



Identification of cloud-to-ground lightning and intra-cloud lightning based on their radiated electric field signatures using different types of neural networks and machine learning classifiers

G. Karnas^{*}, G. Dralus, G. Maslowski

Rzeszow University of Technology, Department of Electrical and Computer Engineering Fundamentals, Rzeszow, Poland

HIGHLIGHTS

- Research based on electric field data for CG and IC lightning from Rzeszow.
- CG/IC lightning discrimination was verified in presence of a small dataset.
- MLP, RBF, LSTM and convolutional neural networks as well as ML were verified.
- 3-layer MLP neural networks showed the best performance overall.
- Analysis results can be adapted for a real-time lightning location system purposes.

ARTICLE INFO

Keywords:

Lightning
Neural networks
Lightning location systems
Lightning electric field registration
Machine learning

ABSTRACT

The scope of this paper is to test the efficiency of three different neural networks in the area of cloud-to-ground/intra-cloud lightning classification using a not very large data set of electric field data. Effective neural network learning verified for a small lightning event dataset but with whole 2-seconds particular record length is an important feature of fast and efficient procedures implemented into lightning location systems among which discrimination between cloud-to-ground and intra-cloud lightning is one of the most important task. Measurement data obtained at the Lightning Observatory in Rzeszow were used for the analysis. About 100 extremely low/medium frequency bandwidth electric field of cloud-to ground lightning registrations and the same number of intra-cloud lightning events were taken into account. During the study 81% of those registrations were dedicated for learning of neural networks while the remaining 19% were put into the testing set. All data were preselected manually with consideration of the electric field variation and information from the lightning location system database. Lightning events were selected uniformly with the well-pronounced presence of their basic components and reported distance to the registration station which ensured a wide variation of expected lightning electric field signatures. The multilayer perceptron neural network, the radial basis function neural network and the convolutional neural network were tested and compared in different configurations while distinguishing lightning. The results were compared with other conventional classification approaches, such as traditional machine learning methods as well as one a little more contemporary architecture, the long short-term memory neural network. The multilayer perceptron neural network achieved the best detection accuracy overall. Presented research is an extension of lightning identification studies available so far mainly based on the convolutional neural network by the results achieved using the multilayer perceptron neural network, the radial basis function neural network, the long short-term memory neural network and several machine learning-based classification methods.

* Corresponding author.

E-mail address: gkarnas@prz.edu.pl (G. Karnas).

1. Introduction

Lightning, being a powerful electrical discharge occurring in nature, constitutes a serious threat for structures at the ground, air vehicles as well as human beings. There are two primary types of lightning those which reach the earth surface called cloud-to-ground (CG) lightning and those which develop only in the thunderstorm cloud and therefore termed as intra-cloud (IC) lightning [1]. Discrimination between cloud-to-ground and intra-cloud lightning is one of the most important tasks during operation of lightning location systems (LLS) [2–4]. A significantly higher occurrence of IC lightning can take a lot of computer resources and therefore significantly reduce the performance of LLS for less frequent CG events [1]. CG/IC classification of a particular lightning stroke is conventionally based on the lightning stroke source height parameter which is estimated from simultaneous multi-station LLS registrations [5]. Most of LLS does not report in their databases the CG/IC lightning stroke type classifier but if this feature is implemented then the other detection techniques can also be used. One possibility involve a high-speed video image processing which is one of the most reliable sources of information about development nature of a particular lightning event [6,7]. However, this method is limited to the bottom part of the thunderstorm cloud where the lightning channel is not covered to camera view. Therefore, the high-speed video registration is not commonly used by commercial LLS but rather implemented into specialized lightning research stations. An alternative approach to the CG/IC discrimination is satellite observation [8–10] which covers a wide area of the Earth but has a problem with identification of the lightning activity near the ground surface.

On the other hand, the process of CG/IC lightning classification can be effectively supported by application of neural networks (NN). This NN application can be motivated by a little better detection accuracy and much faster discrimination time (for pretrained NN) than in case of conventional multi-station lightning electric field (EF) algorithms [2–5]. Recent increase of computing power of modern computers enabled to run many hardware demanding artificial neural network algorithms and techniques which highly gained their popularity among industrial, medical as well as household applications. Currently, it is hard to find a branch of live where a NN cannot be implemented [11]. The common examples can be: voice, image or electrocardiography signals recognition [12,13]. In the field of lightning phenomena analysis the artificial neural networks were commonly used, among others, for forecasting and prediction of lightning occurrence [14,15]. Interestingly, there is still a plenty of room for lightning detection with more contemporary neural networks such as long short-term memory (LSTM) architectures [16,17], for example, in order to improve the CG/IC lightning event classification based on their corresponding electromagnetic field waveforms recorded by LLS.

One of the first approaches to classification of CG/IC lightning events using NN involving the registered lightning electric field was made by Peng et al. [18]. In their work one dimensional convolutional neural network (1D-CNN) using a rectified linear unit (ReLU) as activation function was implemented for classification of lightning signatures in the low and very low frequency bandwidth (LF/VLF). The classification had five categories including both polarities of CG and three types of IC lightning. The network was trained with over 50,000 of events. Only 10% of entire dataset was selected randomly for network validation. The approach described in [18] gained average detection accuracy more than 95%.

Complete study in the topic of CG/IC lightning identification was also made by Wang et al. [19]. Each lightning event was assigned by the NN algorithm to one of 10 categories including 3 cloud-to-ground and 7 intra-cloud lightning types. The amount and type of the NN learning process input data were similar as in [18] while the ratio of training/testing amount of data was 80–20%. Discrimination algorithm was based on a multi-layer 1D-CNN. Input lightning electric field waveforms were preliminary filtered to reduce noise level and further normalized to

improve NN teaching process of target analysis. Overall gained detection accuracy was better than 97%.

Another artificial neural network approach to lightning stroke discrimination was made by Zhu et al. [20]. His model was based on Support Vector Machines (SVM) method being one of the machine learning (ML) classifiers [21,22]. That algorithm did not require extraction of individual pulses in their electric field signatures which is a real advantage over traditional multi-parameter methods. The total dataset in this study was 543 events, wherein 80% of strokes were used for training and the remaining 20% for validation. The classification accuracy was 97%.

Pu et al. [23] developed a neural network framework to identification of so called terrestrial gamma-ray flashes (TGF) being the high-energetic electromagnetic field effects radiated from lightning thunderstorms [24]. Frequency bandwidth of TGF-s is above X-ray level therefore consists a serious threat because of strong voltage induced effects in ground structures as well as flying vehicles. The NN framework of Pu et al. merging unsupervised and supervised machine learning to classify only a specific type of most energetic IC lightning TGF signals called energetic in-cloud pulses (EIP) [24]. The clustering results based on 10,000 event data set were further used to train supervised convolutional network. Average detection accuracy was 95.2% and maximum 98.7% for +EIPs with electrical currents above 130 kA. Next, the pretrained CNN classifier was used to identify lower peak value EIPs (>50 kA) from larger dataset of 30 000 events.

Generally, there are quite a lot of papers in the field of lightning phenomena analysis supported by neural networks but the vast majority of those articles are published in the context of weather predictions as a meteorological issue. In the above-mentioned works, the authors concentrated on the comparison of effectiveness and performance of different variations of convolutional neural networks. Other NN structures were not effectively applied to such discrimination task. The aim of this research is to verify performance and effectiveness of several other artificial neural networks applied to CG/IC lightning classification based on the same lightning event dataset. Special emphasis was put into testing of detection accuracy for the multilayer perceptron (MLP) and radial basis function (RBF) networks when only a small dataset with 2-seconds long whole lightning EF waveforms recorded during an ongoing storm is available. Then, the presented results were compared with those obtained for CNN, LSTM NN and ML classifiers.

