



Article

A Deep-Learning Scheme for Hydrometeor Type Classification Using Passive Microwave Observations

Ruiyao Chen * and Ralf Bennartz

Department of Earth and Environmental Sciences, Vanderbilt University, Nashville, TN 37215, USA

* Correspondence: ruiyao.chen@vanderbilt.edu

Abstract: This paper proposes a novel approach for hydrometeor classification using passive microwave observations. The use of passive measurements for this purpose has not been extensively explored, despite being available for over four decades. We utilize the Micro-Wave Humidity Sounder-2 (MWSH-2) to relate microwave brightness temperatures to hydrometeor types derived from the global precipitation measurement's (GPM) dual-frequency precipitation radar (DPR), which are classified into liquid, mixed, and ice phases. To achieve this, we utilize a convolutional neural network model with an attention mechanism that learns feature representations of MWSH-2 observations from spatial and temporal dimensions. The proposed algorithm classified hydrometeors with 84.7% accuracy using testing data and captured the geographical characteristics of hydrometeor types well in most areas, especially for frozen precipitation. We then evaluated our results by comparing predictions from a different year against DPR retrievals seasonally and globally. Our global annual cycles of precipitation occurrences largely agreed with DPR retrievals with biases being 8.4%, −11.8%, and 3.4%, respectively. Our approach provides a promising direction for utilizing passive microwave observations and deep-learning techniques in hydrometeor classification, with potential applications in the time-resolved observations of precipitation structure and storm intensity with a constellation of smallsats (TROPICS) algorithm development.

Keywords: hydrometeor classification; deep learning; passive microwave observations; GPM DPR; FY-3C



Citation: Chen, R.; Bennartz, R. A Deep-Learning Scheme for Hydrometeor Type Classification Using Passive Microwave Observations. *Remote Sens.* **2023**, *15*, 2670. <https://doi.org/10.3390/rs15102670>

Academic Editor: Arnab Muhuri

Received: 19 March 2023

Revised: 17 May 2023

Accepted: 18 May 2023

Published: 20 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Precipitation is one of the three fundamental components of the water cycle, and it plays a key role in weather and climate systems as well as energy-transfer processes [1–3]. The study of precipitation is highly valuable in many aspects. The spatial-temporal distribution of precipitation and its variation have wide-ranging effects on human well-being and ecosystems [4–6]. For example, changes in the intensity, frequency, and distribution of precipitation can have a direct impact on crop water supply and soil water balance, and further have an impact on atmospheric balance and water resources management [7–9]. Precipitation studies are also critical for regulating the global energy flow, i.e., the movement of heat, and further adjusting the numerical weather-prediction models. In addition, there is high climatological significance because studying the long-term variation in global precipitation can inform scientists of how precipitation responds to a changing climate [10,11].

The effects of precipitation on the hydrological cycle, energy balance budget, land–air interactions, and weather-prediction models further vary depending on the forms of precipitation. In general, there are three types of precipitation that reach the Earth's surface: liquid-phase (rain), mixed-phase (sleet), and solid-phase (snow) [12,13]. While snowfall events can have a dramatic impact on the energy budget by increasing surface albedo by over 50% [14,15], rainfall can quickly raise soil moisture, replenish groundwater, and form surface runoff [16,17]. In light of the significance of hydrometeor classification for improved understanding in various earth science fields, including weather forecasting,

climate-change dynamics, and the usage of water resources, an increasing number of studies have been focusing on the identification of hydrometeor types (rain/sleet/snow).

Most existing hydrometeor classification schemes rely on remote sensing measurements from ground-based, airborne, and spaceborne active radars. As listed in Table 1 of the literature review summary, among the various classification algorithms, the fuzzy logic approach has gained considerable attention due to its ability to characterize different ranges of polarimetric measurements using empirically derived membership functions. The fuzzy logic approach was first proposed by Vivekanandan et al. [18] and has since undergone several improvements with more sophisticated membership functions and decision criteria [19,20]. Later, the National Severe Storms Laboratory at the National Oceanic and Atmospheric Administration designed another fuzzy logic algorithm for the operation of the U.S. NEXRAD, which has been upgraded over the years [21–23].

Table 1. Literature review on past work related to hydrometeor classification.

Papers	Summary	Active/Passive Data
Vivekanandan, J., et al. [18]	One of most fundamental algorithms of hydrometeor classification, known as fuzzy logic	Active
Liu, Hongping, and V. Chandrasekar [19]; Lim, S., et al. [20]	Improvement algorithms based on [18]	Active
Ryzhkov, Alexander V., et al. [21]; Park, Hyang Suk, et al. [22]; Scharfenberg, Kevin A., et al. [23]	Hydrometeor classification algorithm used for U.S. NEXRAD system	Active
Yang, Ji, et al. [24]; Lukach, Maryna, et al. [25]	Recent examples of continued efforts on further improved hydrometeor classification methods	Active
Klaes, K. Dieter, et al. [26]	Spaceborne passive microwave data have been available for five decades	Passive
Dolant, Caroline, et al. [27]; Bennartz, Ralf, and Grant W. Petty [28]; Petty, Grant W., and Ke Li [29]; Skofronick-Jackson, Gail M., and James R. Wan [30]; Wilheit, Thomas T [31]; Kadem, Benjamin, et al. [32]	Passive microwave measurements are particularly useful for global precipitation study	Passive
Bennartz, Ralf, and Grant W. Petty [28]	Simulated passive microwave data are responsive to hydrometeor scatter to different degrees between 19 and 150 GHz	Passive
Leppert, Kenneth D., and Daniel J. Cecil [33]; Chen, Ruiyao, and Ralf Bennartz [34,35]	Studies using spaceborne and airborne microwave data agree with [28]	Passive

Although efforts have been continuously carried out by many researchers to improve hydrometeor classification, most studies have focused on using polarimetric radar measurements [24,25]. In this paper, we aim to explore the potential of passive microwave measurements, which have not been fully explored for hydrometeor classification. Passive microwave radiation measurements have proven useful for retrieving unique signatures that identify Earth surface features and obtain atmospheric temperature and composition. Microwave measurements interact with various types of hydrometeors through the vertical columns of precipitating clouds, making them particularly beneficial for global precipitation study [26–31]. The first spaceborne radiometer, the electrically scanning microwave radiometer (ESMR), was aboard NASA’s Nimbus-5 launched in 1972. Since then, various microwave radiometers have flown aboard satellites worldwide, collecting an increasingly large volume of radiation measurements emitted from the Earth at selected frequencies between 6 and 190 GHz, such as the advanced microwave sound unit-A and unit-B (AMSU-A and AMSU-B) and the microwave humidity sounder (MHS) [32].

Previous research has indicated that the relationship between hydrometeor scattering intensity and passive microwave brightness temperature varies across the frequency range of 19 to 150 GHz. In a study by Bennartz and Petty [27], the correlation between hydrom-

eteor scattering intensity and passive microwave brightness temperature was examined using radiative transfer modeling data, revealing variations in the relationship for the four frequencies studied within the aforementioned range. Later studies that explored the relationship between airborne and spaceborne microwave data and signatures of hydrometeor types have supported this finding [33–35]. To augment existing precipitation algorithms and to exploit passive microwave observation capabilities to their full potential, herein we explore a deep-learning approach which uses passive microwave radiance to diagnose hydrometeor types from liquid, mixed, and ice phases in an unprecedented manner. McCulloch and Pitts first proposed and developed the conceptual model of an artificial neural network in 1943. Since then, its application has expanded tremendously over the past few decades [36]. Recently, convolutional neural networks (CNN) have emerged as one of the most popular deep-learning approaches to deep in various fields, including meteorological studies that involve remote sensing observations [37]. Because meteorological applications tend to have large earth observation datasets with spatially and temporally coherent information, conventional statistically based methodologies may not be accurate enough to capture the spatio-temporal patterns in the vast amount of earth observation data, especially for ice particles and snowflakes which are of non-spherical shape and, hence, are more sophisticated and have imperfectly known ice particle scattering properties. The pattern recognition abilities of CNN models fit well with this type of data [38]. They are especially suited to approximate complicated nonlinear relationships between input values and output results through learning phases and to extract information from image-like data and sequential data. Because the idea is relatively new and it is more challenging to classify hydrometeor types using passive microwave observations compared to conventionally used precipitation radar measurements, developing a deep-learning algorithm that eliminates the need for a well-defined function to describe the relationship between the input passive microwave radiance and the output hydrometeor types is desirable. Herein, we leverage a CNN-based model in conjunction with an attention mechanism to learn meaningful feature representations from the spatial and temporal dimension space of a passive microwave data for the task of hydrometeor classification.

