Data Mining Analysis of Lightning Strike Probability to Simple Structures

Aderibigbe Israel Adekitan and Michael Rock

Abstract—**In this data era, data mining and machine learning, in general, have been applied in various research fields. Designing adequate lightning protection systems requires a good understanding of how lightning interacts with grounded structures. The probability of a lightning strike to various points on a structure reflects the associated likelihood for a lightning strike to terminate on such points. In this study, the dataset generated by applying a numerical approach to the computation of the dynamic electro-geometrical model for a cuboid structure was analyzed using data mining techniques. Data classification and regression-based predictive analyses were carried out on the Orange data-mining platform and MATLAB. Cases with 100% classification accuracy were observed using the Random Forest and the AdaBoost algorithms.**

Index Terms—**Data mining algorithms, machine learning, dynamic electro-geometrical model, lightning protection system**

I. INTRODUCTION

LIGHTNING is one of the major natural causes of damages to structures on earth, and these damages come with associated economic consequences [1-3]. Lightning hazards can be in different forms, such as thermal and mechanical damages, insulation failures, lightning-induced fires, electromagnetic compatibility issues, touch and step voltages, death of animals and humans in extreme cases etc. A number of lightning-related disasters such as plane crashes, disruption of power grids and supply outages [4], space rocket launch failures, e.g. Atlas-Centaur, and for humans, cases of skin burns, cardiac and respiratory arrest, eye injury, hearing losses, post-lightning-strike depression, and even death have been recorded [5, 6].

Protecting structures and lives against the risk of lightning strikes requires deploying appropriately sized and technically accurate lightning protection systems, and this should be achieved at a reasonable cost [7, 8]. The intensity of lightning strikes varies in different parts of the world, as expressed by the satellite observed ground flash densities measured in flashes/km^2/year. This implies that the risk of a lightning strike and associated lightning energy varies for similar structures depending on their geographical locations.

On a structure, such as a building, the risk of a strike is generally not the same for all points on the structure. The geometry and nature of points, e.g. flat, curved, edges, corners, etc., influences how lightning interacts with such points [9]. While some points are critically at risk of a direct lightning strike, others are not. In addition, some points on a structure may be struck by high lightning current strokes, while only lightning with low currents may be able to hit other points. These realities imply that structures must be adequately analyzed to determine the suitable type of lightning protection system to be deployed to ensure proper sizing, adequate protection, and design reliability.

The electro-geometrical model is a major technique for evaluating likely strike points on a structure, by applying a rolling sphere of specific radii based on the desired level of protection, in line with IEC 62305 lightning protection standard [10, 11]. The electro-geometrical model helps to determine points to position lightning rods, but it cannot be used to estimate the probability of strike to different points on a structure. This can be achieved by using the dynamic electro-geometrical model (DEGM) [12].

In this study, the probability of a lightning strike to various points on a cuboid structure was computed using numerical simulations based on the dynamic electro-geometrical model. The simulation requires extensive coding and analysis, which may be difficult to implement by engineers in the field, and as such, suitable alternative approaches that are fast to deploy will be more appropriate for the field design of lightning protection systems using the dynamic electro-geometrical model. This study evaluates the possibility of data mining useful knowledge from the simulated results in order to develop a trained data mining model that can be deployed for both structure point-type classification and predictive strike probability analysis.

II. DYNAMIC ELECTRO-GEOMETRICAL MODEL

The concept of the dynamic electro-geometrical model entails simulating possible lightning strikes to a structure of interest within a space volume known as lightning collection volume, from which lightning can strike the structure [13]. The object is meshed into several surface points using numerical techniques, and also, spaces around and above the structure from which lightning can head towards the structure are also meshed into several space points [14], and the potential of a strike from each of the space point to all of the points on the structure is evaluated, and the final point of the strike, based on the shortest geometrical distance is computed iteratively [15, 16]. The overall probability of a lightning strike to any point on the structure can be

---

Manuscript received January 31, 2020; revised April 07, 2021.

