A LSTM Recurrent Neural Network for Lightning Flash Prediction within Southern Africa using Historical Time-series Data

Yaseen Essa
School of Computer Science and Applied Mathematics,
The University of the Witwaterstand,
Johannesburg, South Africa
YaseenEssa@essamail.co.za

Ritesh Ajoodha
School of Computer Science and Applied Mathematics,
The University of the Witwaterstand,
Johannesburg, South Africa
Ritesh.Ajoodha@wits.ac.za

Hugh G.P. Hunt
The Johannesburg Lightning Research Laboratory,
School of Electrical and Information Engineering,
The University of the Witwaterstand,
Johannesburg, South Africa
Hugh.Hunt@wits.ac.za

***Abstract*—We evaluated the prediction ability of a Long-Short-Term-Memory Recurrent Neural Network (LSTM) model to predict short-term lightning flash density within South Africa using historical lightning events. We predicted the lightning flash densities for one-hour, three-hour and 24-hour periods for two areas within Southern Africa using data from the South African Lightning Detection Network. These areas represent a moderate and a high-flash density area. The inputs were (1) historical one-hour lightning density, and (2) a five-day lightning density average for the corresponding period in the previous year. Models were trained using four years of data and predictions were made for every one-hour interval for one-year. Our LSTM model is composed of 1x50 network layer, 2x25 layers, 1 dense layer (activation=Leaky ReLU, optimizer=adam). Models were minimized for Mean Square Error (MSE) but evaluated based on Mean Absolute Error (MAE) and MSE. Our models were tested for repeatability by re-running the model for prediction years 2012, 2013, 2014 and 2015. The results showed a combined MAE of 2.87 flashes.hr^{-1} and combined MSE of 1209 for one-hour predictions. MAE values remained relatively similar but standard deviations of the errors increased with longer forecast periods. This LSTM model may be used with a weather parameter prediction model to manage short-term lightning risk.**

***Index Terms*—lightning forecast, Long Short Term Memory Recurrent Neural Network, weather forecasting, risk management, Southern Africa**

I. Introduction

Lightning is an electrostatic discharge that results in a spectacular display of electromagnetic radiation and a pressure wave called thunder [1]. There are between 30-to-100 lightning discharges every second on earth [2]. About seventy-five-percent of these lightning events originates and ends in the clouds; this type of lightning is called cloud-to-cloud lightning [3]. The remaining lightning originates from the cloud and discharges to the ground; this is called Cloud-to-Ground (CG) lightning.

Lightning is responsible for both human and economic loss. [4] estimates that 264-people on average die from lightning events in South Africa every year. Further, lightning is responsible for about twenty-percent of all outages from electrical distribution in South Africa [4]. Knowing the quantity of lightning helps to reduce human loss and assist in planning for expected lightning damage.

The South African Lightning Detection Network (SALDN) is the most accurate lightning detection tool available for lightning detection in South Africa. It is composed on 24 Vaisala LS7000/01 series sensors that are spread across South Africa [6]. Lightning detection networks detect and geo-locate individual strokes which are then grouped into flashes [6]. The SALDN has a detection efficiency of 90% [8]. It is important to note that the SALDN network is designed to detect cloud-to-ground lightning events, thus the network detects only a portion of cloud-to-cloud lightning [8]. Our study thus focuses on the prediction of CG lightning as this form of lightning is more relevant to society.

Although lightning is familiar and well researched, its prediction remains challenging. Academic studies have used Numerical Weather Prediction (NWP) data to build a weather model for lightning forecasts (Table I). Neural networks have been the most popular tool to predict lightning-flashes in recent years. Neural network models outperformed other models based on predicting the likelihood of lightning occurring (Table I). A neural-network model was not previously applied for forecasting lightning events within South Africa (Table I). It is also important to note that all the studies mentioned in Table I have focused on the likelihood of lightning occurring; none of the studies attempted to predict the quantity of lightning occurring.

