

Lightning Location and Peak Current Estimation From Lightning-Induced Voltages on Transmission Lines With a Machine Learning Approach

Martino Nicora¹, Member, IEEE, Mauro Tucci², Senior Member, IEEE, Sami Barmada³, Senior Member, IEEE, Massimo Brignone⁴, Member, IEEE, and Renato Procopio⁵, Senior Member, IEEE

Abstract—In this article, a machine-learning-based model for the regression of cloud-to-ground lightning location and peak current from time-domain waveforms of lightning-induced voltage measurements on overhead transmission lines is presented. A principal component analysis (PCA) procedure is applied for extracting significant features and decreasing the dimension of the input vector. Then, a shallow neural network is trained with the results of the PCA. The obtained results show that the proposed approach can be the base for a tool able to regress lightning location with an accuracy comparable to or even better than traditional methods [i.e., lightning location system (LLS)] and provide a peak current estimate more accurate than LLS and more actual and widespread than direct tower measurements (which are limited to a reduced number of recorded events in some specific regions). Such a tool would also have significant advantages in terms of costs, since it would not require a dedicated instrumentation.

Index Terms—Lightning-induced effects, lightning location, machine learning (ML), neural networks (NNs), transients on transmission lines.

I. INTRODUCTION

LIGHTNING strokes are natural phenomena representing a significant source of risk for people, structures (e.g., civil buildings and wind turbines), infrastructures (e.g., power transmission and distribution, and telecommunications systems), railway, aviation, and natural environments. Knowing the location of a lightning strike and its channel-base peak current (i.e., the peak value of the return stroke current waveform measured at the base of the lightning channel [1]) is relevant for several geophysical and electrical engineering applications.

Manuscript received 27 September 2023; revised 9 January 2024; accepted 9 February 2024. Date of publication 27 March 2024; date of current version 13 June 2024. This work was supported in part by the Italian Ministry for Education, University, and Research through the project PRIN 2022 under Grant 20224CL7HM and in part by the European Union through the NextGenerationEU project. (Corresponding author: Martino Nicora.)

Martino Nicora, Massimo Brignone, and Renato Procopio are with Department of Naval, Electrical, Electronics and Telecommunications Engineering, University of Genoa, I-16145 Genoa, Italy (e-mail: martino.nicora@edu.unige.it; massimo.brignone@unige.it; renato.procopio@unige.it).

Mauro Tucci and Sami Barmada are with the Department of Energy, Systems, Territory and Construction Engineering, University of Pisa, I-56126 Pisa, Italy (e-mail: mauro.tucci@unipi.it; sami.barmada@unipi.it).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TEMC.2024.3375452>.

Digital Object Identifier 10.1109/TEMC.2024.3375452

Geophysical researchers are interested in lightning as a precursor of severe weather events. Lightning forecasting algorithms [2], [3], [4] can be applied to track the spatial evolution of convective structures in real time [5], [6], with the aim of preventing catastrophic weather effects.

On the other hand, electrical engineers need stroke location and peak current data in procedures aimed at quantifying the lightning risk for a structure or infrastructure and designing the appropriate lightning protection systems (e.g., the lightning performance assessment of an overhead power line [7], [8], [9], [10], [11]).

At present, cloud-to-ground lightning strokes are mainly localized with lightning location systems (LLSs). An LLS is a network of sensors separated by 50–400 km and detecting cloud-to-ground discharge signals in the very low frequency/low-frequency range [12]. These sensors typically measure the radiation electric and/or magnetic field and localize the lightning stroke employing different techniques (magnetic direction finding, time-of-arrival, and hybrid methods) [12]. LLS location accuracy can vary depending on the sensor coverage and technology. For the Italian network *Sistema Italiano Rilevamento Fulmini*, which is part of the European Cooperation for Lightning Detection (EUCLID), the median location accuracy is 250 m or even less in some regions [13], [14].

Furthermore, LLSs infer peak currents from measured peak fields. As discussed in [12], some issues exist and are summarized as follows.

- 1) The field-to-current linear model partially integrated into the lightning location algorithm to account for field attenuation is applied to all types of strokes although it has been validated with rocket-triggered lightning and direct tower measurements only for negative subsequent strokes with peak currents lower than 60 kA.
- 2) LLSs have increased their capability of detecting weaker sources; this increments the possibility of misidentifying cloud discharges. Such a classification error may distort peak current statistics. Different studies have revealed that an increase of the flash (i.e., first stroke) detection efficiency leads to a reduction in mean and median peak current (e.g., according to [15] when the flash detection efficiency raises from 75% to 100%, the mean peak current value decreases by 13.6%).

Current measurements directly performed on instrumented towers are more accurate, but limited to a reduced number of recorded events in some specific regions, e.g., Monte San Salvatore station (Switzerland) [16], Gaisberg Tower in Salzburg (Austria) [17], Hoher Peissenberg tower (Germany) [18], CN tower in Toronto (Canada) [19], CSIR research mast (South Africa) [20], and Morro do Cachimbo station (Brazil) [21], among others.

However, for both lightning protection and lightning research, the current parameters provided in [16] are still used to a large extent as the primary reference source [11].

In this context, the aim of this article is to present an alternative method for cloud-to-ground lightning location and peak current estimation with a machine learning (ML) model trained with lightning-induced voltage measurements on overhead transmission lines. This is intended to be the first step of a broader project whose final outcome is hopefully a tool able to: 1) locate lightning strokes with an accuracy comparable to or even better than LLS and 2) perform a peak current estimate more accurate than LLS and more actual and widespread than direct tower measurements. Such a tool would also have significant advantages in terms of costs, since it would not require a dedicated instrumentation.

The first attempt to implement this method can be found in [22], whose authors presented an ML-based lightning location algorithm relying on data from preinstalled voltage measurement systems on power transmission lines. In spite of obtaining promising results in terms of location error, the main drawback of [22] is that the model of the physical system (lightning+power network) is really simplified. Indeed, the lightning-induced voltage database required for the ML model training and testing is generated with the Rusck's formula [23] (which is valid for a lossless single-wire transmission line above a perfectly conducting ground), and the excitation source is a step current propagating upward along the lightning channel.

Therefore, the aim of this work is to test the validity of the idea in a more realistic situation. To this extent, the Lightning Power Electromagnetic Simulator for Transient Overvoltages (LIGHT-PESTO) [24] code is used to simulate the voltage waveforms measured on the line. LIGHT-PESTO allows us to consider: 1) more complex and realistic lines and a detailed model of the power systems connected at the lines ends; 2) any value of soil conductivity and permittivity; and 3) different channel-base current waveforms and different propagation models, considering the most accurate literature references.

Furthermore, in [22], the voltage signals are normalized, thus losing information about the peak current. On the contrary, in this work, the peak current is inserted in the target of the ML model, allowing a peak current estimate from voltage measurements.

From a methodological point of view, the ML model has the task to predict three numerical values that represent the target, i.e., the (x, y) coordinates of the stroke impact point and the peak current I . A simulation with LIGHT-PESTO provides the input features to the ML model, which are represented by the induced voltage waveforms in some points of the line in the selected time samples (i.e., simulating the presence of voltage sensors).