This is to appreciate the support of the Nigerian Petroleum Technology Development Fund (PTDF) through the Overseas Scholarship Scheme in partnership with the German Academic Exchange Service (DAAD).

A. I. Adekitan (MIAENG) is with the Group for Lightning and Surge Protection, Technische Universität Ilmenau, Germany, (phone: +49 3677 69-1878; e-mail: aderibigbe-israel.adekitan@tu-ilmenau.de).

M. Rock is with the Group for Lightning and Surge Protection, Technische Universität Ilmenau, Germany, (e-mail: michael.rock@tu-ilmenau.de).

determined using lightning current probability density functions,

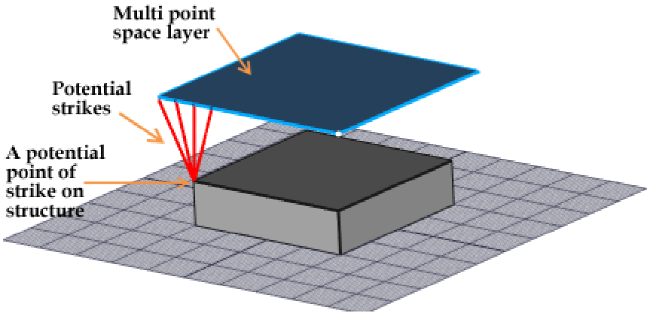

Fig. 1. Potential lightning strikes from points in space above the structure

The higher the probability of a lightning strike to any point on the structure, the higher the need for adequate air termination at that point. The relative differences in the probability of lightning strike among the points are indicative of the level of exposure of each point to a direct lightning strike.

The dynamic electro-geometrical model is computer-resources intensive, as it requires several thousands of iterations, which take hours to compute and extensive programming skills, are essential to implement. Fig. 1 demonstrates how potential lightning strikes from points in space to points on the structure were analyzed.

Fig. 2 displays the shape of the lightning collection volume around a cuboid structure with increasing height of space points. This creates a limit of the attractive zone for each height considered around and above the structure.

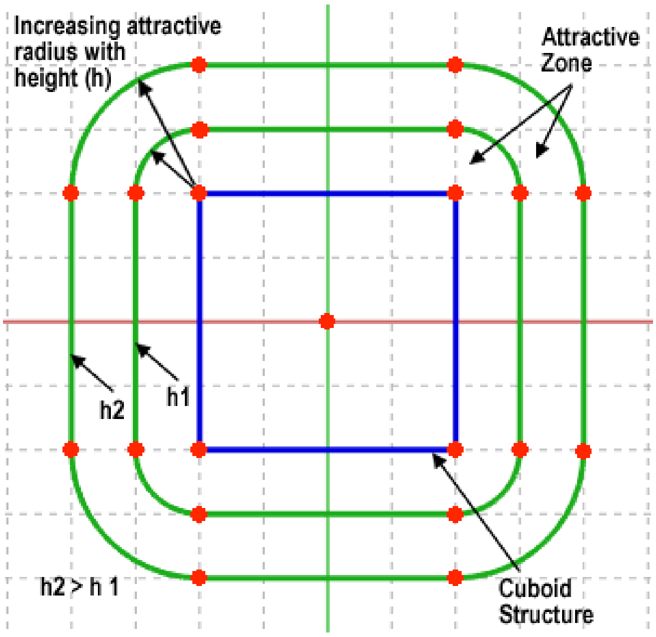

Fig. 2. The attractive zone around the cuboid structure

III. THE CASE STUDY

In this study, two types of cuboid structures were analyzed. The first cuboid shown in Fig. 3 has a square surface of 40 m by 40 m, and it is hereby referred to as cuboid A. The second cuboid shown in Fig. 4 has a surface area of 20 m by 50 m, and it is referred to as cuboid B. For both cuboid A, and cuboid B, three cuboid height H_c were considered, and these are 10 m, 30 m, and 50 m.

As shown in Fig. 5, the discretized points on the surface of the cuboid were classified into five: corner, roof edge, wall edge, inner roof, and sidewall. Based on the surface point classification, four-parameter values were created as features to be used as inputs in the data mining analyses using the probability of lightning strike to each point, which was obtained from the dynamic electro-geometrical model as the desired model target.

