Using Traditional Machine Learning Algorithms to Classify Lightning Occurrence

Ndaedzo Makgatho
School of Computer Science and Applied Mathematics
The University of Witwatersrand
Johannesburg, South Africa
1843599@students.wits.ac.za

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

***Abstract*—The study proposes a method for predicting the occurrence of lightning on an hourly basis using historical meteorological data. The meteorological data used is provided by South African Weather Services (SAWs). Humidity, temperature and pressure are the meteorological variables studied. Two machine learning algorithms have been evaluated namely Logistic regression and Neural networks. The false alarm rate and the probability of detection were used to evaluate the lightning forecasting models. The Neural network has indicated better accuracy of 95% with an area under the graph of 98% whereas Logistic regression performed with an accuracy of 87% with an area under the graph of 87%.**

***Index Terms*—Neural network, Logistic regression, lightning forecast**

I. Introduction

Lightning is natural phenomena that most people do not understand fully. It is responsible for the cause of death or injuries to animals and humans [6].It also causes wildfires, damage to infrastructure and it has a significant effect on the economy [8]. Lightning is an event where an electrostatic discharge is released from the clouds. This occurs due to the friction between negatively charged and positively charged areas in the clouds, cloud-to-air or cloud-to-ground [7]. Most of the lightning events originate and end in the clouds (cloud-to-cloud lightning).

The earth experiences 30-100 lightning events every second [2], [3]. There is an increase in lightning events due to global warming. As the global temperature rises by a degree, lightning events increases by approximately 5-6% [6]. Worldwide the world is suffering from global warming and climate change and this has made weather to be unpredictable. Lightning is commonly associate with rainfall and thunderstorms, but lightning can also occur without the aforementioned weather conditions.

One of the biggest cause of death from natural disaster is lightning. In South Africa, there are about 264 people who die each year from lightning events [2], [3]. It also affects 20% of the power outages distribution in south Africa [2], [3]. In order to lessen the damage caused by lightning strikes to the distribution networks, shielding wires and surge protectors can be used to provide adequate lightning protection. In addition to installing preventive measures, it is critical to respond appropriately by moderating lightning incidents and forecasting lightning.

The authors in [2] proposed an approach for forecasting short term lightning flash densities using historical data. They assessed Long Short Term Memory Recurrent Neural Network and they found that it predicts about 30% of lightning events. In [5], the authors proposed a method of forecasting lightning by under-sampling meteorological data. They used Random forest and Support vector machines (SVMs) and SVMs were found to outperform Random forest model. Another study by [8] used temporal parameters and meteorological data to forecast lightning occurrence using ANN and Decision trees. From this study, Decision trees were found to outperform ANN.

In this study, we offer a method for predicting the occurrence of lightning using the machine learning algorithms: Logistic regression and Neural Network. In specifically, we develop a method for predicting whether or not lightning will occur based on an hourly basis by using meteorological data.
This study contributes to literature as it extends the study of lightning occurrence using historical meteorological data. It implemented a logic method for lightning systems applications since it is able to imitate the conditions for the occurrence of lightning events.

The remaining part of this study is laid out as follows: A summary pf related works is offered in Section II. The methodology that has been proposed is provided in Section III. Section IV presents the results and discussion. The conclusion and suggestions for future projects are presented in Section V.

II. Related Work

The following table I summaries the recent studies conducted to predict lightning occurrence. Most of these studies uses meteorological data to forecast lightning occurrence using the machine learning algorithms.

III. Methodology

IV. Data

Historical meteorological data from South African Weather Services (SAWs). The data captured is an hourly data for the duration of 2018 to 2019. The data contains Pressure, Humidity and Temperature. To design a logic method for determining whether lightning is imminent, the following conditions we set:

1) If the pressure is low, the humidity is high, and the temperature is high, lighting is imminent.
2) If the pressure is low, the humidity is low, and the temperature is high, lighting is imminent.
3) If the pressure is low, the humidity is high, and the temperature is low, lighting is imminent.
4) If the pressure is low, the humidity is low, and the temperature is low, lighting is imminent.
5) If the pressure is high, the humidity is high, and the temperature is high, lighting is imminent.

V. Pre-processing

The input and output parameters for the proposed approach is summarized in Table II below.