Even in the case of a practical implementation of the procedure, where the induced voltage waveforms are measured, several hundreds of time samples would be required to accurately represent the waveforms. As a consequence, we are approaching a difficult regression problem with a high-dimensional input, and many ML methods, such as fully connected neural networks (NNs), are known to suffer the curse of dimensionality in such cases.

For this reason, in order to reduce the number of inputs, a principal component analysis (PCA) procedure is set up, with the aim of extracting significant features. Reducing the dimension of the input vector allows the use of a shallow NN that is trained with the results of the PCA applied to the simulated induced voltages. The results of the NN approach are encouraging, outperforming other ML methods such as support vector regression (SVR) and k -nearest neighbors (K-NN).

The rest of this article is organized as follows. Section II provides details on the properties of the lightning-induced voltage simulator LIGHT-PESTO [24], defines the physical system model, and describes the generated voltage database. In Section III, this voltage database is used to train an ML-based model able to regress the lightning location and peak current. Section IV shows the obtained results. Finally, Section V concludes this article.

II. DATA ACQUISITION

The ML model training and testing are performed with a database of simulated lightning-induced voltage measurements. Section II-A discusses the features of the employed simulator. Section II-B describes the simulation setup, i.e., the power system model, the lightning model, numerical assumptions, simulation parameters, and the generated voltage database.

A. Lightning-Induced Voltage Simulator

Lightning-induced voltage measurements are simulated with the dedicated tool LIGHT-PESTO. This simulator originated from a numerical code proposed in [25] that solves the field-to-line coupling problem in a PSCAD-EMTDC environment by means of the second-order finite-difference time-domain scheme of the Agrawal's model [26]. The approach of [25] allows us to model a realistic network, automatically distinguish between direct and indirect lightning strokes, and account for power system precontingency conditions. Validation has been performed against experimental results obtained at the high-voltage facility of the University of São Paulo, Brazil [27], [28]. Several experimental measurements proved the accuracy of the model in different realistic scenarios, i.e., single-phase straight distribution line (see [28, Figs. 10–14]), in the presence of discontinuities (see [28, Figs. 18 and 19]), and in the presence of surge arresters and multilateral configuration (see [28, Figs. 22 and 23]).

Finally, the LIGHT-PESTO interface was developed in the MATLAB-Simulink environment and presented in [24]. The interface allows the user to select: 1) the lightning parameters, i.e., the channel-base current waveform and peak, the channel height, the return stroke speed, and the model of the propagation

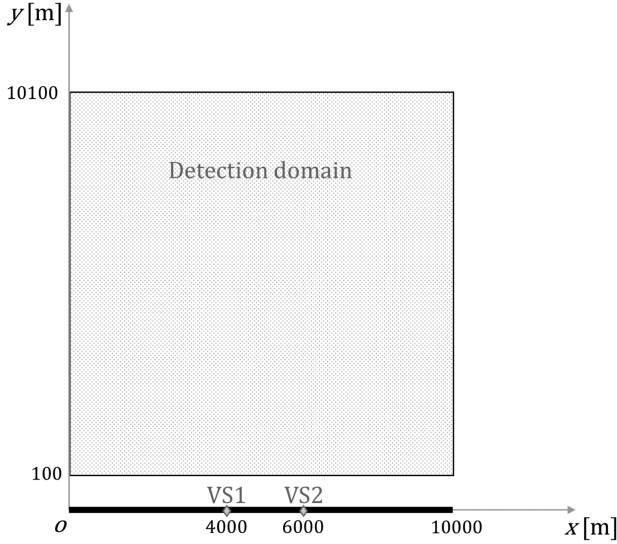


Fig. 1. Top view of the detection domain. The power system is represented by the bold solid line along the x -axis. The two voltage sensors VS1 and VS2 are located on the line 2 km from each other.

of the current along the channel; 2) any ground permittivity and conductivity; 3) the simulation parameters (duration, time step, and space step); and 4) the power system parameters, allowing multiline systems and discontinuities.

B. Lightning-Induced Voltage Simulation

1) *Power System Model*: The configuration of the power system consists in a 10-km straight single line, with a conductor diameter of 1 cm and height from the ground of 10 m. It is assumed that termination resistances at the line ends are equal to the characteristic impedance of the line (i.e., 498 Ω).

2) *Detection Domain*: A 10×10 km² area is considered (see Fig. 1). The transmission line is located 100 m beside the detection domain. Two voltage sensors (named VS1 and VS2) are located on the line at 2 km from each other ($x = 4$ km and $x = 6$ km, respectively). An ambiguity in lightning location may be caused by considering only two sensors. Indeed, signals from (x, y) and $(x, -y)$ are the same. However, in applications, the ambiguity can be solved by signals from a sensor installed on another line, or by taking into account the terrain topography, or the presence of objects and structures making the system asymmetric [22]. A number ($N = 2000$) of lightning events are simulated into the detection domain. The coordinates of the lightning strokes are extracted according to uniform distributions [7], [8], [9], [10].

3) *Lightning Model*: A typical straight vertical lightning channel with height $H = 8$ km is considered. The return stroke propagation speed is assumed to be $0.4c_0$, where c_0 is the speed of light in vacuum (3×10^8 m/s). For the propagation of the current along the channel, the well-known modified transmission line with exponential decay (MTLE) model is selected [29]. The channel-base current waveform is modeled with the Heidler function [30] with typical parameter values of the first stroke [31], except for the peak current I , which is

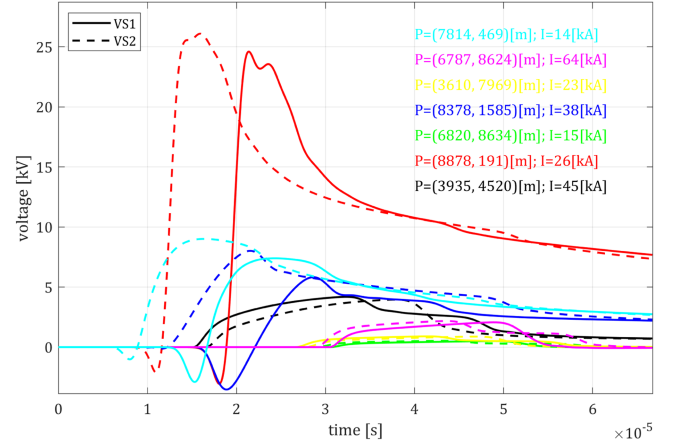


Fig. 2. Extract of the simulated lightning-induced voltage measurement database for Case 2 (random I , finite ground conductivity of 10 mS/m). Solid lines represent signals at the voltage sensor VS1 ($x = 4$ km) and dashed lines represent signals at VS2 ($x = 6$ km). The location of the stroke and the peak current are reported for each lightning event.

extracted according to a log-normal distribution with median 31.1 kA and logarithmic standard deviation 0.485 kA [11].

4) *Ground*: The ground is supposed to be lossy, but uniform, with conductivity equal to 10 mS/m, whereas the ground relative permittivity is assumed to be 10.

5) *Simulation Parameters*: A time step $dt = 0.1 \mu\text{s}$ is chosen. The considered time frame is $[0, T]$, where $T = \frac{H}{0.4c_0} = 66.7 \mu\text{s}$ is the limit of validity of the current propagation model [29]. Thus, the simulated voltage measurements are arrays with $n = 1 + \frac{T}{dt} = 668$ elements. The line discretization step is selected with a formula that combines the Courant's stability condition with the method of characteristics [32], i.e., $dx = 3c_0 dt = 90$ m. Computational efficiency is assured by the use of LIGHT-PESTO (a Microsoft Windows 10 PC equipped with 16 GB of RAM and Intel Core i7-2600 CPU at 3.4 GHz needs about 3 s for the simulation of a single random stroke and the computation of the lightning-induced voltage on the power system [24]).