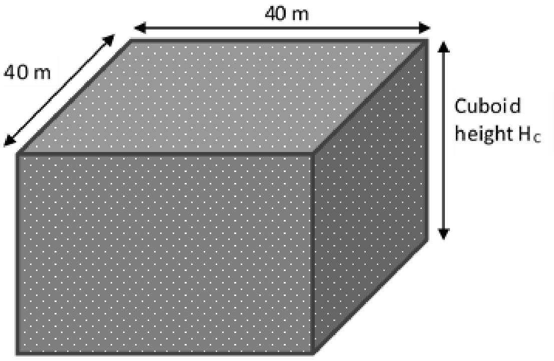

Fig. 3. Cuboid A

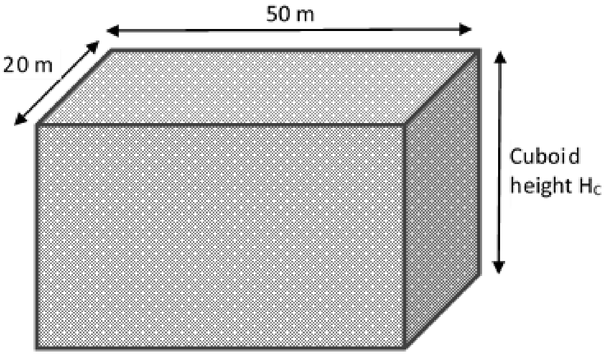

Fig. 4. Cuboid B

IV. METHODOLOGY

Data mining involves an intuitive search through a dataset in order to identify hidden knowledge and patterns by analyzing the dataset using purpose-specific algorithms. Data mining has been applied in various engineering studies, in medicine for identifying unhealthy organ test images, and for cell sample analysis. It has also been applied for educational data mining, for fault classification in electrical engineering [17], for internet traffic analysis [18], for lightning hazard classification [4] etc. It is a field that involves both computer science and statistics-based knowledge.

The probability of strike to discretized points on cuboid structures A and B, of height 10 m, 30 m, and 50 m, were determined using the dynamic electro-geometrical model. The simulated probability is the target of the data mining analysis. Four input features were created as follows:

- Feature A – the distance between the center of the cuboid at ground level and every other discretized point on the surface of the cuboid.
- Feature B – the angular exposure of each surface point on the structure on a flat plane, as shown in Fig. 6. For corners and wall edges, it is 90°, but for other types of points, it is 1°.

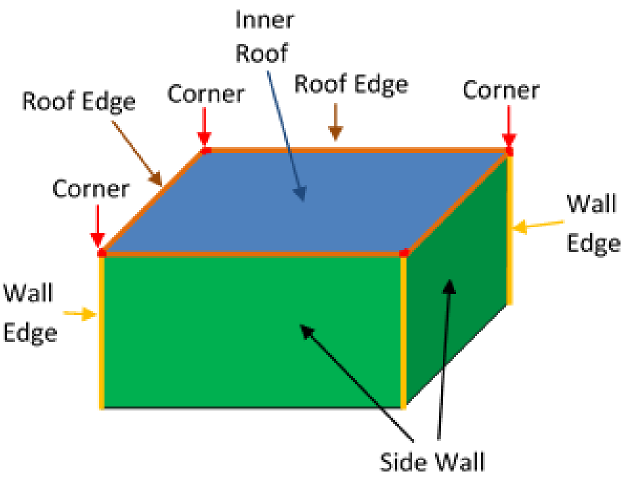

Fig. 5. Classification of surface points into five types

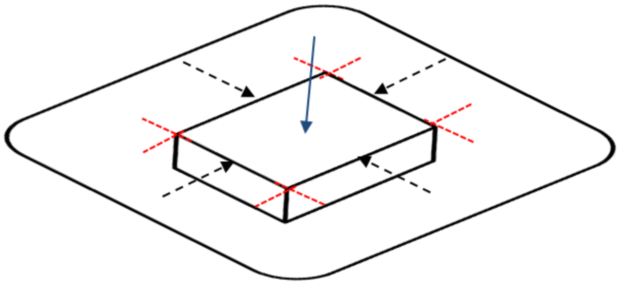

Fig. 6. The planar angular exposure of the classified surface points

- Feature C – the height of each discretized surface point above the ground.
- Feature D – the type of surface point coded as corner (C), roof edge (RE), wall edge (WE), inner roof (IR), and sidewall (SW).