2. Measurement system and lightning electric field data

The primary source of electric field data used for training and testing of neural networks was the station operating at the Lightning Observatory in Rzeszow (LOR) belonging to Rzeszow University of Technology (RUT) (Fig. 1). Detailed information about the LOR was given in [25] with further improvements described in [26]. Lightning electric field signatures were recorded using the so-called fast electric field antenna having the frequency bandwidth from 0.5 Hz up to 3 MHz [27,28]. This sensor was initially supplied with the CG trigger circuit but since 2019 the registration triggering technique at the station was improved. This caused a significant increase of a number of IC lightning in overall amount of registrations. The lightning EF data was further recorded by the analog-to-digital (ADC) converter and saved in the hard drive in the form of 2 second time frames with the native sampling frequency of 25 MS.

For the purpose of the conducted research a total of 100 cloud-to-ground and the same number of intra-cloud events were preselected from the LOR database covering period of 2014–2020 thunderstorm seasons. Note that the term “event” used in this paper means one specific case of CG or IC lightning discharge being recorded during one EF registration. Consequently, the particular NN input data for one event corresponds to a one-dimensional vector of lightning electric field samples. For neural network training, 81% of entire CG and IC lightning event dataset was used. Out of this number there were only 21 CG events

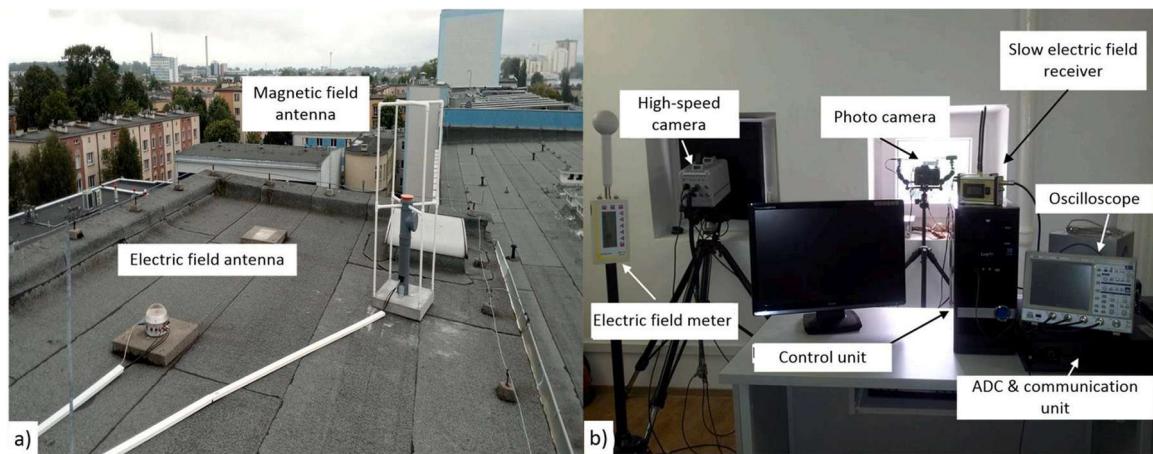


Fig. 1. General view of the Lightning Observatory in Rzeszow. (a) lightning electromagnetic field antennae on the roof of the building located on the campus of the Rzeszow University of Technology, (b) signal acquisition equipment and control CPU device which is situated in the room just under the roof.

and 23 IC events which significantly reduced quality of NN learning process. Data files were selected to include unequivocal CG or IC lightning and exclude a mixed type of CG/IC events obtained during one registration. Several representative CG and IC lightning E-field registrations were presented in Fig. 2 and Fig. 3.

In order to improve the understanding of preselected EF lightning signatures a demonstrative representation of a typical lightning current waveform and the corresponding EF waveforms depending on the distance of the lightning channel from the station was given in Fig. 2a. Typical cloud-to-ground lightning discharge might consist of one or more so called return strokes (RS) occurring in average time intervals of several tens of milliseconds (see upper-left side of Fig. 2a). First return stroke is denoted as 1RS, while subsequent: 2RS, 3RS, ..., or more generally as SS (subsequent strokes). Note that the electric field waveforms of CG lightning are varied and strongly dependent on registration distance (see upper right side of Fig. 2a) and the content of other lightning components such as a preliminary breakdown (PB) or continuing current (CC) (see bottom side of Fig. 2a). One of the most detailed and comprehensive description of CG lightning was given in [1]. Change of the lightning electric field waveform with distance is especially observed in leader stage and at the return stroke tail (Fig. 2a). For close distances to the lightning channel the EF leader stage polarity is usually opposite to the initial ramp of return stroke and with increasing distance the EF polarity tend to change. The EF tail corresponding to the return stroke is much longer in case of close distance because of electrostatic component dominance [1]. A similar longer tail effect can be observed in EF signatures while the continuing current initiated by the preceding RS is flowing in the lightning channel (see bottom right side of Fig. 2a). For distant return strokes the registered electric field is dominated by the radiated component [1] and therefore its tail decays very fast. A significant amount of information for NN teaching can be also supplied by presence of a preliminary breakdown stage in form of fast changing impulse bursts which is characteristic initial part of CG lightning occurring within thunderstorm cloud usually several hundreds of milliseconds before RS. However, those bursts are not always visible on the electric field signatures mainly due to the greater attenuation of their high-frequency components with the distance and their statistically lower peak value. All mentioned above CG lightning components were depicted in panels (b)-(k) of Fig. 2. When analyzing the lightning electric field, it is also necessary to take into account the possibility of overlapping EF waveforms from different CG flashes occurring almost simultaneously in different locations as in Fig. 2c.

Pre-selection of CG lightning events (Fig. 2) was made on the basis of the strictly defined criteria: CG events with a well-pronounced preliminary breakdown and the following stepped leader or dart leader,

return-stroke together with the following continuing current stage; multiple CG lightning; close, middle and distant CG lightning; polarity of CG waveform. Additionally, 10 events were selected randomly from the entire LOR database and added to the training set [29]. Entire preliminary selection was made by the dedicated automatic procedure developed in Matlab.

All CG lightning signatures presented in Fig. 2. differ considerably each other. Manual pre-selection for some part of the CG lightning E-field waveforms can be difficult without using of the corresponding data from the LLS. The main source of CG lightning misclassification and their incorrect interpretation as IC events can be an intense MF noise and an opposite polarization of the stepped leader stage preceding the return stroke (see panels d and e in Fig. 2). LLS information about the return stroke distance and lightning current polarization was necessary for correct distinction between CG lightning registered in close (Fig. 2h), middle (Fig. 2i) and far (Fig. 2j) distance to the lightning channel. A manual pre-selection was also applied because during a random pre-selection procedure it would be difficult to provide enough quantity of the positive CG lightning data due to less occurrence of this type of natural lightning comparing to IC ones [1]. Thus, the number of positive (Fig. 2g) and negative (Fig. 2e) lightning events was equal each other to improve learning quality for the former CG lightning type. Additionally, the number of multiple RS registrations in NN analysis was increased. This improved a NN performance during CG/IC classification for datasets involving multiple stroke CG flashes (Fig. 2b, c and k) and K-change IC lightning activity (Fig. 3a) which both have characteristic multiple stepped E-field waveform type. A significant part of the electrostatic component of lightning electric field observed for most of E-field signatures in Fig. 2 is an unique feature of this study. Lightning electric field waveforms used to teach the neural networks in other studies [18, 19] tend to have much higher contribution of radiated component. The reason is a considerably higher bottom cut-off frequencies of their systems which were 800 Hz [18] and 3 kHz [19] in respect to 0.5 Hz for the LOR system. A discussion which type of NN input data could be better for learning process can be an interesting topic of further research because some ways of the LLS acquisition device development can be established.

