Ukrainian Journal
of Educational Studies and
Information Technology

ISSN: 2521-1234 online
Ukr. J. of Educ. Stud. and Inf. Technol.
11(4), 227 – 241
doi: 10.32919/uesit.2023.04.01

Vol. 11, Issue 4, 2023

Techniques for lightning prediction: A review

Babatunte Dauda Raheem, Emeka Ogbuju, Francisca Oladipo, Taiwo Abiodun

Submitted:
2023 – 08 – 11
Accepted:
2023 – 12 – 16
Published online:
2023 – 12 – 30

Abstract
Lightning is a natural occurrence which is created through the mixture of hot and cold air in the cloud. Sudden occurrence of lightning has caused damages to many lives and properties, for this reason; there is a need to develop a system that can predict lighting occurrence for people to take necessary precaution. However, accurately predicting lightning has been a challenge among researchers, as they find it difficult to select the right approach and algorithms to use when predicting lightning. Thus, this paper presents a systematic literature review on the best techniques for lightning prediction by reviewing relevant papers that are systematically collected based on the inclusion and the exclusion criteria from four different academic databases which includes Scopus, IEEE Xplore, Science direct, and SpringerLink. The findings from the review shows that the Random Forest algorithm is mostly used for lightning prediction and has generally out performed all other algorithms that have been used in lightning prediction in remote region. Also the review finds out that there is an inverse relationship between predicting system accuracy and lead time. Another observation in the research is that numerical weather prediction predicts more accurately compare to geo satellite prediction.
Keywords: lightning prediction; systematic literature review; numerical weather prediction; machine learning and algorithms.

INTRODUCTION

Lightning is a natural phenomenon that is created through the mixture of hot and cold air in the cloud (Pakdaman et al., 2020). There are basically three types of lightning; which are cloud to itself, cloud to another cloud, and cloud to ground (Yao et al., 2021). The formal exists between the cloud and the earth while the later exists within the cloud. Cloud to ground is the major concern of this paper, this is in view of its great destruction to lives and property. The aftermath of lightning strikes when it happens is destruction of lives and properties worth millions of Dollars. Researchers have worked hard to find how lightning strike damage can be reduced, probably preventing the damage. It has been proved that as lightning strikes is unpreventable. However, lightning can only be predicted (Essa et al., 2021). This makes researchers to narrow their solutions towards working on different approaches of developing lightning prediction/forecast systems. GEO-Satellites have been the used to predict lightning based on cloud datasets imagery and other cloud related parameters. The lightning predictions are full of flaws and not specific also accuracy of remote regions is very low. The use of Machine Learning techniques employing Numerical Weather Prediction (NWP) has proved to be specific in remote regions as well as giving lightning prediction with high accuracy (Mostajabi, 2021). The purpose of carrying out this systematic literature review is to make fact findings on the best lightning prediction techniques with high accuracy, cost effective as well as very low lead time using Machine Learning Approach. Three research questions were answered in this review with the use of relevant journal and conference papers that were comprehensively and systematically collected from some top academic database. The research questions were answered in order to provide a summary of exiting evidence and generate new insight regarding the techniques for lighting prediction.

Some of the terms used in this paper includes; lightning which is the occurrence of a natural electrical discharge of very short and high voltage between a cloud and the ground or within a cloud, accompanied by a bright flash and typically also thunder (Essa et al., 2021). Lightning Prediction is a type of lightning detection equipment that determines when atmospheric conditions are likely to produce lightning strikes and sounds an alarm (Etten-Bohm et al., 2021). Numerical Weather Prediction is the use of weather parameterization for lightning forecast, always in numerical formats (Schultz et al., 2021). Lead Time is the skill of lightning forecast statistically which ranges from minutes to days depending on the types of forecast time of year and region (Tippett & Koshak, 2018). Machine Learning is a branch of AI that enables software applications to become more accurate at predicting outcomes without being explicitly programmed to do so (Schultz et al., 2021).

METHODOLOGY

In this study, we used the systematic literature review method by adopting Host & Orucevic-Alagic (2011) as guideline. Three research questions were formed based on a guideline by Kitchenham et al. (2009). The answer to the research questions were provided by summarizing exiting evidence and generating new insight regarding the techniques for lighting prediction from relevant existing journal and conference papers that were gathered from some top academic databases.

Research question

The research questions are as follows:

(i) What are the most effective Machine Learning algorithms employed in lightning prediction? (RQ1)

(ii) Regardless of the approach employed in lightning prediction system, what is the relationship between lightning prediction system accuracy and lead time? (RQ2)

(iii) What are the comparative advantages between Geo-satellite prediction and Numerical Weather Prediction in terms of Lightning Prediction? (RQ3)

Search string

We collected relevant papers by creating search phrases with the use of keywords related to the previously formulated research questions. We carry out the search in four common academic databases which includes Scopus, IEEE Xplore, Science direct, and SpringerLink. The search sting that were created are “lightning prediction”, lightning prediction algorithms”, “lightning prediction systems”, “Geo-satellite prediction”, “Numerical weather prediction”, and “lightning prediction approaches”. Boolean operators like “AND”, and “OR” were used to connect the search strings.

