exploiting the daily cycle of land surface temperatures. *Remote Sens. Environ.* **2008**, *112*, 249–258. [[CrossRef](#)]
16. Barton, I.J.; Prata, A.J. Difficulties associated with the application of covariance–variance techniques to retrieval of atmospheric water vapor from satellite imagery. *Remote Sens. Environ.* **1999**, *69*, 76–83. [[CrossRef](#)]
17. Knabb, R.D.; Fuelberg, H.E. A comparison of the first-guess dependence of precipitable water estimates from three techniques using GOES data. *J. Appl. Meteorol.* **1997**, *36*, 417–427. [[CrossRef](#)]
18. Nielsen, M.A. *Neural Networks and Deep Learning*; Determination Press: San Francisco, CA, USA, 2015; Available online: <http://neuralnetworksanddeeplearning.com/> (accessed on 29 December 2017).
19. Wang, W.; Sun, X.; Zhang, R.; Li, Z.; Zhu, Z.; Su, H. Multi-layer perceptron neural network based algorithm for estimating precipitable water vapour from MODIS NIR data. *Int. J. Remote Sens.* **2006**, *27*, 617–621. [[CrossRef](#)]
20. Zhang, S.L.; Xu, L.S.; Ding, J.L.; Liu, H.L.; Deng, X.B. Precipitable Water Vapor Retrieval Using Neural Network from Infrared Hyperspectral Soundings. *Key Eng. Mater.* **2012**, *500*, 390–396. [[CrossRef](#)]
21. Lee, Y.-K.; Li, Z.; Li, J. Evaluation of the GOES-R ABI LAP Retrieval Algorithm Using the GOES-13 Sounder. *J. Atmos. Ocean. Technol.* **2014**, *31*, 3–19. [[CrossRef](#)]
22. Dee, D.P.; Uppala, S.M.; Simmons, A.J.; Berrisford, P.; Poli, P.; Kobayashi, S.; Andrae, U.; Balmaseda, M.A.; Balsamo, G.; Bauer, P.; et al. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **2011**, *137*, 553–597. [[CrossRef](#)]
23. Basili, P.; Bonafoni, S.; Mattioli, V.; Pelliccia, F.; Ciotti, P.; Carlesimo, G.; Pierdicca, N.; Venuti, G.; Mazzoni, A. Neural-network retrieval of integrated precipitable water vapor over land from satellite microwave radiometer. In Proceedings of the 2010 11th Specialist Meeting on Microwave Radiometry and Remote Sensing of the Environment IEEE, Washington, DC, USA, 1–4 March 2010; pp. 161–166. [[CrossRef](#)]
24. Bonafoni, S.; Mattioli, V.; Basili, P.; Ciotti, P.; Pierdicca, N. Satellite-based retrieval of precipitable water vapor over land by using a neural network approach. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 3236–3248. [[CrossRef](#)]
25. Ingleby, B. On the Accuracy of Different Radiosonde Types—ECMWF. TECO-2016 Madrid, 30 September 2016. Available online: [https://www.wmo.int/pages/prog/www/IMOP/publications/IOM-125\\_TECO\\_2016\\_Session\\_4/O4\(8\)\\_pres\\_Ingleby\\_TECO\\_types\\_4\\_8.pdf](https://www.wmo.int/pages/prog/www/IMOP/publications/IOM-125_TECO_2016_Session_4/O4(8)_pres_Ingleby_TECO_types_4_8.pdf) (accessed on 7 May 2019).