6) *Voltage Database*: As specified before, 4000 simulated voltage measurements with $n = 668$ elements are provided for the training and testing phases of the ML model. Moreover, to make a fair comparison with the results of [22], another voltage measurement database is generated considering a simplified case where a constant peak current of 30 kA and a perfect electric conductor (PEC) ground are assumed (the other characteristics remain unchanged with respect to what is described in the previous paragraphs). Therefore, in the following section, two different ML models will be presented, one for each case, i.e., the simplified one (constant $I = 30$ kA, PEC ground) and the advanced one (random I , finite ground conductivity of 10 mS/m), which are labeled Case 1 and Case 2, respectively. Fig. 2 reports an extract of the database generated for Case 2.

III. ML MODEL

As explained in Section II, two different databases have been created: for Case 1, only the stroke location is unknown, while for Case 2, the peak current should be determined too. The

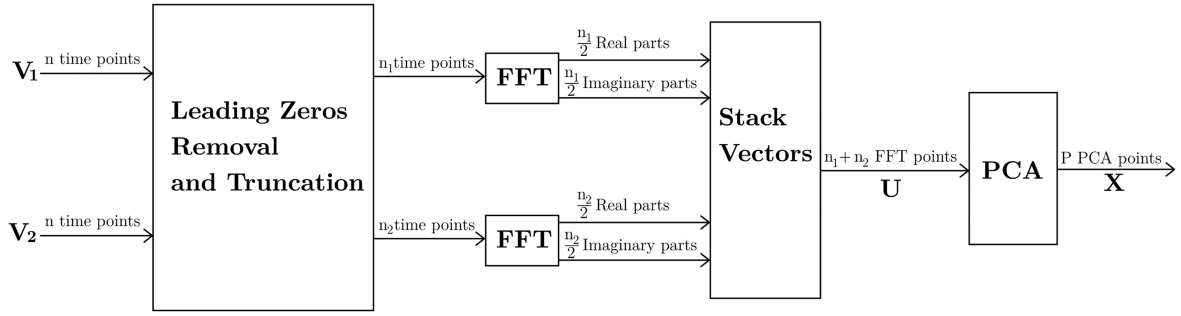


Fig. 3. Preprocessing procedure of the acquired waveforms.

criteria and choices made by the authors in the definition of the preprocessing and dimensionality reduction procedures are unique; however, different NNs have been created for different cases. In the following subsections, the main rationale behind the authors' choices is described in detail.

A. Waveform Preprocessing and Feature Extraction

Looking at the waveforms shown in Fig. 2, there are several issues that need to be addressed in order to build an efficient and accurate ML based tool. One aspect is related to the fact that the time-domain signals carry redundant information (for instance, the waveforms tails), and a more efficient and compact representation of the signals could be achieved. A second aspect is related to the so-called curse of dimensionality, which should be avoided when an NN is used.

A more physically related consideration is relative to the delay time between the induced voltage waveforms. Lightning strokes are simulated one by one, and all starting at time $t = 0$ (see Fig. 2). Then, the induced voltage starting time depends on the distance between the lightning impact point and the closest voltage sensor; this is a meaningful information that, however, could not practically be used by a real LLS since the time $t = 0$ could not easily be measured with the required degree of accuracy.

This is overcome by removing from both waveforms v_1 and v_2 the number of leading zero values present at the beginning of the closest waveform before performing the analysis; in this way, all the impulses start at the same time. This action creates vectors of different dimension; consequently, all the waveforms are afterward truncated, removing point from the tail, to the length of the shortest one, i.e., to $n_1 < n$ time samples for waveforms acquired by VS1, and to $n_2 < n$ time samples for waveforms acquired by VS2. After this initial processing, the discrete Fourier transform (DFT) is applied to each vector v_1 and v_2 resulting in a set of complex values; considering only values up to the Nyquist frequency, a vector $U \in \mathbb{R}^{n_1+n_2}$ is created by stacking the real part $\Re\{\text{DFT}(v_i)\}$ and the imaginary part $\Im\{\text{DFT}(v_i)\}$ of both vectors $i = 1, 2$. As mentioned before, in order to avoid the curse of dimensionality, the input vector dimension ($n_1 + n_2$ after the truncation described above) must be reduced, and to reach this goal, a PCA has been set up. Among different approaches, the authors have verified that a PCA applied to the frequency content of the voltage signals gives good results and low computational costs (low number of principal

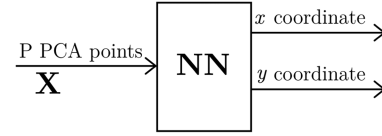


Fig. 4. NN for predicting the strike position in Case 1: all the events have the same peak current.

components). In particular, the application of the PCA to the frequency-domain dataset, consisting in the N observations of vectors U , results in P principal components capturing the $\Gamma\%$ of the total variance. This yields a feature vector $X \in \mathbb{R}^P$, which is the representation of U in the principal component space. The complete preprocessing procedure is depicted in Fig. 3.

B. Case 1: NN for the Lightning Strike Location Regression

A shallow feedforward NN is chosen as the ML paradigm, with a sigmoidal activation function in the hidden layer and a linear activation function in the output layer. As shown in Fig. 4, the extracted PCA features are given as an input to the NN, whereas the output of the NN is the position vector $p = (x, y) \in \mathbb{R}^2$. In this simplified case, all the events share the same peak current; therefore, the focus is on the prediction of the lightning strike coordinates.

For Case 1, we compared the NN approach with other ML algorithms, such as SVR and K-NN, obtaining the best results using NNs. For this reason, we only adopt NNs in Case 2.

C. Case 2: NN for the Lightning Strike Location Regression

In the second and more advanced case, the events are simulated assuming a random peak current with log-normal distribution [11], and the extracted peak current values can be conveniently used as ground truth for training the regression model. Since the NN algorithm has shown good performance for Case 1, the same approach is selected also for Case 2. However, to better exploit the knowledge of peak currents at prediction time, the scheme depicted in Fig. 5 has been implemented.

A first NN, NN_1 , is trained to predict the lightning peak current associated with each voltage waveform. Then, at prediction time, the feature vector X is normalized by the predicted peak current and given as input to a second NN, NN_2 that shall predict the coordinates of the point of impact. The training of NN_2 is performed with a dataset where each waveform is normalized

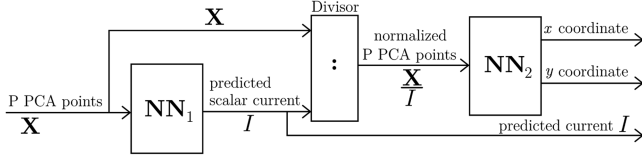


Fig. 5. NN model for predicting the strike position in Case 2. For the training of NN_2 , the PCA features $\mathbf{X}$ are normalized by means of the target current, while at prediction time, the PCA features are normalized by means of the current predicted by NN_1 .

by the corresponding ground truth peak current. It is worth to note that the normalization can be applied to the PCA values $\mathbf{X}$ in place of the original waveforms $\mathbf{v}_1$ and $\mathbf{v}_2$, because of the linearity of the preprocessing steps, consisting in fast Fourier transform and PCA operations.

