

Unsupervised Clustering and Supervised Machine Learning for Lightning Classification: Application to Identifying EIPs for Ground-based TGF Detection

Yunjiao Pu¹, Steven A. Cummer¹, Fanchao Lyu², Yu Zheng², Michael S. Briggs³, Stephen Lesage³, Bagrat Mailyan³, and Oliver J Roberts⁴

¹Duke University

²Nanjing Joint Institute for Atmospheric Sciences

³University of Alabama in Huntsville

⁴USRA/NASA-MSFC

December 31, 2022

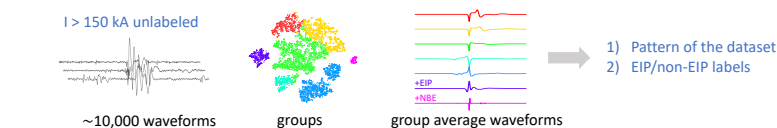
Abstract

We developed a framework merging unsupervised and supervised machine learning to classify lightning radio signals, and applied it to the possible detection of terrestrial gamma-ray flashes (TGFs). Recent studies have established a tight connection between energetic in-cloud pulses (EIPs, >150 kA) and a subset of TGFs, enabling continuous and large-scale ground-based TGF detection. However, even with a high peak current threshold, it is time-consuming to manually search for EIPs in a background of many non-EIP events, and it becomes even more difficult when a lower peak-current threshold is used. Machine learning classifiers are an effective tool. Beginning with unsupervised learning, spectral clustering is performed on the low-dimensional features extracted by an autoencoder from raw radio waveforms, showing that +EIPs naturally constitute a distinct class of waveform and 67% of the total population. The clustering results are used to form a labeled dataset ($\sim 10,000$ events) to further train supervised convolutional neural network (CNN) that targets for +EIPs. Our CNN models identify on average 95.2% of true +EIPs with accuracy up to 98.7%, representing a powerful tool for +EIP classification. The pretrained CNN classifier is further applied to identify lower peak current EIPs (LEIPs, >50 kA) from a larger dataset ($\sim 30,000$ events). Among 10 LEIPs coincident with Fermi TGF observations, 2 previously reported TGFs and 2 unreported but suspected TGFs are found, while the majority are not associated with detectable TGFs. In addition, unsupervised clustering is found to reflect characteristics of the ionosphere reflection height and its effect on radio wave propagation.

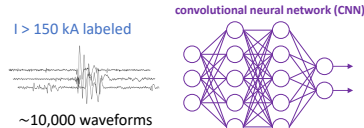
Hosted file

951907_0_art_file_10525651_rnbmhp.docx available at <https://authorea.com/users/569546/articles/615348-unsupervised-clustering-and-supervised-machine-learning-for-lightning-classification-application-to-identifying-eips-for-ground-based-tgf-detection>

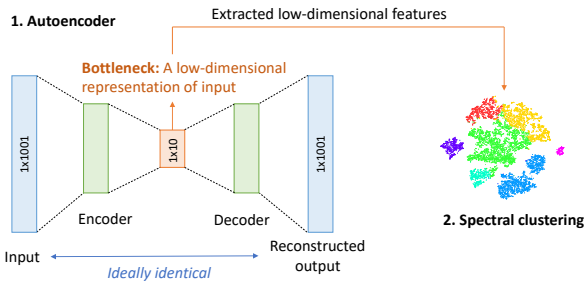
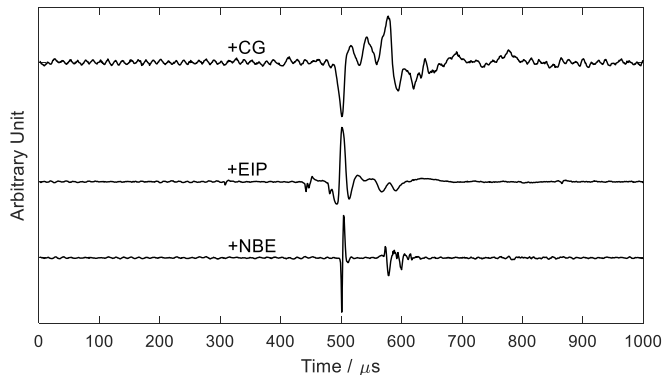
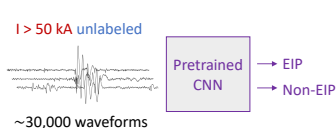
1. Unsupervised clustering, $I > 150$ kA