LSTM Recurrent Neural Networks (RNN) networks are well-suited to classifying, processing and making predictions based on time series data since there can be lags of unknown duration between important events in a time series [10]. Unlike standard feed-forward neural networks, RNN has feedback connections. LSTM models have a unit called a cell that is composed on an input gate, an output gate and a forget gate. The cell remembers values over arbitrary time intervals and the three gates regulate the flow of information into and out of the cell.

Our aim is to evaluate the use of an LSTM neural network model to predict short-term (1HR, 3HR and 24HR) lightning flash densities for two areas within South Africa. These areas are Bloemfontein, a moderate lighting-flash dense area, and Piet Retief, a high lightning flash density area. Our study contributes to current literature in two main ways (1) this is the first time a LSTM RNN model has been applied to historical lightning data events from the SALDN dataset, and (2) it is the first attempt to quantify lightning density rather than forecasting the likelihood of lightning occurring. This study may set the foundations of a more accurate lightning forecast model that may also incorporate weather data variables.

II. Methods

Historical Lightning Flash Data. The South African Lightning Detection Network consists of 24 Vaisala CG lightning sensors distributed through-out South Africa. Lightning detection networks consist of sensors which measure magnetic and/or electric fields propagated from lightning events. These lightning events are time correlated to derive a location of a lightning event [17]. The network can detect lightning with a location accuracy of ±0.5 km and an estimated detection efficiency of ±90% over most of South Africa [7]. The SALDN, which is run by the South African Weather Service, captured each lightning flash between June 2007 to March 2016. This consists of in excess of 100 million lightning flashes. The data is in comma-separated-value format and includes time and GPS location and peak for each lightning stoke.

Preprocessing. We collated data into 1HR, 3HR and 24HR intervals, and categorized each flash into grids (27.76km x 27.76km grids) based on location. Data was scaled between 0 and 1 to assist for model optimization.

Inputs. (1) One-hourly flash density data for 30-days (720-timesteps) for the predicted grid and 48-surrounding grids and (2) a 5-day average for the same period for the last year for the predicted grid.

Long-Short-Term-Memory Recurrent Neural Network Model. Our network model was developed with python (v 3.6.11) under Jupyter notebook (v 5.7.10) using Keras. Our network has 1x50layer, 2x25layer 1x dense network an activation function of a leaky Rectified Linear Unit (alpha=0.02) to reduce negative forecast values. We ran the model for 35 epochs with a batch size of 1644. The LSTM Keras module version 2.3.1 was used for LSTM RNN Network. used the Adaptive Moment Estimation (Adam) [18] method to adapt learning rates. Adam essentially combines RMSProp optimization technique with momentum.

Outputs. 1HR, 3HR and 24HE lightning flash densities for predicted sites.

Sites/Predicted Grids. Piet Retief (27.0184° S, 30.8092° E) and Bloemfontein (29.0852° S, 26.1596° E).

Computer Hardware. We used Amazon Web Services ml.m54xlarge Sagemaker instances, which corresponds to 16x vCPU processor and 64GiB RAM.

Model Repeatability. We assessed model repeatability by running the model four times for each grid; thereby forecasting for years 2012, 2013, 2014 and 2015 as shown in Table II.

A. Analysis

1) Evaluation of Lightning Flash Prediction Densities: We assessed model performance based on Mean Absolute Error (MAE) and Mean Squared Error (MSE).

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \tag{1}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2 \tag{2}$$

Let n = the number of data points, y_i is prediction value and x_i is actual value.