D. Training and Model Selection

The selection of the significant hyperparameters of the various models, such as the best number of neurons in the hidden layer H , is performed by a grid search based on a tenfold cross validation, i.e., cyclically training on 90% of the data and testing on the resulting 10%. This method can provide an insight into how the model generalizes to an unknown dataset and contributes to improve model prediction performance. Applying the PCA to the training set results in $P = 10$ principal components capturing $\Gamma = 99.52\%$ of the total variance. The average location error e of a given model is calculated as the tenfold cross-validation mean value of the Euclidean distance (denoted by $\|\cdot\|$) between the predicted location $\hat{\mathbf{p}}$ and the real location $\mathbf{p}$:

$$e = \frac{1}{N} \sum_{k=1}^N \|\hat{\mathbf{p}}_k - \mathbf{p}_k\| \quad (1)$$

where $N = 2000$ is the number of observations. In particular, using cross validation means that each point is used in test once; then, the cross-validation error (1) is computed averaging the errors in all test folds.

The NN was the best performing method in our experiments, among SVR and K-NN, and a special attention to the training parameters was needed to obtain the best results. In particular, the NNs were trained using Bayesian regularization within the Levenberg–Marquardt algorithm [33], where the weight vector $\mathbf{w}$ of the network is updated as

$$\mathbf{w}_{t+1} = \mathbf{w}_t - (\mathbf{J}^T \mathbf{J} + \mu \mathbf{J})^{-1} \mathbf{J}^T \mathbf{e} \quad (2)$$

where $\mathbf{J}$ is the Jacobian matrix that contains first derivatives of the network errors with respect to the weights and biases, and $\mathbf{e}$ is a vector of network errors. The Jacobian matrix can be computed through a standard backpropagation technique [33]. The positive scalar parameter μ controls the applied method: when μ is zero; (2) is just Newton's method. On the other hand, when μ is large, (2) becomes gradient descent with a small step size. Newton's method is faster and more accurate near a minimum of the error; thus, the adaptive value μ is decreased by μ_{dec} after each successful step (reduction in performance

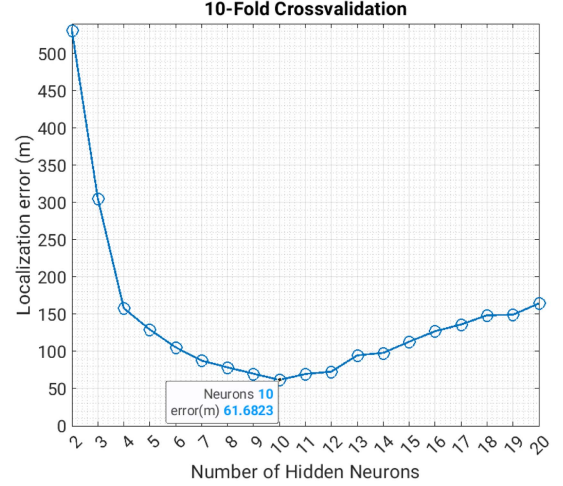


Fig. 6. Location error as a function of the number of hidden neurons for Case 1.

function) and is increased by μ_{inc} only when a tentative step would increase the performance function. In our application, the choice of the two hyperparameters μ_{dec} and μ_{inc} has proved to be fundamental, especially for Case 2. In fact, being the learning a difficult task, it was important to avoid early stopping, therefore choosing not to use a validation set. As a result, the most relevant condition that stops the training was that μ exceeds a maximum allowed value μ_{max} , giving to this parameter a crucial role. For these reasons, the relevant hyperparameters μ_{dec} , μ_{inc} , and μ_{max} were selected and fixed using a preliminary search based on cross validation.

IV. RESULTS

A. Case 1: Constant Peak Current and PEC Ground

The results of the calculation of the error defined in (1) using a tenfold cross validation as a function of the number of hidden neurons for Case 1 are shown in Fig. 6. The proposed approach results in a very low location error of about 62 m, based on the averaging of $N = 2000$ test samples using cross validation, where each sample is used in the test once. This prediction is better than the median location accuracy of the European LLS EUCLID, which is 250 m or slightly less in some regions [13], [14]. It is worth to note that these results have been obtained for Case 1 where different simplifying assumptions are considered. However, the average value of this prediction is comparable with the results of [22]. Regarding the distribution of the location error, starting from the same system considered in Case 1, Karami et al. [22] obtained that in more than 65% of the cases, their ML model was able to predict the source with less than 100-m location error. In our case, the percentage of prediction errors below 100 m is 89%, indicating that the variance of the error of our model is much lower than the one presented in [22], despite that a similar average error is reported. Table I resumes the comparison between the two models. The promising result is a strong indication of the goodness of the preprocessing procedure described in Fig. 3 and of the prediction capabilities of the NN. Regarding the other methods applied, SVR gave

TABLE I
CASE 1 MODEL ASSUMPTIONS AND RESULTS COMPARED WITH THE METHOD OF KARAMI ET AL. [22]

Model	Source current	Current propagation	Voltage simulation	R_x^2	R_y^2	e	e_{100}^b
Case 1	Heidler's function [30] ^a	MTLE model [29] ^a	LIGHT-PESTO [24]	0.993	0.993	62 m	89%
Karami et al. [22]	Step function	(none)	Rusck's formula [23]	0.9999	1	65 m	65%

^a The approach can be applied with any source current and any current propagation model.

^b Percentage of cases with location error lower than 100 m.

TABLE II
TENFOLD LOCATION ERROR e RESULTS

	NN	SVR	K-NN
Case 1	62 m	350 m	560 m
Case 2	290 m	780 m	1200 m

TABLE III
NN COORDINATE LOCATION MAE RESULTS

	Coordinate x	Coordinate y
Case 1	44 m	32 m
Case 2	109 m	245 m

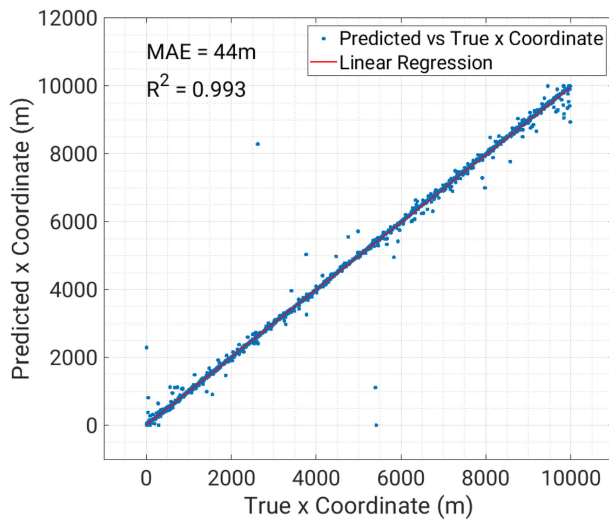


Fig. 7. Regression scatter plot of the predicted x -coordinate for Case 1, showing 2000 test points.

