

## Article

# An Ice Microphysics-Based Machine Learning Approach to Classify Precipitation Type over Land from Global Precipitation Measurement Microwave Imager (GPM-GMI) Measurements

Spandan Das<sup>1,†</sup>, Yiding Wang<sup>2,†</sup>, Jie Gong<sup>3,4,8,\*</sup>, Leah Ding<sup>2,\*</sup>, Stephen J. Munchak<sup>4,6</sup>, Chenxi Wang<sup>4,5,8</sup>, Dong L. Wu<sup>4</sup>, Liang Liao<sup>4,7</sup>, William Olson<sup>4,5,8</sup> and Donifan Barahona<sup>4</sup>

<sup>1</sup> Department of Computer Science, Carnegie Mellon University, Pittsburgh, PA

<sup>2</sup> Department of Computer Science, American University, DC

<sup>3</sup> GESTAR/Universities Space Research Association, Columbia, MD

<sup>4</sup> NASA Goddard Space Flight Center, Greenbelt, MD

<sup>5</sup> JCET/University of Maryland at Baltimore County, MD

<sup>6</sup> Now at Tomorrow.io

<sup>7</sup> GESTAR-II/Morgan State University, MD

<sup>8</sup> Now at GESTAR-II/University of Maryland at Baltimore County, MD

\* Correspondence: Jie.Gong@nasa.gov; Ding@american.edu

† These authors contributed equally to this work.

**Abstract:** Precipitation type is a key parameter used for better retrieval of precipitation characteristics as well as to understand the cloud-convection-precipitation coupling processes. Ice crystals and water droplets inherently exhibit different characteristics in different precipitation regimes (e.g., convection, stratiform), which reflect on satellite remote sensing measurements that help us distinguish them. The Global Precipitation Measurement (GPM) Core Observatory's Microwave Imager (GMI) and Dual-Frequency Precipitation Radar (DPR) together provide ample information on global precipitation characteristics. As an active sensor, DPR provides an accurate precipitation type assignment, while passive sensors like GMI are traditionally only used for empirical understanding of precipitation regimes. Using collocated precipitation type flags from DPR as the "truth", this paper employs machine learning (ML) models to train and test the predictability and accuracy of using passive GMI-only observations together with ancillary information from reanalysis and GMI surface emissivity retrieval products. Out of six ML models, four simple ones (Support Vector Machine, Neural Network, Random Forest, and Gradient Boosting) and the 1-D convolutional neural network (CNN) model are identified to produce 90% - 94% prediction accuracy globally for 5 types of precipitation (convective, stratiform, mixture, no precipitation, and other precipitation), which is much more robust than previous similar effort. One novelty of this work is to introduce data augmentation (subsampling and bootstrapping) to handle extremely unbalanced samples in each category. Careful evaluation of Impact matrices demonstrate that polarization difference (PD) and surface emissivity at high-frequency channels dominate the decision process, which are consistent with the physical understanding of polarized microwave radiative transfer over different surface types, as well as in snow and liquid clouds with different microphysical properties. Furthermore, the view-angle dependency artifact that DPR precipitation flag bears with does not propagate into the conical-viewing GMI retrievals. This work provides a new and promising way for future physics-based ML retrieval algorithm development.

**Keywords:** machine learning/artificial intelligence; precipitation type classification; passive microwave; precipitation radar; retrieval algorithm

## 1. Introduction

Surface precipitation comes with different dynamical and microphysical mechanisms. Take two major types of precipitation - convective and stratiform - as an example, convective precipitation is characterized by strong upward motion, high intensity, large hydrometeor particles and small areal coverage. On the other hand, stratiform precipitation usually

forms in relatively stable environment, low intensity, containing large snow aggregates, and large areal coverage. Observations and global model simulations suggest they contribute to about 60% and 40% of the global total precipitation amount, respectively (e.g., [24], [25]). There also exist mixtures of both precipitation types and precipitation that can be classified as neither of the two (such as anvil cloud precipitation).

Being able to discriminate between these two types of precipitation has many applications in weather forecasting and climate research. From the modeling perspective, separating precipitation types correctly can help better understand the diurnal cycles of each one, which allows us to conduct direct comparisons to global climate model simulations ([32]). From remote sensing point of view, a high-quality classification of precipitation scenes is the first step for precipitation rate retrieval from measurements. This is because different microphysics assumptions need to be made to convert the measured variables (e.g., radar reflectivity, radiance, etc.) to the physical quantity (e.g., precipitation rate). Both active and passive microwave remote sensing measurements are sensitive to cloud and precipitation microphysical properties. For active sensors such like the DPR onboard NASA's Global Precipitation Measurement mission's Core Observatory (GPM), the operational algorithm first determines whether its profile has precipitation hydrometeor or not, and then further determines if a brightband exists or not, which is equivalent for identification of stratiform precipitation. Different coefficients are then applied to retrieve precipitation rate as the former is a necessary input to the hydrometeor melting model ([6], [30]). For passive microwave (PMW) sensors such like the Microwave Imager (GMI) onboarding the same GPM core observatory satellite, although precipitation type is not yet in its GPM-GMI Radiometer Precipitation Profiling (GPROF) product yet, some recent research indicate that pre-grouping according to different weather systems or atmospheric stabilities (e.g., convective available potential energy, or CAPE) could potentially improve the precipitation rate retrieval ([28], [29]).

Machine learning/Artificial Intelligence (ML/AI) has gained a lot of attention over the last decade with the recent boom of big data. ML techniques are able to tackle higher-dimensional non-linear problems without requiring explicit supervisions. ML/AI approach has becoming trending in atmospheric science applications in recent years. This trend has been spread to the retrieval algorithm development for spaceborne passive sensors after the deployment of spaceborne active sensors such like cloud profiling radar (CPR) on CloudSat satellite, Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) on CALIPSO (Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations) satellite, and GPM-DPR. For example, [17] developed a random forest (RF) model for predicting MODIS ice cloud flag using collocated MODIS-CALIPSO observations as the training dataset and CALIPSO ice cloud mask as the "truth" flag. [18] recently developed a neural network (NN) based cloud-aerosol discrimination algorithm for Advanced Himawari Imager (AHI) on board Himawari-8 geostationary satellite using collocated AHI-CALIPSO observations as the training dataset and CALIPSO cloud/aerosol type mask as the "truth" flag. [19] developed an ice cloud a random forest model for predicting ice cloud flag from multiple passive microwave sensors using collocated PMW-CloudSat observations as the training dataset and CloudSat ice cloud mask as the "truth" flag.

Compared to cloud, aerosol and atmospheric gas that are ubiquitous globally, precipitation happens much less frequently. Treating the highly imbalanced training dataset imposes a challenge to this work. Similar effort has been carried out by [27] before using a deep neural network (DNN), where they arrived at 98% accuracy for stratiform precipitation prediction using GMI, but only 39% correction rate for convective precipitation prediction. Inspired by that pioneer study, this work evaluates extensively different ML models and different data augmentation techniques as well as different training attributes (or features). The goal is to generate a GMI-only precipitation type retrieval product that is not only consistent with DPR retrievals in 5 precipitation type classes (see the next Section for details of these 5 classes), but also consistent with physical understandings of radiative transfer processes at MW frequency.

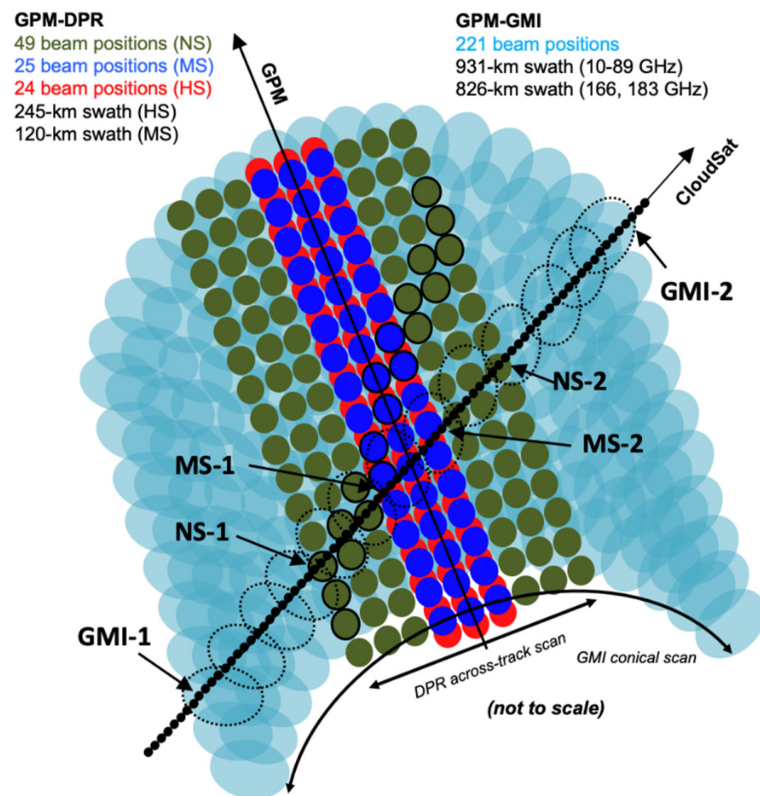
## 2. Data, Models and Methodology

### 2.1. Data and Pre-processing

Data used for the project comes from NASA's Global Precipitation Measurement (GPM) mission's Core Observatory. The Core Observatory has two instruments: GPM Microwave Imager (GMI) and Dual-frequency Precipitation Radar (DPR). As a passive sensor, GMI data makes up the features for training, while active sensor DPR provides the "truth" label for precipitation type classification. The goal is to develop a GMI-only precipitation type product.

Since they are on board the same satellite, GMI and DPR are always collocated and hence it's easy to produce large amount of training and validation data records. However, it is important to acknowledge the differences in swath width, viewing geometry and footprint size between them. As shown in Fig. 1, DPR's scan is cross-track, and has three modes, which are normal scan (NS; green), matched scan (MS; blue) and high sensitivity scan (HS; red). NS and HS are from Ku-DPR and Ka-DPR, respectively, while MS is the mode of dual-scan. Because of the Nyquist scan design, DPR footprint size is kept at  $\sim 5\text{km}$ . Correspondingly there are three sets of precipitation type flag products. For this current project, NS precipitation type retrieval product (Version 05B) is used as the "truth". The detailed retrieval procedure can be found in [4]. To summarize, A unified V- and H- method is employed for Ku-DPR precipitation flag assignment to make the best use of brightband feature and convective feature from the radar reflectivity profiles ([2]; [3]). Three types of precipitation flags are retrieved in the NS product: stratiform, convective and other. Although theoretically speaking, MS mode should produce the highest quality precipitation flag, we have three reasons to justify using the NS mode product. First, NS has the widest cross-track coverage ( $245\text{km}$  versus  $120\text{km}$  in HS/MS scan), hence produces more collocated samples with GMI. Secondly, the minimal detection threshold for Ku-DPR is  $15.46\text{ dBZ}$ , which is only slightly higher than  $13.71\text{ dBZ}$  threshold for MS ([6]). Thirdly, Ku-band is more sensitive to precipitation-sized hydrometeors than Ka-band. Its precipitation flag is hence likely more closely associated with ground precipitation per se. Nevertheless, we will extend similar strategy to HS/MS in the future.