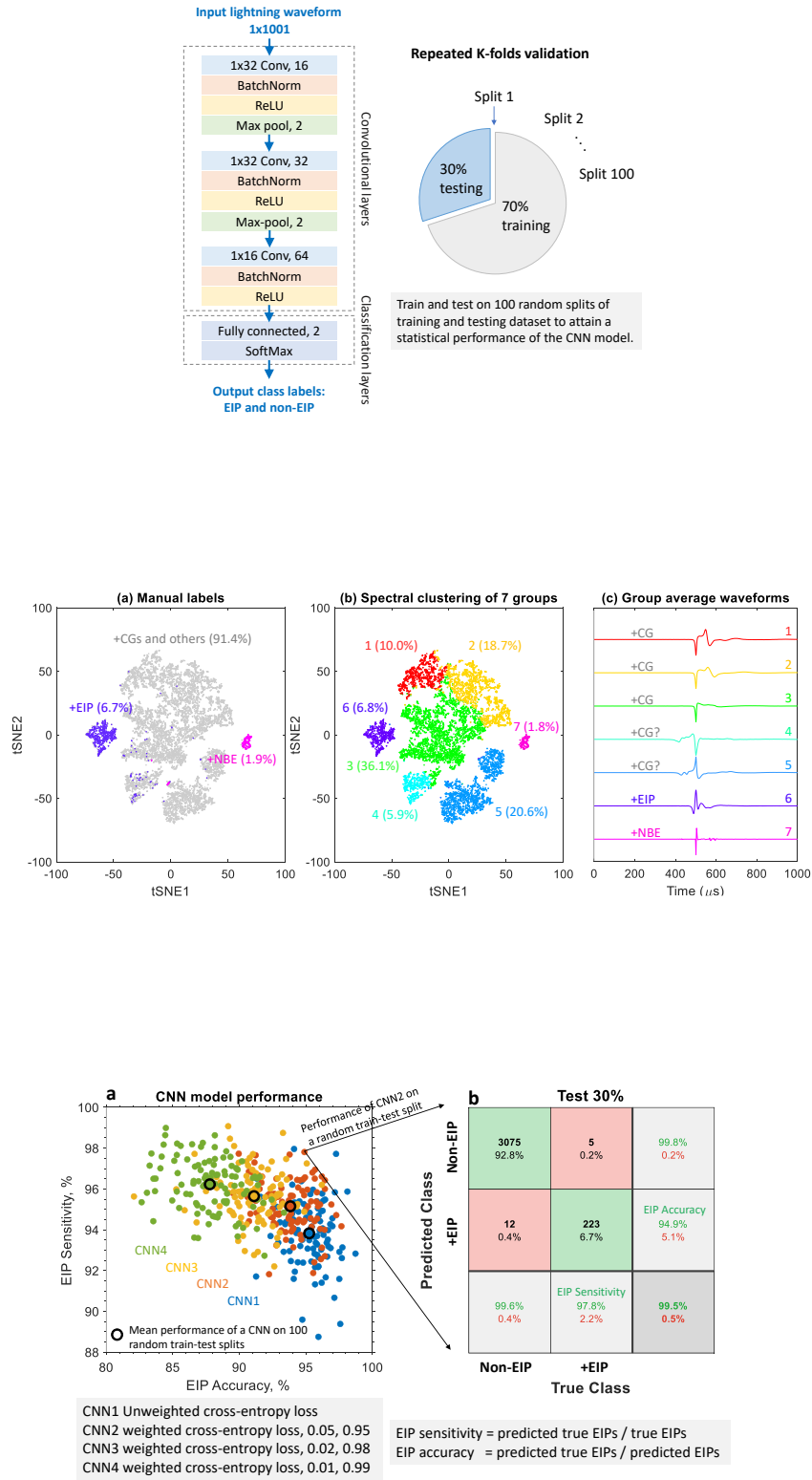


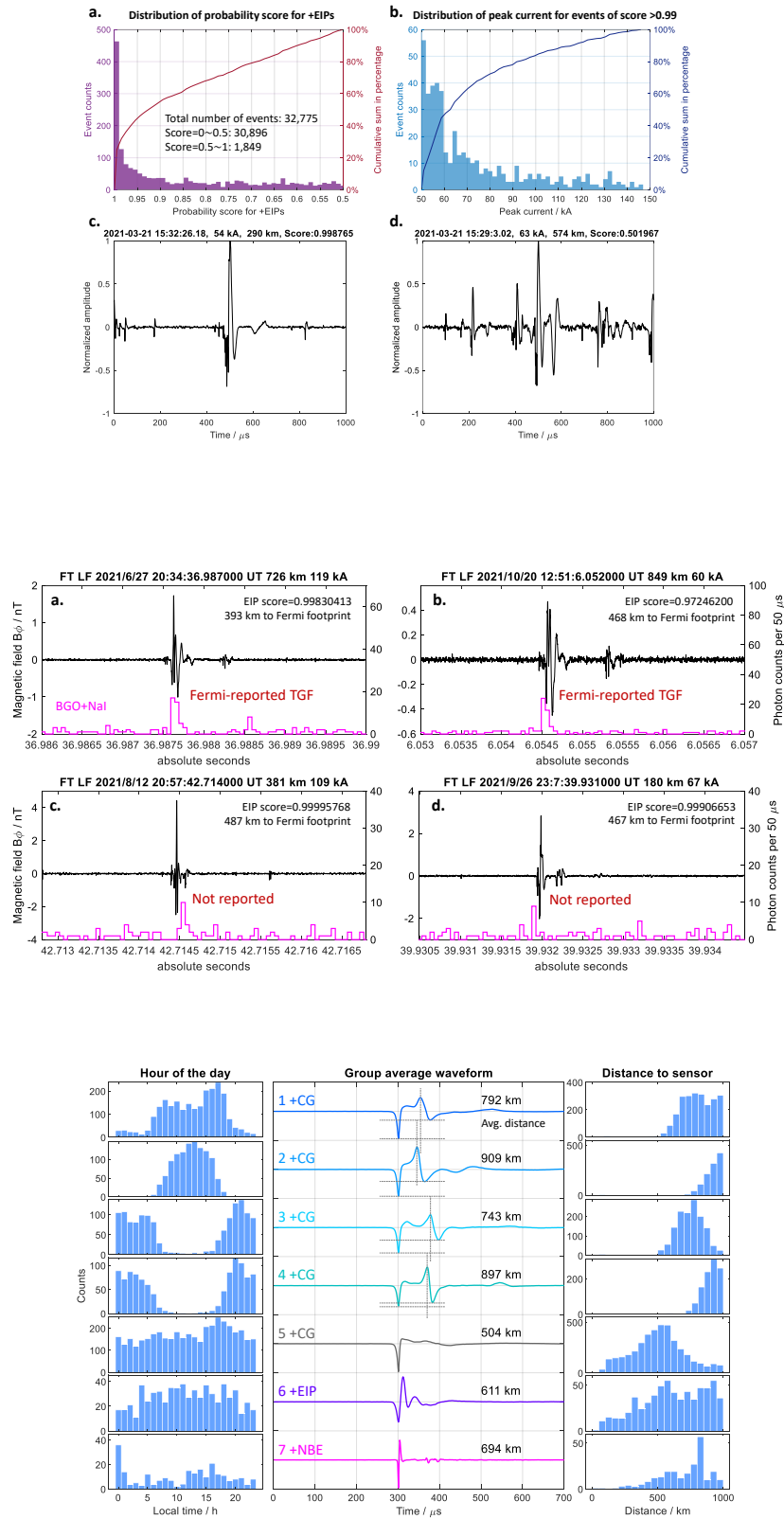
2. Supervised classification, train CNN, $I > 150$ kA



3. Predict using pretrained CNN, $I > 50$ kA







Hosted file

951907_0_supp_10525662_rmpjg2.docx available at <https://authorea.com/users/569546/articles/615348-unsupervised-clustering-and-supervised-machine-learning-for-lightning-classification-application-to-identifying-eips-for-ground-based-tgf-detection>