Data collection

Quality papers were collected papers from four major academic databases which are Scopus, IEEE Xplore, Science direct, and SpringerLink. A tabula representation of these papers is presented in Table 1.

Table 1. **Tabular representation papers collected from different databases**

S/N	Database	Number of papers
1	Scopus	6
2	IEEE Xplore	7
3	Science direct	6
4	SpringerLink	5
Total numbers of papers collected		24

Inclusion criteria

Journal and conference papers that are publish in English language from 2014 to 2023 were included in this research. However, in some cases where we have identical studies, we chose the most current paper.

Exclusion criteria

Papers written in other languages aside English language, papers published before 2014, and papers whose abstract does not explicitly state its contribution to the work were excluded from this study.

RESULT AND DISCUSSION

RQ1: What are the most effective Machine Learning algorithms employed in lightning prediction?

Adekitan & Rock (2021) published a paper on data mining analysis of lightning strike probability to simple structure. In their work, dataset was produced by using numerical procedure to calculate dynamic electro-geometrical system for a cuboid structure. It was synthesized employing data mining procedures using ML. Regression and classification were employed on the orange data mining involving MATLAB environment. The accuracy achieved using both Random Forest and Ada Boost algorithm is 100%.

Using machine learning techniques in Switzerland, Mostajabi (2021) reported on nowcasting lightning probability from easily available meteorology variables. They built a four-parameter model based on these four commonly used weather numerical parameters; air, temperature, relative humidity, atmospheric pressure, and wind speed. The produced signal is then confirmed using lightning tracking mechanisms. They were subjected to the random forest machine learning algorithm. Based on the evaluation's conclusions, its lead time has been between 0 and 10 minutes.

However, according to Romps et al. (2018), lightning for Australia might be projected using multiple land-scale atmospheric variables. This paper investigated how well six statistical and machine learning classification methods could discriminate between days containing and excluding lightning at the broad temporal and spatial scales of the latest broadly available models and reanalysis. Six sites in Australia from 2004 to 2013 were categorized using an integration of essential element analysis and logistic regression, regression trees and classification, linear discriminant analysis, random forests, quadratic discriminant analysis, and logistic lightning count. Performance on the classification assignment was measured via tenfold cross-validation. The outcomes show that logistic regression was superior to other classifiers and that applying climatological values for prediction is massively inferior.

Pakdaman et al. (2020) reported on lightning prediction using an ensemble learning approach for North of Iran. Lightning dataset imbalance problem was solved by using under sampling to get balanced datasets, neural network and decision tree algorithm were used for lightning prediction. The result shows decision tree outperforms neural network.

La Fata et al. (2021) presented a paper on cloud to ground lightning nowcasting using machine learning procedure. A specific datasets of eighteen (18) years geo environmental properties were used to predict a lead time of 1 hour in a three month. The result using Random Forest shows better accuracy in predicting a better lead time.

A presentation was made by Leal at al. (2022) on short term lightning forecast in the Amazon region using ground based weather apparatus and machine learning procedures was employed. A ground based datasets such as humidity, temperature, pressure and wind speed were subjected to ML algorithm. The result proved an accuracy of 71% in Amazon region.

Schon et al. (2019) published a work on how to forecast lightning using model prediction error. The properties of ML such as error-dimensional optical flow process was employed to extract images of meteorological features which is the main pointer to thunderstorm and lightning. Decision Tree and neural network classifiers were trained to forecast lightning in the next hours. The findings show that an accuracy of 96% gotten over 15 minutes, also as the lead time increases the accuracy decreases.

Coughlan et al. (2021) published a paper on using ML to predict fire ignition occurrences from lightning prediction. They stated clearly that there are relationship between fuel moisture and the ignition. Three ML algorithms such as decision tree, random forest and AdaBoost were employed involving ensemble procedure. Findings proved that Random Forest and AdaBoost has 78% accuracy. Also 71% of similar cases has the tendency of lightning occurrence.

Lightning forecast and environment factor analysis employing Random Forest algorithm was published by Yao et al. (2021). They proved lightning nowcasting model performs better when the lead time decreases. The accuracy of lightning prediction model for 0-6 hours are stated as 0.827-0.864 and the impact rate is 0.677-0.851.

Essa et al. (2021) published a work on short term prediction of lightning. They made use of Autoregressive ML Technique; models were subjected to train-test Auto Regressive, Auto Regressive Integrated Moving Average (ARIMA) and long term memory recurrent neural network (LSTM). The result shows the LSTM outperforms other models

Brocco et al. (2018) released an article on enhancing climate with data mining knowledge discovery. The benefits and challenges of mining enormous climate datasets are presented in this paper, in addition to the potential for KDD (Knowledge

Discovery by Data Mining) to discover patterns pertinent to climate. This work focuses on a complicated and dynamic analysis's beta MAPS. It is used to highlight the links between recent statistical results that are local and non-local. This has included contrasting and comparing the known teleconnections between the outputs of climate reanalysis and climate model.