A simplified graphical description of EF components in case of IC lightning is much more complicated because of numerous complex physical processes taking place in electrically charged clouds. There are only a few defined components of IC lightning such as a J-stage and K-changes which can be clearly detected through ground stations. Some of the most extensive descriptions of EF signatures for those components were given in [1,24]. In many cases only the manual IC lightning discrimination is possible using the supplementary LLS data and such approach was applied in preselection of lightning events presented in

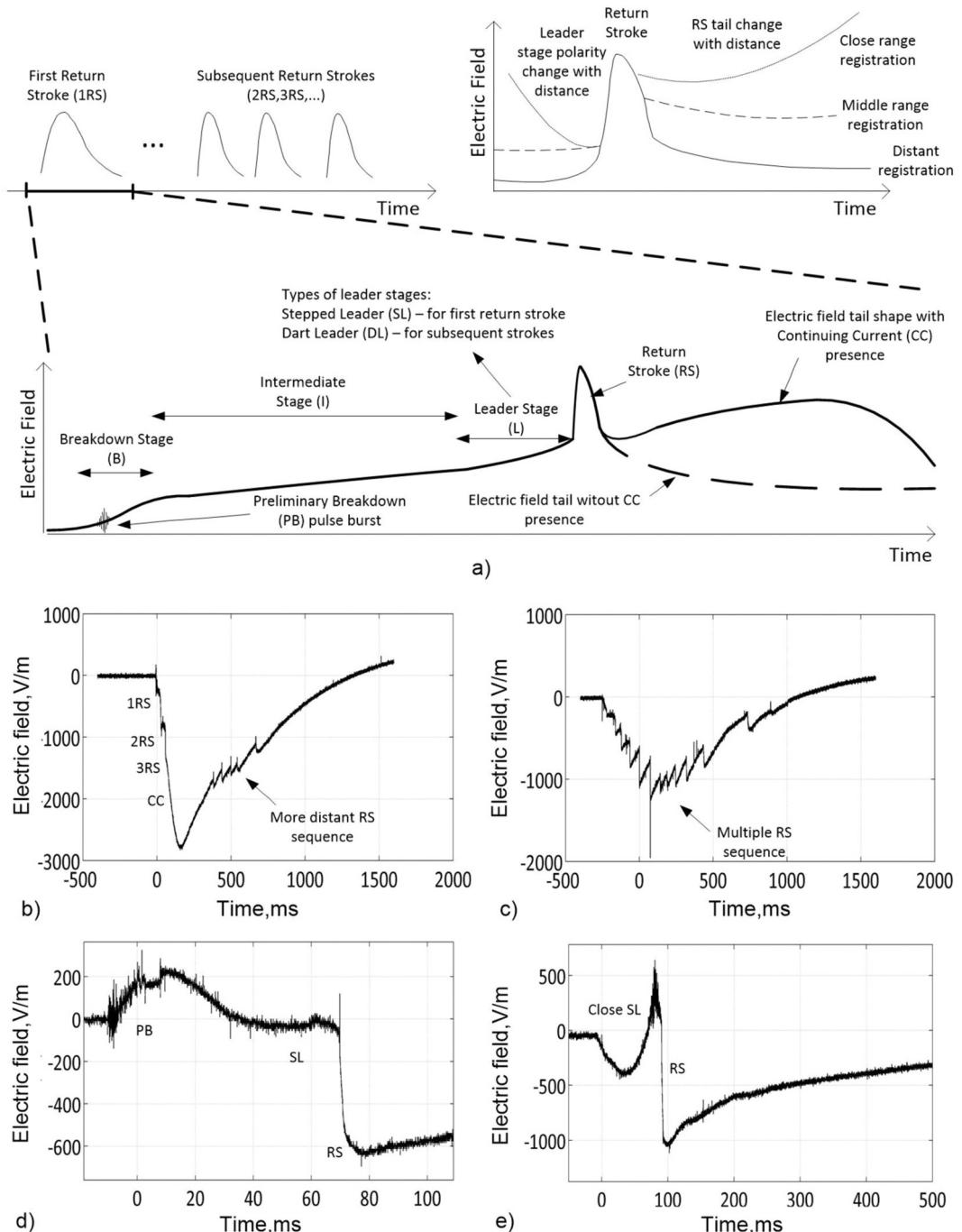


Fig. 2. Representative part of CG lightning electric field signatures selected from the Lightning Observatory in Rzeszow database for the neural network analysis (panels (b)-(k)) supplied with definition of depicted CG lightning features (panel (a)). Panel (a): demonstrative representation of typical lightning current waveform and corresponding EF waveforms depending on the distance of the lightning channel from the station. Panel (b): multiple RS with a well pronounced CC stage. Panel (c): multiple RS lightning event. Panel (d): an intensive PB stage. Panel (e): close distance stepped leader followed by negative RS. Typical negative (panel f) and positive (panel g) single RS CG lightning. Panels (h-k): close, middle and distant electric field signatures of CG flashes. Panel (h): CG registration selected randomly from entire LOR database. Note that further description of the electric field waveform stages depicted in panels (b)-(k) was given in panel (a) as well as in the paper text.

this paper to ensure that all lightning data is properly interpreted. Note that the manual preselection of IC lightning events was much more difficult than for CG lightning. The reason was an usual high variation and irregularity of their E-field signatures (see Fig. 3). An electric field recording was clearly selected as IC when there was a respective IC lightning stroke reported in LLS database or when there was any other information about this stroke. The second condition which should be

met for IC type of lightning activity was a lack of radiated component in the initial part of its corresponding E-field signature. In fact, this was the main E-field feature during the CG/IC lightning selection, especially in case of very similar bipolar IC changes and close distance CG lightning signatures, e.g. those presented in Fig. 2e and Fig. 3b as well as for distinction between positive IC changes and positive CG lightning (Fig. 2g and Fig. 3c).