GMI is a conical-scanning passive microwave (PMW) imager with a fixed scan angle of  $48.5^\circ$ . Each scan produces 221 footprints covering a swath width of  $931\text{ km}$  for low-frequency channels ( $10 - 89\text{ GHz}$ ) and  $826\text{ km}$  for high-frequency channels, as shown by the light blue ovals in Fig. 1. GMI's Level-1CR (Version 04) product contains calibrated and co-registered brightness temperature ( $T_c$ ) measurements which resolves the footprint size discrepancies between low-frequency and high-frequency channels ([7]). Therefore, in Level-1 CR data, the effective footprint size for GMI is  $4.4\text{km} \times 7.2\text{km}$ , and the swath width is  $826\text{ km}$ , which is still  $\sim 4$  times as wide as the NS scan. This means that GMI has a much higher chance in capturing a full weather system in its different evolution stages compared to DPR (Fig. 12 as an example).



**Figure 1.** Illustration of GPM-GMI, GPM-DPR and CloudSat radar viewing geometry. Adapted from Fig. 1 in [5]. See text for detail.

As shown in Fig. 1, each GMI footprint overlaps with multiple NS footprints. In this study, we further loosen the collocation and coincidence threshold to 6km and 90second to allow some fuzzy logic (FL) learning. A fourth precipitation type category called "mixed" is hence created when multiple DPR flags inside one GMI collocation footprint disagree with one another. Therefore, this category actually includes various scenarios (e.g., non-precip + convective, stratiform + convective, etc.). A further deep learning algorithm is in development to further learn how to predict sub-category flags for "mixed" precipitation scenes. In this study, we target at five main precipitation type categories, which are enumerated as (0) **non-precipitating**; (1) **stratiform**; (2) **convective**; (3) **other** and (4) **mixed**. One year of collocated GMI-DPR data (2016) is used for training and another year (2017) used for independent validation. Every three days are sampled for training and testing due to limited computational resources. As one can clearly see from Fig. 1, training is only performed on the narrow swath (green circles), while prediction happens for every GMI L1-CR footprint (blue ovals). Yet the training sample size is about 37.8 million, large enough to cover all weather scenarios. We then use the entire 2017 GMI-only prediction to study the climatology of precipitation flag distribution (Section 4). Since the training only utilized one year of data, we do not expect the prediction to capture the inter-annual variability, which is also beyond the scope of this research.

Although DPR is used as the "truth", it inherently has a view-angle dependent artifact at off-nadir views due to surface cluttering effect ([22]). This is a known issue for spaceborne radar. We will show in Section 3.4 that this artifact didn't propagate into GMI-only predictions because of its conical scan geometry. This new finding adds more merits to GMI-only predictions.

## 2.2. Data Augmentation

On top of using GMI's 13 channel brightness temperature ( $T_c$ ) measurements and their location (i.e., latitude/longitude) and month of the year in order to learn the seasonality



information, we also used some auxiliary information as input features, such like cloud liquid water path (CLWP), total column water vapor (TWV), and 2-meter temperature (T2m). They come from collocated and coincident MERRA-2 reanalysis dataset using nearest neighborhood method. There are two unique variables that are introduced in this study. They are polarization difference (PD, Section 2.2.1) and surface emissivity (Emis, Section 2.2.2), totaling 18 new input features. They are listed in Fig. 2. All variables are normalized and unitless before training.

Feature	Name	No. of Channels	Channel Info	Data Source	Reference
Tc	GMI brightness temperature	13	10V, 10H, 18V, 18H, 23, 36V, 36H, 89V, 89H, 166V, 166H, 183/3, and 183/7 GHz	L1-CR Observation	L1-C ATBD
*PD	GMI polarization difference	5	10, 18, 36, 89 and 166 GHz	L1-CR Observation	Gong and Wu (2017)
**Emis	Surface Emissivity	13	Same as first row	Retrieval	Munchak et al. (2020)
CLWP	Cloud liquid water path	1		MERRA-2	Auxiliary
TWV	Total column water vapor	1		MERRA-2	Auxiliary
T2m	2m temperature	1		MERRA-2	Auxiliary
Lat/Lon	Latitude/Longitude	2		L1-CR Observation	
Month	Month of the year	1			

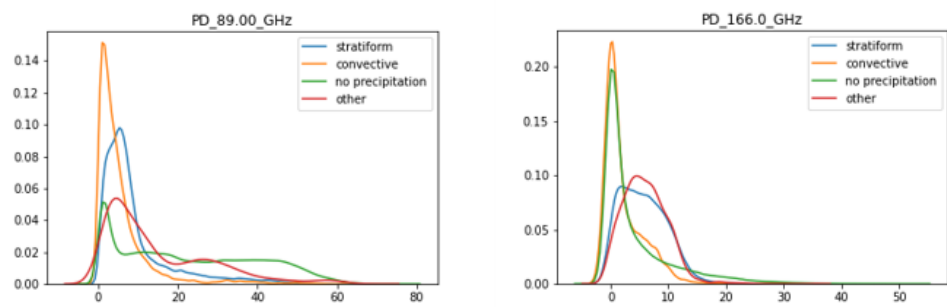
**Figure 2.** Features used as input (\*=hand-engineered feature; \*\*=retrieval using GMI TB[11]).

### 2.2.1. Polarization Differences

The main motivation behind creating hand-engineered features is to try to manually construct associations between variables which may prove to be useful in training the machine learning model. Polarization difference, or PD, is the most important hand-engineered feature for this training exercise. PD is defined as

$$PD = T_{cV} - T_{cH}|_{channel\ frequency}$$

where  $T_{cV}$  ( $T_{cH}$ ) is the Level-1CR brightness temperature observed at vertically (horizontally) polarized channel. For GMI, there are a total of 5 channel pairs that have PDs, which are 10.65, 18.7, 36.5, 89 and 166.5 GHz. An example of PD distribution collected from a random day is given in Fig. 3 for 89 and 166 GHz, where we saw considerable differences between distributions for different precipitation types. The disparity is especially pronounced between  $PD_{166GHz}$  distributions under stratiform (blue) and convective (orange) scenarios. [8] previously found that large PD values at these two channels with medium cold  $T_c$  values are explained by scattering from a layer composed of large horizontally oriented snow, which could only happen in the stratiform regime. [9] further identified PD's positive association with stratiform precipitation occurrence as well as precipitation strength. [12] in TRMM (Tropical Rainfall Measurement Mission; the predecessor of GPM mission) era has had explored the use of 85 GHz PD on detecting stratiform rain from TRMM-Microwave Imager (TMI; the predecessor of GMI) observations. As will be shown in Section 3.3,  $PD_{89GHz}$  and  $PD_{166GHz}$  are proved again from a ML/AI perspective to be substantial factors in accurate precipitation type prediction.



**Figure 3.** Global distribution of PDs by precipitation types at (left) 89 GHz, and (right) 166 GHz in a randomly sampled day in 2017.

### 2.2.2. Surface Emissivity

Surface Emissivity is a notoriously known headache to "contaminate" precipitation retrievals for PMW sensors as MW can penetrate through a top layer of land surfaces. It is particularly challenged to retrieve precipitation over snow-covered surfaces or certain desert surface because the signals they emit often exhibit similar features to precipitation ([33]). Even for ocean surface where emissivity is relatively well understood, physics-based emissivity models often require a variety of inputs (e.g., ocean surface wind) that are usually not readily available or of poor quality (e.g., reanalysis data).

GPM team currently uses an empirical surface emissivity model called TELSEM (the Tool to Estimate Land Surface Emissivities at Microwave frequencies) ([13]) for their precipitation product. This model uses a clustering approach to identify self-similar adjacent pixels on spatial and temporal grids from Special Sensor Microwave/Imager (SSM/I) observations to empirically retrieve surface emissivity from 19 to 89 GHz. For GMI, this means 10, 166 and 183 GHz emissivities are extrapolated; further, the spatial and temporal resolution is too coarse for GPM's precipitation retrieval.

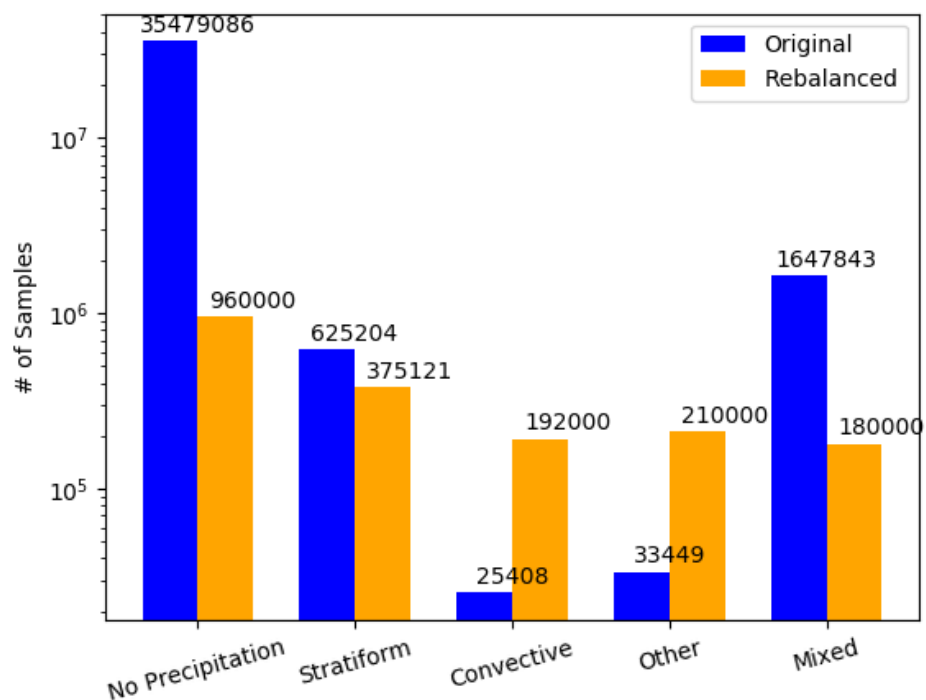
A recent work by [11] applied an optimal estimation (OE) approach to retrieve clear-sky emissivity from GMI clear-sky  $T_{cs}$ . The clear-sky flag is jointly determined by GMI, DPR and MERRA-2 data. This emissivity database provides pixel-by-pixel land surface (including snow-covered and sea ice-covered surfaces) emissivity retrievals at every GMI channel, and are hence easily adaptive features in our study. For precipitation and/or ice cloud affected pixels, a larger  $\chi^2$  value is produced that indicates a larger error in the emissivity retrieval. Liquid cloud also brings negative impact on the emissivity retrieval error, so MERRA-2 cloud liquid water path (CLWP) is also included in this product for users to screen out potential liquid cloud impacted scenes. Note that CLWP is included as one of the 37 features as shown in Fig. 2, which should work as a liquid-cloud screening flag within the machine learning process.