We trained neural networks using observations from the Fengyun-3C (FY-3C) Microwave Humidity Sounder-2 (MWS-2) at frequencies between 89 and 190 GHz. The training was supervised by “ground truth” data of hydrometeor types derived from measurements of the global precipitation measurement (GPM) mission’s core observatory which carries two critical instruments: GPM microwave imager (GMI) and dual-frequency precipitation radar (DPR). It is worth noting that MWS-2 carries five channels that are centered at the 118 GHz oxygen line, which is of tremendous significance because it is the first instrument measuring Earth’s radiance from space at 118 GHz. The thermal emission spectrum of the atmosphere near 118 GHz provides us with exceptional data to probe the atmosphere at 118 GHz with the combination of other channels at 90, 150, and 183 GHz. To best prepare datasets for training and testing models, we performed a series of data preprocessing, such as removing biases from observations, collocating observations and hydrometeor types as well as data sub-setting (which will be explained in detail in Section 2). We used the trained model to predict the type of hydrometeor given an input of MWS-2 observations.

Our model is capable of learning spatial and temporal feature representations from satellite observations through sophisticated neural networks [39]. We also innovated and enhanced its functionality with two mechanisms, context and channel-attention networks. These mechanisms allow us to exploit contextual information around each channel feature and emphasize their contributions to the output of the classification task. We will elaborate on each of those components in the following sections.

This study presents a novel approach to combining 118 GHz channels with other conventional channels between 89 and 190 GHz. By doing so, this study provides new insights into the distribution and variation of global hydrometeor types. Moreover, this is the first study to use spaceborne passive microwave observations to classify various types of hydrometers globally over ocean. This approach represents a significant advancement

in the application of deep-learning techniques to investigate hydrometeor characteristics using passive microwave observations. Furthermore, the proposed deep-learning scheme has significant implications for algorithm development in future missions. Specifically, the scheme can pave the way for innovative algorithm development for the forthcoming time-resolved observations of precipitation structure and storm intensity with a constellation of smallsats (TROPICS) mission and its pathfinder in hydrometeor classification work [40].

The remainder of this article is structured as follows. In Section 2, we provide information about the instrument and data used in our study. Section 3 elaborates on the data-driven deep-learning model mechanism we proposed. We present the experimental evaluation of our model in Section 4. Finally, in Section 5, we discuss the obtained classification accuracy and describe future work.

2. Instruments and Data

2.1. Input Feature

The MWHS-2 is an improved total power cross-track millimeter-wave radiometer onboard FY-3C midmorning polar-orbiting operational satellite launched by the China Meteorological Administration/National Satellite Meteorological Center (CMA/NSMC) in September 2013. Carrying 15 channels ranging from 89 to 190 GHz listed in Table 2 along with polarizations, the MWHS-2 has proven to be successful in observing global atmospheric thermodynamics information and monitoring storms in all-weather conditions as well as providing observations for improved numerical weather-prediction modeling. There are two window channels at 89 and 150 GHz, and five sounding channels around 183.31 GHz. In addition, eight sounding channels near 118.75 GHz were added to MWHS-2 based on its earlier version onboard Fengyun-3A, which makes MWHS-2 the first spaceborne microwave radiometer that carries 118 GHz channels whose content information is critical for improved understanding atmospheric composition. This unique channel feature allows us to perform an unprecedented global assessment of such channels in combination with other traditional sounding channels from the perspective of hydrometeor type identification. The MWHS-2 has a swath width of 2700 km, and its spatial resolution is about 16 km at nadir for 183 GHz and 29 km for 89 GHz. In this paper we used MWHS-2 Level-1B files downloaded from the CMA/NSMC (<http://satellite.nsmc.org.cn/portalsite/default.aspx>, accessed in 31 March 2018). These data were preprocessed using the methods described in Sections 2.2–2.5 and then used as input features to the deep-learning model. We can simply consider one column of the preprocessed MWHS-2 data to be one feature and the number of features are the data dimensions.

Table 2. MWHS-2 channel frequencies and polarization at nadir used in RTTOV.

Channel Number	Frequency (GHz)	Polarization at Nadir Used in RTTOV
1	89	H
2	118.75 ± 0.08	V
3	118.75 ± 0.2	V
4	118.75 ± 0.3	V
5	118.75 ± 0.8	V
6	118.75 ± 1.1	V
7	118.75 ± 2.5	V
8	118.75 ± 3.0	V
9	118.75 ± 5.0	V
10	150	H
11	183.31 ± 1.0	V
12	183.31 ± 1.8	V
13	183.31 ± 3.0	V
14	183.31 ± 4.5	V
15	183.31 ± 7.0	V

2.2. Ground Truth

We used the GPM Combined Precipitation L2B (2BCMB) product at matched scans (MS swath) to acquire the precipitation structure as the ground truth to train the model. The 2BCMB product uses data from DPR at Ku and Ka bands, and GPM microwave imager (GMI) [41]. The MS swath we used contains 25 rays per scan that match the 25 Ka-band precipitation radar (KaPR) rays. The resolution of the 2BCMB product is 5 km horizontally and 250 m vertically. The GPM core observatory has an orbit inclination of 65° which it allows the orbit to cut across the orbits of sun-synchronous satellites, including FY-3C. It also has the advantage of sampling at various times of the day at latitudes where most high-accumulation-precipitation events occur.

Each of the precipitation profiles in the GPM 2BCMB product is classified as liquid-phase, mixed-phase, or ice-phase precipitation. To determine the falling hydrometeor type, we first define the two different precipitation quantities as ice water content or liquid water content. The vertical integral of each of these two quantities represents the total mass of the corresponding precipitation phase in a vertical column, known as ice water path and liquid water path. For each profile, the falling hydrometeor is considered as liquid-phase (rain) if there is zero ice water path, and ice-phase (snow) if there is zero liquid water path. Otherwise, it is mixed-phase (sleet). To ensure the input data are as unbiased as possible, we performed several conventional preprocesses to clean and prepare datasets, including collocate coincidental measurements, bias removal, and noise cleaning that will be described in detail in Sections 2.3–2.5, respectively.

2.3. Collocated and Coincidental Measurements

Due to the more complex simulation of surface emissivity over land, in this study we focus on oceans where the surface emissivity is generally homogenous and well-characterized. To project the MWHS-2 observations and DPR profiles into the same geographical fields, we divide the whole globe into individual grids with each grid being 0.25° latitude $\times$ 0.25° longitude. As a result, we count a collocated and coincidental measurement, or a matchup when a MWHS-2 observation and a DPR profile fall into the same grid and their observing time difference is within 15 min time window. This process is summarized in Table 3 along with the data sources and preprocessing techniques, which were elaborated previously in Sections 2.1 and 2.2. Herein, we use Figure 1 as an example to show single orbital track overlap between MWHS-2 (gray) and DPR (black). The color bar is the time difference in between ranging from 0 to 15 min. In other words, only the colorful overlap can be considered as useful collocations before applying other criteria. The DPR footprint is 5 km which is much smaller than the grid size, therefore, dozens of DPR observations fall into one 25 km grid. As a result, we average all the DPR profiles within one grid to represent the mean over this grid. On the contrary, the MWHS-2 footprint is about 29 km for channels 1–9 at nadir, which is comparable to the preselected grid size of 25 km. Consequently, only a few, typically less than three, MWHS-2 pixels fall into the same grid. To avoid the sparse sampling issue, instead of averaging as for DPR, we take the MWHS-2 measurements from the pixel closest to the center of the grid in distance to represent that grid. The cross-tracking scanning mechanism of the MWHS-2 provides measurements with varying footprint size and Earth incidence angle across the scanline. The accuracy at the two outer edges of a scan from a cross-scan radiometer such as MWHS-2 is usually not as high as the rest of the scanline due to the high zenith angles and the low spatial resolutions, to ensure the high training-data quality, the five outermost scan positions on each side of the scanline are disregarded from the collocated dataset. By collocating the whole year of 2017 data, we eventually obtained a total of more than 1.5 million samples over oceans. We used a commonly used data-splitting ratio of 8:1:1 to divide this dataset, which means 80% of the data was for training, 10% for testing, and 10% for validation.

Table 3. Data sources and preprocessing.