VI. Models

A. Logistic Regression

Logistic regression is binary classification model used to predict a certain event or class [4]. The logistic model is obtained from the logistic function (probability function) expressed as follows,

$$f(y) = \frac{1}{1 + e^{-y}} \tag{1}$$

where,

$$y = \alpha + x_1\beta_1 + ... + x_k\beta_k \tag{2}$$

α and β are the regression coefficients which are estimated from the data obtained [4]. The variable x is the independent variables. $f(y)$ is the probability of the outcome which ranges between 0 and 1

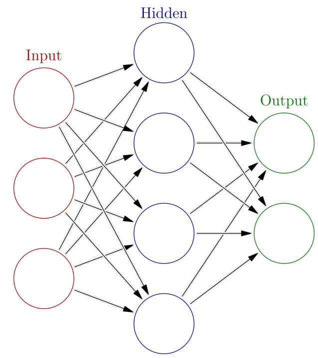

Fig. 1. Neural Network

B. Neural Network

From the preceding figure 1, the first layer of neurons collects the input data, then consumed by the hidden layer which then provides the final output. The layers are made up of one or more neurons and each neuron computes an activation function. The connection between neurons of the successive layers has weight associated to it. The weight defines the relative importance between the input and the output.

$$\sum_{i=1}^{m} w_i x_i + bias = w_1 x_1 + ... + w_m x_m \tag{3}$$

$$f(x) = \begin{cases} 1 & if \sum w_i x_i + b \geq 0 \\ 0 & ,if \sum w_i x_i + b < 0 \end{cases} \tag{4}$$

where $f(x)$ is the output.

VII. Evaluation Techniques

The model prediction ability is evaluated using the false alarm rate (FAR), Receiver operating characteristic (ROC) curve and the probability of detection (POD). The FAR measures the rate of false-positives lightning events. The ROC curve is curve that is used in binary classification to study the output of the classifier and the POD measures true-positive lightning events.

$$POD = \frac{T_p}{T_p + F_p} \tag{5}$$

$$FAR = \frac{F_p}{T_p + F_n} \tag{6}$$

where F_p is the false positives, F_n is the false negatives and T_p is true positives.

VIII. Results and Discussion

The results in Table III and IV shows that the logistic regression model predicted 51% of the lightning events and the neural network predicted 79% of the lightning events. The accuracy of logistic regression and the neural network are 87% and 95% respectively. The area-under-curve ratio of 2 and 3 shows that neural network has a higher AUC of 98% whereas

TABLE I
RECENT STUDIES OF LIGHTNING PREDICTION.

Authors	Data	Features	Model	Significance	Outcome
[5]	Meteorological data from European Centre for Medium-range weather forecasts (ECMWF)	Relative humidity, temperature, K-index, wind speed, CAPE and showalter stability index	Random forests and Support Vector Machines (SVMs)	Forecsting lightning by undersampling meteorological data	SVMs outperforms Random forest
[2]	Historical lightning data from South African Lightning Detection Network (SALDN)	Lightning flash	Long Short Term Memory Recurrent Neural network	Evaluating LSTM Neural Network to forecast short term lightning flash densities	LSTM RNN is good tool to forecast lightning in SA
[9]	Global Lightning Network and WSI Corporation and meteorological data	Images of Meteorological satellite	Decision trees, Random forest, Gradient boosting and adaBoost	Using satellite image to investigate a new approach to predict thunderstorms	Gradient boosting performed better than rest of model
[8]	Iranian meteorological data	Temporal parameters, visibility, cloud cover, wind speed, dewpoint, wind direction pressure, humidity	ANN and Decision trees	Identifying best way to predict lightning occurrence	Decision tree outperforms Neural Network
[7]	Malaysian meteorological data	Temperature, humidity and dewpoint	To design and develop lightning prediction system using fuzzy logic technique	Fuzzy logic technique	Model predicts more than 90% of positive results
[6]	Meteorological data from Switzerland	Wind speed, relative humidity, air temperature and air pressure	Decision trees	Examining the correlation between lightning and atmospheric data	POD of 78%
[10]	Jiangxi meteorological Administration	Meteorological data	SVMs, Rough set SVMs(RS-SVM), and Rough set(RS)	Forecasting thunderstorm using Rough SVMs	RS-SVM is more suitable
[3]	Historical cloud-to-ground data from SALDN	Lightning flashes	Auto regressive integrated moving average (ARIMA), LSTM RNN and AUto Regressive (AR)	To assess and compare the ability to forecast short-term lightning	LSTM RNN outperforms AR and ARIMA
[1]	Global Lighntning Network and WSI Corporation, Meteorological data from Malaysian Meteorological Services	Rainfall, Temperature, and humidity	Neural network with gradient descent backpropagation and Neural Network with Scaled conjugated gradient backpropagation	Determining the severity of lightning using Neural network	Gradient descent outperforms scaled conjugated gradient backpropagation

TABLE II

INPUT AND OUTPUT DATA.