One could argue that this emissivity database is a repetitive use of  $T_c$  information and hence should not be included in the features. Our rationale to include emissivity is as follows. Firstly,  $\chi^2$  are not included in the input features, so we are not repetitively using any existing precipitation or cloudy-sky flag information that is otherwise indicated by large  $\chi^2$  values. Secondly, surface induced PD signals are often hard to be differentiated from cloud-induced PD signals, especially for low-frequency channels. Emissivity data over clear-sky pixels could help on differentiating them. We will elaborate more in Section 3.3 and Section 3.2 regarding reasons why emissivities are among the top factors in determining a precipitation flag. Furthermore, a sensitivity experiment is carried out in Section 3.2 to remove emissivity as the input feature. As will be shown in Section 3.2, removing emissivity has some marginal benefit for simple ML models to elevating the accuracy rate of "other" type but decreasing the accuracy of "convective" as a trade-off. For a more sophisticated model such like CNN, the impact is rather trivial.

### 2.2.3. Sample Balancing

A challenge we have in this work is that some classes (i.e., no precipitation and mixed) have a notably higher number of data samples than other classes. The total number of samples under each precipitation type is shown in Fig. 4 as blue bars (note that the vertical axis is in log-scale to amplify small sample size for "convective" and "other" types). This class imbalance can significantly compromise the process of learning because the model tends to focus on the prevalent class and to ignore the small classes with much less data samples. The scarcity of data from small classes results in poor estimates of the model's accuracy of those classes.

To address the class imbalance problem, we use the random majority undersampling method ([15]) to reduce the class size by randomly removing data samples from the big classes (i.e., no precipitation and mixed). In addition, we use a smoothed bootstrap re-sampling based technique ([16]) to rebalance the class distribution for this imbalanced dataset. Essentially, we draw from the training dataset an example from each small class (i.e., "convective" and "other") to generate a new example in its neighborhood, and retrain the models until the recall of the small class is maximized. The rebalanced sample size is shown as orange bars in Fig. 4. Note that this rebalancing process is tuned to holistically maximize the accurate prediction rate for all 5 classes for the CNN model, while it is not necessarily the best choice for other simple ML models.



**Figure 4.** Number of samples under each precipitation type from the training dataset from the original (blue) and after rebalancing (orange) for the entire training dataset.

### 2.3. Machine Learning Models

For this project, we trained and evaluated 6 different machine learning models, which are Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), Gradient Boosting (GB), Neural Network (NN) and Convolutional Neural Network (CNN). The former 5 models were trained, tuned, and evaluated using the Python scikit-learn package, while CNN is constructed using the Python tensorflow package. The working mechanism/philosophy behind these models are briefly described below.

The Support Vector Machine (SVM) model uses a subset of the training data, known as support vectors, and applies the combination of a kernel with a modified loss function to make predictions. The SVM was originally designed for binary classification, but can be extended to multi-class classification as in our case via the one-versus-the-rest approach to maximize the margin from each to the remaining classes. The Logistic Regression model uses a logistic function to model the features and class variables by minimizing the sum of the squares of the residuals for each data sample in the training set. With a goal to reduce the variance of the prediction, Random Forest, proposed in ([34]), trains different decision trees on different subsets of the training data, chosen randomly with replacement, and then computes the ensemble. It tries to decorrelate the base learners based on a randomly chosen subset of input variables. Such models often have very good predictive accuracy ([35]). Moreover, we trained a Multilayer Perceptron model, a class of feedforward neural networks that use back-propagation for training in this work. We also trained an XGBoost model, a parallel gradient-boosted decision tree model that uses decision trees as weak prediction models in gradient boosting ([36]).

Convolutional neural networks (CNN) are specific types of neural networks that contain convolutional layers. It is based on the idea of transforming input data with a set of differentiable operations with feature extraction, which allows them to learn meaningful representations of the data. In this work, we use a CNN composed of three types of layers: convolutional layer, max-pooling layer, and fully-connected layer. The high-dimensional input data is fed to the convolutional layers and max-pooling layers. The last pooling layer replaces the output of the network at certain locations by deriving a summary statistic of the nearby outputs. This helps in reducing the spatial size of the representation, which decreases the required amount of computation and weights. Then the low-dimensional features are fed to the fully connected layers. Neurons in this layer have full connectivity with all neurons in the preceding and succeeding layer as seen in regular feed-forward neural networks.

In recent years, ML approach is trending in atmospheric retrieval algorithm development, which can optimize the retrievable information content from joint observations taken at multiple platforms. RF and CNN are among the most popular ML models in application (e.g., [17], [18]).

### 3. Results

In this section, statistical metrics from the validation datasets are first presented, where two models with highest overall accuracy - GB and CNN - are selected for generating the final predictions. The final predictions from full-swath GMI observations are evaluated from multiple perspectives, including their consistency with physical understandings, their view-angle dependencies, their performance on a single weather event, and their climatological distributions.

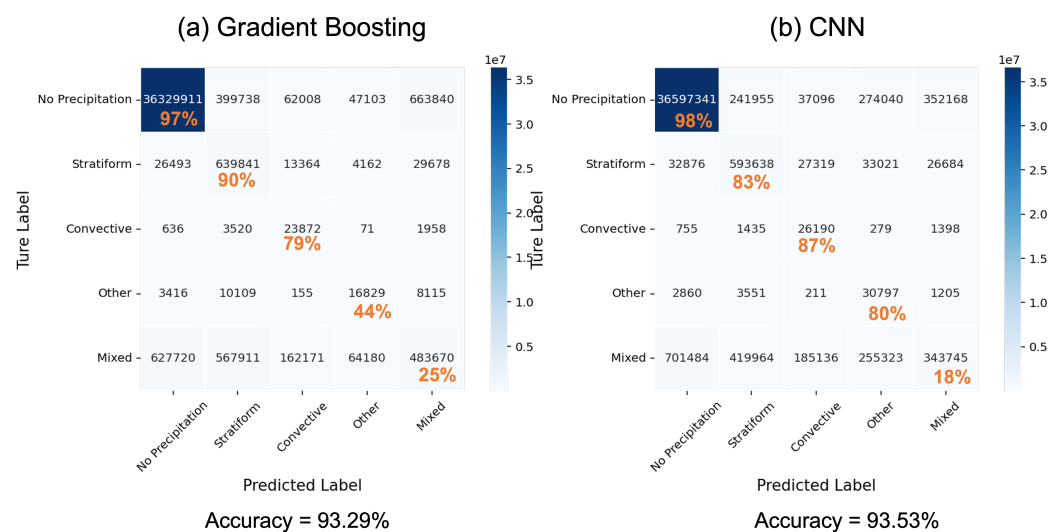
#### 3.1. Prediction Accuracy

Two metrics were used for evaluating the model performance, which are overall accuracy and area under the ROC curve (AUC ROC). The overall accuracy is the sample-size weighted averaged accuracy rate (%) from each of the five type of precipitation flag predictions. AUC ROC score is a metrics to better evaluate the performance of a classifier, especially for highly imbalanced data ([37]). The AUC ROC score for a classifier ranges from 0 to 1, where a score of 0.5 is given to a random classifier. From Fig. 5, we can clearly see that CNN produces overall the highest accuracy (93.53%). 4 out of the 5 relatively simple ML models achieve similar accuracy at  $\sim 90\%$ , while the LR model performance is the lowest. As GB model comes with explicit ranking of feature importance (Fig. 9) and it produces the second highest accuracy score, CNN and GB are selected for producing further statistics for comparison.

Classifier	Overall Accuracy (%)	AUC Score
Support Vector Machine (SVM)	91.15	N/A
Logistic Regression (LR)	76.07	0.8995
Gradient Boosting (GB)	93.31	0.9672
Random Forest (RF)	89.99	0.9594
Neural Network (NN)	93.56	0.9661
Convolutional Neural Network (CNN)	93.53	0.9678

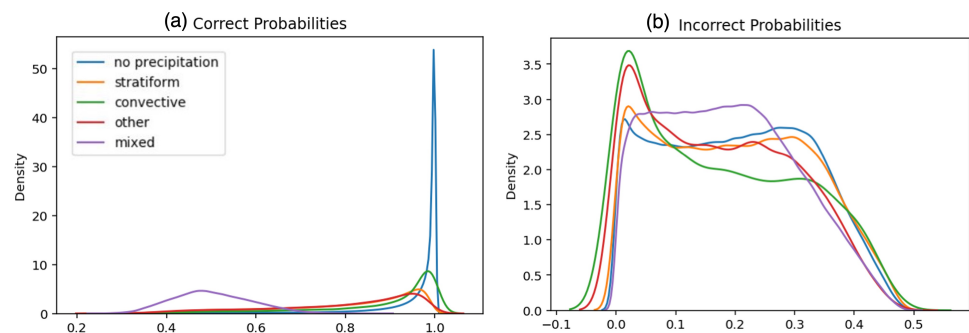
**Figure 5.** Comparison of prediction accuracy (%) and area under ROC curve (AUC ROC) from different ML/AI models tested. Two models that achieved the highest accuracies are highlighted in yellow, and all results shown below are from these two models.

Further evaluation of the distribution of accuracy and mis-classification rate in each precipitation category reveals that the accuracy rate is the highest for non-precipitation scenes (top left cells in Fig. 6). This is understandable from the physics point of view, as it's a relatively easier task to separate apart precipitating and non-precipitating scenes, especially considering that more than half of the latter contains clear-sky pixels that reflect by warmer  $T_{cs}$  at high frequency channels. This fact is also expected from computer science point of view, as ML model evaluations are carried out on sample-size weighted metrics in every layer, where non-precipitation pixels still occupies the majority of samples even after rebalancing (Fig. 4). We can also notice that although the deep learning model CNN produces only a slightly higher overall accuracy rate, it outperforms other simple ML models in the "other" precipitation category significantly.



**Figure 6.** Confusion matrices to show the numbers of samples that are accurately predicted (along the diagonal) or miss-classified using (a) GB model; (b) CNN model.





**Figure 7.** Distribution of probabilities for each precipitation type class for (a) correct predictions, and (b) incorrect predictions. Density corresponds to the absolute occurrence rate among all situations.

Among the four types of precipitation, three mono-type ones - stratiform, convective and other - have comparable accuracies (80-87%) for CNN. Fig. 6 only showcases results from GB and CNN, but RF and NN models produce similar statistics (not shown). The robustness of single-type precipitation accuracy using GMI-only data reveals that PMW like GMI can differentiate precipitation features at about 80-87% of the scenes, as long as the footprint is occupied by the same type of precipitation. Hence, it is important to have small footprint size for PMW sensors, especially for convective precipitation scenes as the area for convective precipitation is usually small and highly inhomogeneous. For the rest 16 - 20% scenes that are mis-classified, features from them are likely ambiguous and a confident classification cannot be made. For example, mis-classified convective scenes are often assigned to "stratiform" or "mixed" classes, which likely happen at the boundary of convective cell and stratiform layer for weather systems like mesoscale convective systems (MCSs) or winter frontal systems, while it's unlikely to be assigned to the "other" type, as "other" precipitation is mainly associated with anvil clouds that are high in altitude and relatively thin in thickness, which hence usually produces a less depression in the  $T_c$  signal for GMI channels ([9]).