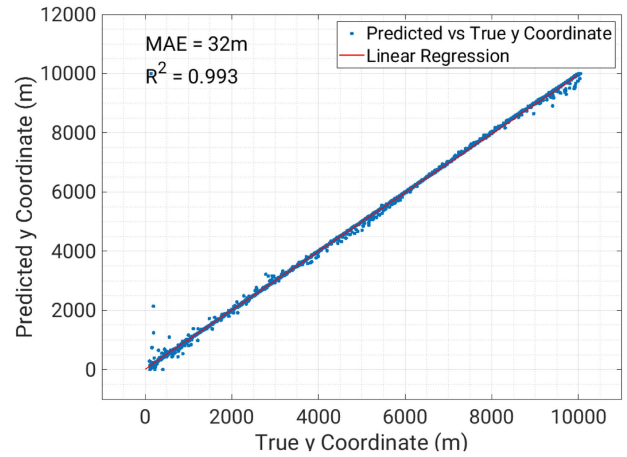


Fig. 8. Regression scatter plot of the predicted y -coordinate for Case 1, showing 2000 test points.

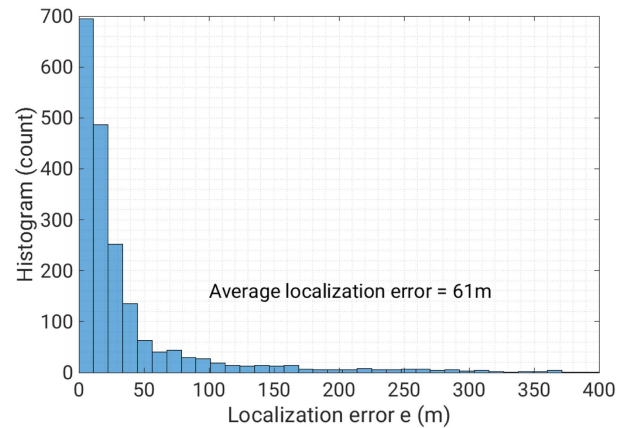


Fig. 9. Histogram of the location error for the 2000 simulated events (Case 1).

an average location error of 350 m and K-NN of 560 m. We used fine-tuning approaches to optimize the hyperparameters of both SVR and K-NN, but their performance lagged behind. The main results of the experiment performed for Case 1 are shown in Tables II and III and in Figs. 7–10. The good R^2 values found in the regression analysis of the single x and y coordinates shown in Figs. 7 and 8 attest both the goodness of the model and the fact that the simplified assumptions made in Case 1 make the problem easier to solve with respect to Case 2. The histogram shown in Fig. 9 suggests that the points for which the location error is high are statistically few and not able to affect the result. Indeed, the 95th percentile of the model location error is 220 m. On the other hand, the 95th percentile of LLS location accuracy is typically higher (e.g., 1600 m in [13]). The scatter plot of the stroke events in the spatial domain (see Fig. 10) shows that greatest errors are

made in the proximity of the line. This behavior is obtained also in Case 2, and it is discussed in detail in the following sections.

B. Case 2: Variable Peak Current and Non-PEC Ground

Regarding Case 2, we performed a cross validation of the model shown in Fig. 5, with a number of neurons fixed to 10, which was optimal for Case 1, both for NN_1 and NN_2 . In fact, a grid search conducted on the advanced dataset confirmed that 10 was the optimal number of hidden neurons for all the networks in Case 2 too. Special attention was paid to avoid data leakage, i.e., using training data during the test or test data during the training. In particular, NN_1 and NN_2 are trained simultaneously using training data, while the test is done in

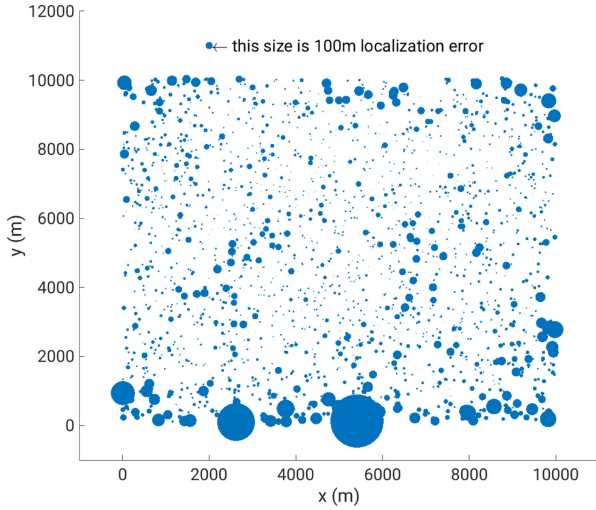


Fig. 10. Scatter plot of the 2000 simulated events in the detection domain. The size of the marker is proportional to the location error (Case 1).

series, following the flow of information, as shown in Fig. 5. As a result, the tenfold cross-validation location error obtained for Case 2 is 290 m. The degradation of the performance with respect to Case 1 was expected, because of the more accurate simulation and the variable peak current. However, the result is encouraging because it is comparable with LLS location accuracy [13], [14]. LLSs typically have a better accuracy near the sensors. However, the work presented in this article is intended as the first step of a broader project, which aims at developing a lightning detection system integrated with overhead power systems and exploiting measurements from voltage sensors. Since for Italy the density of power lines is very significant (14 and 136 km/km² for transmission lines and distribution lines, respectively, as declared by the Italian transmission system operator [34]), in principle, one can expect the presence of voltage sensors sufficiently close to the lightning strike. This would allow a prediction with performances of the order of magnitude of that obtained in the present work.

The average error $e_I\%$ of the peak current $\hat{I}$ predicted by NN_1 , with respect to the real current I , was calculated as a mean absolute percentage error (MAPE)

$$e_I\% = \frac{100}{N} \sum_{k=1}^N \left| \frac{\hat{I}_k - I_k}{I_k} \right| \quad (3)$$

resulting in an average value of $e_I\% = 12.5\%$. The model of Fig. 5 was also evaluated substituting NN_2 with an SVR model and a K-NN model, while keeping NN_1 to predict the peak currents. The results of the experiments performed are shown in Table II. The mean absolute errors (MAEs) of the location prediction along the x and y coordinates are shown in Table III. With the two sensors aligned along the x -direction, a better location performance is expected along the x axis.

The prediction error of the proposed model in Case 2 is further analyzed in Figs. 11–16. As has been observed for Case 1, the model location error distribution represented by the histogram of Fig. 13 significantly decreases at high values: the 95th percentile

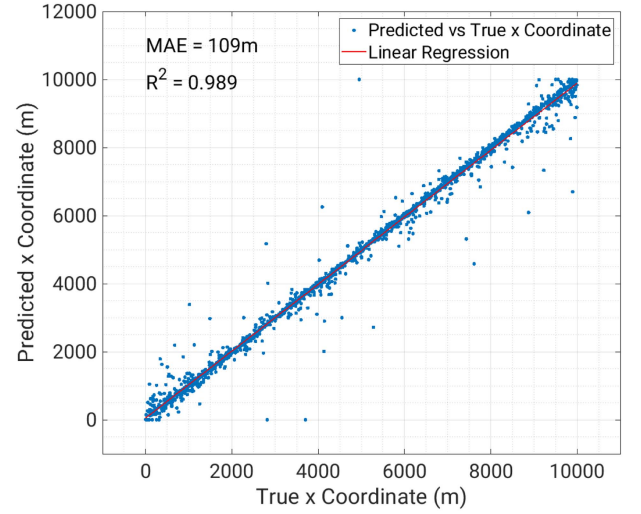


Fig. 11. Regression scatter plot of the predicted x -coordinate for Case 2, showing 2000 test points.