2) Evaluation for Predicting Major Lightning Events: We used a confusion matrix to evaluate the prediction efficiency using two metrics: (1) probability of detection (POD) and the false alarm ratio (FAR). These performance metrics have been widely applied in weather prediction [19]. POD is a measure of accurately predicted lightning events, and FAR is a measure of false-positive events.

$$POD = \frac{n_1}{n_1 + n_2} \tag{3}$$

$$FAR = \frac{n_2}{n_1 + n_3} \tag{4}$$

Let n_1 = total number of true-positive grids, n_2 = total number of false-positive grids and n_3 = total number of false-negative grids

III. Results

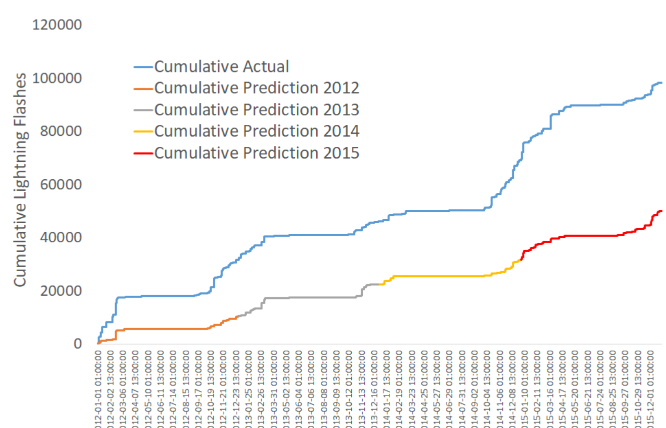

Fig. 1. Cumulative Actual vs Cumulative Predicted 1HR Lightning Flashes for Piet Retief between 2012 to 2015.

A. 1HR Predictions

Figure 1 and Figure 2 shows the cumulative LSTM lightning flash predictions and actual lightning flashes between 2012 and 2015. Both graphs indicate good correlation but the model tended to under-predicted lightning flashes for Piet Retief.

TABLE I
RECENT LIGHTNING PREDICTION STUDIES. LAST TWO ROWS REFLECT SOUTHERN AFRICA LITERATURE.

Authors	**Title**	**Datasets**	**Features**	**Method**	**Outcome** Outcome
Geng et al. (2019) [16]	LightNet: A Dual Spatiotemporal Encoder Network Model for Lightning Prediction	China National Lightning Detection Network, NWP data	NWP and recent lightning observation data	Convolutional long short-term memory Neural Network with Spatio-Temporal dual encoders	Threefold improvement in for six-hour prediction compared with three established forecast methods (PR92, LF1, LF2)
Gharaylou et al. (2018) [11]	Numerical study of performance of two lightning prediction methods based on: Lightning Potential Index (LPI) and electric POTential difference (POT) over Tehran area	Lightning Imaging Sensor and Lightning Detection Network Data	Numerical weather variables	Numerical Weather Prediction Model	Does not outperform competing methods
Moon, SH & Kim, YH (2020) [12]	Forecasting lightning around the Korean Peninsula by postprocessing ECMWF data using SVMs and undersampling	Weather station data, lightning detection networks	Temperature, wind speed, relative humidity, convective available potential energy (CAPE), K-index and the Showalter stability index	Support Vector Machines and Random Forests	Support Vector Machines and random forests had equitable threat scores of 0.0885 and 0.0828
Mostajabi et al. (2019) [13]	Nowcasting lightning occurrence from commonly available meteorological parameters using machine learning techniques	Lightning Detection Networks, Weather Station Data	Air pressure at station level, air temperature, relative humidity, and wind speed)	Decision Tree	80%+ HSS Score (Outperforms CAPE and Persistence Model)
Simon et al. (2019) [14]	NWP-based lightning prediction using flexible count data regression	NWP data, Lightning Detection Network	Numerical Weather variables and Lightning detection network	Count-Data Regression	Outperforms a reference climatology method up to a forecast horizon of 5 days
Speranza, DV (2019) [15]	Lightning Prediction Using Recurrent Neural Networks	Weather Station Data, Lightning Detection Network	-	Long Short-Term Memory (LSTM) Neural Networks	Max 84% accuracy
Booysens et al. 2014 [9]	Detection of Lightning Pattern Changes Using Machine Learning Algorithms	Lightning Imaging Sensor on Tropical Rainfall Measuring Mission	Historical lightning event data	Data from Satellite Lightning Imaging sensor, various machine learning models	Decision Tree: 97% spatial global pattern detection rate. Naïve Bayes achieved temporal 31% global pattern detection rate
Gijben et al. 2017 [7]	A statistical scheme to forecast the daily lightning threat over Southern Africa using the Unified Model	SALDN, Numerical Weather Parameter from SAWS	Precipatable water, Lifted index, Equivalent temperature potential, Relative humidity, Average temperature, Convective available potential energy	Stepwise Logistic Regression	Area Under Curve value of (AUC/ROC Curve) $\pm$ 0.91