"Mixed" type is the hardest to predict. This is mainly because we assign all GMI footprint with multiple DPR precipitation flags to this category, which is apparently not an optimal way for classification. The confidence of prediction in this category is also the lowest (Fig. 7). Even if the prediction was correct for this class, the confidence is still very low (peaking at 0.5, or 50% probability) and essentially in-separable from the confidence level for a wrong prediction (10-30% probability). We are currently exploring the optimal way of sub-grouping under this "mixed" class.

To summarize, GMI-only prediction works exceptionally well at distinguishing non-precipitation and precipitation scenes (>97% accuracy rate for predicting a non-precipitating scene). For precipitation scenes, GMI-only prediction works for 80 - 87% of the cases when its footprint is occupied by one single type of precipitation if a CNN model is employed. It works poorly when the footprint has large in-homogeneity or at the transition boundary between two major types. The latter point will be demonstrated later by a weather event study in Section 4.

### 3.2. Sensitivity to Surface Emissivity

As emissivity is a new hand-engineered feature included in the training, it is necessary and beneficial to test how sensitive our trained models are with respect to these variables. The confusion matrices for GB and CNN models without emissivity are shown in Fig. 8. The overall accuracy change for every of the 6 ML models is less than 1% (only GB, RF and CNN results are shown). However, we can see from the statistics that excluding emissivity particularly further downgrades the capability at differentiating the signature for the "mixed" class category. For CNN, prediction accuracy for all rest 4 mono-type classes remains robust. Interestingly, however, for the other 5 simple ML models, prediction for

the "other" type is significantly improved if surface emissivity database is excluded for the training, while the rest 3 mono-types remain robust.

Classifier	Non-Precip (%)	Stratiform (%)	Convective (%)	Other (%)	Mixed (%)	Overall Accuracy (%)
GB + emis	97	90	79	44	25	93.29
GB - emis	97	87	83	76	20	92.78
RF + emis	92	85	74	43	45	89.99
RF - emis	94	86	73	66	36	91.29
CNN + emis	98	83	87	80	18	93.53
CNN - emis	97	86	86	80	15	92.68

**Figure 8.** Accuracy rate (%) with including emissivity and without emissivity in the training process for each precipitation class for GB, RF and CNN model from independent validations.

Although it is a simple sensitivity test and we didn't try to exclude emissivity at each channel one-by-one, the results are indicative of at least two important implications. Firstly, the fact that all 6 ML models remain robust in producing a high accuracy rate for "non-precipitation", "stratiform" and "convective" classes indicates that these three mono-types are well separated in terms of  $T_c$  features in the multi-dimensional space. Such a separation is hard to model with radiative transfer models (RTMs) as hydrometeor scattering and absorption involves great variabilities and uncertainties; instead, even simple ML models can capture the feature separation easily. Second of all, the benefit of including surface emissivity in the training process is rather hard to determine. There is enough hint that including surface emissivity helps better identify the "mixed" class from the other types. However, the degradation and sensitivity of the "other" class from all 5 simple ML model to surface emissivity features suggests that the emissivity product itself might not be good enough to separate light precipitation against surface emissions. The deep learning CNN model is the only one among the 6 ML models that stays robust against adding or removing emissivity in the training features.

3.3. Rank of Importance and Corresponding Physics Mechanisms

As described in Section 2.2, one of the most important novelties of this work is to include PDs and retrieved surface emissivity for training and prediction, both of which come with solid radiative transfer theoretical basis. In this section, we demonstrate from ML point of view how its results are consistent with the physical basis. This is a critical step to build-up physics-consistent retrieval algorithms using ML/AI approach in the future.

For simple ML models like GB and RF, the rank of feature importance is provided together with the trained models. Feature importance refers to a class of techniques to assign a score to each input feature that indicates the marginal importance of each feature when a prediction is made. RF model is essentially composed of a set of decision trees. For each decision tree, it is made from a set of internal nodes and leaves. The nodes determines how to divide the dataset into two different branches. The internal nodes are the most important features, and ranking of the feature importance helps elucidating the important factors in the decision (or voting) process. GB model employs a similar decision tree process, except it tries to minimize the cost function in each layer, rather than voting. This feature does not present for NN or CNN models. Here only the top 15 most important features are listed in Fig. 9 for GB and RF models, with emissivity included in the training (+emis) and without (-emis).

Feature Importance Rank	GB + emis	RF + emis	GB - emis	RF - emis
1	Emis_166H	Emis_10V	CLWP	TWV
2	Emis_10V	Emis_10H	Tc_166H	Ts
3	CLWP	Emis_166H	Tc_166V	Tc_89V
4	Emis_10H	PD_89	Tc_36V	Tc_183/7
5	Tc_166H	Tc_89V	Tc_89V	CLWP
6	Emis_89H	CLWP	TWV	PD_89
7	TWV	Emis_89V	Tc_183/7	Tc_166H
8	PD_89	TWV	PD_89	Tc_166V
9	Tc_89V	Ts	Tc_183/3	Tc_183/3
10	Tc_166V	Tc_183/7	Ts	Tc_89H
11	Emis_89H	Emis_166V	PD_166	PD_166
12	Ts	Emis_183/7	Tc_89H	Tc_36V
13	Emis_36V	Emis_183/3	Tc_10V	Tc_24
14	PD_166	PD_166	Tc_36H	Tc_10V
15	Tc_166H	Tc_166H	Tc_24	Tc_36H

**Figure 9.** First 15 most important features listed for the GB and RF models with (+emis) and without (-emis) emissivity in the training process. High-frequency channel  $T_c$ s are highlighted by orange letters, and  $PD$ s are further highlighted by yellow backgrounds.

The first difference between "+emis" and "-emis" experiments we can observe is the dominance of surface emissivities at 10, 89 and 166 GHz in the rank of importance if emissivity is included. In the meantime,  $T_c$  values at 89 and 166 GHz also play some of the most dominant roles across the board. This indicates that there is a redundancy of input information at least for 89 and 166 GHz. The original  $T_c$  observations at these two channels contain enough distinctive surface emission and cloud scattering/emission signatures that can help us separate the 5 precipitation classes. When emissivity is removed from the input features,  $T_c$  at 183/3 and 183/7 GHz channels show up among the tops. The high-frequency channels ( $> 85$  GHz) are much more sensitive to cloud ice and snow scattering rather than rain/liquid emission ([38]). In contrast, classical PMW rainfall or liquid water path (LWP) retrieval channels (18 - 36 GHz; [20],[21]) are not highly ranked in any of these models. All these findings strongly suggest that the precipitation type is more closely related to the cloud regime (frozen or liquid) above rather than surface precipitation characteristics.

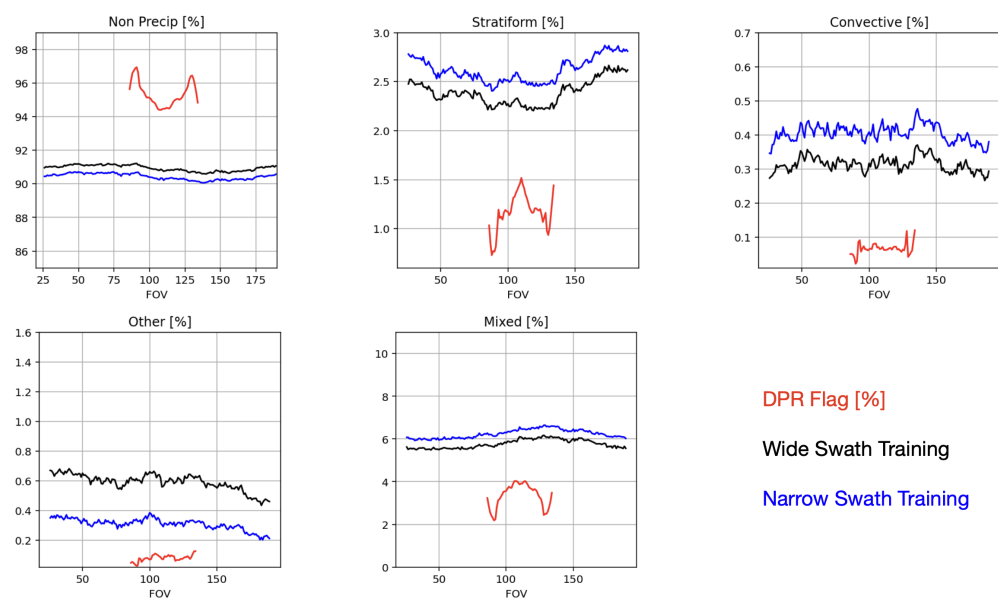
The second commonality we can easily observe is that  $PD$  at 89 and 166 GHz are ranked highly on the list for all four experiments, especially when surface emissivities are removed from the training. Furthermore,  $PD_{89\text{GHz}}$  ranks higher than  $PD_{166\text{GHz}}$  across the board. 85 - 89 GHz is a band that is sensitive to both liquid water emission and ice scattering. This complicated non-linear feature making it particularly challenging to simulate cloudy-sky  $T_c$  signal at this band for RTMs. However, ML models successfully identified the connection between 89 GHz  $PD$  and precipitation type, which is consistent to the previous findings in [12] and [9], where they identified the association between snow aggregates and 85 - 89 GHz  $PD$  in the stratiform layer using radiative transfer theories.

To summarize, the fact that  $T_c$ s at 89, 190 and 166 GHz are the top three direct radiance observations in determining the precipitation type suggests that precipitation type is tightly correlation with middle troposphere snow amount and distribution, while the high ranks of  $PD$ s at 89 and 166 GHz indicate that snow microphysics details are closely related to

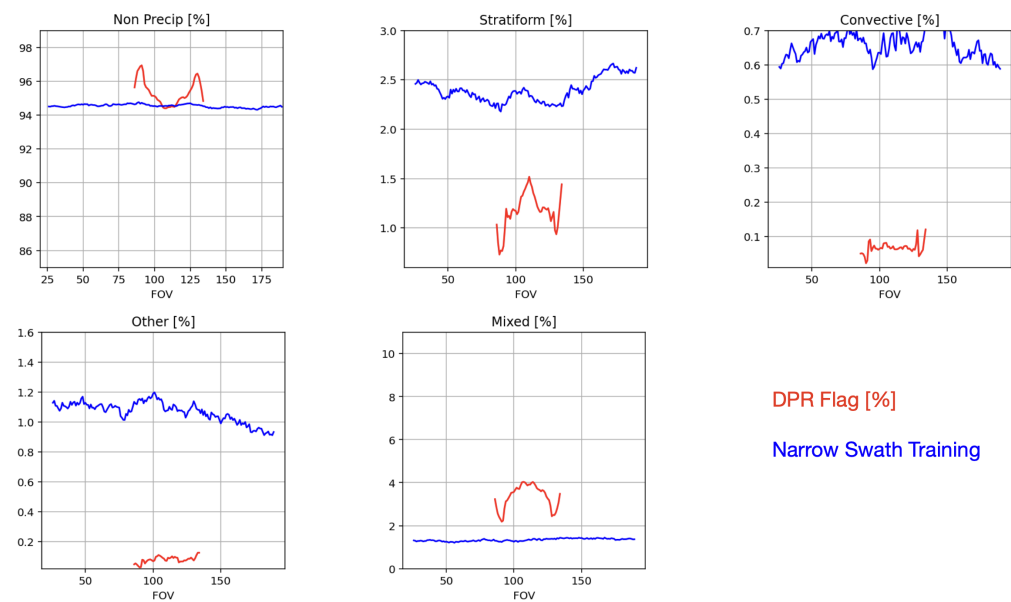
precipitation type as well. It should be noted that these two variables are also useful to differentiate surface and cloud/precipitation when  $T_c$  is relatively warm ([9]). Interestingly, liquid cloud water amount (CLWP) from auxiliary MERRA-2 reanalysis data ranks among the top factors from all experiments, which fits into the common sense as liquid cloud layer is closer in the vertical distance to the surface precipitation. The importance of total water vapor (TWV) and surface temperature ( $T_s$ ) are also learned by the ML models to differentiate clear-sky and cloudy-sky. Surface emissivities at 89, 166 and 10 GHz, when included, play a major role in determining a prediction. However, as the overall prediction accuracy barely drops or even increases for certain type (Fig. 8) when emissivity is excluded, it is temporarily concluded here that there is some redundant information between raw  $T_c$  observations and surface emissivity at these channels. However, further thorough investigations is needed, for example, to understand whether including emissivity could help improve "mixed" class prediction, as suggested in Fig. 8.