Data Source	Passive Measurements from MWHS-2	Retrieval Profiles from DPR	Simulated TBs from RTTOV
Resolution (km)	16–29 depend on channels	5	N/A
Gridding method	Measurements of the pixel closest to the center of the grid	Average of all profiles over the grid	N/A
Sampling resolution (km)		25	
Time difference (min)		15	
Year	2017 for model training/testing; 2016 for model validation		

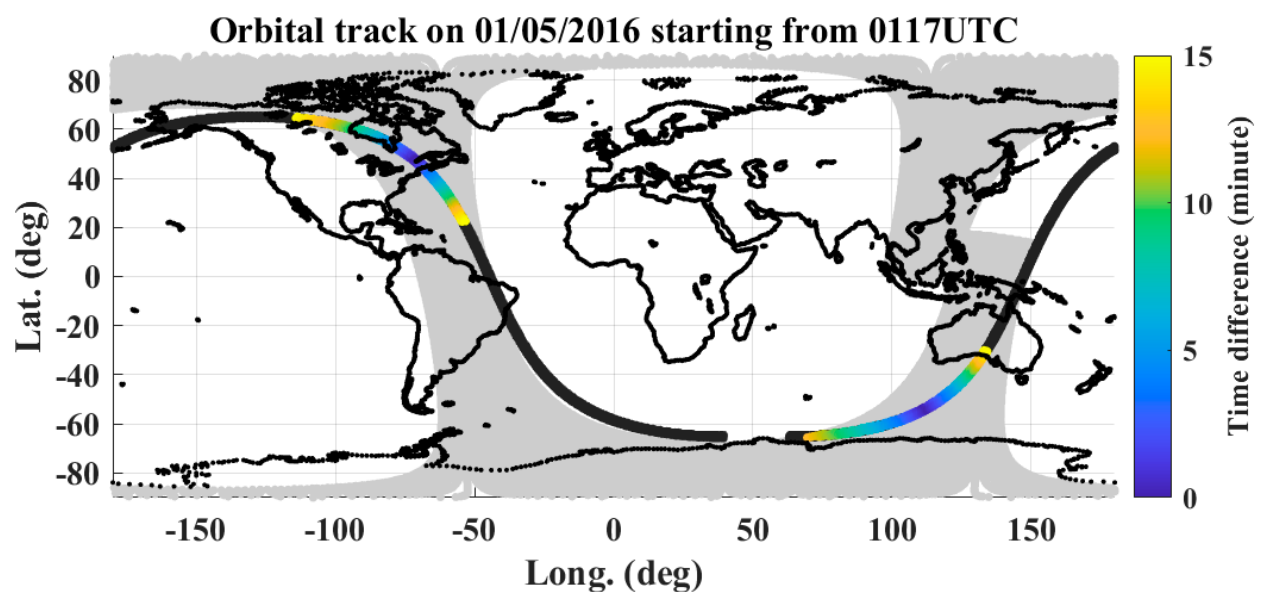


Figure 1. Comparison of Orbit Tracks between MWHS-2 and DPR on 5 January 2016. The gray line represents one orbit track of MWHS-2, while the black line represents that of DPR. The color bar indicates the time difference between the two, ranging from 0 to 15 min.

2.4. Bias Correction

It is important for any model training that the input data are bias-free to ensure that the trained model is as unbiased as possible. In consequence, we applied a bias-correction method developed by Chen and Bennartz [35] to correct the MWHS-2 observations based on clear-sky radiative transfer simulations from RTTOV (radiative transfer for TOVS). This bias-correction method is based on the fact that most observations are non-precipitating and being impacted at a minimum level. Therefore, the data value that occurs the most often in the data set, which is the mode of the histogram of the brightness temperature (TB) differences between observations and clear sky simulations, can be treated as an estimated bias.

2.5. Data Sub-Setting

To ensure the best possible dataset for model training, we implemented a data-cleaning process. This process is critical in creating a reliable and high-quality dataset for accurate relationships acquisition. There are two aspects need to be considered: noisy and duplicated data. Bennartz and Bauer previously found that passive microwave radiances at 150 GHz channel are highly sensitive to changes in ice scattering using radiative transfer simulations at 90, 150, and 183 GHz [42]. Based on the DPR retrievals, the vast majority (84%) of the over 1.5 million collocated data are non-precipitating. Therefore, this largely skewed non-precipitating data can be considered as noise and should be removed as fully as possible.

To do this, we defined a scattering index as the difference between MWHS-2 observed and simulated TBs at 150 GHz and any data with a scattering index larger than 2 K are considered as non-precipitating and, therefore, noisy. As a result, 56% of the data were removed, of which 88% were found to be non-precipitating.

Despite the data-cleaning process, a significant proportion of the remaining data are still non-precipitating, accounting for 80% of the dataset. This severe class imbalance is a common challenge in classification problems such as spam detection and fraud detection, where the minority class (liquid, mixed, or ice hydrometeor) is of more interest. A naïve model tends to focus on exclusively learning the characteristics of the excessive observations (non-precipitating observations) and underemphasizes the cases from the minority class that are of more interest and whose projections are more needed. This class imbalance poses a difficult problem for accurate classification, and must be addressed on a case-by-case basis. To address this, we randomly selected 10% of the non-precipitating data (removing 90%), resulting in a final training dataset composed of 20% liquid, 35% mixed, 13% ice, and 32% non-precipitation.

3. ResNet-18 Network by Attention Mechanism

In this section, we present the enhanced ResNet-18 architecture using stacked convolutional and self-attention modules for hydrometeor classification. While there are several other CNN architectures, such as VGG16, Densenet, and Mobilenet, we conducted extensive experimentation and evaluation and found that ResNet-18 provided the best performance for our specific problem of classifying hydrometeors using passive microwave observations. Furthermore, recent research has shown that ResNet variants with self-attention have achieved state-of-the-art results in various classification tasks [43–46]. Figure 2 illustrates the entire workflow of our proposed architecture, where ResNet-18-based neural networks are employed, comprising stacked convolutional layers with varying filter sizes. ResNet-18 is a variant of ResNet families, and also an ensemble of 18 layers of residual nets to reduce prediction error on the ImageNet test set [39]. As shown in Figure 3a, the input of MWHS-2 passive microwave observation features is converted into vector representations by a layer of convolutional embedding, and then fed into a stack of bottleneck blocks to exploit spatial and temporal knowledge in MWHS-2 observations for predicting hydrometeors types. To capture spatial knowledge of observations between a channel and its neighboring ones, we extend the functionality of the vanilla bottleneck block in ResNet-18 by (1) using a context-attention layer to replace spatial convolution, (2) embedding a channel-attention module before the last down-sampling convolution layer. To improve the performance of the model and stabilize its output, we repeat the modified bottleneck layer multiple times, in which an output of one block is fed into the next one. Unless otherwise specified, we use the notation of “ $N \times N$ ” to specify the size of a filter in a convolutional layer in the following sections and figures. For example, “ 3×3 ” in Figure 3 indicates a convolutional module with a 3×3 filter. Finally, the output probabilities of hydrometeor types are produced by a generator model comprising a linear neural network and a softmax activate function. In the implementation process, we split the data into training, testing, and validation in a ratio of 8:1:1 to build and evaluate our proposed model. Furthermore, we leverage the early stopping strategy to avoid issues with model overfitting during training process [47]. We will depict all sub-components in the remainder of this section.

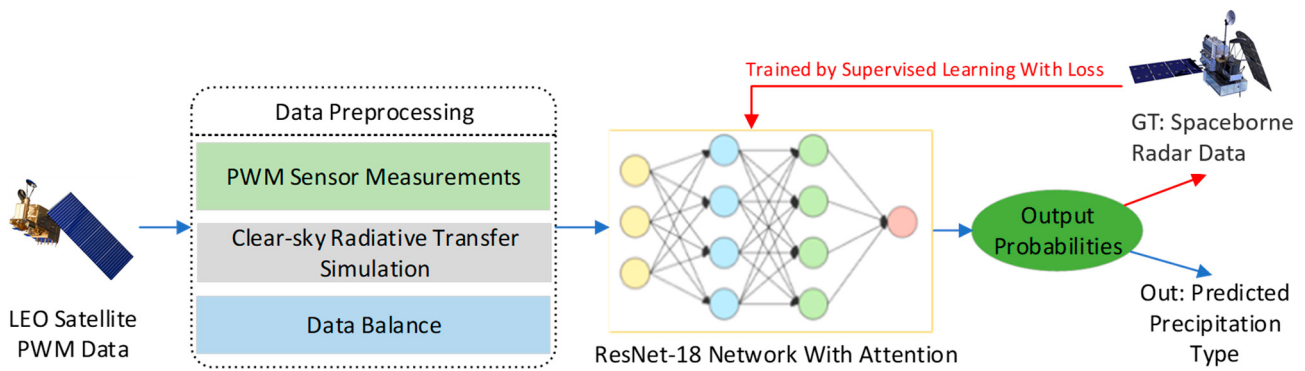


Figure 2. Proposed model architecture for hydrometeor classification using satellite-borne passive microwave observations.

