IEICE TRANS. INF. & SYST., VOL.E105–D, NO.9 SEPTEMBER 2022

LETTER

A Novel Method for Lightning Prediction by Direct Electric Field Measurements at the Ground Using Recurrent Neural Network

Masamoto FUKAWA[†a)], **Xiaoqi DENG**[†], **Shinya IMAI**[†], **Taiga HORIGUCHI**[†], **Ryo ONO**[†], **Ikumi RACHI**[†], **Sihan A**[†], **Kazuma SHINOMURA**[††], **Shunsuke NIWA**[††], **Takeshi KUDO**[††], ***Nonmembers***, **Hiroyuki ITO**[†], **Hitoshi WAKABAYASHI**[†], ***Members***, **Yoshihiro MIYAKE**[†], ***and*** **Atsushi HORI**[†], ***Nonmembers***

SUMMARY A method to predict lightning by machine learning analysis of atmospheric electric fields is proposed for the first time. In this study, we calculated an anomaly score with long short-term memory (LSTM), a recurrent neural network analysis method, using electric field data recorded every second on the ground. The threshold value of the anomaly score was defined, and a lightning alarm at the observation point was issued or canceled. Using this method, it was confirmed that 88.9% of lightning occurred while alarming. These results suggest that a lightning prediction system with an electric field sensor and machine learning can be developed in the future.

key words: *lightning, prediction, LSTM, electric field*

1. Introduction

Lightning is a meteorological phenomenon that can cause enormous damage to humans and electric equipment [1]–[10]. To realize an "internet of things" society, the development of a lightning prediction system to prevent damage caused by lightning is critical. Previous research has proposed a lightning prediction method based on a threshold of atmospheric electric field data, which is not pragmatic. The lightning analysis system of the Japan Meteorological Agency (JMA), Lightning Nowcast [11], uses radar and electromagnetic radiation from lightning to observe thunderclouds and lightning on a 1 km mesh at 10 min intervals. However, it cannot detect thunderclouds and lightning in advance [12]. By utilizing Machine learning technology, alarm and alarm canceled system has been obtained with electric field data from a single field mill for the first time. As shown in Table 1, It is part of the papers we surveyed on using machine learning to predict lightning.

The purpose of using machine learning in this study is to predict the inference result from a large amount of time-series electric field data. Our model consisted of four layers and was composed of time-series electric field data recorded at observation points during fair-weather. The time-series electric field data at the time of occurrence of lightning was input into the model, and output values are used in lightning prediction calculations. The output values were compared with the actual electric field data, and the anomaly level was calculated. The threshold value of the anomaly level was determined using a confusion matrix. If the anomaly level exceeds the threshold value, the system issues a lightning alarm.

Table 1 Comparison of lightning prediction methods.

Author	Variables	Method	Field mill	alarm
[Tao *et al.*, 2016 [13]]	electric field	RNN	5	alarm
[Karen *et al.*, 2016 [14]]	multivariate	random forest	several	alarm
[Daniel *et al.*, 2018 [15]]	electric field	CRNN	30	alarm
[Bryson *et al.*, 2018 [16]]	multivariate	logistic regression	6	alarm
This paper	electric field	CRNN / Decision tree	1	alarm / canceled

Table 2 Dates when data were collected in 2017.

Propoerties of data		Month	Date
Fair-weather	Training data	July	13th, 14th, 15th, 20th, 21st
		Aug	2nd, 3rd, 12th, 13th, 14th, 16th, 19th, 20th, 24th, 27th, 28th
	Test data	July	3rd, 6th
		Aug	30th, 31st
Lightning-weather	Verification data	July	9th, 17th, 18th, 28th, 30th
		Aug	4th, 5th, 18th, 23th, 26th

2. Experiments

In experiments, the electric field existing between clouds and the ground were recorded at 1s intervals using a field mill placed in Amagasaki City Hyogo, Japan [17]. The field mill can record the electric field change induced by lightning 15 km away. The lightning data is provided by the JMA. Evacuation action is required within a radius of 10 km from where lightning occurs.

As shown in Table 2, the dataset was created from 16 days of fair-weather over a period of 2 months.

The training data consisted of 300 s of time-series electric field data as explanatory variables and (300+1) s of electric field data as response variables. The machine learning layer was designed with a many-to-one model to determine a single predicted value from continuous time-series electric field data. The model architecture is shown in Fig. 1.

