A data-driven approach for lightning nowcasting with deep learning

Amirhossein Mostajabi[1], Ehsan Mansouri[1], Pedram Pad[2], Marcos Rubinstein[3], Andrea Dunbar[2], and Farhad Rachidi[1]

[1]Electromagnetic Compatibility Laboratory, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland

[2]Centre Suisse d'Electronique et de Microtechnique (CSEM), Neuchâtel, Switzerland

[3]Institute for Information and Communication Technologies, University of Applied Sciences of Western Switzerland (HES-SO), Yverdon-les-Bains, Switzerland

Lightning is responsible directly or indirectly, for significant human casualties and property damage worldwide. [1,2] It can cause injury and death in humans and animals, ignite fires, affect and destroy electronic devices, and cause electrical surges and system failures in airplanes and rockets.[3–5] These severe and costly outcomes can be averted by predicting the lightning occurrence in advance and taking preventive actions accordingly. Therefore, a practical and fast lightning prediction method is of considerable value.

Lightning is formed in the atmosphere through the combination of complex dynamic and microphysical processes, making it difficult to predict its occurrence using analytical or probabilistic approaches. In this work, we aim at leveraging advances in machine learning, deep learning, and pattern recognition to develop a lightning nowcasting model. Current numerical weather models rely on lightning parametrization. These models suffer from two drawbacks; the sequential nature of the model limits the computation speed, especially for nowcasting, and the recorded data are only used in the parametrization step and not in the prediction.[6,7]

To cope with these drawbacks, we propose to leverage the large amounts of available data to develop a fully data-driven approach with enhanced prediction speed based on deep neural networks. The developed lightning nowcasting model is based on a residual U-net architecture.[8] The model consists of two paths from the input to the output: (i) a highway path copying the input to the output in the same way as the persistent baseline model does, and (ii) a fully convolutional U-net which learns to adjust the former path to reach the desired output. The U-net itself consists of a contracting part with alternating convolution, and max pooling layers followed by an expanding part of alternating upsampling, convolution, and concatenation layers.[9–11]

Our dataset consists of post-processed data of recorded lightning occurrences in 15-minute intervals over 60 days obtained from the GOES satellite over the Americas. We have optimized the model using data from the northern part of South America, a region characterized by high lightning activity. The model was then applied to other regions of the Americas. We are using 70-15-15% separation for training, validation, and test datasets. Upon completion of the training process, the model can achieve an overall F1 score of 70% with a lead time of 30 minutes over

South America in fractions of a second. This is more than 25% increase in the F1 score compared to the persistent model which is used as our baseline forecast method.

To the best of our knowledge, our model is the first data-driven approach for lightning prediction. The developed model can pave the way to large-scale, efficient, and practical lightning prediction, which in turn can protect lives and save resources.