Machine learning technique was used by Coughlan et al. (2021) to determine the possibility of fire ignition from predictions of lightning. One of the most unpredicted elements of the fire zone remains to be lightning ignitions. In light of this, this research examines datasets on lightning ignition to predict the occurrence of lightning using machine learning. It would be advantageous to establish a lightning-ignition relationship in order to construct a system that would support timely warning systems planned for fire containment and prevention. Based on lightning forecasts and unforeseen conditions, a machine learning (ML) method was used to construct a predictive system for wildfire ignition. Three different binary classifiers AdaBoost, Decision Tree, and Random Forest—were applied. These three generated positive outcomes, with both ensemble approaches (AdaBoost and Random Forest) getting an out-of-sample accuracy of 78%. Over 145 lightning-ignited wildfires in various parts of Australia in 2016 which led to the use of statistics from a Western Australia wildfire database. This demonstrated that the ML models successfully forecasted the experience of an ignition when a fire was prompted in at least 71% of the cases.

On the basis of internet-of-things technology, Zhou et al. (2020) advanced the research on intelligent lightning protection. The safety and interests of people are closely tied to the industry of lightning protection. It is necessary to connect the Internet of Things with traditional lighting control. The architecture-level view of the zonal lightning intelligent observing and early alert system is used to present the framework. The Internet of Things and machine learning techniques were utilized to forecast the visitors' quality of life. Make intelligent lightning protection a reality.

For the aim of assessing earth system model error, Silva et al. (2022) applied an interpretable machine learning approach. Modern Artificial Intelligence (A.I.) algorithms have revealed that these errors can be successfully predicted. Investigating the weaknesses in the NASA GEOS model's forecasting of the incidence of lightning, a popular Earth System Model, utilizing XGBoost classification trees and SHapley Additive exPlanations (SHAP) analysis. The model error can be adequately projected by this interpretable error prediction method, which also shows that the errors are highly correlated to convective processes and the properties of the land surface.

Table 2. **Summary of most effective Machine learning algorithms employed in lightning prediction**

Authors	Algorithms	Key Findings
Adekitan & Rock (2021)	Random Forest and Ada Boost	Random Forest and Ada Boost gave 100% accuracy
Mostajabi (2021)	Random Forest	Random Forest yielded highest accuracy
Romps et al. (2018)	Logistic Regression	Logistic Regression performs best
Pakdaman (2020)	Decision Tree, and Neural Network	Decision Tree Algorithm outperforms Neural Network
La Fata et al. (2021)	Random Forest	Random Forest shows better performance
Leal & Matos (2022)	ML	ML Algorithm proved 71% accuracy

Authors	Algorithms	Key Findings
Schon et al. (2019)	Decision Tree	The DT achieved 90% accuracy
Coughlan et al. (2021)	Random Forest and Ada Boost	Random Forest has 78% accuracy, Ada Boost 71% accuracy
Yao et al. (2021)	Random Forest	Random Forest performed better
Essa et al. (2021)	LSTM	LSTM models performed better
Coughlan et al. (2021)	AdaBoost and Random Forest	AdaBoost and Random Forest perform well

Based on Research question RQ1, different ML techniques and algorithms were used to achieve high accuracy in the course of developing an accurate lightning prediction model, these are expatiated below:

Mostajabi (2021) and Leal & Matos (2022) made use of meteorological datasets employing ML Random Forest algorithm. Mostajabi obtained highest accuracy with Random Forest while Leal & Matos (2022) got 71% accuracy.

Coughlan et al. (2021) and Coughlan et al. (2021) made use of ignition dataset considering ML algorithm. Coughlan et al. (2021) employed Random Forest Algorithm and got highest accuracy of 78% while Coughlan et al. (2021) with both AdaBoost and Random Forest produced out of sample accuracy.

Adekitan & Rock (2021) used geo-magnetic cubic structure using both Random Forest and Adaboost algorithm both came out with 100% accuracy

In view of the above findings, it has evidently shown that Machine Learning Random Forest algorithm with different datasets has outperforms all other ML algorithms and techniques.

RQ2: Regardless of the approach employed in lightning prediction system, what is the relationship between lightning prediction system accuracy and lead time?

Using machine learning techniques in Switzerland, Mostajabi (2021) reported on nowcasting lightning probability from easily available meteorology variables. They built a four-parameter model based on these four commonly used weather numerical parameters; air, temperature, relative humidity, atmospheric pressure, and wind speed. The produced signal is then confirmed using lightning tracking mechanisms. They were subjected to the random forest machine learning algorithm. Based on the evaluation's conclusions, its lead time has been between 0 and 10 minutes.

La Fata et al. (2021) presented a paper on cloud to ground lightning nowcasting using machine learning procedure. A specific datasets of eighteen (18) years geo environmental properties were used to predict a lead time of 1 hour in a three month. The result using Random Forest shows better accuracy in predicting a better lead time.

Schon et al. (2019) published a work on how to forecast lightning using model prediction error. The properties of ML such as error-dimensional optical flow process was employed to extract images of meteorological features which is the main pointer to thunderstorm and lightning. Decision Tree and neural network classifiers were trained to forecast lightning in the next hours. The findings show that an accuracy of 96% gotten over 15 minutes, also as the lead time increases the accuracy decreases.