Equation (1) is the Mean Absolute Percentage Error (MAPE). Since the model's output in this study is numeri-

Manuscript received March 14, 2022.
Manuscript revised April 25, 2022.
Manuscript publicized June 8, 2022.

[†]The authors are with the Tokyo Institute of Technology, Tokyo, 152–8550 Japan.
[††]The authors are with the OTOWA ELECTRIC CO, LTD., Amagasaki-shi, 661–0976 Japan.
a) E-mail: fukawa.m.aa@m.titech.ac.jp

DOI: 10.1587/transinf.2022EDL8026

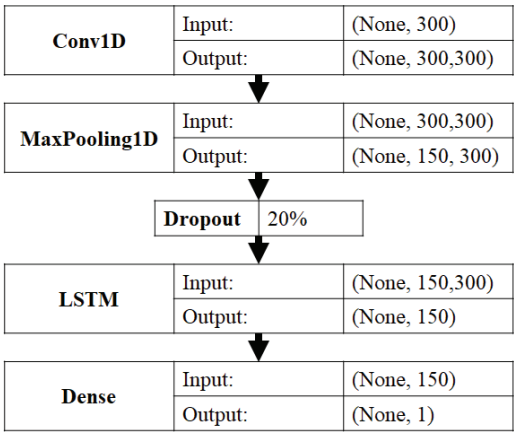

Fig. 1 Machine learning layer structure.

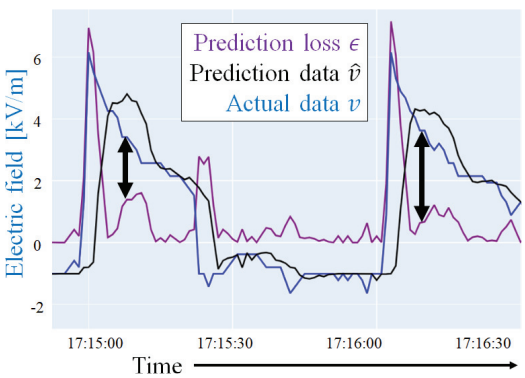

Fig. 2 The difference between the predicted data and the actual data is the predicted loss.

cal, MAPE was applied to measure accuracy.

$$MAPE = \frac{100}{n}\sum_{n=1}^{n}\left|\frac{\hat{v}_i - v_i}{v_i}\right| \tag{1}$$

$$= 24.00\%$$

Equation (2) is the relationship between ϵ, $\hat{v}$ and v. If an actual data v is input that the model cannot predict, the prediction loss ϵ returns a very large value. As shown in Fig. 2, if the prediction data $\hat{v}$ is accurate, then the prediction loss ϵ can be viewed as a residual of the actual data v. When v changes in a manner that the model can predict, the residual ϵ converges by $\hat{v}$ and approaches 0.

$$\textit{Prediction Loss} \quad \epsilon = |\hat{v} - v| \tag{2}$$

Equation (3) is the anomaly score proposed in the present study. The anomaly score is the sum of the difference 300 s between the model output and the actual electric field values. The 300 s was determined by evaluating several intervals. Equation (4) is the value obtained using only the electric field values via the conventional method. This model is intended to accurately predict the electric field value in a fair-weather. Since Eq. (3) is the sum of Eq. (2), the ideal anomaly score in fair-weather should be 0. In the case of lightning, the model is exposed to sudden changes in the actual electric field value. The discrepancy between the model's predicted value and the actual electric field value increases, resulting in a rise in anomaly score.

$$\textit{Anomaly score} \quad \alpha = \log_2\left(\sum_{i}^{i+300}\epsilon\right) \tag{3}$$

$$\textit{Electric field sum} \quad \beta = \log_2\left(\sum_{i}^{i+300}|v|\right) \tag{4}$$

A confusion matrix was used for the simulation to determine the threshold. The variables in the confusion matrix are as follows. The number of true positives (TP) is the number of correct answers where the system predicted lightning and it occurred. The number of false negatives (FN) is the number of incorrect answers where the system predicted that lightning would not occur, but it occurred. The number of false positives (FP) is the number of incorrect answers where the system predicted that lightning would occur, but one did not occur. The number of true negatives (TN) is the number of correct answers for which the system predicted that lightning would not occur, and it did not occur. The parameters used to evaluate the confusion matrix include the accuracy, precision, recall, and the F1 score; the formulas are as follows.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{5}$$

