

Category: Computer Science / IT

Machine Learning Approach to Classify Precipitation Type from A Passive Microwave Sensor

Spandan Das¹, Jie Gong²

¹Thomas Jefferson High School for Science and Technology, 6560 Braddock Rd, Alexandria, VA, 22312

²Climate and Radiation Laboratory, NASA Goddard Space Flight Center, Mail Code 613, Greenbelt, MD, 20771, United States

Precipitation flag (precipitating or not; stratiform or convective) is a key parameter used for better retrieval of precipitation characteristics as well as to understand the cloud-precipitation physical processes. 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 flag assignment, while passive sensors like GMI are traditionally believed not to be differentiate between precipitation types. Using collocated precipitation flags from DPR as the “truth”, this project employs machine learning models to train and test the predictability and accuracy of using passive GMI-only observations together with ancillary atmosphere information from reanalysis. The satellite data used spans the course of a year, reducing the model’s potential bias towards any season. Precipitation types are classified into the following classes: convective, stratiform, convective-stratiform mixture, no precipitation, and other precipitation. To overcome the inherent imbalance among the precipitation categories, a simple dynamic sub-sampling technique is employed. A variety of classification algorithms are tested, including Support Vector Machines, Naive Bayes, Random Forests, Gradient Boosting, and Neural Networks (Multilayer Perceptron Network), and their results are evaluated and compared. The highest-performing trained models (Neural Network, Random Forest, and Gradient Boosting) have around 85% prediction accuracy for every type of precipitation when tested on holdout data. High-frequency channels (166 GHz and 183 GHz channels) and 166 GHz polarization difference are found among the most important factors that contribute to the model performance, which shed light on future instrument channel selection.