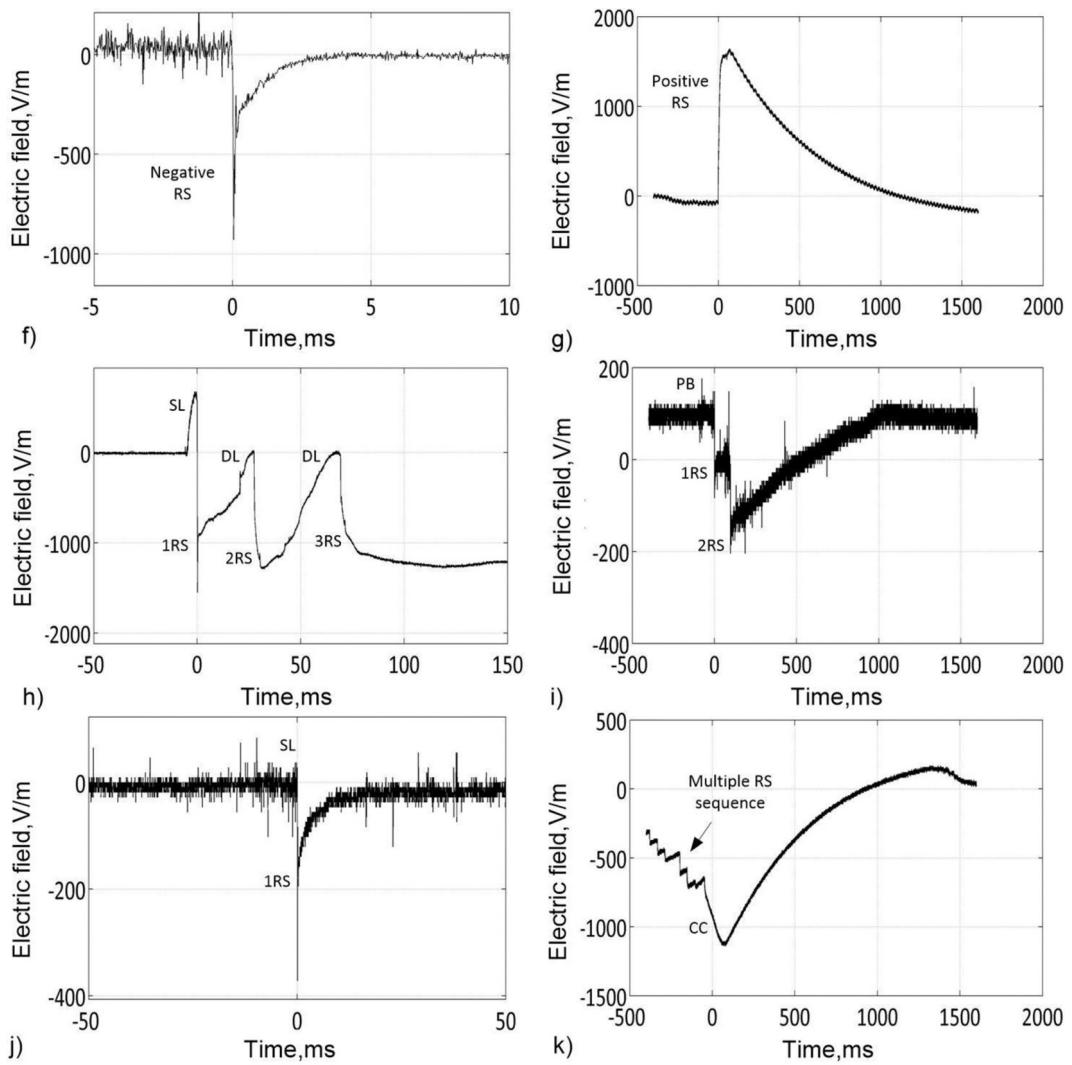


Fig. 2. (continued).

Finally, the preselection of IC lightning events was similar as in case of CG ones regarding to distance and polarity variation criteria (Fig. 3). The bipolar option was also added to possible stroke polarity categories. In case of IC lightning components the selection was made for discharges with the well-pronounced J and K processes [1], and additionally, 30 IC lightning events were selected randomly from the entire LOR database to achieve the same number of total IC and CG lightning registrations for the target NN simulation. It should be highlighted that among all IC lightning events there were not only IC strokes reported in the LLS database but also the ELF/LF IC lightning electric field changes without stroke-type structure. Significant amount of such kind of IC lightning E-field signatures can be potentially recorded by a typical LLS station, and in consequence, they need to be classified as well.

The time interval of analysis in each data file was 2 seconds. The sampling frequency was reduced from 25 MS/s to 5 kS/s which allowed to get 10,000 sample length of data frame for particular CG or IC lightning event. It was the greater length than in other studies [18,19]. Note that such wide time window of analysis was necessary to take into account all major components of lightning discharges during NN learning process. The reason that a particular CG/IC lightning signature had been measured for such a long time was the ELF/MF electric field antenna time constant equal to 2 seconds was greater than in other studies [18,19]. Note that several different sampling frequencies of input electric field waveforms were tested but the frequency of 5 kHz turned out to be the most optimal for the reported analyses. In fact,

setting the sampling frequency to 5 kHz allowed to reduce the computation time significantly without any noticeable reduction of the NN learning process quality.

2.1. CG/IC lightning classification using MLP and RBF neural networks

CG/IC lightning identification was made using three main types of artificial neural networks commonly used to classification tasks. Under analysis were: two- and three-feedforward multilayer perceptron network (MLP) and radial basis functions neural network (RBF) in several different variations of their structure [11]. The detailed architecture of the MLP and RBF neural networks used in conducted research is shown in Table 1. For the study, 5 selected MLP neural network architectures were used, varying the number of neurons in the hidden values. Thus, 2-layer MLP neural networks have 10, 25, 100, 150, 225 neurons in the (first) hidden layer. 3-layer neural networks have the following number of neurons in the first and second hidden layers 9–3, 25–5, 100–10, 150–13, 225–15, respectively. In RBF neural networks, the number of neurons in the hidden layer is determined during the learning process of the network, so 5 different values of the spread parameter were used to obtain the five variants of the RBF network in our testing (see Table 2). In estimation phase each model of each type was subjected to a learning procedure (parameter's tuning) 10 times, for 100 epochs. From among them, one best model was selected.