Lightning forecast and environment factor analysis employing Random Forest algorithm was published by Yao et al. (2021). They proved lightning nowcasting model performs better when the lead time decreases. The accuracy of lightning prediction model for 0-6 hours are stated as 0.827-0.864 and the impact rate is 0.677-0.851.

Bao et al. (2022) released a paper on Artificial Neural Network for Lightning Forecasting based on Atmospheric Electric Field Observation. The properties of time series datasets from multiple electric field site were gotten, Sparse Auto Encoder (SAE) and image conversion were employed. Prediction of lightning in a given interval obtained. The findings yielded better performance with 82% accuracy and 92% precision.

Smith et al. (2020) published a research paper on the measures they obtained using machine learning on semantically segmented high-speed lightning video. An assessment of the reliability of a semantic segmentation system in detecting the number of individual strokes, leaders' directions, and striking points in high-speed lightning footage. The last layers of the model are retrained on lightning imagery, and the work uses a pre-trained DeepLabv3+ network, which was chosen to allow for the lowest processing needs. The network creates conceptually segmented images in which every pixel has a numerical label that is evaluated to figure out the number of strokes. Per strike, regions of interest are made and used to dampen noise. When tested, the system's stroke detection performance is 70.1%, its direction accuracy is 80%, and its strike point accuracy is 89.5%.

A baseline for the prediction of cloud-to-ground lightning in the United States was constructed by Bates et al. (2018). As for cloud-to-ground lightning in the United States, convective potential energy and precipitation rate products were used. This provides an easy technique for determining the threat of cloud-to-ground lightning by proxy computations from the output of numerical weather forecast models. Depending on the sort of forecast, this has generated a lead time of 15 days.

A paper was presented by Geng et al. (2020) on a heterogeneous spatiotemporal network for lightning prediction. The use of data driven model cannot handle a complex prediction with spatiotemporal datasets. The introduction of modules such as Guassian Difusion Module, ST encoder and ST decoder. This method is called heterogeneous spatiotemporal network (HSTN). The combination of these modules brings about better accuracy in both space and time domain.

Yucelbas et al. (2021) study concentrated on the use of suitable meteorological parameters which was before distance-based lightning. The aim of this effort is to develop a warning that will enable lightning to be foreseen ahead and safety precautions to be taken. By employing meteorological information obtained from a unique meteorological station, this work aims to anticipate lightning incidents one hour in advance. Ten atmospheric variables collectively make up the datasets used in this study. This is accomplished by looking at three groups while recording the lengths [DG-1, DG-2, DG-3] -2km, 2-4km, and 4-6km from the station, respectively. For each data group, the probability of lighting was calculated using the sequential forward selection (SFS) procedure. With a distance of 0–2 km, DG-1 generated

Meng et al. (2019) developed an architecture, a modular structure, and integrated warning approaches into the system to detect lightning a few minutes before it impacts. The system is capable of identifying, detecting, extrapolating, and distributing lightning and warning services in regions where lightning danger may occur. These products not only automatically reflect the changing trend of the lightning activity area but also the likelihood of lightning in significant locations. The system gives parameter interfaces and human-machine interaction capabilities. It can be used in many circumstances and locations.

The Schrodinger-Electrostatic Algorithm was used in a study by Emetere et al. (2014) to simulate lightning threat forecasts. The mathematical model showed considerable improvements in its use of micro-scale plasmas to produce the macro-scale atmospheric plasma, which has a main impact on lightning. With this approach, lightning may be detected more efficiently and precisely. The simulation also revealed that the primary factor influencing lightning forecast is air density in the high atmosphere.

Table 3. **Summary of relationship between lightning prediction system accuracy and lead time**

Author	Datasets	Method	Accuracy	Lead Time
Mostajabi (2021)	Meteorological datasets	ML Random Forest Algorithm	Higher Accuracy	0-10 Minutes
La Fata et al. (2021)	Not stated	ML Random Forest	Better Accuracy	1 hour
Christian (2019)	Images of meteorological features	ML Decision Tree and Neural Network	96%	15 minutes
Han (2021)	Not stated	ML Random Forest	82.7% to 86.4%	0-6 hours
Bao et al. (2022)	Time series datasets	Not stated	82% accuracy and 92% precision	Lead Time given at interval
Smith et al. (2020)	Strokes, Leaders direction and striking point	ML Algorithm	Striking point accuracy is 89.5%	
Bates et al. (2018)	CAPE and Precipitation	NWF Model	Not stated	15 days
Geng et al. (2020)	Spatiotemporal datasets	Gaussian Model	Better accuracy	Accurate time domain
Emetere et al. (2014)	Macroplasma and microplasma datasets	Schrodinger Mathematical Model	Higher accuracy	Low Lead Time

Based on research question RQ2, as shown in the table above. We observed that whenever the lead time is high, the accuracy is always low. On the other hands, when the lead time is low, then accuracy is always high. Additionally, the findings of Christian et al. (2019), Han et al. (2021) and Bao et al. (2022) shows that accuracy of lightning prediction models have inverse relationship with the lead time regardless of the techniques or datasets employed.