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall\ (TPR) = \frac{TP}{TP + FN} \tag{7}$$

$$F1Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{8}$$

The success of lightning prediction was evaluated for each occurrence of lightning. It is necessary to determine the evaluation radius when evaluating the success because lightning can be somewhere on a global scale. Therefore, we defined lightning within 15 km of the field mill as actual lightning and 15 to 21 km as no lightning. If the threshold was too low, alarms were issued even though no lightning occurred, resulting in numerous false alarms. Therefore, the confusion matrix was calculated for each change in the threshold to find the optimal threshold. Although recall and precision are both accuracy indices, they complement each other; the appropriate threshold should be considered, and the highest value of the F1 score should be observed, which is the harmonic mean. As shown in Fig. 3, we determined that the threshold for the anomaly score was 6.6 and that the threshold for the electric field sum was 5.1. Table 3 showed the confusion matrix when the threshold for the anomaly score was set to 6.6.

As shown in Fig. 4, the same trend was confirmed for 10 days of lightning-day data, one of which is shown below. An alarm can be issued for both the anomaly score and

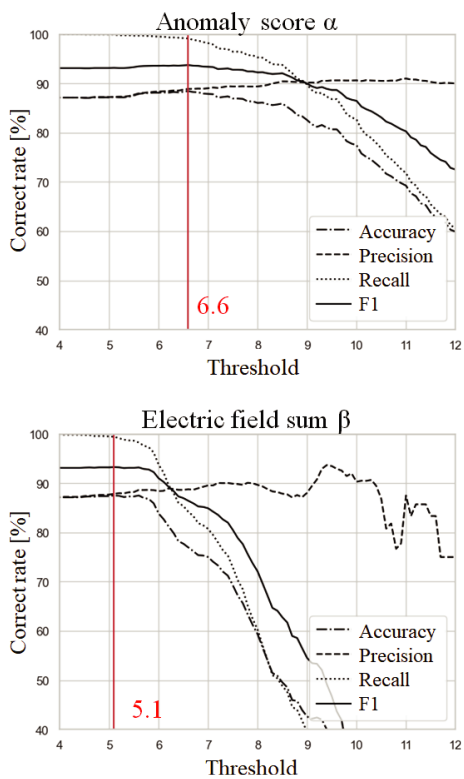

Fig. 3 We looked for the point of maximum F1 score while varying the threshold.

Table 3 Confusion matrix of anomaly score with optimal F1 score.

Anomaly score α Threshold = 6.6		Actual: Lightning	Actual: No lightning
Predicted	Lightning	(TP) 1368	(FP) 171
	No lightning	(FN) 12	(TN) 33

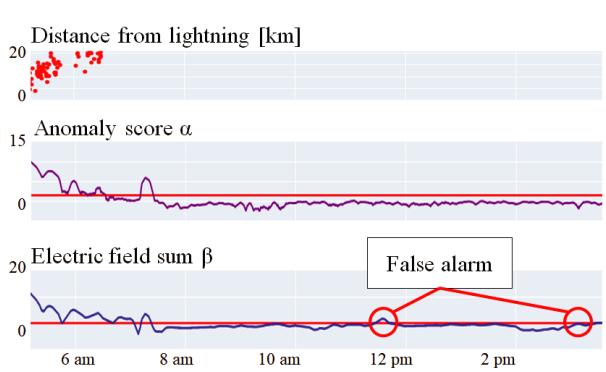

Fig. 4 Comparison of this study and conventional methods for lightning.

the electric field value when the score exceeds the threshold. For the anomaly score, when lightning passed through, the score was well below the threshold and did not result in a false alarm. However, in the electric field value, the score was often above the threshold when lightning passed through, resulting in a false alarm.

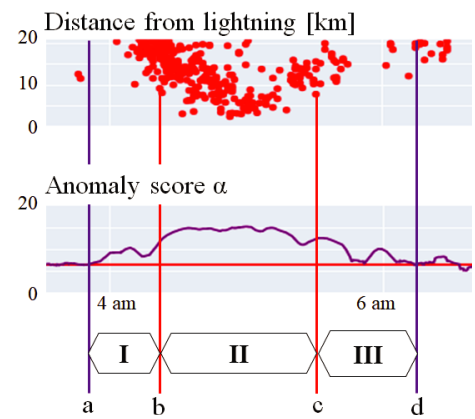

Fig. 5 Lightning alarm and alarm canceled based on the Anomaly score.

Table 4 The alarm is 32 m and 10s before the lightning, and the alarm cancels at 42 m and 15 s after the lightning passes.