Influence of neural network type, input data sampling time and NN

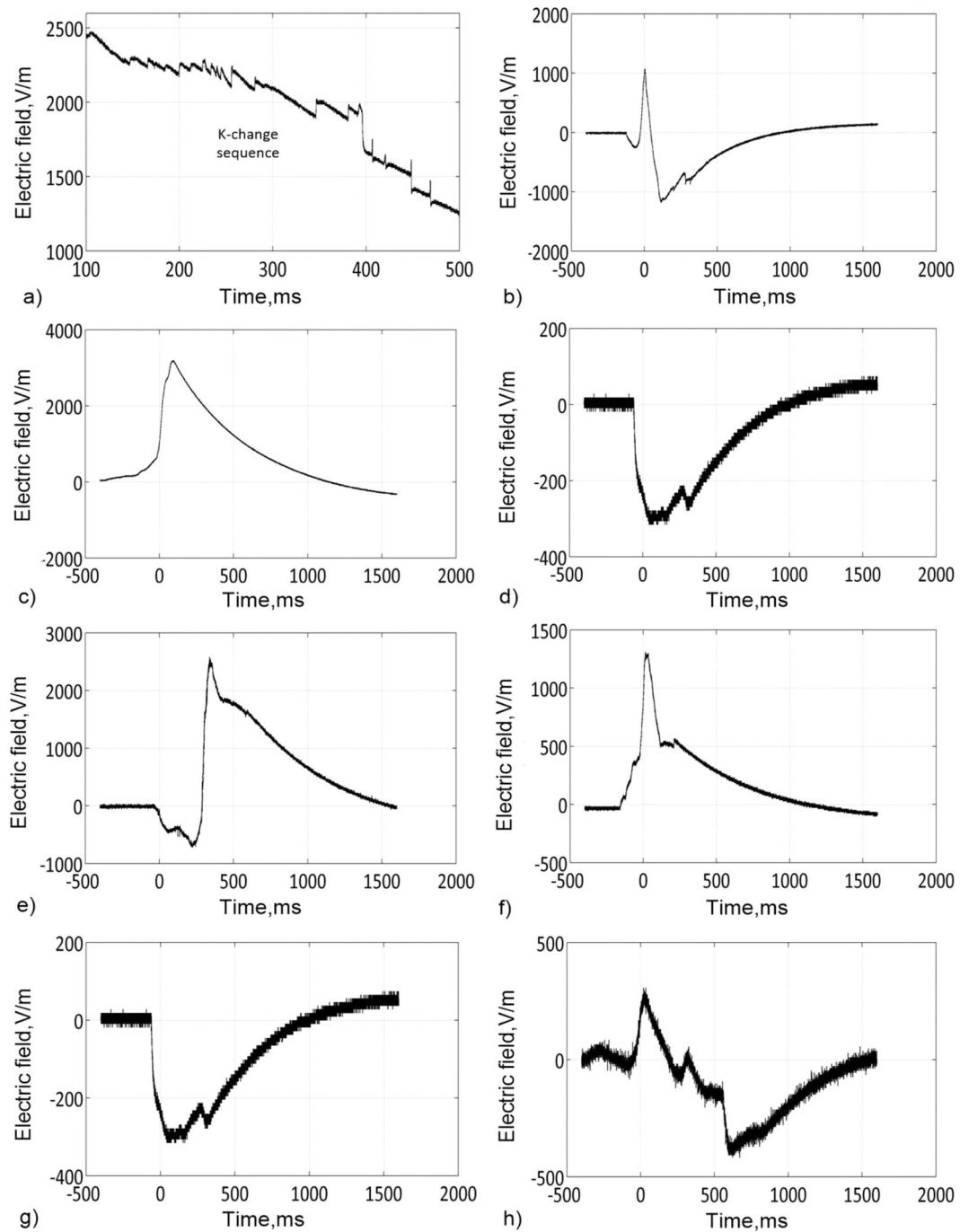


Fig. 3. Basic types of IC lightning events recorded at the Lightning Observatory in Rzeszow. Panel (a): a part of IC activity with superimposed K-changes. Bipolar (panel b), positive (panel c) and negative (panel d) IC E-field changes. Panels (e-g): close, middle and distant range IC lightning events. Panel (h): a randomly selected IC event.

Table 1
Detailed architecture of neural networks applied to lightning classification.

Layer Number	Range of neurons in First Hidden Layer, Activation function	Range of neurons in Second Hidden Layer, Activation function	Range of neurons in Output Layer, Activation function
2-layer	10–225	-	1
MLP	'tansig'	-	'purelin'
3-layer	9–225	3–15	1
MLP	'tansig'	'tansig'	'purelin'
2-layer	2–150	-	1
RBF	'rbf'	-	'purelin'

hidden neuron structure on the final effect of CG/IC lightning classification was tested. Summary of simulation results including accuracy and performance for each NN was presented in Table 2. The detection accuracy was calculated in accordance to the similar procedure given in [19]:

$$\text{Accuracy} = \frac{\text{Correct Predictions}}{\text{Total Samples}} \times 100\% \quad (1)$$

The performance parameter was also defined in compliance with the common neural network analysis terminology. The performance of each neural network was defined as a mean squared error (MSE) of data fitting process obtained at specified epoch number and was given in the

Table 2

Structure complexity and performance of MLP and RBF neural networks used for CG/IC lightning classification.

Neural Network Type	Number of Hidden Neurons				
<i>2-layer MLP</i>					
Hidden Neuron Number	10	25	100	150	225
Accuracy, (%)	93.3	93.3	90.0	86.6	86.6
Mean Square Error	0.425	0.268	0.379	0.164	0.115
<i>3-layer MLP</i>					
Hidden Neuron Number ^{a)}	9–3	25–5	100–10	150–13	225–15
Accuracy, (%)	100	100	100	100	100
Mean Square Error	0.270	0.195	0.091	0.015	0.092
<i>2-layer RBF</i>					
Hidden Neuron Number	62	62	62	62	62
Spread	10	20	30	40	50
Accuracy, (%)	50.0	50.0	50.0	50.0	50.0
Mean Square Error	0.061	0.001	0.001	0.001	0.001

^{a)} The number of neurons within the first and second layer were separated by “–” sign

bottom panel of Figs. 4–6 [30]. The lower value of the MSE can be interpreted as the better quality of the NN teaching process.

The MLP-type neural networks become more efficient than RBF NN structures during CG/IC lightning discrimination. The minimum neuron number within a 2-layer MLP NN was 10 and the maximum was 225,

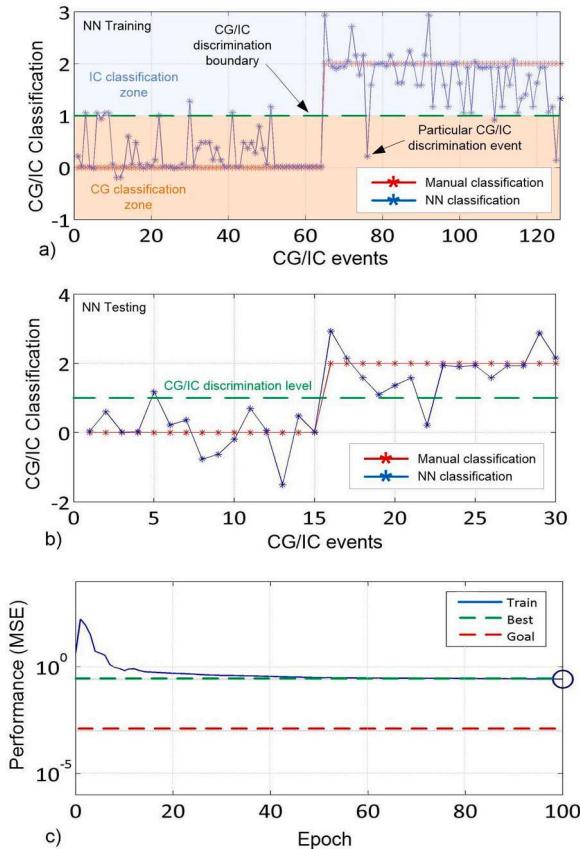


Fig. 4. Results of CG/IC lightning discrimination for 25 hidden neuron 2-layer MLP NN model using 5 kS/s input data file sampling frequency. (a) training process, (b) testing process and (c) root mean square error change during simulation. Note the meaning of panels (a) and (b) is similar and was explained graphically in the panel (a). The green dashed line shares discrimination area into IC and CG classification zones. If the lightning event estimation was below 1 then this particular case was classified as CG, else as IC lightning type. The legend in panel (c) means: MSE variation (blue curve), the best achieved NN performance during simulation (green dashed line) and a primary set goal for simulation (red dashed line).

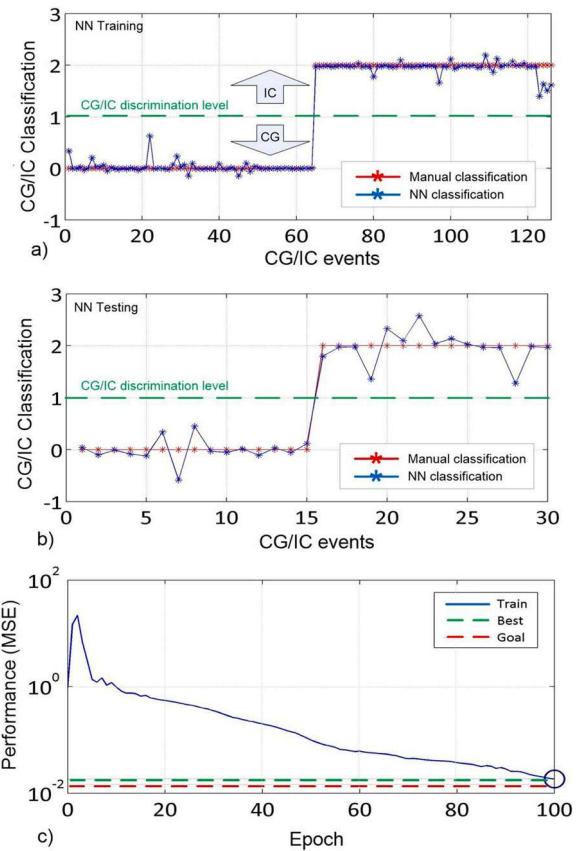


Fig. 5. Results for training (a), testing (b) and root mean square error (c) for 150–13 hidden neuron 3-layer MLP model using 5 kHz sampled data file sampling frequency. Plot interpretation was given in the caption and panel (a) of Fig. 4.

and for the 3-layer MLP NN it was 9–3 and 225–15 respectively. All MLP structures were trained by 100 epochs using *Rprop* algorithm. The target MSE error which was stopping the training and validation procedure was set to 0.001. Note that the neural network classifiers had a much lower values of the MSE for data sampled at 5 kHz than for other sampling frequencies which were not further analyzed. The best accuracy was obtained for the simplest structures of 2-layer MLP NN with the lowest number of hidden neurons (Fig. 4). However, the highest learning quality (lowest MSE) was observed for 3-layer MLP networks when the number of their hidden neurons was between 100 and 150 (Fig. 5).