RQ3: What are the comparative advantages between Geo-satellite prediction and Numerical Weather Prediction in terms of Lightning Prediction?

A presentation was made by Leal & Matos (2022) on short term lightning forecast in the Amazon region using ground based weather apparatus and machine learning procedures was employed. A ground based datasets such as humidity, temperature, pressure and wind speed were subjected to ML algorithm. The result proved an accuracy of 71% in Amazon region.

Schon et al. (2019) published a work on how to forecast lightning using model prediction error. The properties of ML such as error-dimensional optical flow process was employed to extract images of meteorological features which is the main pointer to thunderstorm and lightning. Decision Tree and neural network classifiers were trained to forecast lightning in the next hours. The findings show that an accuracy of 96% gotten over 15 minutes, also as the lead time increases the accuracy decreases.

Coughlan et al. (2021) published a paper on using ML to predict fire ignition occurrences from lightning prediction. They stated clearly that there are relationship between fuel moisture and the ignition. Three ML algorithms such as decision tree, random forest and AdaBoost were employed involving ensemble procedure. Findings proved that Random Forest and AdaBoost has 78% accuracy. Also 71% of similar cases has the tendency of lightning occurrence.

Using machine learning techniques in Switzerland, Mostajabi (2021) reported on nowcasting lightning probability from easily available meteorology variables. They built a four-parameter model based on these four commonly used weather numerical parameters; air, temperature, relative humidity, atmospheric pressure, and wind speed. The produced signal is then confirmed using lightning tracking mechanisms. They were subjected to the random forest machine learning algorithm. Based on the evaluation's conclusions, its lead time has been between 0 and 10 minutes.

However, according to Romps et al. (2018), lightning for Australia might be projected using multiple land-scale atmospheric variables. This paper investigated how well six statistical and machine learning classification methods could discriminate between days containing and excluding lightning at the broad temporal and spatial scales of the latest broadly available models and reanalysis. Six sites in Australia from 2004 to 2013 were categorized using an integration of essential element analysis and logistic regression, regression trees and classification, linear discriminant analysis, random forests, quadratic discriminant analysis, and logistic lightning count. Performance on the classification assignment was measured via tenfold cross-validation. The outcomes show that logistic regression was superior to other classifiers and that applying climatological values for prediction is massively inferior.

Brocco et al. (2018) released an article on enhancing climate with data mining knowledge discovery. The benefits and challenges of mining enormous climate datasets are presented in this paper, in addition to the potential for KDD (Knowledge Discovery by Data Mining) to discover patterns pertinent to climate. This work focuses on a complicated and dynamic analysis's beta MAPS. It is used to highlight the links between recent statistical results that are local and non-local. This has included contrasting and comparing the known teleconnections between the outputs of climate reanalysis and climate model.

The INR-ELEC model was used by Ghavaylou et al. (2020) to evaluate lightning activity. The comparison of the INR-ELEC model with the World Wide Lightning Location Network (WWLLN) and satellite-based lightning data from the Lightning Imaging Sensor is the main objective of this study (LIS). Considered are the severity frequency as well as some of the physical and dynamic features of the lightning that occurred during an 11-year span of time (2004-2014). The observations and numerical simulations are juxtaposed both quantitatively and qualitatively. The study showed that there is relatively good agreement between the locations of lightning occurrence of WWLLN data as well as LIS observations and the simulated tune-averaged horizontal patterns of the lightning potential index (LPI) derived from ERA-Interim-based experiments. Furthermore, GFS-based simulations beat FNL (Final Analysis) and ERA-Interim simulations quantitatively in terms of NOL prediction, based on values of standard deviation (SD) and centered Root Mean Square Error (RMSE). Additional statistical analysis using numerous variables has shown that the ERA-Interim initialization has the highest performance for predicting lightning activity.

John et al. (2022) presented a work on a deep learning system for satellite based lightning nowcasting. Convolutional neural network architecture used to develop a model called lightningCast, it was trained with Geo-stationary satellite. The model is compactible will all Geostationary satellite. It has the capacity to convert enormous volume of satellite imagery to required objectives.

For the aim of assessing earth system model error, Silva et al. (2022) applied an interpretable machine learning approach. Modern Artificial Intelligence (A.I.) algorithms have revealed that these errors can be successfully predicted. Investigating the weaknesses in the NASA GEOS model's forecasting of the incidence of lightning, a popular Earth System Model, utilizing XGBoost classification trees and SHapley Additive exPlanations (SHAP) analysis. The model error can be adequately projected by this interpretable error prediction method, which also shows that the errors are highly correlated to convective processes and the properties of the land surface.

For a multi-source spatiotemporal lightning forecast, Geng et al. (2020) used a deep learning system. Studying a wide range of meteorological data in-depth is necessary for accurate weather forecasting. Using a lightning scenario, a deep neural network-based data-driven forecasting tool called LightNet+. Our framework's framework supports LightNet+ to foresee by mining additional data from various sources, which may be diverse in terms of space (continuous versus discrete) and time (past observations versus future simulations). Using actual climatic data from North China, we assess LightNet+. The research results indicate that: (a) LightNet+ significantly outperforms three established lightning forecasting methods; and (b) the amount of data supplied into LightNet+ enhances the predicting quality.