### 3.4. View-angle Dependency

The DPR conducts cross-track scan consisting of 49 incidence angles in the range from -170 to 170 relative to nadir in achieving three-dimensional precipitation measurement. However, the measurements from off-nadir incidence angles are subject to contamination of surface clutter, leading to the DPR blind zone in which precipitation measurements are unavailable near surface ([22]). The degree of impact of surface clutter on the DPR precipitation measurements near surface depends on incidence angle and surface roughness. In general, near-surface range contaminated by surface clutter increases with an increase of the off-nadir incidence angles. This caveat poses a practical problem in detecting shallow rain and classifying precipitation type when the bright-band/melting layer is close to surface ([31]). Moreover, the precipitation retrieval cannot be made without assumption in the blind zone. As influence of surface clutter on the DPR measurements depends on the incidence angle, it is expected that some of the DPR product might exhibit angle dependence to various degrees. Therefore, it is intriguing and beneficial to study the view-angle dependency of GMI-only retrieval product to understand whether the view-angle dependent artifact in the DPR product is propagated toward GMI or not.



**Figure 10.** The view-angle dependency of occurrence frequency of each precipitation type from DPR (red), wide-swath training result using RF model (black), and narrow-swath training result using RF model (blue). – needs to update to comply the same format with Fig. 11



**Figure 11.** Same with Fig. 10, except using CNN model, and only wide-swath training result is shown. – needs to change y-axis to percentage, remove legend from sub-panels and only one legend at bottom right panel.

The GMI and DPR precipitation occurring frequency (OF) as a function of view-angle is shown for each precipitation type are shown in Fig. 10 from the RF model and Fig. 11 from the CNN model outputs. "Wide swath training" (Black lines in Fig. 10) corresponds to the entire NS swath, while "narrow swath training" (blue lines in Fig. 10) corresponds to only the near-nadir 20 Ku-DPR footprints. The latter was a sensitivity experiment to test whether mitigating the DPR precipitation flag artifact at off-nadir angles would or would not affect the training results.

Evidently, DPR's off-nadir artifact is clearly seen for the non-precipitating, stratiform and mixed classes. These are three most frequently occurring types globally. Luckily, GMI didn't learn this view-angle dependent artifact because view-angle is not an input feature for training. Rather, for this 1D retrieval, GMI treats each DPR view independently and hence didn't propagate this error to a larger oblique view angles. As a sensitivity study, we limit the training "truth" to only the near-nadir 20 DPR footprints, but the prediction results barely change in terms of view-angle dependency, although there is a hint that narrow-swath training tends to result in predicting more stratiform and convective precipitations, while the "other" class tends to be less predicted (Fig. 10).

Comparing the results generated from RF versus CNN, CNN's prediction for non-precipitation scenes are closer to that of DPR. However, if we closely inspect the breakout from Fig. 10 and Fig. 11, there is not much difference between them for stratiform, convective and other precipitation types, except CNN model predicts 50% more convective and 60% more other types of precipitation compared to RF and GB (not shown) results. All three models overestimate the mono-type precipitations, which may likely not indicate a bias, but rather sensitivity of GMI to light precipitation that DPR is not sensitive to. This is also found previously in [11] and [5]. Future sensitivity study using the MS as the "truth" could help better elucidate this point and quantify their difference. RF results over-predict the occurrence chance for "mixed" class while CNN shows opposites result. Since the accuracy rate for the "mixed" class is low for all the models, it is not very meaningful to compare their prediction.

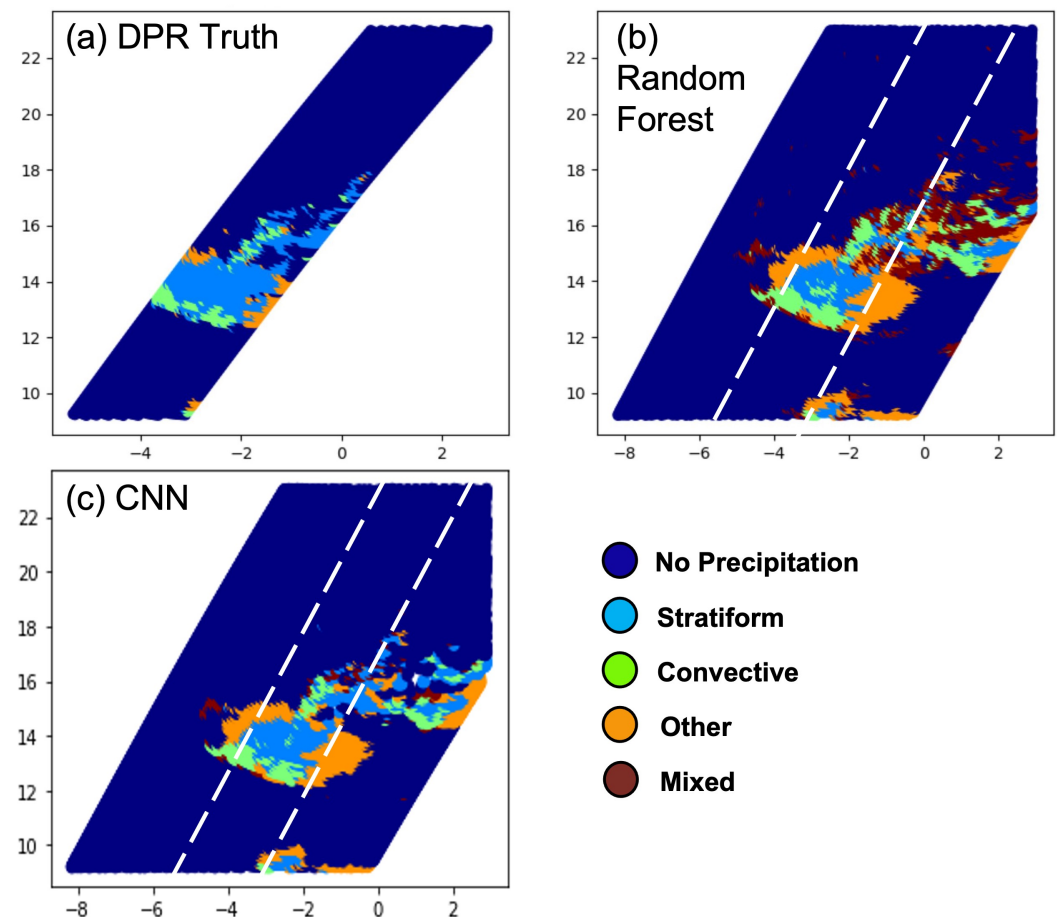
#### 4. Application of GMI-only Prediction on Weather and Climate Studies

The spatial and temporal distribution of each precipitation type is not only critical for making good weather forecast and potentially useful for enhancing weather-related



warning products, its occurring frequency (OF) is also a straightforward matrix that can be used for comparing with GCM outputs, as GCM produces convective, stratiform and anvil precipitation separately in each grid box ([32]). In this section, we will showcase the power of using GMI-only precipitation type classification for weather and climate research.

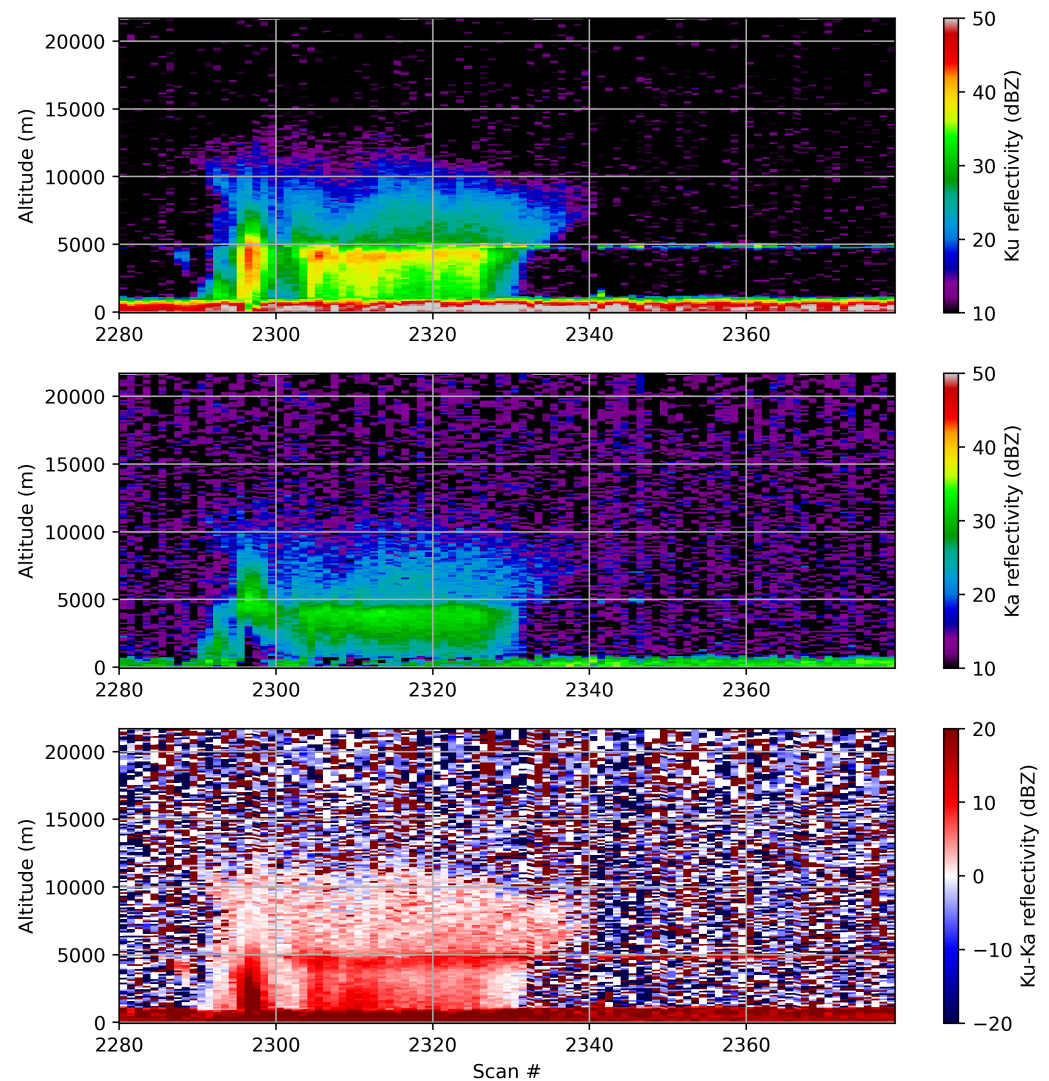
A squall line case passed by Mali, Africa on June 06, 2017, which was captured by DPR and GMI. On the Ku-DPR retrieval shown in Fig. 12a, the major squall line body, including the front line of convective cells (green), the large stratiform tail (blue) and the periphery anvil precipitation (orange), are well captured. We can further see another line of convection forming at its tail that seems to propagate northwestward, and the tip of another smaller squall line at the bottom right corner of the DPR swath. The full-swath (850 km) GMI-only prediction using the RF and CNN models are shown in Fig. 12b and c, respectively, while the DPR swaths are bounded by the white dashed lines to ease visual comparison. One can clearly see the overall structure agree very well along the overlapped swaths, but GMI also captures the long and complicated "goldfish tail" structure, and the other squall line system in the lower-bottom corner. Further, GMI prediction captures the anvil head associated precipitation signal ahead of the convection line (other, orange) which DPR missed probably due to its low reflectivity.