The data mining analysis was performed on Orange data mining software and MATLAB using the classification learner and the regression learner. A stratified 10-fold cross-validation sampling was applied. First, the dataset was classified to identify the suitability of data mining techniques for this type of dataset. Using Feature A, Feature B, Feature C, and the Target (DEGM strike probability) as inputs, the samples were classified based on the different types of points, i.e. classes in Feature D. Further, a prediction of the strike probability, i.e., the Target was also performed using regression learners.

V. RESULTS

The results of the data mining analysis, both for classification and prediction, are presented in this section.

A. Classification analysis

On the Orange platform, the following four algorithms were applied: Tree, Random Forest, Neural Network, and the AdaBoost algorithm. 36900 samples were analysed in the data mining experiment. Table I presents the percentage of the true positive (TP) and the false positive (FP) predictions for the four data mining algorithms.

From Table 1, it was observed that the Random Forest, Neural Network, and the AdaBoost algorithm had zero false predictions. All the samples were accurately classified. The Tree algorithm had the only false classification of 1.99%, of which falsely classified sidewall (SW) points contributed 1.8%. The sidewall points generally have a low probability of strike, with values that are relatively close, which affected the ability of the algorithm to classify all the SW samples accurately. The receiver operating characteristic (ROC) curve for the classification of the SW data class is shown in Fig. 7.

The overall performance of the data mining classifiers was evaluated according to their Precision rate, the Classification Accuracy (CA), the F1 score, the Recall, and the Area under ROC Curve (AUC) as presented in Table II. The Random Forest, Neural Network, and the AdaBoost algorithm had a perfect classification accuracy of 1.0, i.e., 100%.

TABLE I
COMPARISON OF THE FALSE AND TRUE PREDICTIONS OF THE ALGORITHMS

Classes	Tree (%)		Random Forest (%)		Neural Network (%)		AdaBoost (%)	
	TP	FP	TP	FP	TP	FP	TP	FP
C	0.00	0.00	0.20	0.00	0.20	0.00	0.20	0.00
IR	19.93	0.00	19.93	0.00	19.93	0.00	19.93	0.00
RE	4.75	0.20	4.75	0.00	4.75	0.00	4.75	0.00
SW	71.22	1.80	71.22	0.00	71.22	0.00	71.22	0.00
WE	2.10	0.00	3.90	0.00	3.90	0.00	3.90	0.00
Overall	98.01	1.99	100.00	0.00	100.00	0.00	100.00	0.00

TABLE II
PERFORMANCE COMPARISON OF THE ALGORITHMS

Model	AUC	CA	F1	Precision	Recall
Tree	0.9698	0.9801	0.9765	0.9786	0.9801
Random Forest	1.0000	1.0000	1.0000	1.0000	1.0000
Neural Network	1.0000	1.0000	1.0000	1.0000	1.0000
AdaBoost	1.0000	1.0000	1.0000	1.0000	1.0000