26. Ebell, K.; Orlandi, E.; Hünerbein, A.; Löhnert, U.; Crewell, S. Combining ground-based with satellite-based measurements in the atmospheric state retrieval: Assessment of the information content. *J. Geophys. Res. Atmos.* **2013**, *118*, 6940–6956. [[CrossRef](#)]
27. Lee, S.J.; Ahn, M.H.; Ha, S. Total Column Ozone Retrieval From the Infrared Measurements of a Geostationary Imager. *IEEE Trans. Geosci. Remote Sens.* **2019**, 1–9. [[CrossRef](#)]
28. Zhou, L.; Pan, S.; Wang, J.; Vasilakos, A.V. Machine learning on big data: Opportunities and challenges. *Neurocomputing* **2017**, *237*, 350–361. [[CrossRef](#)]
29. Karahoca, A. *Advances in Data Mining Knowledge Discovery and Applications*; InTech: Rijeka, Croatia, 2012; ISBN 978-953-51-0748-4. [[CrossRef](#)]
30. Yu, L.; Wang, S.; Lai, K.K. An integrated data preparation scheme for neural network data analysis. *IEEE Trans. Knowl. Data Eng.* **2006**, *18*, 217–230. [[CrossRef](#)]
31. Amani, M.; Salehi, B.; Mahdavi, S.; Granger, J.; Brisco, B. Wetland classification in Newfoundland and Labrador using multi-source SAR and optical data integration. *GISci. Remote Sens.* **2017**, *54*, 779–796. [[CrossRef](#)]



32. Liu, T.; Abd-Elrahman, A.; Morton, J.; Wilhelm, V.L. Comparing fully convolutional networks, random forest, support vector machine, and patch-based deep convolutional neural networks for object-based wetland mapping using images from small unmanned aircraft system. *GISci. Remote Sens.* **2018**, *55*, 243–264. [[CrossRef](#)]
33. Guo, Z.; Du, S. Mining parameter information for building extraction and change detection with very high-resolution imagery and GIS data. *GISci. Remote Sens.* **2017**, *54*, 38–63. [[CrossRef](#)]
34. Richardson, H.J.; Hill, D.J.; Denesiuk, D.R.; Fraser, L.H. A comparison of geographic datasets and field measurements to model soil carbon using random forests and stepwise regressions (British Columbia, Canada). *GISci. Remote Sens.* **2017**, *54*, 573–591. [[CrossRef](#)]
35. Forkuor, G.; Dimobe, K.; Serme, I.; Tondoh, J.E. Landsat-8 vs. Sentinel-2: Examining the added value of sentinel-2's red-edge bands to land-use and land-cover mapping in Burkina Faso. *GISci. Remote Sens.* **2018**, *55*, 331–354. [[CrossRef](#)]
36. Georganos, S.; Grippa, T.; Vanhuyse, S.; Lennert, M.; Shimoni, M.; Kalogirou, S.; Wolff, E. Less is more: Optimizing classification performance through feature selection in a very-high-resolution remote sensing object-based urban application. *GISci. Remote Sens.* **2018**, *55*, 221–242. [[CrossRef](#)]
37. Zhang, C.; Smith, M.; Fang, C. Evaluation of Goddard's LiDAR, hyperspectral, and thermal data products for mapping urban land-cover types. *GISci. Remote Sens.* **2018**, *55*, 90–109. [[CrossRef](#)]
38. Sonobe, R.; Yamaya, Y.; Tani, H.; Wang, X.; Kobayashi, N.; Mochizuki, K.I. Assessing the suitability of data from Sentinel-1A and 2A for crop classification. *GISci. Remote Sens.* **2017**, *54*, 918–938. [[CrossRef](#)]
39. Santos, L.D. GPU Accelerated Classifier Benchmarking for Wildfire Related Tasks. Ph.D. Thesis, NOVA University of Lisbon, Lisbon, Portugal, 2018. Available online: <https://run.unl.pt/handle/10362/61547> (accessed on 11 May 2019).
40. Nisa, I.; Siegel, C.; Rajam, A.S.; Vishnu, A.; Sadayappan, P. Effective Machine Learning Based Format Selection and Performance Modeling for SpMV on GPUs. In Proceedings of the 2018 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), Vancouver, BC, Canada, 21–25 May 2018; pp. 1056–1065.
41. Babajide Mustapha, I.; Saeed, F. Bioactive molecule prediction using extreme gradient boosting. *Molecules* **2016**, *21*, 983. [[CrossRef](#)] [[PubMed](#)]
42. Yang, J.; Guo, A.; Li, Y.; Zhang, Y.; Li, X. Simulation of landscape spatial layout evolution in rural-urban fringe areas: A case study of Ganjingzi District. *GISci. Remote Sens.* **2019**, *56*, 388–405. [[CrossRef](#)]
43. Omrani, H.; Tayyebi, A.; Pijanowski, B. Integrating the multi-label land-use concept and cellular automata with the artificial neural network-based Land Transformation Model: An integrated ML-CA-LTM modeling framework. *GISci. Remote Sens.* **2017**, *54*, 283–304. [[CrossRef](#)]
44. Antonanzas, J.; Urraca, R.; Aldama, A.; Fernández-Jiménez, L.A.; Martínez-de-Pisón, F.J. *Single and Blended Models for Day-Ahead Photovoltaic Power Forecasting*; Springer International Publishing: La Rioja, Spain, 2017; pp. 427–434, ISBN 978-3-319-59649-5.