Input	Description
H	Humidity (%)
T	Temperature ($^{\circ}C$)
P	Pressure (mbar)
Output	**Description**
0	No lightning occurrence
1	Lightning occurrence

TABLE III

CONFUSION MATRIX: LOGISTIC REGRESSION.

Occurrence	POD	FAR
0	0.88	0.98
1	0.51	0.16

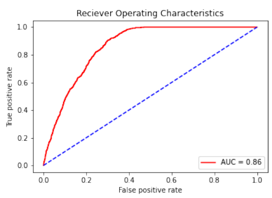

Fig. 2. Receiver Operating Characteristic ROC

TABLE IV

CONFUSION MATRIX: NEURAL NETWORK.

Occurrence	POD	FAR
0	0.98	0.97
1	0.79	0.86

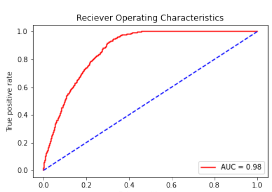

Fig. 3. Receiver Operating Characteristic ROC

logistic regression has AUC of 87%. This study used a similar approach as the work done in [7]. The authors used a fuzzy logic technique to predict lightning using meteorological data and they produced an accuracy of 95%.

IX. CONCLUSION

In this study, we showed that meteorological parameters such as humidity, temperature, rainfall and pressure can be used to predict lightning events. A logic technique was implemented to predict whether lightning was imminent or not for Johannesburg conditions. Since the logic technique is built upon the weather conditions in Johannesburg, the lightning prediction parameters that are identified at other areas in the world are not guaranteed to be applied to Johannesburg. Thus, these results highlight the significance of determining weather parameters based on specific geographical location. It was found that the neural network performed better at predicting lightning events as compared to logistic regression.

REFERENCES

[1] M. Azhar Omar, M. Khair Hassan, Azura Che Soh, and M.Z.A Ab. Kadir.Lightning severity classification utilizing the meteorological parameters: A neural network approach. In 2013 IEEE International Conference on Control System, Computing and Engineering,pages 111–116, 2013.

[2] Yaseen Essa, Ritesh Ajoodha, and Hugh G.P. Hunt. A lstm recurrent neural network for lightning flash prediction within southern africa using historical time-series data. In 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), pages 1–6, 2020.

[3] Yaseen Essa, Hugh G.P. Hunt, and Ritesh Ajoodha. Short-term prediction of lightning in Southern Africa using autoregressive machine learning techniques. In 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), pages 1–5, 2021.

[4] Morńe Gijben, Liesl L. Dyson, and Mattheus T. Loots. A statistical scheme to forecast the daily lightning threat over Southern Africa using the unified model.Atmospheric Research, 194:78–88, 2017.

[5] Seung-Hyun Moon and Yong-Hyuk Kim. Forecasting lightning around the korean peninsula by postprocessing ecmwf data using svms and undersampling.Atmospheric Research, 243:105026, 2020.

[6] Amirhossein Mostajabi, Declan L Finney, Marcos Rubinstein, and Farhad Rachidi. Nowcasting lightning occurrence form commonly available meteorological parameters using machine learning techniques.npj Climate and Atmospheric Science, 2, 2019.

[7] Lemuel Chung Chen Nyap, Waddah Waheeb, and Jacqueline Luokse. Lightning prediction using fuzzy logic technique.Journal of Applied Technology and Innovation (e-ISSN:2600-7304), 4(3):1, 2020.

[8] Morteza Pakdaman, Sina Samadi Naghab, Leili Khazanedari, Sharare Mal-bousi, and Yashar Falamarzi. Lightning prediction using an ensemble learning approach for northeast of Iran.Journal of Atmospheric and Solar-Terrestrial Physics, 209:105417, 2020.

[9] Christian Schön, Jens Dittrich, and Richard Müller. The error is the feature: How to forecast lightning using a model prediction error. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery Data Mining, page 2979–2988. Association for Computing Machinery, 2019.

[10] Qiu Taorong, Zhang Shanshadan, Zhou Hou, Bai Xiaoming, and Liu Ping. Application study of machine learning in lightning forecasting.Information Technology Journal, 12:6031–6037.