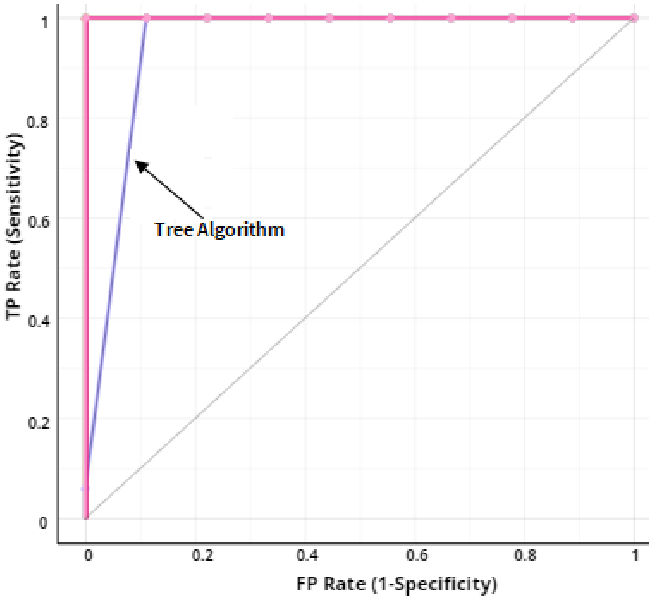

Fig. 7. ROC curve for the SW classification

A similar analysis to that performed on Orange was repeated on the MATLAB classification learner app. A 10-

fold cross-validation analysis was performed using 4 predictors, 5 response classes, and 36900 observations. Fine Tree, Fine k-Nearest Neighbors (KNN), Weighted KNN, and the Ensemble (Bagged Trees) algorithms were found suitable for the dataset. For the Fine Tree and the Fine KNN, the Euclidean was applied as the distance metric using equal distance weight. For the Weighted KNN, the squared inverse was applied as distance weight, while for the Ensemble learner, 30 decision tree learners were bagged together for the analysis. All 4 algorithms on MATLAB had an accuracy of 100%. The similar area under the ROC curve for the classification is shown in Fig 8.

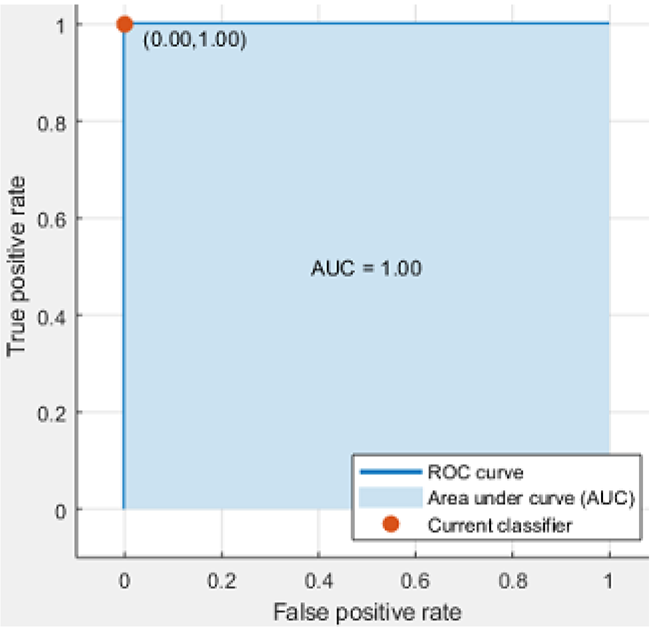

Fig. 8. The area under the ROC curve for the classification on MATLAB

B. *Strike probability (Target) prediction*

A technique for predicting the probability of strike to structures; in this case, cuboid structures without having to conduct data-intensive numerical simulations is very much desired. In an effort towards achieving this, a data mining analysis for developing trained regression-based models for predicting the dynamic electro-geometrical model probability (i.e., Target) using the four input features was performed. For this analysis, the sidewall (SW) was coded as 1, the wall edge (WE) as 2, corner (C) as 3, roof edge (RE) as 4, and inner roof (IR) as 5.

For the Orange software, the result of the four algorithms is presented comparatively in Table III, in terms of the Mean Square Error (MSE), the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the R-squared values of each model.

TABLE III
PERFORMANCE COMPARISON OF THE ALGORITHMS

Model	MSE	RMSE	MAE	R^2
Tree	0.0000	0.0008	0.0001	1.0000
Random Forest	0.0000	0.0008	0.0001	1.0000
Neural Network	0.0011	0.0333	0.0133	0.9973
AdaBoost	0.0000	0.0008	0.0001	1.0000

The results in Table IV show the predicted probability of each of the four data mining algorithms in comparison with the actual expected values (Target) for a few selected sample surface points on the cuboid structures. Although the predictions are not 100% perfect, they give reasonable values that are sufficient for guiding decisions when designing air termination systems.