45. Pan, B. Application of XGBoost algorithm in hourly PM2.5 concentration prediction. In Proceedings of the IOP Conference Series: Earth and Environmental Science, Harbin, China, 8–10 February 2018.
46. Just, A.; De Carli, M.; Shtein, A.; Dorman, M.; Lyapustin, A.; Kloog, I. Correcting Measurement Error in Satellite Aerosol Optical Depth with Machine Learning for Modeling PM2.5 in the Northeastern USA. *Remote Sens.* **2018**, *10*, 803. [[CrossRef](#)]
47. EUMETSAT. Product Tutorial on Total Precipitable Water Content Products. 2014. Available online: [http://www.eumetrain.org/data/3/359/print\\_3.htm#page\\_1.0.0](http://www.eumetrain.org/data/3/359/print_3.htm#page_1.0.0) (accessed on 5 May 2019).
48. Hocking, J.; Rayer, P.; Saunders, R.; Madricardi, M.; Geer, A.; Brunel, P.; Vidot, J. RTTOV v11 Users Guide. NWP SAF, Version 1.4. September 2015. Available online: [https://www.nwpsaf.eu/site/download/documentation/rtm/docs\\_rttov11/users\\_guide\\_11\\_v1.4.pdf](https://www.nwpsaf.eu/site/download/documentation/rtm/docs_rttov11/users_guide_11_v1.4.pdf) (accessed on 5 July 2019).
49. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
50. Wright, M.N.; Ziegler, A. Ranger: A fast implementation of random forests for high dimensional data in C++ and R. *J. Stat. Softw.* **2017**, *77*. [[CrossRef](#)]
51. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.

52. Dewancker, I.; McCourt, M.; Clark, S. Bayesian Optimization Primer. 2015. Available online: [https://app.sigopt.com/static/pdf/SigOpt\\_Bayesian\\_Optimization\\_Primer.pdf](https://app.sigopt.com/static/pdf/SigOpt_Bayesian_Optimization_Primer.pdf) (accessed on 11 May 2019).
53. Blackwell, W.J.; Chen, F.W. *Neural Networks in Atmospheric Remote Sensing*; Artech House: Norwood, MA, USA, 2009; p. 234, ISBN 978-1-59693-372-9.
54. Bengio, Y.; Grandvalet, Y. No unbiased estimator of the variance of k-fold cross-validation. *J. Mach. Learn. Res.* **2004**, *5*, 1089–1105.
55. Khalid, S.; Khalil, T.; Nasreen, S. A survey of feature selection and feature extraction techniques in machine learning. In Proceedings of the Science and Information Conference (SAI), London, UK, 27–29 August 2014.
56. Mears, C.A.; Santer, B.D.; Wentz, F.J.; Taylor, K.E.; Wehner, M.F. Relationship between temperature and precipitable water changes over tropical oceans. *Geophys. Res. Lett.* **2007**, *34*. [[CrossRef](#)]
57. Noh, Y.-C.; Sohn, B.-J.; Kim, Y.; Joo, S.; Bell, W. Evaluation of Temperature and Humidity Profiles of Unified Model and ECMWF Analyses Using GRUAN Radiosonde Observations. *Atmosphere* **2016**, *7*, 94. [[CrossRef](#)]
58. Bastani, O.; Ioannou, Y.; Lampropoulos, L.; Vytiniotis, D.; Nori, A.; Criminisi, A. *Measuring Neural Net Robustness with Constraints*; Curran Associates Inc.: Red Hook, NY, USA, 2016; pp. 2613–2621, ISBN 978-1-5108-3881-9.
59. Shalev-Shwartz, S. Online learning and online convex optimization. *Found. Trends<sup>®</sup> Mach. Learn.* **2012**, *4*, 107–194. [[CrossRef](#)]



© 2019 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 (<http://creativecommons.org/licenses/by/4.0/>).