The 3-layer MLP NN had a better accuracy within the entire tested NN structure complexity variation range. In case of the higher-level MLP networks the number of neurons in hidden layers does not significantly influenced the detection accuracy. Detailed results of analysis for the best performing 3-layer MLP network, which was the best performing NN overall, were presented in Fig. 5. Comparing 2- and 3-layer MLP networks given in Fig. 4 and Fig. 5 respectively, the much lower scattering of NN prediction data was obtained for the 3-layer structure. This had a direct impact on the lower RMS error for the more complex MLP neural network.

Results obtained during the RBF NN learning and validation process were shown in Fig. 5. This network was trained using a standard RBF algorithm with target error being set to 0.001. The course of RBF network teaching process (Fig. 6c) was different than for MLP NN structures (Fig. 4c and Fig. 5c). During analysis of particular RBF network structures a *spread* parameter being changed manually. Then the training algorithm was fitting the number of hidden neurons without user interaction. Interestingly, for all proceeded simulations the number of hidden neurons fixed on the same value of 66 (Table 2). The quality of

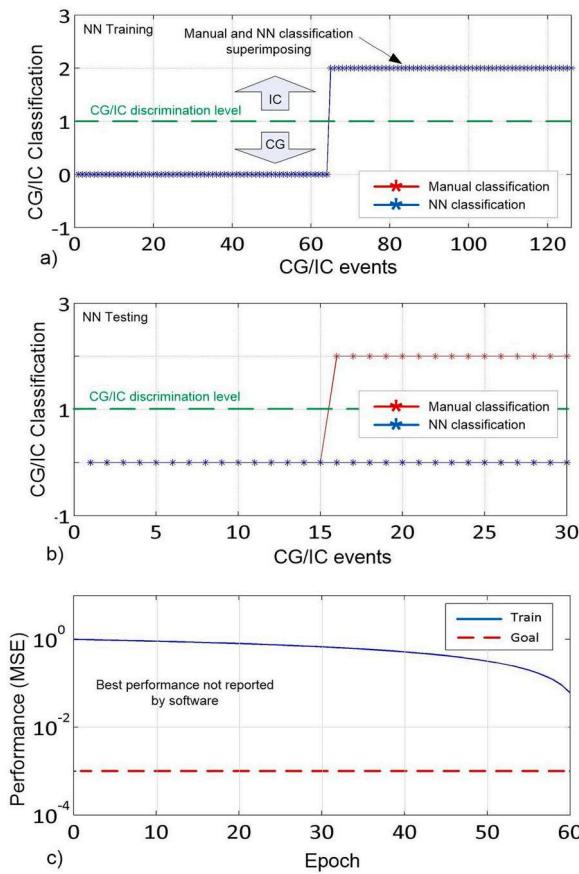


Fig. 6. Learning (a), validation (b) and root mean square error (c) for 62 hidden neuron 2-layer RBF model obtained for parameter *spread* = 10 and using the same data sampling as for Figs. 4 and 5. Plot interpretation was given in the caption and panel (a) of Fig. 4.

teaching process was better than in case of MLP networks but the detection accuracy verified on tested dataset was much lower and equal for all simulated RBF structures (Fig. 6b).

Comparing results for different types of MLP and RBF networks given in Table 2. and Figs. 4–6 the much better CG/IC lightning classification performance of the multilayer perceptron networks can be seen. All variations of RBF networks failed during CG/IC identification. Interestingly, very complex MLP networks with number of layers much greater than 3 were also not as such effective as the simpler NN structures. It suggests that low level NN can be better classifiers while operating on a small lightning event dataset.

2.2. CG/IC lightning classification using convolutional neural networks

The considered CG/IC lightning dataset was also used to verify the performance of one-dimensional convolutional neural networks. The aim of this analysis was to check if the neural network structures presented in section 3.1 can be the better or worse CG/IC lightning classifiers than the convolutional neural networks typically applied to this

task in previous research [18,19]. Two convolutional network structures similar to those adapted by Peng et al. [19] and Wang et al. [19] were tested on 5 kHz sampled lightning event data. Another 1D-CNN based on the neural network used by Wang et al. [19] was tested on 500 Hz sampled lightning events to check the NN performance for the lower quality of input data. The number of convolutional and pooling layers in our study was recalculated and fitted proportionally to the size of our data record which differs from data length used in the research conducted so far (see Table 3).

The major difference between this and previous studies was a completely another approach to the analysis of registered lightning discharges, because, on the one hand, a small number of lightning events were selected to train and validate the neural networks, but on the other hand, each event covered all lightning stages. Such data structure was intentionally chosen as optimal for real-time registration using typical lightning location systems. In our case it was only 200 event and the number of samples per event, which were under analysis, was only slightly higher than in previous studies, even though the observation time was three orders of magnitude longer (see Table 3). As a result, the total number of samples at the frequency of 5 kHz was similar to those which was analyzed in Zhu et al. [20], but two orders less than in research described in Peng et al. [18] and Wang et al. [19]. The wide time range and a very low bottom cut-off frequency of lightning electric field registration allowed to take into account not only particular return strokes but also their electrostatic background, arrangement, time intervals and other lightning components such as a preliminary breakdown, stepped leader and continuing current. Therefore, there was more information in one event in our simulation than in other studies which usually included only single CG/IC lightning stroke signatures. However, during such a comprehensive analysis, it should be remembered that lightning electromagnetic field signatures recorded at LLS stations are dependent not only on the lightning type and distance but also of many local factors such as: terrain, antenna altitude, signal noise and reflections and therefore the obtained measurements cannot be generalized uncritically to other stations, even operating in the same LLS system.

Results of CG/IC lightning identification for all our convolutional networks developed from structures adapted by Peng et al. [18] and Wang et al. [19] were presented in Table 4. The learning (process of the Peng (5 kHz) and Wang (500 Hz) CNNs was carried out in 15 epochs (180 iteration steps for *miniBatchSize* equal to 10). The estimation process in the Wang (5 kHz) CNN was carried out in 25 epochs (300 iteration steps, *miniBatchSize*=10), due to the much larger number of parameters of this network. In all CNN models, the "adam" optimizer was used in the learning process.

Lightning electric field data records sampled with 5 kHz were used from the same reasons as given in Section 2 and also to allow a direct comparison with simulation results already obtained for the MLP and RBF neural networks. The size of 5 kHz input data for our neural network based on Wang's 1D-CNN was two times longer than in case of Wang's study [19]. Therefore, another variation of this 1D-CNN was analyzed for 500 Hz sampled data. It gave 1000 sample length of data record which was five times shorter than in [19]. This auxiliary 1D-CNN structure allowed to check the performance of our convolutional neural networks within a wider range of input data record length for which Wang's study was somewhere in the middle.

Table 3

Comparison of neural network input dataset used in this paper with previous studies.