TABLE IV
COMPARISON OF THE PREDICTED PROBABILITY OF A FEW SURFACE POINTS

	Cuboid Height	Random Forest	AdaBoost	Neural Network	Tree	Target
1	A - 10 m	0.00005	0.00005	0.00114	0.00000	0.00001
2	A - 10 m	0.29900	0.29900	0.31030	0.29900	0.29900
3	A - 10 m	0.00646	0.00640	0.00937	0.00637	0.00637
4	A - 10 m	10.91780	10.91780	11.66230	10.91780	10.91780
5	A - 30 m	0.23580	0.23580	0.23764	0.23580	0.23580
6	A - 30 m	0.00296	0.00296	-0.00241	0.00295	0.00288
7	A - 30 m	14.65170	14.65170	14.98110	14.65170	14.65170
8	A - 50 m	0.20343	0.20343	0.20143	0.20343	0.20343
9	A - 50 m	0.00009	0.00010	0.04605	0.00038	0.00030
10	A - 50 m	16.05320	16.05320	16.60030	16.05320	16.05320
11	B - 10 m	0.00005	0.00005	0.00692	0.00000	0.00001
12	B - 10 m	0.33125	0.33125	0.31122	0.33125	0.33125
13	B - 10 m	12.09540	12.09540	11.62390	12.09540	12.09540
14	B - 30 m	0.00182	0.00181	0.04332	0.00181	0.00219
15	B - 30 m	0.00005	0.00005	0.00526	0.00008	0.00010
16	B - 30 m	15.65920	15.65920	14.95380	15.65920	15.65920
17	B - 50 m	0.01251	0.01260	0.13857	0.01260	0.01260
18	B - 50 m	16.94610	16.94610	16.60570	16.94610	16.94610
19	B - 50 m	0.00005	0.00005	0.00196	0.00023	0.00023
20	B - 50 m	0.21474	0.21474	0.20159	0.21474	0.21474

The use of a predictive model of this nature will eliminate the need for repetitive numerical simulations once a predictive data mining model is appropriately trained with sufficient relevant samples using various dimensions of the target structure.

A 10-fold cross-validation analysis was also performed on the regression learner app on MATLAB for predicting the DEGM-simulated probability (i.e., Target) using the four input features. For this analysis, only the Fine Tree algorithm worked well with the dataset, with a minimum leaf size of 4, and the surrogate decision splits set to OFF. The performance of the model is presented in Table V.

TABLE V
PERFORMANCE OF THE FINE TREE ALGORITHM

Model	MSE
MSE	0.0000
RMSE	0.0015
MAE	0.0004
R^2	1.0000

Fig. 9 shows the residual plot. This indicates the difference between the actual value of the Target and the predicted value. Fig. 10 shows the actual versus response plot of the Fine Tree regression algorithm.

TABLE VI
SELECTED PREDICTIONS OF THE FINE TREE ALGORITHM

	Cuboid Height	Fine Tree	Target
1	A - 10 m	0.00005	0.00001
2	A - 10 m	0.29900	0.29900
3	A - 10 m	0.00663	0.00637
4	A - 10 m	10.91782	10.91780
5	A - 30 m	0.23824	0.23580
6	A - 30 m	0.00246	0.00288
7	A - 30 m	14.65173	14.65170
8	A - 50 m	0.20691	0.20343
9	A - 50 m	0.00196	0.00030
10	A - 50 m	16.05324	16.05320
11	B - 10 m	0.00005	0.00001
12	B - 10 m	0.33125	0.33125
13	B - 10 m	12.09545	12.09540
14	B - 30 m	0.00196	0.00219
15	B - 30 m	0.00005	0.00010
16	B - 30 m	15.65919	15.65920
17	B - 50 m	0.00196	0.01260
18	B - 50 m	16.94614	16.94610
19	B - 50 m	0.00005	0.00023
20	B - 50 m	0.20691	0.21474

The predicted probability values for the Fine Tree algorithm using the regression learner on MATLAB is presented in Table VI for selected data samples. The prediction shows a reasonable level of accuracy when compared with actual values.