Period	Time [hh:mm:ss]	Description
I	03:49:50 to 04:22:00	Period of need to evacuate immediately. In this case for 32 m 10 s.
II	04:22:00 to 05:31:00	Period of dangerous with lightning. In this case for 1 h 9 m 0 s.
III	05:31:00 to 06:13:15	Period of check if no danger. In this case for 42 m 15 s.

Table 5 Lightning after the anomaly score exceeds the threshold and it falls below the threshold when lightning passes.

Moment	Time [hh:mm:ss]	Description
a	03:49:50	Anomaly score is exceed the threshold. It is an evacuation alarm can be issued.
b	04:22:00	First lightning 9.28-km distance.
c	05:31:00	Final lightning 7.88-km distance.
d	06:14:15	Anomaly score is fall the threshold. It is the evacuation alarm can be canceled.

3. Results and Discussion

As shown in Fig. 5, the same trend was confirmed for 10 days of lightning-day data, one of which is shown below. Tables 4 and 5 show the specific time of issuance and cancellation of the alarm.

Equations (9), (10), (11), and (12) can be used to measure the accuracy of lightning predictions. The true positive rate (TPR) is the correct rate when the predicted lightning occurs. The probability of false detection (POFD) is a false alarm rate when lightning is predicted even though there is no lightning. The probability of false alarm (POFA) is a false alarm rate when the predicted lightning did not occur. True Skill Statistics (TSS) is the difference between TPR and POFD, a rating index sometimes referred to as Hanssen-

Table 6 Comparison of lightning prediction accuracy between related studies and this paper.

	TPR % (Recall)	POFD %	POFA %	OUI %
[Mazany *et al.*, 2002]	87.5	23.1	30.0	60.2
[Kehrer *et al.*, 2006]	95.0	47.0	45.3	45.0
[Murphy *et al.*, 2008]	37.7	N / A	71.0	N / A
[Da Silva Ferro *et al.*, 2011]	60.0	N / A	41.0	N / A
[Daniel *et al.*, 2018]	77.6	83.0	48.1	53.9
This Paper	88.9	83.8	11.1	44.3

Kuiper Skill Score. The operating utility index (OUI) is a statistic from a related paper by Kehrer *et al.* [18] that is used to weigh the importance of TPR. This paper follows the method used by Daniel *et al.* [15] to evaluate the lightning prediction accuracy. However, a simple comparison is difficult because each study uses different prediction methods. In this study, we confirmed that the anomaly score α calculated from the prediction loss ϵ responds to steep fluctuations in the electric field due to lightning. We have shown that it is possible to distinguish between fair-weather and lightning using only electric field data. As shown in Table 3, it is possible to suppress the FN values by distinguishing between the states. As shown in Table 5, the measured accuracy based on TPR is 88.9% probability of alarm in actual lightning. This is the second-highest accuracy after that achieved by Kehrer *et al.* [18]. The POFD is higher because the present study evaluates the success prediction for each lightning and only evaluates alarms for lightning that are likely to cause false alarms over short distances of 15 to 21 km. The weighted accuracy of OUI was 44.3%.

$$POFD = \frac{FP}{FP + TN} \tag{9}$$

$$POFA = \frac{FP}{TP + FP} \tag{10}$$

$$TSS = TPR - POFD \tag{11}$$

$$OUI = \frac{3(TPR) + 2(TSS) - (POFA)}{6} \tag{12}$$

4. Conclusion

Focusing on TPR, the result of this paper, the lightning prediction accuracy is 88.9%. The alarm was 32 m and 10 s before the lightning approached 10 km, and the alarm canceled 42 m and 15 s after the lightning passed. The proposed model can be trained using only fair-weather data that does not require observations of lightning-day and is expected to be used in regions with little lightning. Since lightning can be predicted using only a tiny amount of electric field data, the proposed method is expected to improve real-time performance and reduce calculation costs. Furthermore, since only one observation point is needed, the mesh can be refined and more accurate by increasing the number of observation points. Future work is expected further to deepen the architecture of the machine learning model structure and add electric field data to improve the accuracy of electric field predictions. This will improve the accuracy of lightning predictions and enable alarm earlier.

Acknowledgments

This work was partly supported by JST COI Grant Number JPMJCE1309; the RNN program was supported by NEUSOFT Japan.

References

[1] C.V. Wetli, "Keraunopathology: An analysis of 45 fatalities," The American Journal of Forensic Medicine and Pathology, vol.17, no.2, pp.89–98, 1996.