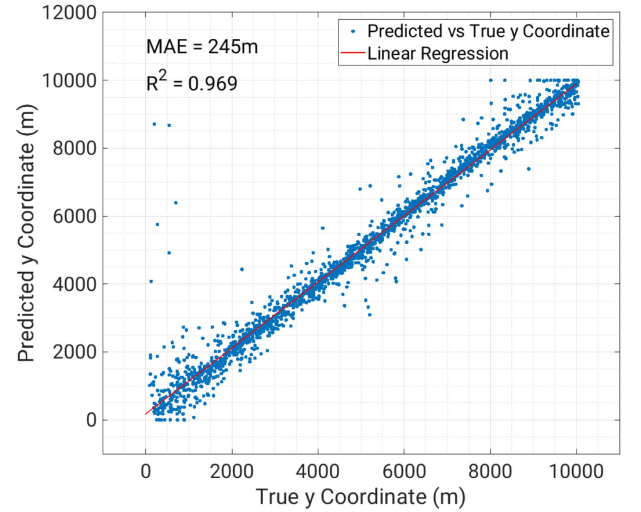


Fig. 12. Regression scatter plot of the predicted y -coordinate for Case 2, showing 2000 test points.

is 930 m, which is higher with respect to Case 1 (220 m), but lower than that of typical LLS location accuracy (e.g., 1600 m for [13]). It can be observed from Fig. 14 that a small number of large errors are distributed for lightning strokes located near the line where the voltage sensors are placed, hence with low y values. This small number of large errors, that represent only the 0.45% of the total number of events, have a large impact on the global error e , which would reduce by 7% from 290 to 270 m if one excluded those samples. This behavior is due to physics of the system. Indeed, the exact analytical dependence of the lightning-induced voltage on the distance between the point of impact and the line is difficult to evaluate since it requires to compute lightning electromagnetic fields and then the solution of the Maxwell's equations describing the field-to-line coupling problem. However, simplified approaches have verified that, at least, the induced voltage peak can be approximated with an expression that is linear [23] or quadratic [35] with respect to the inverse of such a distance. Hence, near the line, small y variations

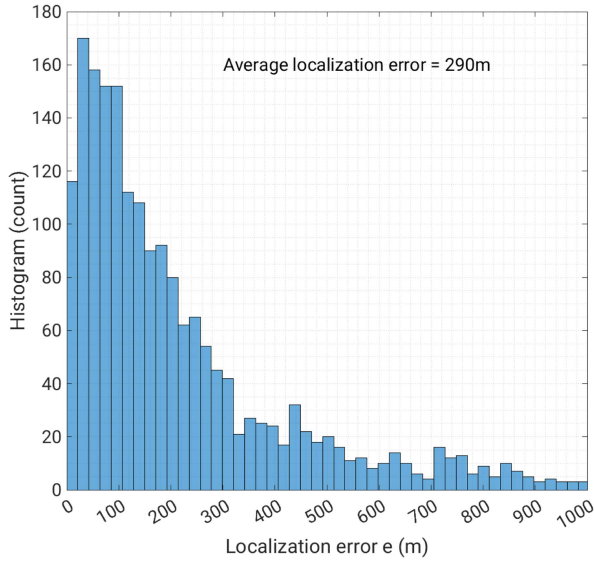


Fig. 13. Histogram of the location error for the 2000 simulated events (Case 2).

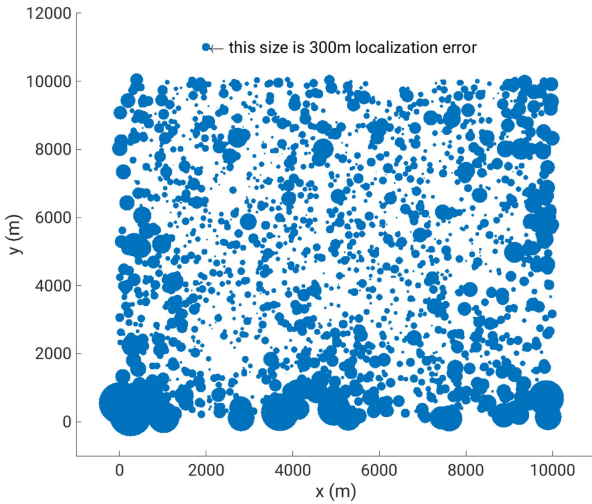


Fig. 14. Scatter plot of the 2000 simulated events in the detection domain. The size of the marker is proportional to the location error (Case 2).

can produce significant variations in the induced voltage, and this may affect the capability of the ML model to regress the y coordinate.

Different cutoff values of the y -coordinate prediction for Case 2 are considered. As reported in Table IV, the average location error slightly decreases with increasing cutoff value, but does not approach the average location error obtained for Case 1.

Particular attention shall be paid to the fact that for Case 2, the output of NN_1 affects the input of NN_2 ; thus, error propagation may occur. In Fig. 15, a plot of the relative error on the peak current prediction for each simulated event is shown. The strong correlation with Fig. 14 can be clearly observed. In particular, the correlation coefficient between the pointwise location error and the corresponding current prediction error is 81%, indicating that a better peak current estimation could

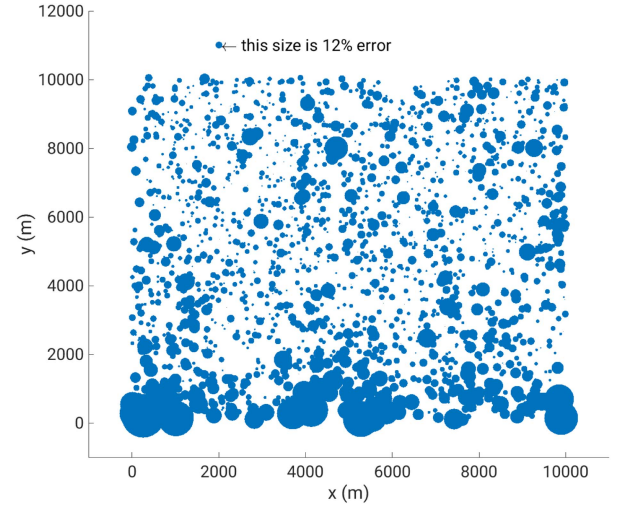


Fig. 15. Scatter plot of the 2000 simulated events in the detection domain. The size of the marker is proportional to the relative error of the predicted peak current (Case 2).

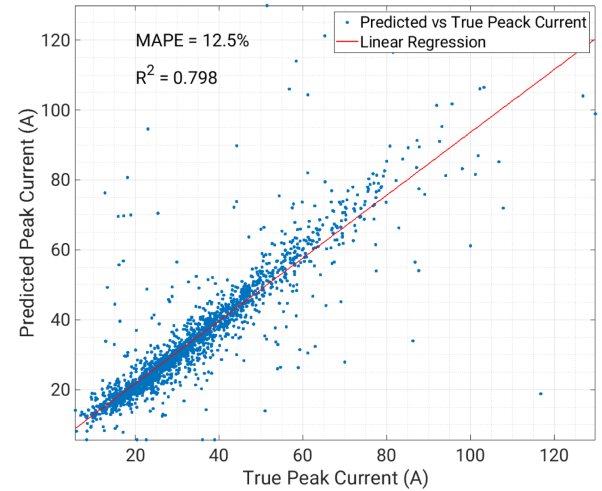


Fig. 16. Regression scatter plot of the predicted peak current for Case 2, showing 2000 test points.