Study	Neural Network Type	Number of Events	Data Record Length ^{a)}	Frequency Bandwidth	Detection Accuracy, %
Peng et al. (2019) [16]	Convolutional	50,000	8000 spl./1.6 msec.	800 Hz – 400 Hz	95
Wang et al. (2020) [17]	Convolutional	50,000	5000 spl./1 msec.	3 kHz – 400 kHz	97
Zhu et al. (2021) [18]	SVM	31,005	101 spl./0.1 msec.	10 Hz – 250 kHz	97
In this study	MLP, RBF, Convolutional	200	10,000 spl./2 sec.	0.5 Hz – 5 kHz	100

^{a)}) shortcut "spl." from samples

Table 4

Detection accuracy and major parameters of different convolutional neural networks simulated in this paper.

Neural Network Type ^{a)}	Data Record Length	Number of Convolutional Layers	Number of Pooling Layers	Learning Accuracy after 15 Epochs ^{b)} , %	Detection Accuracy, %
CNN based on Peng, 5 kHz	10,000	7	5.	88.1	96.7 ^{b)}
CNN based on Wang, 5 kHz	10,000	18	9	84.1	96.7 ^{c)}
CNN based on Wang, 500 Hz	1000	10	5	81.8	96.7 ^{b)}

^{a)} Frequency given in NN description referenced to the sampling frequency of data record applied during analysis^{b)} Detection accuracy calculated according to Eq. (1) achieved after 180 iterations^{c)} Detection accuracy achieved after 300 iterations

Table 4 showed that the detection accuracy for our convolutional neural networks was very close to those reported in the similar studies [18–20] (see **Table 3**). This result was obtained despite the different computer interfaces used in all analyzes. Simulation presented in this paper was done in Neural Network Toolbox Matlab 2022b, on the computer Intel(R) Xeon(R) Platinum 8375 C CPU @ 2.90 GHz, with 8 Cores. Neural network analysis given in [18,19] applied *Keras* and *TensorFlow* libraries operating in Python. Taking into account results from section 3.1, the convolutional neural networks were better CG/IC lightning classifiers than 2-layer MLP structures but not the best overall. 3-layer MLP neural networks remained the most effective NN structures in this study. In the following paragraphs our three 1D-CNN structures were presented in detail.

The first 1D-CNN type was based on Peng et al. [18] structure. The original NN presented in [18] was modified in order to fit to the size of our lightning electric field data record (**Table 5**).

This neural network consisted of 7 convolutional and 5 max pooling layers including one global max pooling layer and the last dense layer. The aim of convolutional layers was to extract features from input CG/IC lightning data while the pooling layers were necessary to decrease the number of those features. This leads to the last dense layer where the final decision was made on the type of particular lightning event. Dense layers, so-called Fully Connected Layers (FC), were used to support convolutional layers in their final classification. Global Max pooling layer was used to vectorization and connection of the preceding layers and feeding its output to Softmax regression layer. ReLu activation function was applied for all convolutional layers and the Softmax function was used for the last dense layer to calculate the prediction probability of each lightning category. Softmax function was defined as:

$$\text{softmax}(\mathbf{u})_i = \frac{\exp(u_i)}{\sum_{j=1}^M \exp(u_j)} \quad (2)$$

Table 5

Detailed structure of convolutional network based on Peng et al. [18] with 5 kHz sampled lightning electric field data.

Layer Number	Layer Type	Number of Filters	Filter Size	Pooling Window Size	Padding ^{a)}	Stride	Activation Function
1	1D-Conv	16	8	–	causal	1	ReLu
2	Normalization	16	–	–	[1 1]	–	–
3	Max Pooling	–	–	2	–	2	–
4	1D-Conv	32	4	–	causal	1	ReLu
5	Max Pooling	–	–	2	[1 1]	2	–
6	1D-Conv	64	4	–	causal	1	ReLu
7	1D-Conv	128	4	–	causal	1	ReLu
8	Max Pooling	–	–	2	[1 1]	2	–
9	1D-Conv	256	4	–	causal	1	ReLu
10	Normalization	256	–	–	–	–	–
11	Max Pooling	–	–	2	[1 1]	2	–
12	1D-Conv	512	4	–	causal	1	ReLu
13	1D-Conv	1024	2	–	causal	1	ReLu
14	Global Max Pooling	–	–	–	–	–	–
15	Dense	–	–	–	–	–	Softmax

^{a)} Parameters “causal” and [1 1] define computation mode of the padding size for convolutional layers (see Matlab 2022b documentation for the Neural Network Toolbox)

where: u_i is the input signal for i -th Softmax neuron from the preceding layer. $M=2$ is the number of CNN classification categories being CG or IC lightning type. Padding and stride convolutions were used to enable more control over the size of the output for particular layers. The same analysis approach was adapted for all following 1D-CNN presented in this paper.

Fig. 7 presents simulation results from training and validation

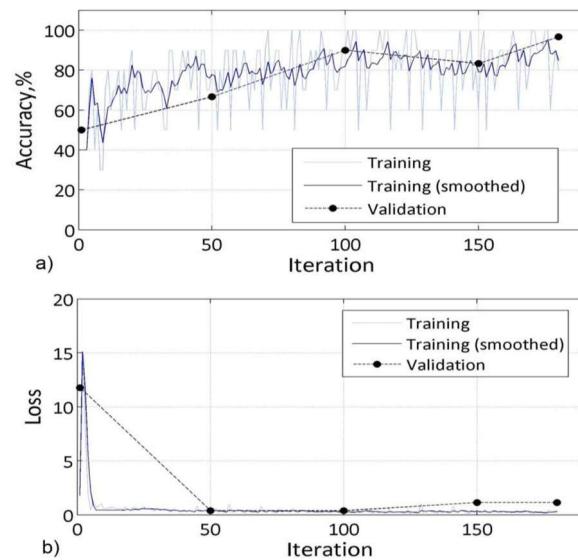


Fig. 7. Detection accuracy (a) and Loss function (b) for 1D-CNN based on Peng 5 kHz NN model [18] using the same data sampling as in Figs. 4 and 5. Plot interpretation was given in the corresponding text in the paper.

process for the neural network structure given in [Table 5](#).

In the upper panel of [Fig. 7](#) the detection accuracy of NN training and validation was shown. This parameter was computed according to the definition (1) given in section 3.1. The Loss function, so-called cross entropy function, was presented in the bottom panel of [Fig. 7](#) and defined as:

$$\text{loss} = -\frac{1}{N} \sum_{n=1}^N \sum_{i=1}^K w_i t_{ni} \ln(y_{ni}) \quad (3)$$

where: N is the number of samples, K is the number of classes, w_i is the weight for class i , t_{ni} is the indicator that the n -th sample belongs to the i -th class, and y_{ni} is the output for sample n for class i which in this case is the value from the softmax function (2). In other words, y_{ni} is the probability that the neural network associates the n -th input with class i . The Loss function is a measure how well the neural network models the training data. Loss functions are used to determine the error between the output of our model and the given target value. The NN performance parameter as defined in section 3.1 (see [Table 2](#)) was not reported by the CNN Matlab interface. In both panels of [Fig. 7](#) the smoothed waveforms can also be seen. The smoothed training curves were computed by Matlab during CG/IC lightning classification procedure and had a lower data spread than original training results. In case of all 1D-CNN simulations given in the paper there were 12 iterations per epoch and this was the constant proportion.

The low number of iterations below 50 did not allow to achieve a satisfying level of detection accuracy (see [Fig. 7](#)). The Loss function had its maximum within first 10 iterations. The highest detection accuracy was achieved between 50 and 100 iterations. After this interval there was no significant improvement of CG/IC lightning classification. Indeed, the detection accuracy even dropped to 80% at 150 iteration. This also influenced the Loss function after 100 iteration.