[2] J. Montanya, D. Aranguren, N. Pineda, G. Sola, D. Romero, and V. March, "Total lightning, electrostatic field and meteorological radar applied to lightning hazard warning," Proc. 20th International Lightning Detection Conference, 2008.

[3] M. Ströhle, B. Wallner, M. Lanthaler, S. Rauch, H. Brugger, and P. Paal, "Lightning accidents in the Austrian alps–A 10-year retrospective nationwide analysis," Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine, vol.26, no.1, pp.1–9, 2018.

[4] A.S. Cruz-Bernal, H. Torres-Sánchez, H. Aranguren-Fino, and J.C. Inampués-Borda, "Lightning mortality rate in Colombia for the period 1997-2014," Revista UIS Ingenierías, vol.17, no.2, pp.65–74, 2018.

[5] J. Hinkelbein, O. Spelten, and W.A. Wetsch, "Lightning strikes and lightning injuries in prehospital emergency medicine. relevance, results, and practical implications," Der Unfallchirurg, vol.116, no.1, pp.74–79, 2013.

[6] D. Pan, F. Wu, Y. Liu, T. Zhang, J. Gao, and H. Cai, "Design and implementation of IoT system and equipment for lightning parameter monitoring," 2017 IEEE 5th International Symposium on Electromagnetic Compatibility (EMC-Beijing), pp.1–5, IEEE, 2017.

[7] A. Necci, G. Antonioni, V. Cozzani, E. Krausmann, A. Borghetti, and C.A. Nucci, "A model for process equipment damage probability assessment due to lightning," Reliability Engineering & System Safety, vol.115, pp.91–99, 2013.

[8] M.M. Lengyel, H.E. Brooks, R. Holle, and M.A. Cooper, "Lightning casualties and their proximity to surrounding cloud-to-ground lightning," Ph.D. thesis, University of Oklahoma, 2004.

[9] M. Gharaylou, M.M. Farahani, M. Hosseini, and A. Mahmoudian, "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, 105067, 2019.

[10] M.A.D.S. Ferro, J. Yamasaki, D.R.M. Pimentel, K.P. Naccarato, and M.M.F. Saba, "Lightning risk warnings based on atmospheric electric field measurements in Brazil," Journal of Aerospace Technology and Management, vol.3, no.3, pp.301–310, 2011.

[11] E. Yoshikawa, S. Horishita, and T. Ushio, "Aircraft lightning potential area detection enhanced by echo top height and its evaluation with winter lightning cases," Journal of Atmospheric Electricity, vol.39, pp.70–78, 2021.

[12] W.H. Beasley, D.E. Williams, and P.T. Hyland, "Analysis of surface electric-field contours in relation to cloud-to-ground lightning flashes in air-mass thunderstorms at the Kennedy space center," 2nd International Lightning Meteorology Conference, pp.24–25, 2008.

[13] T. Guo, R. Liu, H. Yang, L. Shi, F. Li, L. Zhang, Y. Chen, Z. Liu, and F. Luo, "Predict atmosphere electric field value with the LSTM neural network," 2017 International Conference on Computer Systems, Electronics and Control (ICCSEC), pp.263–266, 2017.

[14] K.D. Blouin, M.D. Flannigan, X. Wang, and B. Kochtubajda, "En-

semble lightning prediction models for the province of Alberta, Canada," International Journal of Wildland Fire, vol.25, no.4, pp.421–432, 2016.

[15] D.E. Hill, R.G. Hefron, and R.S. Seymour, "Evaluating the efficacy of electric field mills to predict lightning events near Cape Canaveral using convolutional recurrent neural networks," 25th International Lightning Detection Conference, Florida, USA, 2018.

[16] B.C. Bates, A.J. Dowdy, and R.E. Chandler, "Lightning prediction for Australia using multivariate analyses of large-scale atmospheric variables," Journal of Applied Meteorology and Climatology, vol.57, no.3, pp.525–534, 2018.

[17] W. Rison and G. Chapman, "Lightning protection for high explosives and instrumentation using a field mill system," NOAA, International Aerospace and Ground Conference on Lightning and Static Electricity pp.289–293 (SEE N 89-10429 01-47), 1988.

[18] K. Kehrer, B. Graf, and W.P. Roeder, "Global positioning system (GPS) precipitable water in forecasting lightning at spaceport Canaveral," Weather and Forecasting, vol.23, no.2, pp.219–232, 2008.