TABLE IV
AVERAGE LOCATION ERROR FOR CASE 2 WITH DIFFERENT PREDICTED y -COORDINATE CUTOFF VALUES

Predicted y cut-off filter	Average location error e
$y < 0$ m	290 m
$y < 250$ m	285 m
$y < 500$ m	282 m
$y < 750$ m	278 m
$y < 1000$ m	277 m

greatly improve location prediction. Unfortunately, predicting the current is a difficult task, and this can be observed by the large dispersion of the regression error shown in Fig. 16. To assess the maximum theoretical improvement that could be accomplished with a perfect current estimation, the test phase of Case 2 was repeated by normalizing each test event with the corresponding

real current, instead of the predicted one, and in this manner, the results were very close to those of Case 1. In particular, the average location error reduced from 290 to 68 m, indicating that the performance of Case 1 can be interpreted as theoretical minimum performance for more practical applications where the current has to be predicted. On the other hand, the increased error in the region near to the line shall not be imputed to error propagation between NN_1 and NN_2 as it is also manifested in Case 1, where the peak current is not predicted, indicating that this behavior is somehow intrinsic in the data-driven approach.

Finally, some considerations on the effects of ground conductivity σ , noise, and sampling time are proposed. The ground conductivity mostly affects the radial component of the lightning electric field, which is the source term of the field-to-line coupling differential equations. To assess how this may affect the ML-based prediction, we retrained and retested the algorithm of Case 2 with a database of simulated voltage waveforms computed assuming $\sigma = 5$ mS/m. The average location error when $\sigma = 5$ mS/m is 270 m and, thus, slightly lower than that obtained with $\sigma = 10$ mS/m (290 m).

In order to accommodate the unknown and variable value of ground conductivity, we performed another training and cross-validation round using the composed dataset obtained by combining and shuffling the two aforementioned datasets with $\sigma = 10$ mS/m and $\sigma = 5$ mS/m. The average location error resulted in 280 m, suggesting that even in the case of variable ground conductivity, the prediction is accurate.

As far as the sampling frequency is concerned, from 10 to 5 MHz led to an unchanged location error of 290 m for the $\sigma = 10$ mS/m case. This result is supported by a frequency-domain analysis that indicates that most of the energy of the signals (99%) lies below 2 MHz. This consideration can affect the choice and the core of the measurement system, requiring at least a 5-MHz sampling frequency.

Regarding the effect of the noise, we conducted two experiments by injecting an additive white Gaussian noise (AWGN) in the time-domain signals, considering first a signal-to-noise ratio (SNR) of 76 dB, corresponding to a noise of $1V_{RMS}$, and second an SNR of 56 dB, corresponding to a noise of $10V_{RMS}$. The location error with added noise and SNR equal to 76 dB was 324 m, while with the SNR of 56 dB, the location error was 520 m. Both of these noises levels, especially in the second case, are expected to be much larger than a realistic measurement noise; then, the performance in the presence of real noise is likely to be between 290 m (absence of noise) and 324 m (large noise). To perform properly in the presence of noise, the NNs needed to be trained with the noisy data. All of the aforementioned tests and analyses were performed using a tenfold cross-validation approach, and the reported errors are the cross-validation errors.

Future works will extend the proposed approach to more realistic configurations, and this will allow us to test the effect of some devices here neglected (e.g., voltage transformers and their high-frequency behavior, and nonlinear devices like arresters among others) and to assess multiple-sensor systems. In particular, a more realistic configuration with multiple lines equipped with voltage measurement systems and protections devices against overvoltages will be the object of a future

work. For what concerns a larger domain of localization, this does not appear an issue, as in Italy the actual density of power lines (14 km/km² and 136 km/km² for transmission lines and distribution lines, respectively, as declared by the Italian transmission system operator) would guarantee the presence of voltage sensors sufficiently close to the lightning strike. This would allow a prediction with expected performances of the order of magnitude of the one obtained in the present work.

V. CONCLUSION

This article presented an ML-based model for the regression of cloud-to-ground lightning location and peak current from lightning-induced voltage measurements on overhead transmission lines. A shallow feedforward NN was chosen as the ML paradigm. After a first methodological case that allowed to choose the NN hyperparameters and to compare the proposed approach with the existing literature, a more realistic situation is considered, in which a first NN is trained to regress the lightning peak current; then, the feature vector is normalized by the predicted peak current and given as input to a second NN that shall predict lightning location. As far as localization is concerned, the tenfold cross-validation location error obtained was 290 m, which is comparable with the state of the art of LLS: the main advantage is that, contrarily to the traditional LLS, the proposed approach does not require any dedicated sensors' network. As regards the peak current prediction, a 12.5% MAPE was obtained outperforming traditional systems that have been validated against direct measurements only for negative subsequent strokes. Furthermore, an 81% correlation with the location error was found, suggesting that location regression can be improved with better peak current estimation. Future works will extend the proposed approach to more realistic configurations, and this will allow us to test the effect of some devices here neglected (e.g., voltage transformers and their high-frequency behavior, and nonlinear devices like arresters among others) and to assess multiple-sensor systems. Moreover, the possibility of developing a deep learning algorithm will be investigated.

REFERENCES

- [1] J. R. Dwyer and M. A. Uman, "The physics of lightning," *Phys. Rep.*, vol. 534, no. 4, pp. 147–241, 2014.
- [2] Y. Yair et al., "Predicting the potential for lightning activity in mediterranean storms based on the weather research and forecasting (WRF) model dynamic and microphysical fields," *J. Geophys. Res.: Atmos.*, vol. 115, no. D4, pp. 1–13, 2010.
- [3] E. Fiori, L. Ferraris, L. Molini, F. Siccardi, D. Kranzlmüller, and A. Parodi, "Triggering and evolution of a deep convective system in the mediterranean sea: Modelling and observations at a very fine scale," *Quart. J. Roy. Meteorol. Soc.*, vol. 143, no. 703, pp. 927–941, 2017.
- [4] M. Lagasio, A. Parodi, R. Procopio, F. Rachidi, and E. Fiori, "Lightning potential index performances in multimicrophysical cloud-resolving simulations of a back-building mesoscale convective system: The Genoa 2014 event," *J. Geophys. Res.: Atmos.*, vol. 122, no. 8, pp. 4238–4257, 2017.
- [5] C. Price, "Lightning applications in weather and climate research," *Surv. Geophys.*, vol. 34, pp. 755–767, 2013.
- [6] M. Taszarek et al., "Severe convective storms across Europe and the United States. Part 1: Climatology of lightning, large hail, severe wind and tornadoes," *J. Climate*, vol. 33, pp. 10239–10261, 2020.
- [7] M. Nicora et al., "Estimation of the lightning performance of overhead lines accounting for different types of strokes and multiple strike points," *IEEE Trans. Electromagn. Compat.*, vol. 63, no. 6, pp. 2015–2023, Dec. 2021.