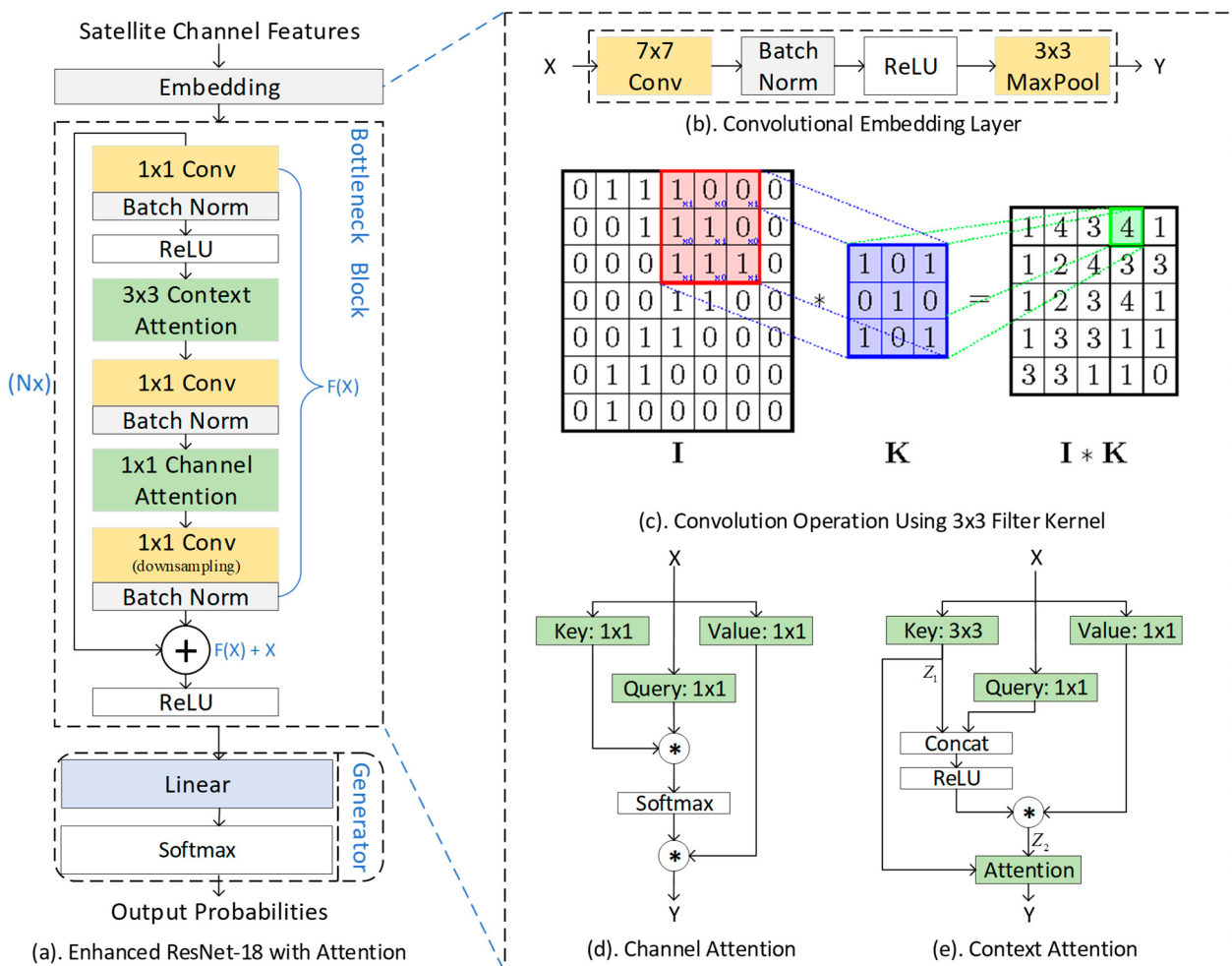


Figure 3. A ResNet-18 model enhanced by context and channel self-attention.

3.1. Model Configuration

Typically, according to Equation (1) and softmax-based weight function $\alpha(Q, K) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)$, the matrix production QK^T has a cost $O(n^2 \times d)$ if assuming $Q, K \in \mathbb{R}^{n \times d}$. Therefore, the complexity of the self-attention layers becomes $O\left(\left(\frac{n}{k}\right)^2 \times d\right)$ in the context and channel self-attention modules, where k is the kernel size of the convolution layer. In addition, the complexity of the initial convolution is $O(n \times d^2)$. Hence, the

overall complexity becomes $(n \times d^2 + (n/k)^2 \times d)$. In this paper, we implemented the algorithm using Pytorch 1.7.1 and trained it using Adam optimizer [48,49]. During training, we leveraged early stopping with 10 patience to avoid overfitting according to model performance on the validation dataset. To be consistent with the experimental settings of baselines, we conducted both training and testing on NVIDIA RTX 2080Ti GPU. The host server is configured with Intel(R) Xeon(R) Gold 6130 CPU @ 2.10GHz and 503 GB memory.

In addition, we leveraged the Adam optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.98$ to compute and update gradients, and cross-entropy is used as loss function. Notably, instead of using hard label directly, we used a regularization technique called label smoothing with ratio 0.1. It computes cross-entropy, not with hard labels, but with a weighted mixture of hard labels with a uniform distribution, to weight average of the hard labels and the uniform distribution on the labels. In addition, we set up the learning rate and dropout as 0.001 and 0.2.

3.2. Convolutional Embedding Layer

As discussed in Section II, MWHS-2 TBs of 12 channels and the corresponding relative airmass were combined and represented as a low-level input feature with original values to the proposed model. Technically, we needed to convert the input matrix to high-level space vectors of 3D dimension with the following considerations: (1) the proposed model works on vectors of 3D dimension directly, rather than 2D matrix; (2) transforming the inputs into a new and larger high-level space without losing too much primary features can result in a better performance. To this end, we followed the similar strategies used in ResNet-18 to process the input matrix, as shown in Figure 3b. More specifically, we first slid (more precisely, convolved) a filter with size of 7×7 across the width of the input (the height is equal to 1 in our case) and performed a convolution operation between the entry of the filter and the input at the corresponding position to summarize the presence of those features in the input which are known as feature maps. An example convolution operation using 3×3 filter is shown in Figure 3c. Typically, a sliding box with the size of 3×3 over the input matrix filters an input block with identical size (blue color area in the input), and then we performed dot products with the entry of the filter to produce an output of feature maps (the output of first cell is 16 in this case). Similarly, the sliding box was moved across the width with a specified step size (the value of stride was equal to 2 in this case) to obtain all values of feature maps. With consideration of updating the gradients of previous layers in a backward way, it was a challenging and complicated task to train a deep neural network since parameters of previous layers changed during training. Therefore, we adopted several strategies to reduce the number of training epochs and stabilize the learning process. For example, batch normalization technique was used to standardize the inputs to a layer for each mini batch [50].

Once a feature map was created and standardized by batch normalization technique, a nonlinearity function was applied on the feature map to approximate such a nonlinear relationship in the underlying data. We used a rectified linear activation function (ReLU), much as we do for the outputs of a fully connected layer [50]. Technically, the ReLU is defined as a max selector between 0 and x : $f(x) = \max(0, x)$. In other words, it returns 0 if it receives any negative input, otherwise returns any positive value x . Usually, the output feature maps were sensitive to the location of the features in the input. In the domain of convolutional network, pooling layers were utilized to address this sensitivity by down-sampling the feature maps. There are two practical pooling methods: average and max pooling layers. The former summarizes the average of a feature while the latter chooses the most activated presence of a feature. In this paper, we applied a max pooling layer using a 3×3 filter on the output of ReLU to reduce the size of each feature map by a factor of 3.

3.3. Bottleneck Residual Block

In this module, we depicted and designed a novel bottleneck block for hydrometeor classification. As shown in Figure 3a, there was a stack of five layers for each residual

function F : 1×1 , 3×3 , 1×1 , 1×1 , and 1×1 convolutions. The first and third layers were convolutional layers using 1×1 filters to reduce and then increase (restore) dimensions. The second layer was designed to capture spatial knowledge of observations at each channel using a 3×3 convolutional layer. In our implementation, we also expanded its capability to exploit contextual information around each channel by using attention mechanism (explained in the next subsection). The fourth layer performed a global attention mechanism among channels to emphasize the relationships among each channel. In the last layer, we conducted down-sampling directly by convolutional layers with 1×1 filter. In order to address the degradation problem of a complicated and deeper neural network and improve accuracy, we simply performed an identity mapping between the input and the end of the stacked layers, and added their yields to the outputs of the stacked layers: $F(X) + X$. Then, a layer of ReLU was followed up to approximate such a nonlinear relationship in the underlying data. We repeated the modified bottleneck layer multiple times (in our case), with which the output of one block was being used as the input of the next block.

3.4. Attention Mechanism

There is a growing number of attention-style neural designs with competitive results in numerous tasks of various fields, such as the domain of natural language processing and computer vision [51,52]. Recently, earth observation studies have also benefited from its success in enhancing model prediction accuracy. Qiao et al. proposed an novel algorithm that combines an attention mechanism with recurrent neural networks to predict future sea-surface temperature (SST) using historical SST data, and experimental results showed that it outperformed other SST prediction approaches [37]. Nevertheless, none of existing deep-learning-based algorithms employ attention over satellite passive microwave observations to exploit contexts among neighboring channels for improving accuracy of the hydrometeor classification task. In this paper, we designed two novel context and channel-attention modules and orchestrated them with the core bottleneck block in ResNet-18 elaborately to capture spatial knowledge of the microwave radiances around neighbor channels.