**Figure 12.** Precipitation types from (a) DPR "truth"; (b) predicted from RF model; (c) predicted from CNN model for a squall line event on June 6, 2017. The DPR swath is marked by the two white dashed lines in (b) and (c).

Subtle differences are identified between the two model outputs. RF produces some convective flags in the middle of the large area of stratiform trailing layers which DPR didn't observe and CNN model didn't predict. A further inspection of the DPR vertical structure along the nadir path of DPR swath suggests that CNN model prediction might be correct. As shown in Fig. 13a and b, the brightband signature corresponding to the

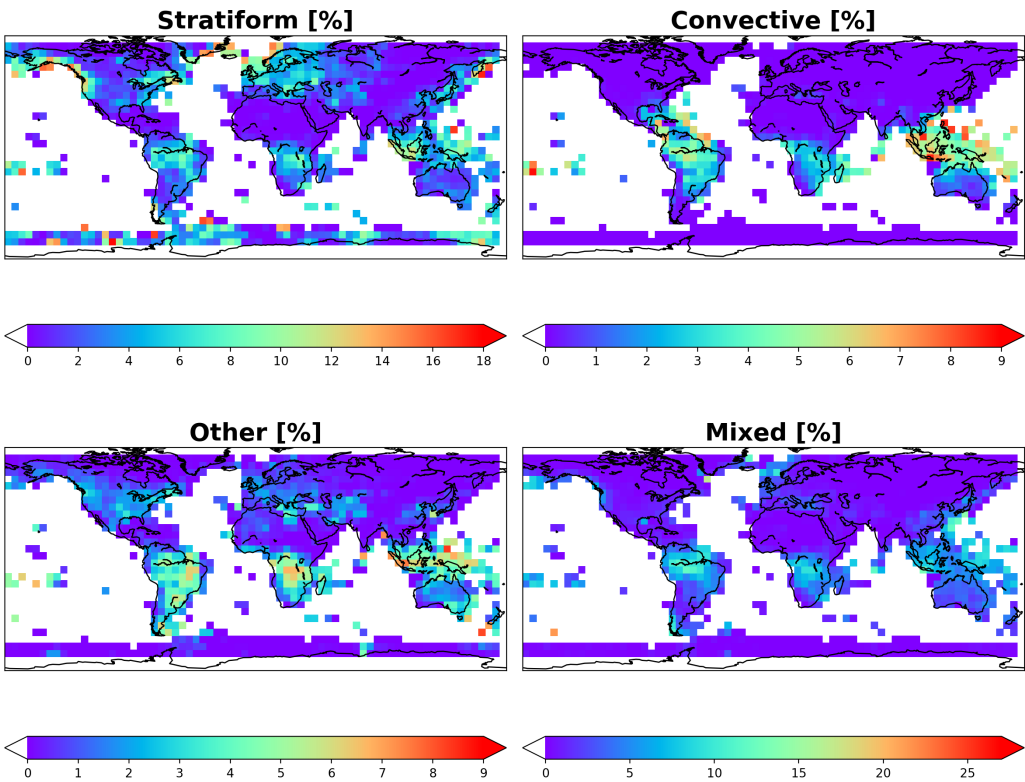
stratiform precipitation is evident between scan # 2305 - 2330. Yet, the enhancement of Ku-DPR reflectivity at scan #2305 indicates larger hydrometeor sizes or stronger melting signature, which GMI may likely capture and RF model learnt. Besides this disparity, CNN barely predicts any mixed precipitation types while RF model predicts quite amount of mixed precipitation pixels in the long tail. Overall, we can conclude again from this case study that the three mono-types (non-precipitating, stratiform and convective) stay robustly against which ML model to choose. There are subtle differences, but the underlying physics play the dominant role. For "other" and "mixed" classes, as DPR has difficulties to sense light precipitation, and our "mixed" class is not well defined, we don't expect our model to produce very good results; nevertheless, they are still useful especially for the 'other' type using the CNN model.



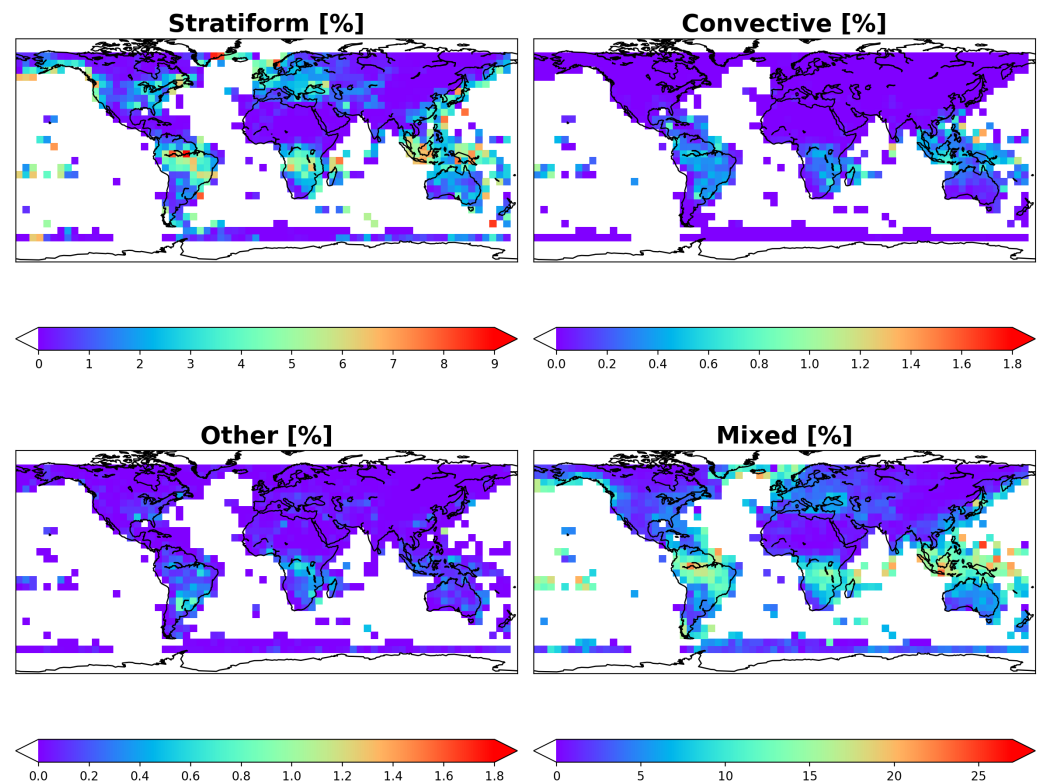
**Figure 13.** Cross-section along the center of the swath in Fig. 12 from (a) Ku reflectivity; (b) Ka reflectivity; (c) Ku-Ka difference.

The global distribution of OF for each precipitation type is shown in Fig. 14 for CNN prediction for boreal winter 2017 (January, February and December). The geographic distribution of stratiform and convective precipitation OFs agrees very well with previously reported features using TRMM-PR and GPM-DPR in the tropics and sub-tropics ([24]; [26]; and Fig. 15 below). As expected, stratiform precipitations are found over tropical convection active centers as well as along the boreal winter storm tracks, while convective precipitations are found to frequently occur in the tropical convection active centers

(i.e., Amazonia, Central Africa and West Pacific/Marine time continents). Because of the connical-scan geometry, GMI covers a little bit more high-latitude areas than DPR, and it is noticeable that stratiform precipitation occurs frequently over Southern Ocean above the sea ice covered area (i.e., where MERRA-2 has a land surface temperature retrieval), especially near the Antarctica Peninsula and McMurdo station where persistent mixed-phase cloud and associated pockets of ice precipitation are reported in several ground measurement works ([1]; [10]). Although all passive sensors have issues at snow/ice covered surfaces, the good comparison between DPR and GMI-only prediction over the Antarctic sea ice suggests that high-frequency polarized passive MW measurements may provide some capability at distinguishing surface ice from frozen hydrometeors aloft in the air.



**Figure 14.** Geographic distribution of precipitation type occurring frequency (%) over land and sea ice during January, February and December of 2017 from CNN model predictions using full-swath GMI L1-CR data for (a) stratiform, (b) convective, (c) other and (d) mixed.