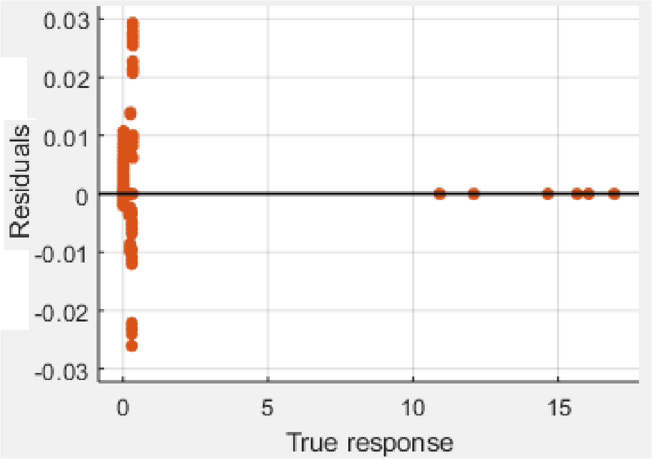

Fig. 9. Residual plot

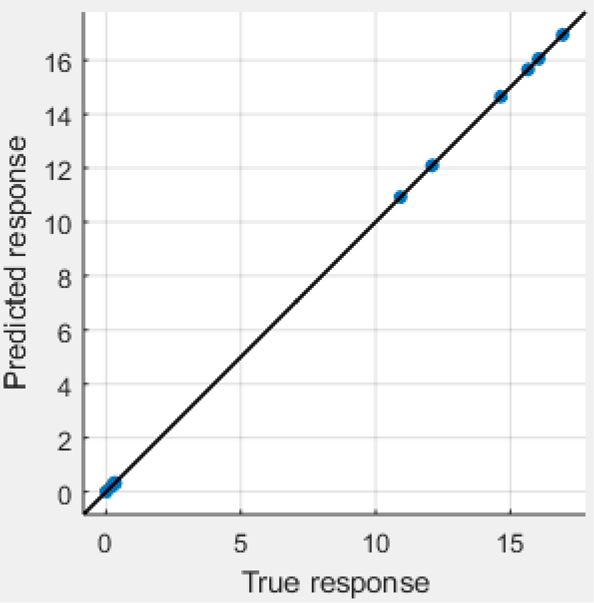

Fig. 10. Predicted vs actual plot

VI. CONCLUSION

Ensuring safety during lightning strikes requires a detailed and accurate design of lightning protection systems, especially the air termination system that acts as sacrificial attachment points. In this study, data mining was applied for the novel analysis of a simulated dataset, compiled from the numerical simulation of lightning strikes to cuboid structures using the dynamic electro-geometrical model. Classification and predictive regression-based models were developed on the Orange data mining software and MATLAB (classification and regression learner). An accuracy of 100% was observed for some of the classification and regression analysis. This is indicative of the suitability of data mining techniques for lightning strike probability predictions for simple structures. The implication of the results is that, with a trained data mining model, the need for extensive numerical DEGM coding and simulations, which may run for days in some cases, can now be avoided. Although, with various structural shapes, other than cuboids, it will be necessary to develop new data mining models trained with a relevant dataset.

The accuracy and the overall performance of the data mining models, in general, are functions of the quality of the dataset and the model implemented. Data mining entails learning patterns from previous, historical data, which implies that it is an experience-based model. Therefore, to ensure a robust and well-trained data mining model, it will be necessary to generate an adequate dataset with various dimensions of the structure of interest in order to ensure that the model is versatile and very well exposed to various possibilities with respect to the concept of the dynamic electro-geometrical model.

REFERENCES

[1] Y. Wang, "Multiphysics analysis of lightning strike damage in laminated carbon/glass fiber reinforced polymer matrix composite materials: A review of problem formulation and computational modeling," *Composites Part A: Applied Science and Manufacturing,* vol. 101, pp. 543-553, 2017.