An attention mechanism maps a query Q for a set of key-value pairs (K, V) to produce a sum of weighted values and its formal equation is defined as:

$$\text{Attention}(Q, K, V) = \alpha(Q, K) \times V \quad (1)$$

where $Q, K, V \in N \times d_k$, N is the number of input observations and d_k is the dimension of features. In addition, the weights $\alpha(Q, V)$ for V are computed following a compatibility function of the query with the corresponding key [53–55].

As shown in Figure 3d, we conducted interactions across different spatial feature locations in the channel-attention module. Specifically, the input X was transformed into Q , K , and V using three separated $1D$ 1×1 convolutions, respectively. Next, we performed a dot product between Q and K , divided each resulting element by their dimension size d_k , and then applied a softmax function to obtain the corresponding weights for all values in V : $\alpha(Q, K) = \text{softmax}(QK^T / \sqrt{d_k})$. Eventually, the output Y was achieved using $\alpha(Q, K) \times V$.

There are advantages and disadvantages by applying a channel-attention module over feature maps. While better performance can be achieved, it lacks scalability and does not consider contextual information among neighboring keys because it handles queries and keys as a group of isolated pairs and investigates their pairwise relationships individually without learning the contexts between them. As a result, we additionally designed a novel local attention for inertial navigation, considered as a context attention module, as shown in Figure 3e. We first employed 3×3 group convolution over all the neighbor keys within a grid of 3×3 to extract local contextual representations for each key, denoted by $Z_1 = XW_{K,3 \times 3}$. Then, we investigated a form of concatenation-based weight function:

$$\alpha(Q, Z_1) = \text{ReLU}(W_\alpha[Q; Z_1]) \quad (2)$$

where $[\cdot]$ denotes a concatenation operation on input vectors and W_α was a weight vector that projected the concatenated vector to a scalar [56]. Next, we computed the attended feature map Z_2 using $(Q, Z_1) \times V$, through which it captured the global contextual information among all observations. We adopted an attention mechanism between local context Z_1 and global context Z_2 to produce a result. Further, in order to allow the model to learn and summarize information jointly, we split the input into multi-heads to represent subspace features at different spatial positions.

3.5. Precipitation Generator

We used the usual learned linear transformation and softmax function to convert the bottleneck output to predicted precipitation probabilities [54]. Technically, the softmax function $\sigma: \mathbb{R}^K \rightarrow [0, 1]^K$ takes as input a vector z of K real numbers (i.e., the number of hydrometeor type in the case of generator) and normalizes it into a probability distribution consisting of K probabilities proportional to the exponentials of the input numbers, which is defined as follows:

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \text{ for } i = 1, \dots, K \quad (3)$$

$$z = (z_1, \dots, z_k) \in \mathbb{R}^K \quad (4)$$

The output of the softmax function represents a categorical distribution over hydrometeor class labels, and we can obtain the probabilities of each input element belonging to a label.

4. Experimental Setting and Results

4.1. Model Training

Following the data preprocessing described in Section 2, about 1.5 million collocations were produced for the full year of 2017 which were then further enhanced with data noise cleaning and data balance checking. 80% of the resulting data were used for training to establish the deep-learning-based model describing the relationships between the microwave observations and hydrometeor types, 10% were for testing, and 10% were for validating the model, as illustrated in Section 3. Another full year of data from 2016 was used to validate the model by comparing predicted results with benchmark DPR data.

We used MWS-2 TBs and relative airmass, defined as $1/\cos(\text{zenith angle})$, as input features, and hydrometeor types from DPR as output targets to build a CNN model that projects the relationships between the inputs and outputs. The relative airmass can account for the atmospheric attenuation due to slant path variation. As we needed to understand how much the different features contribute and then choose the optimal combination of features, we first used 10 different sets of MWS-2 channels in the input features to train the model. The details of these feature sets are listed in Table 4 which strategically displays observations from 90, 118, and 183 GHz channels, relative airmass, and freezing level altitude. All 15 MWS-2 channels were used in V1 along with relative airmass. V3–V10 are only different from V1 in their removal of certain channels or relative airmass from input features. As a result, the difference in model accuracies allowed us to determine the impact of the removed features on the model training. For example, compared with V1, V3–V6 miss channels at 89, 150, 118, and 183 GHz, respectively. In addition, the freezing-level altitude was added into V1 and V9, respectively, to create the configurations of V2 and V10, and the resulting two highest average true rates for all three hydrometeor types shown in Figure 4 demonstrate that the freezing-level altitude can increase the model accuracy. Because the focus of this paper is to explore the passive microwave observations, in the rest of this study we leave out the freezing-level altitude and will continue this investigation in our future study by adding ancillary data, including freezing-level altitude. Figure 4 shows the true/false rates of liquid, mixed, and ice precipitation using testing data for 10 different combinations of features. L, M, and I are short forms of liquid, mixed, and ice, respectively. As such, LL is the liquid true positive rate, LI is the liquid false negative rate that is falsely identified as ice, and LM is the liquid false negative rate that is falsely identified

as mixed. The sum of LL, LI, and LM is 100%. Eventually we selected V9 which achieved the optimal combination of model accuracy and performance included all 12 MWS-2 channels (excluding high-peaking channels 2–4 around 118 GHz) and relative airmass. Our feature selection process aligned with previous research findings. For instance, Chen and Bennartz [35] identified that the same set of channels can be utilized for precipitation retrieval as they are responsive to ice particle scattering to varying degrees. These findings are consistent with earlier work by Bennartz and Petty [27] conducted more than 20 years ago when observations were not as readily available as they are now. The prior research on this topic, combined with our current study, further highlights the importance of these specific channels for precipitation retrieval and underscores the potential utility of passive microwave observations for hydrometeor classification.

Table 4. Feature selection of MWS-2 channels, relative airmass, and freezing-level altitude.

Version	Channels of TBs	Relative Airmass	Freezing Level
1	1–15	Y *	N *
2	1–15	Y	Y
3	2–15	Y	N
4	1–9, 11–15	Y	N
5	1, 10–15	Y	N
6	1–10	Y	N
7	2–9, 11–15	Y	N
8	1, 10	Y	N
9	1, 5–15	Y	N
10	1–15	N	Y

* Y denotes the feature is used and N denotes that the feature is not used in this version.

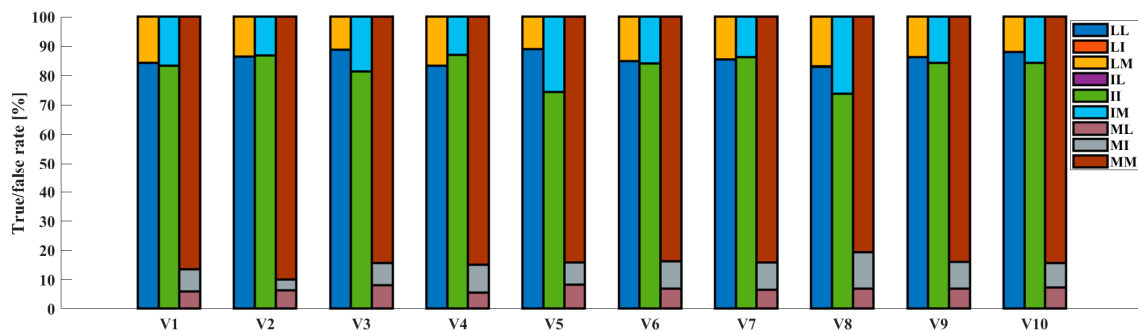


Figure 4. True and false classification rates for liquid, mixed, and ice precipitation using different feature combinations. Here, L, M, and I represent liquid, mixed, and ice precipitation, respectively. For instance, LL denotes the true positive rate for liquid precipitation, LI represents the false negative rate where liquid is mistakenly identified as ice, and LM signifies the false negative rate where liquid is incorrectly labeled as mixed. The sum of LL, LI, and LM totals 100%. It is worth noting that the values of LI and IL, being in close proximity to zero, are not visually discernible in the figure.

The corresponding confusion matrix of the classification results using the testing data on the trained model are shown in Table 5, to visualize the performance of the proposed approach. Each row of the matrix shows actual instances of classification labels while each columns represents the predicted ones, or vice versa. The true positive rates of all three classes, the number of positive samples identified by the proposed approach, are above 83%, and the average accuracy is 84.7%. The false positive cases of both liquid and ice precipitation are all misclassified as mixed with 0% ice or liquid. Mixed precipitation is incorrectly classified into liquid or ice with very close false positive rates (6.0% and 7.5%, respectively).

Table 5. Confusion matrix of the CNN classification results using testing data.