Meng et al. (2019) published a paper on the development of lightning prediction and alerting technique and its uses. The CAM-LNWS was upgraded by introducing two new algorithms related to early alerting and thunderstorm spread. The datasets gotten in Tianji in Beijing also Hebei from 2016-2017 were used to evaluate CAM-LNWS. The result shows more improvement as regards good lightning forecasting and alerting capability.

Kaulfus (2017) used geostationary lightning mapping and satellite technologies to forecast the time of a lightning strike. The GOES-R satellite was launched into orbit with sensors that could detect lightning. (GLM) Geostationary Lightning Mapper This ground-breaking instrument is a paradigm shifter. A computer algorithm will be fed the collected data. This trains computers on the patterns in the atmosphere that occur both before and during lightning events. There is a 75% possibility of lightning on the Geostationary Lightning Mapper (GLM).

Based on the reviewed work or paper on lightning prediction system, Numerical Weather Prediction has a specific and accurate prediction of a remote region as shown in research published by Mostajabi (2021), Leal & Matos (2022) and Romps et al. (2018), while Geo-satellite prediction might not be accurate due to time domain factors and updating as well as being always subjected to continuous upgrading considering the findings of Ghavaylou et al. (2020), Cintineo et al. (2022), Silva et al. (2022) and Geng et al. (2020). In view of the findings above Numerical Weather Prediction produces a better prediction model.

Table 4. **Summary base on Geo-satellite prediction and Numerical Weather Prediction in terms of Lightning Prediction**

Author	Datasets	Method	Status	Accuracy
Leal & Matos (2022)	Meteorological data	ML Algorithm	Remote Area (NWP)	71%
Coughlan et al. (2021)	Fuel moisture and lightning ignition	Random Forest and AdaBoost	Remote (NWP)	Random Forest and AdaBoost have 78% accuracy
Mostajabi (2021)	Meteorological datasets	Random Forest	Remote (NWP)	Random Forest gave very high accuracy
Romps et al. (2018)	Climatology dataset	ML, Logistic Regression Technique	Remote (NWP)	Higher accuracy
Ghavaylou (2020)	Lightning image	Geo-satellite	Global (Geo-satellite)	Higher accuracy
Cintineo et al. (2022)	Satellite imagery	Convolutional neural network	Global (Geo-satellite)	Compatible with all weather prediction
Silva et al. (2022)	Land surface properties	XGBoost classification trees and SHapley Additive exPlanations (SHAP) analysis	Global (Geo-satellite)	Better model error prediction
Geng et al. (2020)	Climatic dataset	Deep Neural Network	Global (Geo-satellite)	Higher performance
Meng et al. (2019)	Not stated	CAM-LNWS Geo-satellite	Global (Geo-satellite)	Improvement in lightning forecast
Kaulfus (2017)	Pattern in the Atmosphere	Geo Lightning Mapper	Global (Geo-satellite)	75% accuracy

TECHNOLOGICAL GAPS DISCOVERED

Researchers have done a lot of work in the area of lightning prediction/forecast as well as achieving many breakthrough. However, there are lots of mystery that needs more revelations as well as a lot of unanswered questions considering tools employed as well as the prediction results. These are well elaborated below:

Mostajabi (2021) and Leal & Matos (2022) made use of meteorological datasets employing Machine Learning procedures as well as different algorithms to predict lightning occurrence while Emetere et al. (2014) considered Schrodinger Mathematical Model using atmospheric microplasma and macroplasma as datasets. We observed that their models proved that there is always an inverse relationship or correlation between model accuracy and lead time. Inview of this, researchers have not come up with a lightning predicting model in which accuracy is lead time independent.

Romps et al. (2018) made use of data from Electric field wave from lightning properties such as shapes, amplitude and time, Meng et al. (2019) made use of thunderstorm datasets while Leal & Matos (2022) and Mostajabi (2021) considered meteorological datasets to develop lightning prediction model. It is observed that no researcher has given consideration to lightning strike incident angle as part of parameters needed for effective lightning prediction. Also, the swampy area parameters not considered as well, it seems lightning with higher magnitude often occur in swampy areas.

The accuracies of lightning prediction model developed by Leal & Matos (2022), Christian et al. (2019) and Coughlan et al. (2021) have 71%, 96%, 78%, and 95% respectively. The model accuracy error has not been used to predict the lightning model accuracy.

Conclusion

In conclusion this systematic literature review has explored various aspect of lightning prediction, focusing on machine learning algorithms, the relationship between lightning prediction accuracy and lead time, and the comparative advantages between Geo-satellite prediction and numerical weather prediction. Regarding the Machine learning algorithms, the review found out that Random Forest algorithm has majorly been used among researchers, and has outperformed any other algorithms used for lightning prediction in terms of accuracy. Regarding the relationship between lightning prediction accuracy and lead time, the review found out that, there is an inverse relationship between them such that when there is a high lead time, the accuracy is always low and vice versa. Finally, in terms of comparative advantages between Geo-satellite prediction model and numerical weather prediction model for lightning prediction, this literature review shows that the numerical weather prediction achieve high accuracies in remote areas compared to Geo-satellite prediction model.