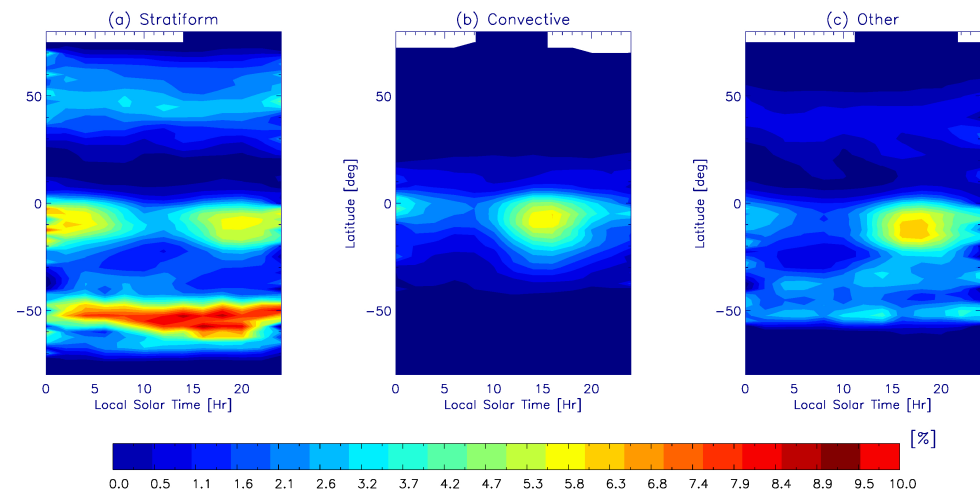


**Figure 15.** Same with Fig. 14, except from the DPR "truth" during DJF, 2016 and 2017.

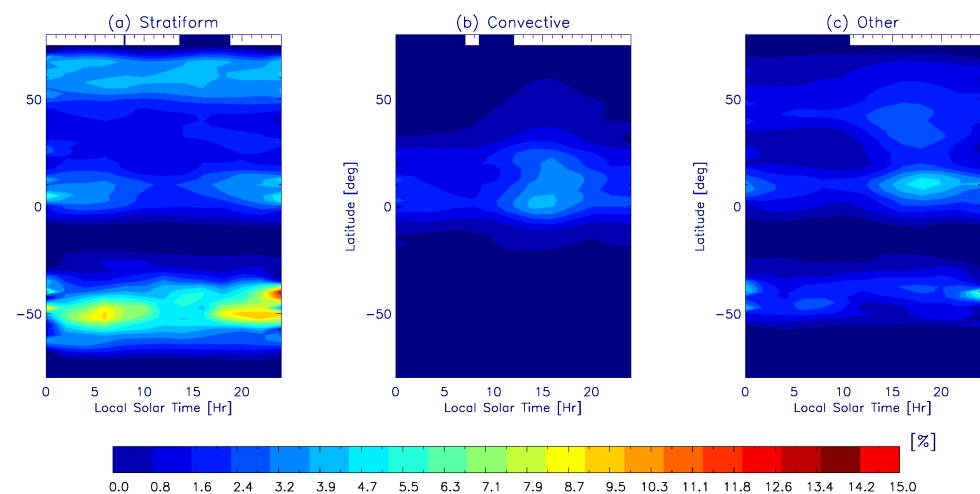
Here we found "other" precipitation also occurred quite often in the similar regions where stratiform and convective precipitations prefer to happen, but the former generally covers a broader area. "Other" precipitation occurs as frequently as convective and stratiform precipitation in most of the tropical land areas. There are significant regional differences in terms of their relative importance. For example, at marinetime continents - west Pacific area, convective precipitation occurs as frequent as stratiform precipitation, while that's not the case at other tropical convection active areas. We can also find convective precipitation rarely occurs over the winter storm tracks. As an extension of the North America storm track, the middle east winter-time precipitation nearly all come from stratiform and "other" type of precipitations. Previous research had never reported distribution of "other" type of precipitation to the authors' best knowledge, probably because it is thought to happen less frequently due to the fact that precipitation radar like TRMM-PR and GPM-DPR often miss to detect them. As also demonstrated in Fig. 15, the "other" type saw by DPR is in general one order of magnitude smaller than what GMI sees. At this point it is still too early to associate "other" class to anvil precipitation, but it is strongly indicated as such, and we will inspect collocated CloudSat radar echos to further study this type of precipitation. Part of the "other" type could also be associated with shallow convections or light precipitation that DPR is not sensitive to.

At last, we would like to share some flavor about using this GMI-only precipitation flag dataset to study the diurnal cycle of precipitation, which is known to be one of the greatest challenges for climate models to represent. At tropics and mid-latitude summer, convection is thought to be responsible for initiating the precipitation processes. However, the time scale it takes for the decaying/dissipation processes determines the spread area and intensity of stratiform precipitation. Hence, it is necessary and a great benefit this product can bring to the precipitation community to separate out the diurnal cycle of different precipitation processes. In the winter hemisphere, it is stratiform precipitation

process that dominates the surface precipitation variation, and hence it is more meaningful to single out this type of precipitation and study it.



**Figure 16.** Diurnal variation of cloud fraction (%) from GMI-only prediction during DJF, 2016 and 2017 for (a) stratiform (b) convective and (c) other precipitation classes, respectively.



**Figure 17.** Same with Fig. 16, except for JJA, 2016 and 2017.

GPM flies on a large inclination angle ( $65^\circ$ ) to cover high-latitudes, hence it takes more than 3 months to sample the entire diurnal cycle at different latitudes. For this consideration, we combine June-July-August (JJA) of 2016 and 2017 GMI-only predictions together to generate the diurnal cycle contours for different latitude bands shown in Fig. 16 for DJF and Fig. 17 for JJA. In the tropics, as well observed and reported before, convection peaks in the late afternoon, but we can still see the peak time tends drifting toward a later time moving away from the equator. The peak and trough of the diurnal cycle of "other" class follows closely with that of the deep convection with a lag of about 2 - 3 hours. This strongly indicates that the "other" class is associated with anvil precipitation which is subject to the development and dissipation of the convective core. However, the decay of stratiform precipitation takes a significantly longer time and has subtle latitudinal variations that do not follow closely with that of the deep convection.

Interestingly enough, the stratiform precipitation over Southern Ocean winter time (Fig. 17) exhibits a distinctive diurnal cycle with peak occurrence at local night to early



morning and a quiet period of precipitation in local afternoon. This is consistent with Macquarie Island ground rain gauge observations ([39]). However, the summer time (Fig. 16) diurnal cycle disagrees with ground measurements in the same paper. Boreal winter storm tracks also exhibit a weak but interesting diurnal cycle. These discrepancies and new features warrant further investigations.

## 5. Conclusion

This work develops a machine learning/artificial intelligence (ML/AI) approach to retrieval precipitation types from passive microwave (PMW) radiometers/imagers. This approach is applied to collocated GPM-GMI and GPM-DPR data, where the latter is used as the "truth" for training and independent validation. In principle this approach can be easily extended to other spaceborne PMW instruments as DPR overpasses frequently with polar orbiting satellites.

Five precipitation types are trained and predicted simultaneously, which are (0) non-precipitating; (1) stratiform; (3) convective; (4) other; (5) mixed. The last class is created when the GMI footprint is filled with multiple DPR footprints that do not share the same precipitation type among the first four. A total of 6 ML models are tested, and only the Logistic Regression model fails to make a good prediction. Gradient Boosting (GB) and convolutional neural network (CNN) produce overall the highest accuracy at 96.7%. Although the overall accuracy is sample size weighted, GMI-only CNN prediction is correct at 80 - 87% chances when it's precipitating and the scene is filled with a single type of precipitation. Challenges are identified at the boundaries of transition of different precipitation types, as well as for mixed-class.

Different from previous ML/AI explorations on the similar topic (e.g.,[27]), this work strives to separate non-precipitation (> 95% of the time) and precipitation scenes simultaneously with predicting the precipitation types, so the training dataset is extremely unbalanced. A data augmentation is subsequently introduced to subset non-precipitation scenes and bootstrap from the "convective" and "other" sample pools. This data augmentation technique is particularly useful for precipitation-related science as data imbalance frequently occurs in this field.

Other than using all 13 GMI channel observations and some ancillary features from the MERRA-2 reanalysis, the most important novelty of this work is to include the hand-engineered features - polarization difference (PD), and emissivity retrieval products. Including these attributes turned out to be extremely useful, as PDs and emissivities at high-frequency (> 89 GHz) channels are ranked among the tops in determining a precipitation type prediction. Simple ML models are sensitive to the emissivity by showing a trade-off of accuracy rate between "other" and "mixed" class. The deep learning CNN model remains robust with or without including the surface emissivity in the training features. We demonstrate in this paper how ML model learning the radiative transfer physics internally and reach microphysics and emissivity consistent results. This finding adds fidelity to use the ML/AI approach for developing retrieval algorithms.

While many of the current spaceborne PMW instruments have high-frequency channels, they barely hold the dual-polarized channel pairs nor have a pixel-by-pixel emissivity retrieval product readily available. This poses challenges on achieving high accuracy in predicting precipitation types from other PMW instruments using this approach, and it will be a good exercise for future works.

Besides demonstrating the advantages of using GMI precipitation type retrievals for weather and climate studies, we identified two particular merits that GMI precipitation type retrieval exhibits while DPR doesn't. Firstly, DPR's view-angle dependent artifact doesn't propagate into GMI's oblique view angles. Combining with the fact that microphysical signals embedded in the PD observation is less complicated in a fixed view-angle conical scanner, this finding implies that conical scanning is a better design for future PMW instruments compared to cross-track scan, if mission budget is not factored in. Secondly, GMI's precipitation type retrieval captures more light precipitation scenes than DPR,

especially for the "other" type. Both findings suggest that ML/AI can not only learn from the "truth", but it also has the capability of unearth embedded information in the training features that is highly intertwined and non-linear that RTM simulations cannot reveal.

**Author Contributions:** Conceptualization, J. G.; methodology, S. D., Y. W., J. G. and L. D.; programing, S. D., Y. W. and J. G.; formal analysis, S. D., Y. W., J. G., L. D. and C. W.; investigation, all authors; resources, J. G., L. D., D. L. W.; data curation, J. G., S. D., Y. W. and S. J. M.; writing—original draft preparation, S. D., J. G. and L. D.; writing—review and editing, all authors; visualization, S. D., Y. W., J. G., C. W., S. J. M. and W. O.; supervision, J. G. and L. D.. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was mainly funded by NASA CloudSat-CALIPSO Science Team (CCST) Program. Mr. S. Das acknowledges funding support from NASA internship program. Computational resources from NASA supercomputer center are highly appreciated. S. J. M., L. L. and W. O. were supported by NASA's Precipitation Measurement Mission (PMM) Grant NNH18ZDA001N-PMMST.

**Data Availability Statement:** GPM data are freely available at <https://gpm.nasa.gov/data/directory>. The ML codes will be provided on 2nd author's Github account once this manuscript is accepted. GMI-only prediction data for 2017 will be provided to interested readers upon request.