Precip. Type	Liquid	Mixed	Ice
Liquid	8902 (84.3%)	1465 (6.0%)	1 (0.0%)
Mixed	1662 (15.7%)	21,124 (86.5%)	2002 (16.7%)
Ice	0 (0.0%)	1829 (7.5%)	10,006 (83.3%)

4.2. Model Validation

It is desirable to have a trained model that allows one to predict the hydrometeor types given a new set of MWS-2 data. Therefore, we fed the full year of 2016 MWS-2 observations into the trained deep-learning model and output the corresponding hydrometeor types into liquid, mixed, and ice phases. These data and the DPR-derived hydrometeor types were then both gridded into 5° latitude $\times$ 5° longitude boxes for comparison. For each grid the occurrences of liquid, mixed, and ice precipitation were counted and the corresponding occurrence fractions of these three types were calculated, which add into one. The top panel in Figure 5a–c are global occurrence fractions of liquid, mixed, and ice precipitation over ocean during January 2016 for MWS-2. Figure 5d–f are the corresponding occurrence fractions during the same time period but derived from DPR.

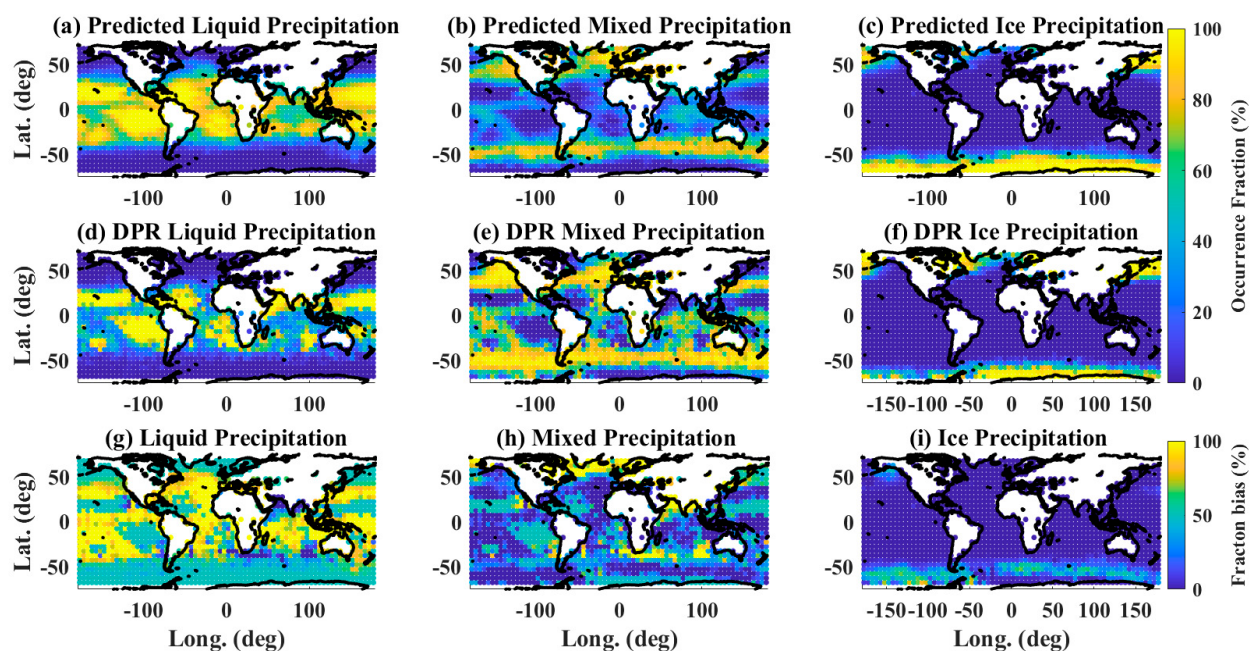


Figure 5. Global distribution of liquid, mixed, and ice precipitation occurrence fractions over the ocean in January 2016. The top panel shows the predicted results using our CNN model, while the middle panel displays the corresponding fractions obtained from DPR data. The bottom panel shows the difference between the two.

Their differences are shown in the bottom panel of Figure 5g–i. The occurrence fractions from MWS-2 are in good agreement with those from DPR. The vast majority of precipitations over tropical and sub-tropical are liquid. Over higher latitudes between 35° and 55° north and south, it is dominated by mixed precipitation, and ice precipitation is more likely to occur elsewhere. The liquid precipitation fractions from MWS-2 show over-estimation in areas such as the intertropical convergence zone (ITCZ) and the southern hemisphere. Reversely, the mixed precipitation shows the opposite trend of an under-estimation in those areas. The DPR-derived occurrence fractions are noisier than the MWS-2 derived ones because of the lower DPR data density due to its narrower swath width which is less than one 10th of MWS-2's (245 km compared with MWS-2's 2700 km

as shown in Figure 1). The biases of averaged occurrence fractions for different hydrometeor types are listed in Table 6 which supports that the accuracy of ice precipitation fraction is the highest among the three hydrometeor types, and that the liquid precipitation fraction is overestimated while the mixed is underestimated. The biases of averaged occurrence fraction of the three hydrometeor types are added into zero.

Table 6. Averaged occurrence fraction bias for different precipitation types.

Metrics	Precip. Type		
	Liquid	Mixed	Ice
Bias	10.5%	−12.1%	1.5%
Variance	10.3%	5.9%	1%

Another evaluation technique compares zonal mean likelihood of different hydrometeor types from MWS-2 with respect to DPR. The likelihood is defined as the average occurrence fraction over a given latitude and it indicates the chance of a certain type of precipitation occurs at that latitude. The mean zonal likelihoods of the three hydrometeors types over ocean for the whole year of 2016 within latitudes (-67.5° , 67.5°) are presented in Figure 6. The comparison between MWS-2 (top) and DPR (bottom) shows that MWS-2-derived hydrometeor types are in line with those from DPR that it is likelier to have liquid precipitation over tropical and subtropical regions and that ice precipitation occurrence is more likely over higher latitudes. The liquid phase is overestimated by averagely 11.1% while mixed is underestimated by 11.3%. The average bias for ice-phase precipitation is 0.15%, which is extremely low. The correlation between MWS-2 and DPR is 0.95, 0.68, and 0.85 for liquid, mixed, and ice phases, respectively. The correlation using all the data is 0.83.

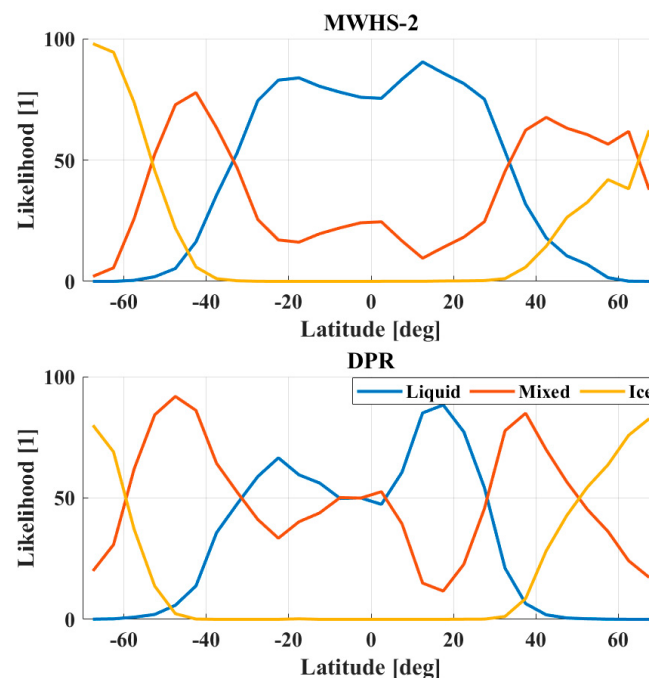


Figure 6. Zonal mean likelihood (occurrence fractions) for liquid, mixed, and ice precipitation over the ocean for the entire year of 2016. The top panel shows results from MWS-2, and the bottom panel shows results from DPR. The blue, red, and yellow lines represent the average occurrence fractions of liquid, mixed, and ice precipitation, respectively.

The annual cycles of the averaged fractions for liquid, mixed, and ice precipitation are also in good agreement with those from DPR, as shown in Figure 6. They both follow

the trend that ice precipitation occurrences decrease in summer months and increase in winter months. Their biases are lower in summer months and higher in winter months. In average, our predicted occurrences for liquid, and ice are overestimated with average biases/uncertainties of 8.5%/3.9% and 3.4%/2.5%. Predicted occurrences for mixed are underestimated with average bias and uncertainty being 11.8% and 6.1%.

5. Conclusions

This study has two major goals:

- (1) Utilizing CNN in conjunction with the attention mechanism to learn meaningful feature representations from spatial and temporal dimension space of passive microwave observations for hydrometeor classification;
- (2) Exploiting the information content of passive microwave observations for the purpose of hydrometeor classification with the unprecedented inclusion of 118 GHz channels.

To achieve these goals, we developed a new deep-learning-based algorithm using coincident MWHS-2 observations and GPM DPR estimates for training. The algorithm was composed of independent modules, in particular, convolutional and attention modules, for learning the non-linear relationships between the input and output and for exploiting contexts among neighboring channels.