Recommendations / feature work

Based on the finding from this research we recommend the use of the a hybrid of Random Forest algorithm and a deep learning algorithm for lightning prediction model as previous research has shown how well the Random Forest algorithm has perform as compared to other machine learning algorithms, and only few researchers have used the deep learning approach which also proven to be effective in predicting lightning occurrence. Also, we recommend a that researcher work on a lightning prediction model that will balance the relationship between the model accuracy and lead time. Finally, we recommend the use of the numerical weather prediction model when predicting lightning in a remote area.

References

Adekitan, A. I., & Rock, M. (2021). Data mining of lightning strike probability to simple structure. In S. I. Ao, L. Gelman, D. WL Hukins, & A. M. Korsunsky (Eds.), *Proceeding of the world congress on engineering (WCE 2021)*, July 7-9, 2021, London, U.K. (pp. 1-6). London, UK: Newswood Limited. Retrieved from https://www.iaeng.org/publication/WCE2021/.

Bao, R, Zhang, Y., Ma, B., Zhang, Z & He, Z. (2022). An artificial neural network for lightning prediction based on atmospheric electric field observations. *Remote Sensing*, *14*(17), 4131. DOI: https://doi.org/10.3390/rs14174131.

Bates, B., C., Dowdy, A., J., & Chandler, R., E. (2018). Lightning prediction for Austrialia using multivariate analyses of large-scale atmospheric variables. *Journal of Applied Meteorology and Climatology*, *57*(3), 525-534. DOI: https://doi.org/10.1175/JAMC-D-17-0214.1.

Brocco, A., Falasca, F., Nenes, A., Fountalis, I., & Dovrolis, C. (2018). Advancing climate science with knowledge-discovery through data mining. *npj Climate And Atmospheric Science*, *1*, 20174. DOI: https://doi.org/10.1038/s41612-017-0006-4.

Cintineo, J. L., Pavolonis, M. J., & Sieglaff, J. M. (2022). ProbSevere LightningCast: A Deep-Learning Model for Satellite-Based Lightning Nowcasting. *Weather and Forecasting*, *37*(7), 1239-1257. DOI: https://doi.org/10.1175/WAF-D-22-0019.1.

Coughlan, R., DiGiuseppe, F., Vitolo, C., Barnard, C., Lopez, P., & Drusch, M. (2021). Using machine learning to predict fire ignition occurrences from lightning forecast. *Meteorological Application*, *28*(1), e1973. DOI: https://doi.org/10.1002/met.1973.

Emetere, E., Akinyemi, M., Uno, U., & Boyo, A. (2014). Lightning threat forecast simulation using the schrodinger-electrostatic algorithm. *IERI Procedia*, *9*(1), 53-58. DOI: https://doi.org/10.1016/j.ieri.2014.09.040.

Essa, Y., Hunt, H. G. P., & Ajoodha, R. (2021). Short-term prediction of lightning in Southern Africa using autoregressive machine learning techniques. In *Proceedings of the 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*, Toronto, ON, Canada (pp. 1-5). DOI: https://doi.org/10.1109/IEMTRONICS52119.2021.9422493.

Etten-Bohm, M., Yang, J., Schumacher C., & Jun, M. (2021). Evaluating the relationship between lightning predictions in global climate models. *Journal of Geophysical Research Atmospheres*, *126*(5), e2020JD033990. DOI: https://doi.org/10.1029/2020JD033990.

Geng, Y., Li, Q., Lin, T., Zhang, J., Xu, L., Yao, W., Zheng, D., Lyu, W., & Huang, H. (2020). A hererogenous spatioemporal network for lightning prediction. In *Proceeding of the 2020 IEEE International Conference on data mining (ICDM)*, Sorrento, Italy, 2020 (pp. 1034-1039). DOI: https://doi.org/10.1109/ICDM50108.2020.00121.

Host, M. & Orucevic-Alagic, A. (2011). A systematic review of research on open source software in commercial software product development. *Information and Software Technology*, *53*(6), 616-624. DOI: https://doi.org/10.1016/j.infsof.2010.12.009.

Kaulfus, A. (2017, March). *Geostationary lightning mapping*. Retrieved from https://ghrc.nsstc.nasa.gov/lightning/overview_glm.html.

Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering – A systematic literature review. *Information and Software Technology*, *15*(1), 7-15. DOI: https://doi.org/10.1016/j.infsof.2008.09.009.

La Fata, A., Amato, F., Bernardi, M., D'Andrea, M., Procopio, R., & Fiori, E. (2021). Cloud-to-Ground lightning nowcasting using Machine Learning. In *Proceeding of the 2021 35th International Conference on Lightning Protection (ICLP) and XVI International Symposium on Lightning Protection (SIPDA)*, Colombo, Sri Lanka, 2021 (pp. 1-6). DOI: https://doi.org/10.1109/ICLPandSIPDA54065.2021.9627428.