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

## References

- Alexander, S. P., G. M. McFarquhar, R. Marchand, A. Protat, E. Vignon, G. G. Mace, A. R. Klekociuk: Mixed-Phase Clouds and Precipitation in Southern Ocean Cyclones and Cloud Systems Observed Poleward of 64°S by Ship-Based Cloud Radar and Lidar, *JGR-Atmosphere*, **2021**, <https://doi.org/10.1029/2020JD033626>.
- Awaka, J., T. Iguchi and K. Okamoto, T. Early results on rain type classification by the Tropical Rainfall Measuring Mission (TRMM) precipitation radar. *Proc. 8th URSI Commission F Open Symp.*, Aveiro, Portugal, **1998**; pp.143-146.
- Awaka, J. and T. Iguchi and K. Okamoto. TRMM PR standard algorithm 2A23 and its performance on bright band detection. *J. Meteor. Soc.*, **2009**, doi:<https://doi.org/10.2151/jmsj.87A.31>.
- T. Iguchi, S. Seto, R. Meneghini, N. Yoshida, J. Awaka and T. Kubota. GPM/DPR Level-2 Algorithm Theoretical Basis Document, Version 06 updates, **2018**; [https://gpm.nasa.gov/sites/default/files/2019-05/ATBD\\_DPR\\_201811\\_with\\_Appendix3b.pdf](https://gpm.nasa.gov/sites/default/files/2019-05/ATBD_DPR_201811_with_Appendix3b.pdf).
- Turk, F. J., Sarah E. Ringerud, Andrea Camplani, Daniele Casella, Randy J. Chase, Ardeshir Ebtehaj, Jie Gong, Mark Kulie, Guosheng Liu, Lisa Milani, Giulia Panegrossi, Ramon Padullés, Jean-François Rysman, Paolo Sandò, Sajad Vahedizade and Norman B. Wood. Applications of a CloudSat-TRMM and CloudSat-GPM Satellite Coincidence Dataset. *Remote Sensing*, **2021**; doi:<https://doi.org/10.3390/rs13122264>.
- Liang Liao and Robert Meneghini. GPM DPR Retrievals: Algorithm, Evaluation, and Validation. *Remote Sensing*, **2022**; doi:<https://doi.org/10.3390/rs14040843>
- GPM Intercalibration (X-CAL) Working Group. Algorithm Theoretical Basis Document (ATBD), NASA Global Precipitation Measurement (GPM) Level 1C Algorithms (Version 1.6), **2016**, [https://gpm.nasa.gov/sites/default/files/2020-05/L1C\\_ATBD\\_v1.6\\_V04\\_0.pdf](https://gpm.nasa.gov/sites/default/files/2020-05/L1C_ATBD_v1.6_V04_0.pdf)
- Gong, J. and Wu, D. L., Microphysical properties of frozen particles inferred from Global Precipitation Measurement (GPM) Microwave Imager (GMI) polarimetric measurements, *Atmos. Chem. Phys.*, **2017**, 17, 2741–2757, doi: <https://doi.org/10.5194/acp-17-2741-2017>.
- Gong, J., Zeng, X., Wu, D. L., Munchak, S. J., Li, X., Kneifel, S., Ori, D., Liao, L., and Barahona, D.. Linkage among ice crystal microphysics, mesoscale dynamics, and cloud and precipitation structures revealed by collocated microwave radiometer and multifrequency radar observations, *Atmos. Chem. Phys.*, **2020**, pp.12633–12653, <https://doi.org/10.5194/acp-20-12633-2020>.
- Mace, G. G., A. Protat, S. Benson, Mixed-Phase Clouds Over the Southern Ocean as Observed From Satellite and Surface Based Lidar and Radar, *J. G. R. - Atmosphere*, **2021**, <https://doi.org/10.1029/2021JD034569>.
- S. J. Munchak, S. Ringerud, L. Brucker, Y. You, I. de Gelis and C. Prigent. An Active–Passive Microwave Land Surface Database From GPM. *IEEE Transactions on Geoscience and Remote Sensing*, **2020**, pp. 6224–6242, doi: 10.1109/TGRS.2020.2975477.
- William S. Olson, Ye Hong, Christian D. Kummerow, and Joseph Turk. A Texture-Polarization Method for Estimating Convective–Stratiform Precipitation Area Coverage from Passive Microwave Radiometer Data. *J. Meteor. and Clim.*, **2001**, doi:[https://doi.org/10.1175/1520-0450\(2001\)040<1577:ATPMFE>2.0.CO;2](https://doi.org/10.1175/1520-0450(2001)040<1577:ATPMFE>2.0.CO;2).
- F. Aires, C. Prigent, F. Bernardo, C. Jiménez, R. Saunders, and P. Brunel. A tool to estimate land-surface emissivities at microwave frequencies (TELSEM) for use in numerical weather prediction, *Quart. J. Roy. Meteorol. Soc.*, **2011**, pp. 690–699.
- Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, and Gang Wang, Recent advances in convolutional neural networks, *Pattern Recognition*, **2018**, vol. 77, pp. 354-377.
- Nathalie Japkowicz, Shaju Stephen, The class imbalance problem: A systematic study, *Intelligent Data Analysis*, **2002**, vol. 6, no. 5, pp. 429-449.

16. Menardi, G. and N. Torelli, Training and assessing classification rules with imbalanced data, *Data Mining and Knowledge Discovery*, **2014**, vol. 28, no. 1, pp.92-122.
17. Wang, C., Platnick, S., Meyer, K., Zhang, Z., and Zhou, Y.. A machine-learning-based cloud detection and thermodynamic-phase classification algorithm using passive spectral observations, *Atmos. Meas. Tech.*, **2020**, 13, 2257–2277, doi: <https://doi.org/10.5194/amt-13-2257-2020>.
18. Robbins, D., Poulsen, C., Siems, S., and Proud, S. Improving discrimination between clouds and optically thick aerosol plumes in geostationary satellite data, *Atmos. Meas. Tech. Discuss.*, [preprint], **2022**, doi: <https://doi.org/10.5194/amt-2021-422>.
19. Gong, J., C. Wang, D. L. Wu, D. Barahona, B. Tian and L. Ding, A GCM-Oriented and Artificial Intelligence Based Passive Microwave Diurnal Ice/Snow Cloud Retrieval Product using CloudSat/CALIPSO as the Baseline, *Atmos. Chem. Phys.*, [submitted], **2022**.
20. Y. Choi, D. -B. Shin, J. Kim and M. Joh, Passive Microwave Precipitation Retrieval Algorithm With A Priori Databases of Various Cloud Microphysics Schemes: Tropical Cyclone Applications, *IEEE Transactions on Geoscience and Remote Sensing*, **2020**, vol. 58, no. 4, pp. 2366-2382, doi: [10.1109/TGRS.2019.2948262](https://doi.org/10.1109/TGRS.2019.2948262).
21. O'Dell, C. W., F. Wentz, and R. Bennartz, Cloud Liquid Water Path from Satellite-Based Passive Microwave Observations: A New Climatology over the Global Ocean, *J. Clim.*, **2008**, doi:[10.1175/2007JCLI1958.1](https://doi.org/10.1175/2007JCLI1958.1).
22. Grecu, M., D. Bolvin, G. M. Heymsfield, S. E. Lang, W. S. Olson, Improved parameterization of precipitation fluxes in the GPM combined algorithm to mitigate ground clutter effects, *AGU fall meeting*, **2021**, <https://agu.confex.com/agu/fm21/meetingapp.cgi/Paper/953816>.
23. Greenwald, T. J., R. Bennartz, M. Lebsock, J. Teixeira, An Uncertainty Data Set for Passive Microwave Satellite Observations of Warm Cloud Liquid Water Path, *J. G. R. -Atm.*, **2018**, doi:<https://doi.org/10.1002/2017JD027638>.
24. Yang, S. and E. A. Smith, Convective–Stratiform Precipitation Variability at Seasonal Scale from 8 Yr of TRMM Observations: Implications for Multiple Modes of Diurnal Variability, *J. Clim.*, **2008**, doi: <https://doi.org/10.1175/2008JCLI2096.1>.
25. Michael G. Bosilovich, Santha Akella, Lawrence Coy, Richard Cullather, Clara Draper, Ronald Gelaro, Robin Kovach, Qing Liu, Andrea Molod, Peter Norris, Krzysztof Wargan, Winston Chao, Rolf Reichle, Lawrence Takacs, Yuri Vikhliav, Steve Bloom, Allison Collow, Stacey Firth, Gordon Labow, Gary Partyka, Steven Pawson, Oreste Reale, Siegfried D. Schubert, and Max Suarez , MERRA-2: Initial Evaluation of the Climate, *NASA/TM–2015-104606/Vol. 43*, <https://gmao.gsfc.nasa.gov/pubs/docs/Bosilovich803.pdf>.
26. Gao, J., Tang, G. and Hong, Y., Similarities and Improvements of GPM Dual-Frequency Precipitation Radar (DPR) upon TRMM Precipitation Radar (PR) in Global Precipitation Rate Estimation, Type Classification and Vertical Profiling, *Remote Sens.*, **2017**, 9(11), 1142; doi:<https://doi.org/10.3390/rs9111142>.
27. Petkovic, V., M. Orescanin, P. Kirstetter, C. Kummerow, R. Ferraro, Enhancing PMW Satellite Precipitation Estimation: Detecting Convective Class, *J. Atm. and Ocn. Tech.*, **2019**, doi:10.1175/JTECH-D-19-0008.1.
28. Petković, V., C.D. Kummerow, D.L. Randel, J.R. Pierce, and J.K. Kodros, Improving the Quality of Heavy Precipitation Estimates from Satellite Passive Microwave Rainfall Retrievals. *J. Hydrometeor.*, **2018**, 19, 69–85.
29. Henderson, D.S., C.D. Kummerow, D.A. Marks, and W. Berg, A Regime-Based Evaluation of TRMM Oceanic Precipitation Biases. *J. Atmos. Oceanic Technol.*, **2017**, 34, 2613–2635.
30. Seto, S., Iguchi, T., Meneghini, R., Awaka, J., Kubota, T., Masaki, T., Takahashi, N. The Precipitation Rate Retrieval Algorithms for the GPM Dual-frequency Precipitation Radar. *J. Meteorol. Soc. Jpn.*, **2021**, 99, 205–237. <https://doi.org/10.2151/jmsj.2021-011>.
31. Awaka, J., Le, M., Brodzik, S., Kubota, T., Masaki, T., Chandrasekar, V., Iguchi, T. Improvements of rain type classification algorithms for a full scan mode of GPM Dual-frequency Precipitation Radar. *J. Meteor. Soc. Jpn.*, **2021**, 99, 1253–1270.
32. Yuan, W., R. Yu, M. Zhang, W. Lin, J. Li, and Y. Fu, Diurnal cycle of summer precipitation over subtropical East Asia in CAM5. *J. Clim.*, **2013**, <https://doi.org/10.1175/JCLI-D-12-00119.1>.
33. Prigent, C., Precipitation retrieval from space: an overview, *Comptes Rendus Geoscience*, **2010**, <https://doi.org/10.1016/j.crte.2010.01.004>.
34. Breiman, L., Random forest, *Machine learning*, **2001**, <https://doi.org/10.1023/A:1010933404324>.
35. Caruana, R. and Niculescu-Mizil, A., An empirical comparison of supervised learning algorithms, *Proceedings of the 23rd international conference on Machine learning*, **2006**, <https://doi.org/10.1145/1143844.1143865>.
36. Friedman, J. H., Greedy function approximation: a gradient boosting machine, *Annals of statistics*, **2001**, <https://doi.org/10.1214/aos/1013203451>.
37. Branco, P., L. Torgo, R. P. Ribeiro, A survey of predictive modeling on imbalanced domain, *ACM Computing Surveys*, **2016**, <https://doi.org/10.1145/2907070>.
38. G. M. Skofronick-Jackson, J.R. Wang, G.M. Heymsfield, R. Hood, W. Manning, R. Meneghini, and J.A. Weinman, Combined Radiometer-Radar Microphysical Profile Estimations with Emphasis on High Frequency Brightness Temperature Observations, *J. of Applied Meteorology*, **2003**, [https://doi.org/10.1175/1520-0450\(2003\)042<0476:CRRMPE>2.0.CO;2](https://doi.org/10.1175/1520-0450(2003)042<0476:CRRMPE>2.0.CO;2).
39. Lang, F.; Huang, Y.; Siems, S.T.; Manton, M.J: Evidence of a Diurnal Cycle in Precipitation over the Southern Ocean as Observed at Macquarie Island. *Atmosphere*, **2020**, 11, 181. <https://doi.org/10.3390/atmos11020181>.