[2] M. A. Cooper and R. L. Holle, "Economic Damages of Lightning," in *Reducing Lightning Injuries Worldwide*, M. A. Cooper and R. L. Holle, Eds., ed Cham: Springer International Publishing, pp. 51-62, 2019.

[3] M. A. Cooper and R. L. Holle, "Current Global Estimates of Lightning Fatalities and Injuries," in *Reducing Lightning Injuries Worldwide*, M. A. Cooper and R. L. Holle, Eds., ed Cham: Springer International Publishing, pp. 65-73, 2019.

[4] Y. He, S. Lindbergh, J. Rakas, and C. M. Graves, "Characterizing Lightning-Strike Hazard to Airport Facilities: A Case Study of Baltimore Washington International Airport," in *2019 Integrated Communications, Navigation and Surveillance Conference (ICNS)*, pp. 1-9, 2019.

[5] A. Yrondi, R. DerKasbarian, A. Gallini, V. Max, C. Pauron, A. Joubin*, et al.*, "Symptoms of depression and post-traumatic stress in a group of lightning strike victims," *Journal of Psychosomatic Research,* vol. 120, pp. 90-95, 2019.

[6] A. Biswas, K. Dalal, J. Hossain, K. Ul Baset, F. Rahman, and S. Rahman Mashreky, "Lightning Injury is a disaster in Bangladesh? - Exploring its magnitude and public health needs," *F1000Research,* vol. 5, pp. 2931-2931, 2016.

[7] C. Karch and C. Metzner, "Lightning protection of carbon fibre reinforced plastics — An overview," in *2016 33rd International Conference on Lightning Protection (ICLP)*, pp. 1-8, 2016.

[8] C. Gomes. (2020) Lightning Protection of Structures: How to do it wrong. *Wattnow* 50-57.

[9] I. Ullah, M. N. R. Baharom, H. Ahmad, H. M. Luqman, and Z. Zainal, "Lightning effect on scaled protected and un-protected building structures and the air terminal bypasses," *Indonesian Journal of Electrical Engineering and Computer Science,* vol. 11, pp. 182-186, 2018.

[10] IEC 62305-1:2010, *Protection against lightning – Part 1: General principles*

[11] F. Heidler, W. Zischank, Z. Flisowski, C. Bouquegneau, and C. Mazzetti, "Parameters of lightning current given in IEC 62305-background, experience and outlook," in *29th International Conference on Lightning Protection*, p. 26, 2008.

[12] P. Moshashaei, S. S. Alizadeh, L. Khazini, and M. Asghari-Jafarabadi, "Investigate the causes of fires and explosions at external floating roof tanks: A comprehensive literature review," *Journal of Failure Analysis and Prevention,* vol. 17, pp. 1044-1052, 2017.

[13] A. Adekitan and M. Rock, "Application of machine learning to lightning strike probability estimation," presented at the International Conference on Electrical Engineering and Informatics, Indonesia, 2020.

[14] A. I. Adekitan and M. Rock, "The impact of space point definition on dynamic electro-geometrical model of lightning strike probability," *Electric Power Systems Research,* vol. 184, p. 106336, 2020.

[15] A. Kern, C. Schelthoff, and M. Mathieu, "Probability of lightning strikes to air-terminations of structures using the electro-geometrical model theory and the statistics of lightning current parameters," presented at the 30th International Conference on Lightning Protection (ICLP), 2010.

[16] A. Kern and R. Brocke, "Planning of Air-Termination Systems with the Dynamic Electro-Geometrical Model – Possible Practical Applications," presented at the 10th Asia-Pasific International Conference on Lightning, Krabi Resort, Thailand, 2017.

[17] A. I. Adekitan and A. Abdulkareem, "Prediction of the voltage status of a three-phase induction motor using data mining algorithms," *SN Applied Sciences,* vol. 1, p. 1675, 2019.

[18] T. Auld, A. W. Moore, and S. F. Gull, "Bayesian neural networks for internet traffic classification," *IEEE Transactions on neural networks,* vol. 18, pp. 223-239, 2007.