TABLE II
MODEL RUNS TO TEST REPEATABILITY

	Model Run 1	**Model Run 2**	**Model Run 3**	**Model Run 4**
Train Years	2008 to 2011	2009 to 2012	2010 to 2013	2011 to 2014
Test Year	2012	2013	2014	2015

TABLE III
PREDICTION ERRORS FOR 1HR LIGHTNING PREDICTIONS AT PIET RETIEF FOR EACH MODEL RUN

	2012	**2013**	**2014**	**2015**	**ALL**
Mean Error	-2.49	**-0.21**	-1.59	-1.22	-1.38
Mean Absolute Error	3.61	**1.73**	2.98	3.49	2.95
Mean Squared Error	2098.39	**552.80**	1273.72	1889.55	1454.06

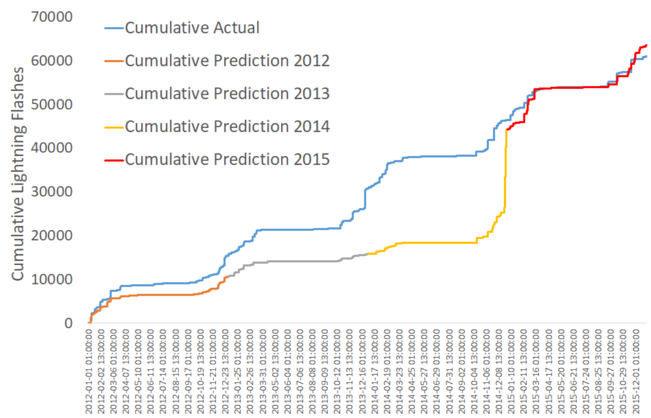

Fig. 2. Cumulative Actual vs Cumulative Predicted 1HR Lightning Flashes for Bloemfontein between 2012 to 2015..

TABLE IV
PREDICTION ERRORS FOR 1HR LIGHTNING-PREDICTIONS AT BLOEMFONTEIN FOR EACH MODEL RUN

	2012	**2013**	**2014**	**2015**	**ALL**
Mean Error	-0.55	-1.15	1.46	0.53	**0.07**
Mean Absolute Error	**1.91**	1.92	4.62	2.72	2.79
Mean Squared Error	**543.13**	1172.46	1462.68	674.75	962.97

Fig. 3. Prediction Error of 1HR lightning-density for Piet Retief during 2012, 2013, 2014, 2015 (Error bars represent 1xStandard Deviation).

B. Model Repeatability

Table III and Table IV indicates the repeatability of the LSTM model training given different sets of data. Yearly MAE values range from 1.73 to 3.49 flashes.hr^{-1} for the Piet Retief LSTM model and 1.91 to 4.62 flashes.hr^{-1} for the Bloemfontein LSTM model. Figure 3 and Figure 4 also shows the repeatability of the LSTM models by visualising prediction errors and standard deviation.