In addition to developing a classification algorithm, this study also investigated the information content of the different microwave channels ranging from 89 to 190 GHz. Of all the channels, the three highest-peaking channels around 118 GHz (channels 2–4 of MWHS-2) were demonstrated to be the least significant which aligns with previous findings. This algorithm has been validated on a different full year of MWHS-2 observations. The prediction of hydrometeor types for this full year shows high agreement with state-of-the-art hydrometeor types from the GPM DPR measurements through the combined algorithm (2BCMB). The global geographical distributions of occurrence fractions for different hydrometeor types show overestimation in some areas, such as the ITCZ for liquid precipitation, and underestimation for mixed precipitation. The differences in zonal mean likelihood of both ice and mixed precipitation occur in higher latitudinal regions.

This work is part of the development of precipitation retrieval algorithms for the upcoming TROPICS mission that consists of a constellation of CubeSats, each of which carries a high-performance radiometer. In particular, the similarities in channel configuration between MWHS-2 and TROPICS radiometers cause the former to be an appropriate substitute to provide passive microwave observations in the higher microwave spectrum (89–190 GHz) for hydrometeor type classification.

Discussions

While the main purpose of this study is to consider the use of spaceborne passive rather than active microwave measurements for hydrometer classification, we acknowledge that the capabilities of the proposed deep-learning-based algorithm are limited without further verification using independent ground-based references, such as radar networks. Despite the potential benefits of ground-based radar measurements, the sparse distribution of such instruments causes global evaluation to be challenging [57,58]. Nevertheless, we remain committed to pursuing this goal in future research. Our future work related to this study will also involve testing the potential of CNN in conjunction with attention mechanism for improved accuracy of hydrometeor classification tasks by incorporating ancillary information into the input data, including freezing-level altitude and temperature and humidity profiles. Preliminary analysis of adding freezing altitude data into model features shows promising results. We are aware of other machine-learning algorithms, such as inductive logic programming and Bayesian networks, which may also be applicable to this topic. Exploring such algorithms is beyond the scope of this paper and may be better suited for future work.

Author Contributions: Conceptualization, R.C. and R.B.; methodology, R.C. and R.B.; software, R.C.; validation, R.C. and R.B.; formal analysis, R.C.; investigation, R.C.; resources, R.C. and R.B.; data curation, R.C. and R.B.; writing—original draft preparation, R.C.; writing—review and editing, R.C. and R.B.; visualization, R.C.; supervision, R.B.; project administration, R.B.; and funding acquisition, R.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by NASA grant number NNX17AJ09G to Vanderbilt University.

Data Availability Statement: Not applicable.

Acknowledgments: The authors thank the NASA Precipitation Processing System (PPS) for providing the GPM 2BCMB data and the China Meteorological Administration/National Satellite Meteorological Center (CMA/NSMC) for providing MWHs-2 L1B data for this research. The authors further acknowledge the support of EUMETSAT and the UK Met Office for providing and maintaining RTTOV.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Rosenfeld, D.; Lohmann, U.; Raga, G.B.; O'Dowd, C.D.; Kulmala, M.; Fuzzi, S.; Reissell, A.; Andreae, M.O. Flood or Drought: How Do Aerosols Affect Precipitation? *Science* **2008**, *321*, 1309–1313. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Trenberth, K.E.; Dai, A.; Rasmussen, R.M.; Parsons, D.B. The Changing Character Of Precipitation. *Bull. Am. Meteorol. Soc.* **2003**, *84*, 1205–1217. [\[CrossRef\]](#)
3. Atlas, D.; Ulbrich, C.W. Path- and Area-Integrated Rainfall Measurement by Microwave Attenuation in the 1–3 cm Band. *J. Appl. Meteorol. Climatol.* **1977**, *16*, 1322–1331. [\[CrossRef\]](#)
4. Crow, W.T.; Bolten, J.D. Estimating Precipitation Errors Using Spaceborne Surface Soil Moisture Retrievals. *Geophys. Res. Lett.* **2007**, *34*, L08403. [\[CrossRef\]](#)
5. Shrestha, B.; Cochrane, T.A.; Caruso, B.S.; Arias, M.E.; Wild, T.B. Sediment Management for Reservoir Sustainability and Cost Implications Under Land Use/Land Cover Change Uncertainty. *Water Resour. Res.* **2021**, *57*, e2020WR02835. [\[CrossRef\]](#)
6. Tuttle, S.; Salvucci, G. Empirical evidence of contrasting soil moisture-precipitation feedbacks across the United States. *Science* **2016**, *352*, 825–828. [\[CrossRef\]](#)
7. Allen, M.R.; Ingram, W.J. Constraints on Future Changes in Climate and the Hydrologic Cycle. *Nature* **2002**, *419*, 228–232. [\[CrossRef\]](#)
8. Gherardi, L.A.; Sala, O.E. Enhanced Interannual Precipitation Variability Increases Plant Functional Diversity that in Turn Ameliorates Negative Impact on Productivity. *Ecol. Lett.* **2015**, *18*, 1293–1300. [\[CrossRef\]](#)
9. Donat, M.G.; Alexander, L.V.; Herold, N.; Dittus, A.J. Temperature and Precipitation Extremes in Century-long Gridded Observations, Reanalyses, and Atmospheric Model Simulations. *J. Geophys. Res. Atmos.* **2016**, *121*, 11174–11189. [\[CrossRef\]](#)
10. Samantaray, A.K.; Ramadas, M.; Panda, R.K. Changes in Drought Characteristics Based on Rainfall Pattern Drought Index and the CMIP6 Multi-model Ensemble. *Agric. Water Manag.* **2022**, *266*, 107568. [\[CrossRef\]](#)
11. Sloat, L.L.; Gerber, J.S.; Samberg, L.H.; Smith, W.K.; Herrero, M.; Ferreira, L.G.; Godde, C.M.; West, P.C. Increasing Importance of Precipitation Variability on Global Livestock Grazing Lands. *Nat. Clim. Change* **2018**, *8*, 214–218. [\[CrossRef\]](#)
12. Hunsaker, C.T.; Whitaker, T.W.; Bales, R.C. Snowmelt Runoff and Water Yield Along Elevation and Temperature Gradients in California's Southern Sierra Nevada1. *JAWRA J. Am. Water Resour. Assoc.* **2012**, *48*, 667–678. [\[CrossRef\]](#)
13. Behrangi, A.; Yin, X.; Rajagopal, S.; Stampoulis, D.; Ye, H. On Distinguishing Snowfall from Rainfall using Near-surface Atmospheric Information: Comparative Analysis, Uncertainties and Uydrologic Importance. *Q. J. R. Meteorol. Soc.* **2018**, *144*, 89–102. [\[CrossRef\]](#)
14. Box, J.E.; Wehrlé, A.; van As, D.; Fausto, R.S.; Kjeldsen, K.K.; Dachauer, A.; Ahlstrøm, A.P.; Picard, G. Greenland Ice Sheet Rainfall, Heat and Albedo Feedback Impacts From the Mid-August 2021 Atmospheric River. *Geophys. Res. Lett.* **2022**, *49*, e2021GL097356. [\[CrossRef\]](#)
15. Loth, B.; Graf, H.-F.; Oberhuber, J.M. Snow Cover Model for Global Cimate Simulations. *J. Geophys. Res. Atmos.* **1993**, *98*, 10451–10464. [\[CrossRef\]](#)
16. Dai, A. Temperature and Pressure Dependence of the Rain-snow Phase Transition over Land and Ocean. *Geophys. Res. Lett.* **2008**, *35*, L12802. [\[CrossRef\]](#)
17. Slater, A.G.; Schlosser, C.A.; Desborough, C.E.; Pitman, A.J.; Henderson-Sellers, A.; Robock, A.; Vinnikov, K.Y.; Entin, J.; Mitchell, K.; Chen, F.; et al. The Representation of Snow in Land Surface Schemes: Results from PILPS 2(d). *J. Hydrometeorol.* **2001**, *2*, 7–25. [\[CrossRef\]](#)
18. Vivekanandan, J.; Zrnica, D.S.; Ellis, S.M.; Oye, R.; Ryzhkov, A.V.; Straka, J. Cloud Microphysics Retrieval Using S-Band Dual-Polarization Radar Measurements. *Bull. Am. Meteorol. Soc.* **1999**, *80*, 381–388. [\[CrossRef\]](#)