Leal, A. F. R. & Matos, W. L. N. (2022). Short-term lightning prediction in the Amazon region using ground-based weather station data and machine learning techniques. In *Proceeding of the 2022 36th International Conference on Lightning Protection (ICLP)*, Cape Town, South Africa, 2022 (pp. 400-405). DOI: https://doi.org/10.1109/ICLP56858.2022.9942500.

Meng, Q., Yao, W., & Xu, L. (2019). Development of lightning nowcasting and warning technique and its application. *Advances in Meteorology*, *2019*, 2405936. DOI: https://doi.org/10.1155/2019/2405936.

Mostajabi, A. (2021). *Augmenting the performance of impoverished sensor networks using machine learning and time reversal: Application to lightning nowcasting and location*.

(Doctoral Thesis). EPFL, Switzerland. Retrieved from https://infoscience.epfl.ch/entities/publication/06e38293-4260-4d5d-ad03-2cebc00e630b.

Pakdaman, M., Naghab, S. S., Khazanedari, L., Malbousi, S., & Falamarzi, Y. (2020). Lightning prediction using an ensemble learning approach for northeast Iran. *Journal of Atmospheric and Solar-Terrestial Physics*, *209*, 105417. DOI: https://doi.org/10.1016/j.jastp.2020.105417.

Romps, D. M., Charn, A. B., Holzworth, R. H., Lawrence, W. E., Molinari, J., & Vollaro, D. (2018). CAPE Times P explains lightning over land but not land-ocean contrast. *Geophysical research Letters*, *45*(22), 12,623-12,630. DOI: https://doi.org/10.1029/2018GL080267.

Schon, C., Dittrich, J., & Muller, R. (2019). The error is the feature: how to forecast lightning using a model prediction error. In *KDD '19: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2979-2988). New York, USA: Association for Computing Machinery. DOI: https://doi.org/10.1145/3292500.3330682.

Schultz, M. G., Betancourt, C., Gong, B., Kleinert, F., Langguth, M., Leufen, L. H., Mozaffari, A., & Stadtler, S. (2021). Can deep learning beat numerical weather prediction? *Philosophical transactions of the Royal Society A*, *379*, 20200097. DOI: https://doi.org/10.1098/rsta.2020.0097.

Silva, S., Keller, C. A., & Hardin, J. (2022). Using an explainable machine learning approach to characterize earth system model errors: Appliction of SHAP analysis to modeling lightning flash occurrence. *Journal of Advances in Modeling Earth Systems*, *14*(4), e2021MS002881. DOI: https://doi.org/10.1029/2021MS002881.

Smith, J. R., Huntt, H. GP, Cross, T., Schumann, C., & Warner, T.A. (2020). Generation of metrics by semantic segmentation of high speed lightning footage using machine learning. In *Proceedings of the 2020 International SAUPEC/RobMech/PRASA Conference*, Cape Town, South Africa, 2020 (pp. 1-6). DOI: https://doi.org/10.1109/SAUPEC/RobMech/PRASA48453.2020.9041123.

Tippett, M. K. & Koshak, W. J. (2018). A Baseline for the predictability of US cloud-to-ground lightning. *Geophysica Research Letters*, *45*(19), 10,719-10,728. DOI: https://doi.org/10.1029/2018GL079750.

Yao, H., Guo, F., Song, L., & Pang, H. (2021). Lightning Prediction and Environment Factor Analysis using Random Forest Algorithm in Shandong Province, China. In *Proceedings of the 2021 IEEE 23rd Int Conf on High Performance Computing & Communications; 7th Int Conf on Data Science & Systems; 19th Int Conf on Smart City; 7th Int Conf on Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys)*, Haikou, Hainan, China (pp. 2227-2232). DOI: https://doi.org/10.1109/HPCC-DSS-SmartCity-DependSys53884.2021.00333.

Yucelbas, S., Erduman, A.,Yucelbas, C., & Yildiz, F. (2021). Pre-estimation of Distance-Based Lightning Using Effective Meteorological Parameters. *Arabian Journal for Science and Engineering*, *46*, 1529-1539. DOI: https://doi.org/10.1007/s13369-020-05257-0.

Zhou, K., Zheng, Y., Dong, W., & Wang, T. (2020). A deep learning network for cloud-to-ground lightning nowcasting with multisource data. *Journal of Atmospheric and Oceanic Technology*, *37*(5), 927–942. DOI: https://doi.org/10.1175/JTECH-D-19-0146.1.

About the authors:

Babatunte Dauda Raheem, Federal University Lokoja, Nigeria. tunderaheem007@gmail.com

Emeka Ogbuju, Federal University Lokoja, Nigeria. emeka.ogbuju@fulokoja.edu.ng

Francisca Oladipo, Federal University Lokoja, Nigeria. francisca.oladipo@fulokoja.edu.ng
Taiwo Abiodun, Federal University Lokoja, Nigeria. taiwo.abiodun-pg@fulokoja.edu.ng