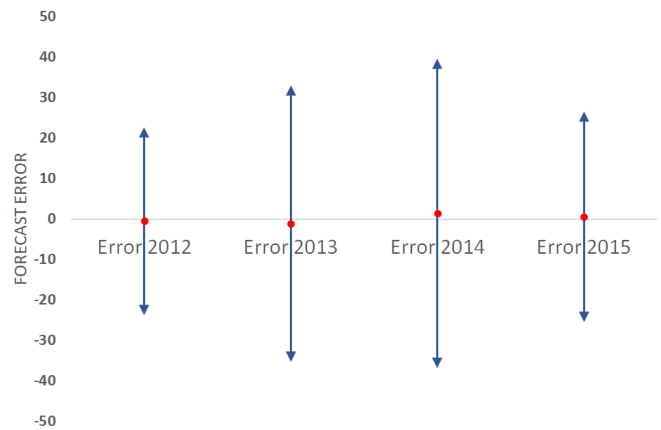

Fig. 4. Prediction Error of 1HR lightning-density for Bloemfontein Retief during 2012, 2013, 2014, 2015. (Error bars represent 1xStandard Deviation).

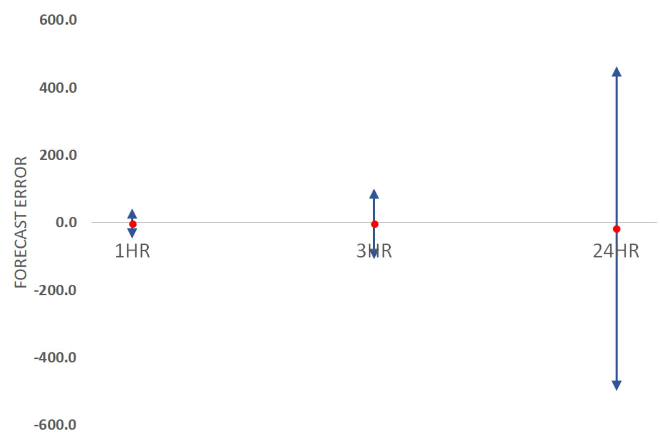

Fig. 5. Prediction Errors for 1HR, 3HR and 24HR lightning-density for Piet Retief during 2012 (Error bars represent 1xStandard Deviation).

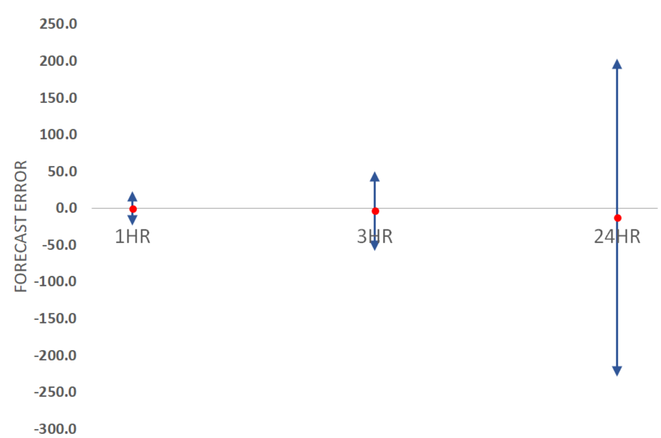

Fig. 6. Prediction Errors for 1HR, 3HR and 24HR lightning-density for Bloemfontein during 2012 (Error bars represent 1xStandard Deviation).

C. Prediction Errors vs. Prediction Length

Figure 5 and Figure 6 illustrates the effects that prediction length on prediction error during 2012 for Piet Retief and Bloemfontein, respectively. The figures indicate that forecast error remains relatively constant but the standard deviaion increases substantially with longer prediction lengths.

D. Confusion Matrices for Impactful-lightning events

Table V and Table VI indicate the percentage of true-positive, true-negative, false-positive and false-negative results for major lightning-events with a confusion table format. Major lightning events were defined as 30 or more lightning flashes .hr-1. This corresponds to a POD and FAR ratio of 32% for and 51% for Piet Retief and a POD and FAR ratio of 27% and 79% for Bloemfontein, respectively.

TABLE V
CONFUSION MATRIX: PIET RETIEF, 1HR-FORECASTS. YES = 30 OR MORE FLASHES PER HOUR.