- [8] M. Nicora et al., "Assessment of the lightning performance of overhead distribution lines based on lightning location systems data," *Int. J. Elect. Power Energy Syst.*, vol. 142, 2022, Art. no. 108230.
- [9] M. Brignone, F. Delfino, R. Procopio, M. Rossi, and F. Rachidi, "Evaluation of power system lightning performance—Part II: Application to an overhead distribution network," *IEEE Trans. Electromagn. Compat.*, vol. 59, no. 1, pp. 146–153, Feb. 2017.
- [10] A. Borghetti, C. A. Nucci, and M. Paolone, "Indirect-lightning performance of overhead distribution networks with complex topology," *IEEE Trans. Power Del.*, vol. 24, no. 4, pp. 2206–2213, Oct. 2009.
- [11] *Guide for Improving the Lightning Perform. of Elect. Power Overhead Distribution Lines*, IEEE Standard 1410-2010 (Revision of IEEE Std 1410-2004), 2011, pp. 1–73.
- [12] G. Diendorfer et al., *Cloud-to-Ground Lightning Parameters Derived From Lightning Location Systems—The Effects of System Performance*. Paris, France: CIGRE, 2009.
- [13] W. Schulz, G. Diendorfer, S. Pedebay, and D. R. Poelman, "The European lightning location system EUCLID—Part 1: Performance analysis and validation," *Natural Hazards Earth Syst. Sci.*, vol. 16, no. 2, pp. 595–605, 2016.
- [14] M. Nicora, D. Mestriner, M. Brignone, M. Bernardi, R. Procopio, and E. Fiori, "A 10-year study on the lightning activity in Italy using data from the SIRF network," *Atmos. Res.*, vol. 256, 2021, Art. no. 105552.
- [15] G. Diendorfer, W. Hadrian, F. Hofbauer, M. Mair, and W. Schulz, "Evaluation of lightning location data employing measurements of direct strikes to a radio tower," *E&I Elektrotechnik Informationstechnik*, vol. 119, pp. 422–427, 2002.
- [16] K. Berger, R. B. Anderson, and H. Kroninger, "Parameters of lightning flashes," *Electra*, vol. 41, pp. 23–37, 1975.
- [17] G. Diendorfer, M. Mair, W. Schulz, and W. Hadrian, "Lightning current measurements in Austria—Experimental setup and first results," in *Proc. 25th Int. Conf. Lightning Protection*, 2000, pp. 44–47.
- [18] F. Fuchs, E. Landers, R. Schmid, and J. Wiesinger, "Lightning current and magnetic field parameters caused by lightning strikes to tall structures relating to interference of electronic systems," *IEEE Trans. Electromagn. Compat.*, vol. 40, no. 4, pp. 444–451, Nov. 1998.
- [19] W. Janischewskyj, A. Hussein, V. Shostak, I. Rusan, J.-X. Li, and J.-S. Chang, "Statistics of lightning strikes to the Toronto Canadian National Tower (1978–1995)," *IEEE Trans. Power Del.*, vol. 12, no. 3, pp. 1210–1221, Jul. 1997.
- [20] A. J. Eriksson, "Lightning and tall structures," *Trans. South Afr. Inst. Elect. Eng.*, vol. 69, no. 8, pp. 238–252, 1978.
- [21] S. Visacro, A. Soares Jr., M. A. O. Schroeder, L. C. L. Cherchiglia, and V. J. de Sousa, "Statistical analysis of lightning current parameters: Measurements at Morro do Cachimbo Station," *J. Geophys. Res.: Atmos.*, vol. 109, no. D1, pp. 1–11, 2004.
- [22] H. Karami, A. Mostajabi, M. Azadifar, M. Rubinstein, C. Zhuang, and F. Rachidi, "Machine learning-based lightning localization algorithm using lightning-induced voltages on transmission lines," *IEEE Trans. Electromagn. Compat.*, vol. 62, no. 6, pp. 2512–2519, Dec. 2020.
- [23] S. Rusck, *Induced-Lightning Overvoltages on Power Transmission Lines With Special Reference to the Overvoltage Protection of Low Voltage Networks*. Stockholm, U.K.: KTH Royal Inst. Technol., 1958.
- [24] L. Farina, D. Mestriner, R. Procopio, M. Brignone, and F. Delfino, "The lightning power electromagnetic simulator for transient overvoltages (LIGHT-PESTO) code: A user-friendly interface with the MATLAB-Simulink environment," *IEEE Lett. Electromagn. Compat. Pract. Appl.*, vol. 2, no. 4, pp. 119–123, Dec. 2020.
- [25] M. Brignone, F. Delfino, R. Procopio, M. Rossi, and F. Rachidi, "Evaluation of power system lightning performance—Part I: Model and numerical solution using the PSCAD-EMTDC platform," *IEEE Trans. Electromagn. Compat.*, vol. 59, no. 1, pp. 137–145, Feb. 2017.
- [26] A. K. Agrawal, H. J. Price, and S. H. Gurbaxani, "Transient response of multiconductor transmission lines excited by a nonuniform electromagnetic field," *IEEE Trans. Electromagn. Compat.*, vol. EMC- 22, no. 2, pp. 119–129, May 1980.
- [27] M. Brignone et al., "Evaluation of lightning-induced overvoltages on a distribution system: Validation of a dedicated code using experimental results on a reduced-scale model," in *Proc. IEEE Int. Conf. Environ. Elect. Eng./IEEE Ind. Commercial Power Syst. Eur.*, 2017, pp. 1–6.
- [28] M. Brignone, D. Mestriner, R. Procopio, M. Rossi, A. Piantini, and F. Rachidi, "EM fields generated by a scale model helical antenna and its use in validating a code for lightning-induced voltage calculation," *IEEE Trans. Electromagn. Compat.*, vol. 61, no. 3, pp. 778–787, Jun. 2019.
- [29] C. Nucci, C. Mazzetti, F. Rachidi, and M. Ianoz, "On lightning return stroke models for LEMP calculations," in *Proc. 19th Int. Conf. Lightning Protection*, 1988, pp. 463–469.
- [30] H. Heidler, "Analytische blitzstromfunktion zur LEMP-berechnung," in *Proc. 18th Int. Conf. Log. Program.*, 1985, pp. 63–66.
- [31] C. A. Nucci and F. Rachidi, "Experimental validation of a modification to the transmission line model for LEMP calculation," in *Proc. 8th Symp. Tech. Exhib. Electromagn. Compat.*, 1989, pp. 389–394.
- [32] C. R. Paul, *Analysis of Multiconductor Transmission Lines*. Hoboken, NJ, USA: Wiley-IEEE Press, 2007.
- [33] M. Hagan and M. Menhaj, "Training feed-forward networks with the Marquardt algorithm," *IEEE Trans. Neural Netw.*, vol. 5, no. 6, pp. 989–993, Nov. 1994.
- [34] *Valutazione Ambientale Del Piano di Sviluppo Della Rete Elettrica di Trasmissione Nazionale 2008*, Terna, Rome, Italy, 2007.
- [35] D. Mestriner et al., "An efficient methodology for the evaluation of the lightning performance of overhead lines," *IEEE Trans. Electromagn. Compat.*, vol. 63, no. 4, pp. 1137–1145, Aug. 2021.