19. Liu, H.; Chandrasekar, V. Classification of Hydrometeors Based on Polarimetric Radar Measurements: Development of Fuzzy Logic and Neuro-Fuzzy Systems, and In Situ Verification. *J. Atmos. Ocean. Technol.* **2000**, *17*, 140–164. [\[CrossRef\]](#)
20. Lim, S.; Chandrasekar, V.; Bringi, V.N. Hydrometeor Classification System using Dual-polarization Radar Measurements: Model Improvements and In Situ Verification. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 792–801. [\[CrossRef\]](#)
21. Ryzhkov, A.V.; Schuur, T.J.; Burgess, D.W.; Heinselman, P.L.; Giangrande, S.E.; Zrnic, D.S. The Joint Polarization Experiment: Polarimetric Rainfall Measurements and Hydrometeor Classification. *Bull. Am. Meteorol. Soc.* **2005**, *86*, 809–824. [\[CrossRef\]](#)
22. Park, H.S.; Ryzhkov, A.V.; Zrnić, D.S.; Kim, K.-E. The Hydrometeor Classification Algorithm for the Polarimetric WSR-88D: Description and Application to an MCS. *Weather Forecast.* **2009**, *24*, 730–748. [\[CrossRef\]](#)
23. Scharfenberg, K.A.; Miller, D.J.; Schuur, T.J.; Schlatter, P.T.; Giangrande, S.E.; Melnikov, V.M.; Burgess, D.W.; Andra, D.L.; Foster, M.P.; Krause, J.M. The Joint Polarization Experiment: Polarimetric Radar in Forecasting and Warning Decision Making. *Weather Forecast.* **2005**, *20*, 775–788. [\[CrossRef\]](#)
24. Yang, J.; Zhao, K.; Zhang, G.; Chen, G.; Huang, H.; Chen, H. A Bayesian Hydrometeor Classification Algorithm for C-Band Polarimetric Radar. *Remote Sens.* **2019**, *11*, 1884. [\[CrossRef\]](#)
25. Lukach, M.; Dufton, D.; Crosier, J.; Hampton, J.M.; Bennett, L.; Neely, R.R., III. Hydrometeor Classification of Quasi-Vertical Profiles of Polarimetric Radar Measurements Using a Top-down Iterative Hierarchical Clustering Method. *Atmos. Meas. Tech.* **2021**, *14*, 1075–1098. [\[CrossRef\]](#)
26. Dolant, C.; Langlois, A.; Montpetit, B.; Brucker, L.; Roy, A.; Royer, A. Development of a Rain-on-snow detection Algorithm Using Passive Microwave Radiometry. *Hydrol. Process.* **2016**, *30*, 3184–3196. [\[CrossRef\]](#)
27. Bennartz, R.; Petty, G.W. The Sensitivity of Microwave Remote Sensing Observations of Precipitation to Ice Particle Size Distributions. *J. Appl. Meteorol. Climatol.* **2001**, *40*, 345–364. [\[CrossRef\]](#)
28. Petty, G.W.; Li, K. Improved Passive Microwave Retrievals of Rain Rate over Land and Ocean. Part I: Algorithm Description. *J. Atmos. Ocean. Technol.* **2013**, *30*, 2493–2508. [\[CrossRef\]](#)
29. Skofronick-Jackson, G.M.; Wang, J.R. The Estimation of Hydrometeor Profiles from Wideband Microwave Observations. *J. Appl. Meteorol. Climatol.* **2000**, *39*, 1645–1656. [\[CrossRef\]](#)
30. Wilheit, T.T. Some Comments on Passive Microwave Measurement of Rain. *Bull. Am. Meteorol. Soc.* **1986**, *67*, 1226–1232. [\[CrossRef\]](#)
31. Kedem, B.; Chiu, L.S.; North, G.R. Estimation of Mean Rain Rate: Application to Satellite Observations. *J. Geophys. Res. Atmos.* **1990**, *95*, 1965–1972. [\[CrossRef\]](#)
32. Klaes, K.D.; Cohen, M.; Buhler, Y.; Schlusser, P.; Munro, R.; Luntama, J.P.; von Engelin, A.; Clerigh, E.O.; Bonekamp, H.; Ackermann, J.; et al. An Introduction to the EUMETSAT Polar System. *Bull. Am. Meteorol. Soc.* **2007**, *88*, 1085–1096. [\[CrossRef\]](#)
33. Leppert, K.D.; Cecil, D.J. Signatures of Hydrometeor Species from Airborne Passive Microwave Data for Frequencies 10–183 GHz. *J. Appl. Meteorol. Climatol.* **2015**, *54*, 1313–1334. [\[CrossRef\]](#)
34. Chen, R.; Bennartz, R. Rainfall Algorithms Using Oceanic Satellite Observations from MWHS-2. *Adv. Atmos. Sci.* **2021**, *38*, 1367–1378. [\[CrossRef\]](#)
35. Chen, R.; Bennartz, R. Sensitivity of 89–190-GHz Microwave Observations to Ice Particle Scattering. *J. Appl. Meteorol. Climatol.* **2020**, *59*, 1195–1215. [\[CrossRef\]](#)
36. McCulloch, W.S.; Pitts, W. A Logical Calculus of the Ideas Immanent in Nervous Activity. *Bull. Math. Biophys.* **1943**, *5*, 115–133. [\[CrossRef\]](#)
37. Qiao, B.; Wu, Z.; Ma, L.; Zhou, Y.; Sun, Y. Effective Ensemble Learning Approach for SST Field Prediction Using Attention-based PredRNN. *Front. Comput. Sci.* **2022**, *17*, 171601. [\[CrossRef\]](#)
38. Sadeghi, M.; Asanjan, A.A.; Faridzad, M.; Nguyen, P.; Hsu, K.; Sorooshian, S.; Braithwaite, D. PERSIANN-CNN: Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks–Convolutional Neural Networks. *J. Hydrometeorol.* **2019**, *20*, 2273–2289. [\[CrossRef\]](#)
39. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Iecognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
40. Blackwell, W.J.; Braun, S.; Bennartz, R.; Velden, C.; DeMaria, M.; Atlas, R.; Dunion, J.; Marks, F.; Rogers, R.; Annane, B.; et al. An Overview of the TROPICS NASA Earth Venture Mission. *Q. J. R. Meteorol. Soc.* **2018**, *144*, 16–26. [\[CrossRef\]](#)
41. Hou, A.Y.; Kakar, R.K.; Neeck, S.; Azarbarzin, A.A.; Kummerow, C.D.; Kojima, M.; Oki, R.; Nakamura, K.; Iguchi, T. The Global Precipitation Measurement Mission. *Bull. Am. Meteorol. Soc.* **2014**, *95*, 701–722. [\[CrossRef\]](#)
42. Bennartz, R.; Bauer, P. Sensitivity of Microwave Radiances at 85–183 GHz to Precipitating Ice Particles. *Radio Sci.* **2003**, *38*, 8075. [\[CrossRef\]](#)
43. Zhang, H.; Wu, C.; Zhang, Z.; Zhu, Y.; Lin, H.; Zhang, Z.; Sun, Y.; He, T.; Mueller, J.; Manmatha, R. Resnest: Split-Attention Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 2736–2746.
44. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A.L.; Zhou, Y. Transunet: Transformers Make Strong Encoders For Medical Image Segmentation. *arXiv* **2021**, arXiv:2102.04306.
45. Srinivas, A.; Lin, T.-Y.; Parmar, N.; Shlens, J.; Abbeel, P.; Vaswani, A. Bottleneck Transformers For Visual Recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Montreal, Canada, 11–17 October 2021; pp. 16519–16529.

46. Li, Y.; Yao, T.; Pan, Y.; Mei, T. Contextual Transformer Networks For Visual Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 1489–1500. [[CrossRef](#)] [[PubMed](#)]
47. Prechelt, L. Early Stopping-but When? In *Neural Networks: Tricks of the Trade*; Springer: Berlin/Heidelberg, Germany, 1998; pp. 55–69.
48. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L. Pytorch: An Imperative Style, High-Performance Deep Learning Library. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 8026–8037.
49. Kingma, D.P.; Ba, J. Adam: A Method For Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
50. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In *Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015*; pp. 448–456.
51. Xu, B.; Wang, N.; Chen, T.; Li, M. Empirical Evaluation of Rectified Activations in Convolutional Network. *arXiv* **2015**, arXiv:1505.00853.
52. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)]
53. Bahdanau, D.; Cho, K.; Bengio, Y. Neural Machine Translation by Jointly Learning to Align and Translate. *arXiv* **2014**, arXiv:1409.0473.
54. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is All You Need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 6000–6010.
55. Wang, X.; Girshick, R.; Gupta, A.; He, K. Non-local Neural Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018*; pp. 7794–7803.
56. Santoro, A.; Raposo, D.; Barrett, D.G.; Malinowski, M.; Pascanu, R.; Battaglia, P.; Lillicrap, T. A Simple Neural Network Module for Relational Reasoning. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4974–4983.
57. Kidd, C.; Becker, A.; Huffman, G.J.; Muller, C.L.; Joe, P.; Skofronick-Jackson, G.; Kirschbaum, D.B. So, How Much of the Earth’s Surface is Covered by Rain Gauges? *Bull. Am. Meteorol. Soc.* **2017**, *98*, 69–78. [[CrossRef](#)] [[PubMed](#)]
58. Chen, H.; Chandrasekar, V.; Cifelli, R.; Xie, P. A Machine Learning System for Precipitation Estimation Using Satellite and Ground Radar Network Observations. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 982–994. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.