		Predicted	
		Yes	No
Actual	Yes	0.355%	0.773%
	No	0.379%	98.488%

TABLE VI
CONFUSION MATRIX: BLOEMFONTEIN, 1HR-FORECASTS. YES = 30 OR MORE FLASHES PER HOUR.

		Predicted	
		Yes	No
Actual	Yes	0.254%	0.673%
	No	0.903%	98.143%

IV. DISCUSSION

A. 1HR Predictions

We found relatively consistent MAE values for both the Bloemfontein and the Piet Retief prediction models. Across all prediction models, we had an MAE of less than four-lightning flashes.hour-1 (Table III and Table IV). Overall, the models performance parameters were similar for each model run. But the model performed better to predict Bloemfontein than Piet Retief.

B. 3HR and 24hour Predictions

Although mean error (and MAE) remain relatively constant, the standard deviations increased considerably with the 24HR prediction period (Figure 5 and Figure 6). The model thus shows considerable prediction decay as the prediction length increases.

C. LSTM Repeatability

We ran the LSTM models four-times for prediction years 2012 to 2015 to test for model repeatability (Figure 3 and Figure 4). We have found that MAE values and standard deviations have remained relatively bounded, thus elucidating good model repeatability.

D. Risk of under-forecasting and over-forecasting Major Lightning Events

We used a confusion matrix to illustrate the models propensity to under- and over-forecast major lightning events (Table V and Table VI). We defined an major lightning event as 30 or more lightning flashes.hr-1 within the predicted area. We found that the LSTM model predicted 30% of major lightning events. For the Piet Retief area, about 50% of predictions came through while only 20% of predictions turned out to be true for Bloemfontein.

Our study's performance is not directly comparable with recent lightning prediction studies. The most recent and thorough study to investigate same day lightning prediction in South Africa is by [8]. In this study, the authors found an Area-Under-Curve ratio of 90% using a stepwise logistic regression mode. But the study predicted the occurrence of a lightning flashes occurring rather than the quantitative number of lightning flashes.

Internationally, our studies methods are similar to the work done by [16]. These authors combined two convolutional long-short artificial neural-network models - one for NWP paramater data and the other for observed lightning data - to produce a relatively good lightning-event prediction accuracy. The model was able to predict up to 89% of lightning events within an hour. But once again, this study did

V. CONCLUSION

In this study, we showed that LSTM RNN models are useful for predicting short-term lightning flash forecasts. Although the model is numerically accurate, the model does not predict about 60% of major lightning events. However, this work has confirmed that a LSTM Recurrent Neural Network tool is valuable for lightning-forecasting in Southern Africa.

A. Future Work

Of the lightning research that we have evaluated, we found that the work by [16] had the most accurate ability to predict the occurrence of lightning. Unlike most lightning prediction methods, which exploit either NWP simulation data or recent lightning observation data, these authors predictions are based on both datasets. The model extracts spatiotemporal features of the simulation and the observation via dual encoders. Then, these features are combined by a fusion module. Finally, the fused features are fed into a spatiotemporal decoder to make forecasts.

We believe that a similar model can be applied for lightning prediction in Southern Africa. The LSTM model from this study may be combined with a LSTM Model using NWP weather data. We believe that such a model will predict lightning with high accuracy.

ACKNOWLEDGMENT

This work is based on the research supported in part by the National Research Foundation of South Africa (Grant number: 121835 and 98244). The authors would also like to thank The South African Weather Service, specifically Morné Gijben and Andrew van der Merwe, for the use of the SALDN data.

REFERENCES

[1] M. A. Uman, “The Lightning Discharge,“ ISSN. Academic Press, ISBN: 9780080959818, 1987, pp. 10—18.

[2] J. R. Dwyer, and M. A. Uman, “The physics of lightning,“ In: Physics Reports 534.4, ISSN: 0370-1573, 2014, pp. 147—241.

[3] V. A. Rakov and M.A. Uman, “Lightning: Physics and Effects,“ Cambridge University Press, ISBN: 9780521583275, 2003, pp. 15–20.

[4] R. L. Holle, and M. A. Cooper, “Lightning Fatalities in Africa From 2010-2017,“ In: 2018 34th International Conference on Lightning Protection (ICLP), 2018, pp. 1–4.

[5] T. B, Andersen, and C. J. Dalgaard, “Power outages and economic growth in Africa,“ In: Energy Economics 38, ISSN: 0140-9883, 2013, pp. 19–23.

[6] H. G. P. Hunt, R. Blumenthal, K. J. Nixon, C. Gomes, ”A multidisciplinary forensic analysis of two lightning deaths observed in South Africa, International Journal of Disaster Risk Reduction,” vol. 51, p101814, , 2020, ISSN 2212-4209, doi: https://doi.org/10.1016/j.ijdrr.2020.101814

[7] M. Gijben, L. L. Dyson, and M. T. Loots, ”A statistical scheme to forecast the daily lightning threat over southern Africa using the Unified Model,” Atmospheric Research, vol. 194, 2017, pp. 78-88.

[8] M. Gijben, ”A lightning threat index for South Africa using numerical weather prediction data,” Thesis: University of Pretoria, 2016

[9] A. Booysens, ”Detection of Lightning Pattern Changes using Machine Learning Algorithms,” Proceeding of International Conference on Communications, Signal Processing and Computers, 2014, pp 78-84

[10] G. Petneházi, ”Recurrent neural networks for time series forecasting,” arXiv preprint arXiv:1901.00069, 2019.

[11] M. Gharaylou et al., ”Numerical study of performance of two lightning prediction methods based on: Lightning Potential Index (LPI) and electric POTential difference (POT) over Tehran area,” Journal of Atmospheric and Solar-Terrestrial Physics, vol. 193, 2019

[12] S. H. Moon and Y. H. Kim. “Forecasting lightning around the Korean Peninsula by postprocessing ECMWF data using SVMs and undersampling,” In: Atmospheric Research 243, 2020, p. 105026.

[13] A. Mostajabi, D. Finney, M. Rubinstein & F. Rachidi, ”Nowcasting lightning occurrence from commonly available meteorological parameters using machine learning techniques,” Climate and Atmospheric Science, vol. 2:41, 2019, doi: https://doi.org/10.1038/s41612-019-0098-0

[14] T. Simon, G. Mayr, N. Umlauf A. Zeileis, ”NWP-based lightning prediction using flexible count data regression. Advances in Statistical Climatology,” Meteorology and Oceanography. vol. 5, 2019, pp 1-16.

[15] D. Speranza. “Lightning Prediction Using Recurrent Neural Networks Thesis,” PhD thesis. Air University, Air Force Institute of Technology, United States of America, 2019

[16] Y. Geng et al. “LightNet: A Dual Spatiotemporal Encoder Network Model for Lightning Prediction,” pp. 2439–2447. ISBN: 978-1-4503-6201-6, 2019, doi: 10.1145/3292500.3330717

[17] H. G. Fensham et al. ”Performance evaluation of the SALDN using highspeed camera footage of ground truth lightning events over Johannesburg, South Africa,” 2018 34th International Conference on Lightning Protection (ICLP), 2018, pp 1-5. DOI:10.1109/ICLP.2018.8503415

[18] D. Kingma and J. Ba. ”Adam: A Method for Stochastic Optimization,” International Conference on Learning Representations, 2014, url: https://arxiv.org/pdf/1412.6980.pdf

[19] A. Clark, W. Gallus, and M. L. Weisman. “Neighborhood-Based Verification of Precipitation Forecasts from Convection-Allowing NCAR WRF Model Simulations and the Operational NAM,” In:Weather and Forecasting, pp. 1495–1509, 2010, doi: https://journals.ametsoc.org/doi/10.1175/2010WAF2222404.1