The next 1D-CNN structure was shown in detail in [Table 6](#).

That deep neural network was the most complex from all three cases

which were analyzed in this paper. It consisted of 32 layers including 18 convolutional, 9 pooling, 4 normalization and one dense layer. The maximum number of filters within one layer was 256 which was four times lower than in case of 5 kHz Peng NN ([Table 5](#)) but the maximum number of filters was over two times greater. Configuration of padding, stride and activation function procedures were the same for both NN structures from [Tables 5 and 6](#).

More complex neural network structure based on Wang approach resulted in the longer time of simulation than that based on Peng approach. However, the detection accuracy achieved at the particular iteration step (see [Fig. 8](#)) was for almost all simulations lower than in case of 5 kHz Peng NN. The training process proceed differently because the detection accuracy was improving continuously without significant decreasing as noticed between 50 and 100 iteration in [Fig. 7a](#). The higher values of detection accuracy can be seen for validation than during training process of 5 kHz Wang NN. The overall shape of the corresponding Loss function in [Fig. 8b](#) was influenced by the mentioned detection accuracy improvement trend. Moreover, the peak of the Loss function in [Fig. 8b](#) was 6 times lower than in case of [Fig. 7b](#). 5 kHz Wang NN was learning much slower than 5 kHz Peng NN but uniform improvement trend was observed within the entire time of simulation. Similar values of the final detection accuracy were achieved for both 1D-CNN structures (see [Table 4](#)), but after almost twice as long simulation in case of the greater 5 kHz Wang NN.

Further variation of Wang's NN [[19](#)] was presented in [Table 7](#).

This time 1D-CNN structure had a total of 18 layers including 10 convolutional, 5 pooling, 2 normalization and one dense layer. This NN was much closer to the structure given in [Table 5](#) than in [Table 6](#). Despite the greater number of layers in this case comparing to 5 kHz Peng NN, the maximum number of filters within one layer was only 128 which was 10 times lower. The reason was much shorter length of input data record which was only 1000 for 500 Hz sampled data and 10,000 for 5 kHz Wang NN. Maximum filter size of 32 was four times higher than for structure presented in [Table 5](#). The remaining parameters of all

Table 6

Detailed structure of convolutional network based on Wang et al. [[19](#)] for 5 kHz sampled lightning electric field data.

Layer Number	Layer Type	Number of Filters	Filter Size	Pooling Window Size	Padding	Stride	Activation Function
1	1D-Conv	16	16	–	causal	1	ReLU
2	1D-Conv	16	16	–	causal	1	ReLU
3	Normalization	16	–	–	–	–	–
4	Max Pooling	–	–	2	[1 1]	2	–
5	1D-Conv	16	32	–	causal	1	ReLU
6	1D-Conv	16	32	–	causal	1	ReLU
7	Max Pooling	–	–	2	[1 1]	2	–
8	1D-Conv	32	32	–	causal	1	ReLU
9	1D-Conv	32	32	–	causal	1	ReLU
10	Max Pooling	–	–	2	[1 1]	2	–
11	1D-Conv	64	16	–	causal	1	ReLU
12	Normalization	64	–	–	–	–	–
13	1D-Conv	64	16	–	causal	1	ReLU
14	Max Pooling	–	–	2	[1 1]	2	–
15	1D-Conv	64	16	–	causal	1	ReLU
16	1D-Conv	64	16	–	causal	1	ReLU
17	Max Pooling	–	–	2	[1 1]	2	–
18	1D-Conv	128	8	–	causal	1	ReLU
19	Normalization	128	–	–	–	–	–
20	1D-Conv	128	8	–	causal	1	ReLU
21	Max Pooling	–	–	3	[1 1]	2	–
22	1D-Conv	128	8	–	causal	1	ReLU
23	1D-Conv	128	8	–	causal	1	ReLU
24	Max Pooling	–	–	3	[1 1]	2	–
25	1D-Conv	256	3	–	causal	1	ReLU
26	Normalization	256	–	–	–	–	–
27	1D-Conv	256	3	–	causal	1	ReLU
28	Max Pooling	–	–	3	[1 1]	2	–
29	1D-Conv	256	3	–	causal	1	ReLU
30	1D-Conv	256	3	–	causal	1	ReLU
31	Global Max Pooling	–	–	–	–	–	–
32	Dense	–	–	–	–	–	Softmax

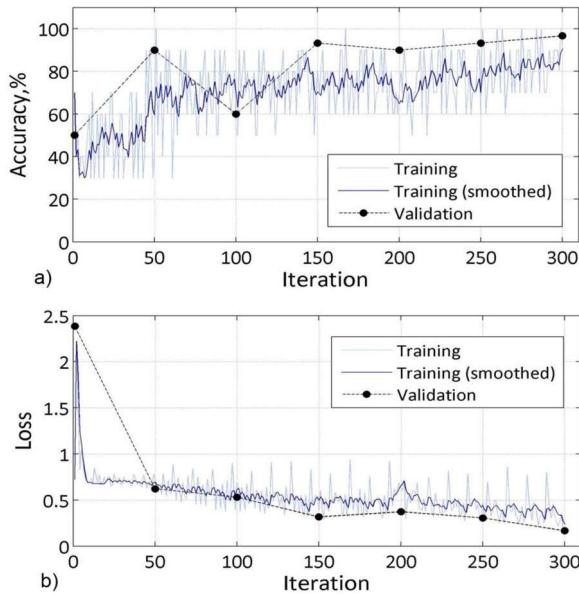


Fig. 8. Detection accuracy (a) and Loss function (b) of 1D-CNN on Wang 5 kHz NN model [19] using the same data sampling as for Figs. 4 and 5. Plot interpretation was given in the corresponding text in the paper.

convolutional NN were very close each other.

Training of 500 Hz Wang NN (Fig. 9a) proceeded almost monotonically as in case of 5 kHz Wang NN (Fig. 8a) but some local decreasing intervals were observed as for 5 kHz Peng NN (Fig. 7a). The detection accuracy achieved at the end of simulation was at the level of 80%. Despite this fact, the CG/IC lightning classification accuracy for validation was much higher being even 100% in the middle of simulation process. The peak value and overall shape of the Loss function was similar as in case of 5 kHz Wang NN.

Summarizing the results obtained for all considered convolutional networks, a very high CG/IC lightning discrimination efficiency was obtained at the end of all simulations. However, some differences were observed in the beginning and in the middle part of training process. In practice, only 100 iterations are necessary to achieve a satisfying 90% level of detection accuracy for all convolutional networks. Results of validation usually differed from teaching process of the CNN based on Wang et al. [19] and were very close for 5 kHz Peng NN. Validation of 500 Hz and 5 kHz Wang NN showed a better detection accuracy in respect to particular iteration number than the corresponding teaching process. Comparing two similar 1D-CNN structures based on Wang et al.

[19], the better CG/IC lightning classification during validation process was obtained for the simpler one (see Figs. 8 and 9). The same trend was noticed for the best overall NN, 3-layer MLP structure, where the highest performance was obtained for the middle size MLP structure.

2.3. CG/IC lightning classification using machine learning methods

CG/IC lightning neural network identification was compared with several conventional classification approaches including traditional machine learning methods such as: decision trees, discriminant analysis, logistic regression and Native Bayes classifiers, support vector machines (SVM), nearest neighbors classifiers (KNN), kernel approximation and ensemble classifiers. All classifier types available in so called Classification Learner, the Matlab module dedicated for classification tasks, were trained and tested using the R2022b version of Matlab.

Machine learning CG/IC lightning classification was done in two steps. The first step was a preliminary analysis based on default setup of all available ML classifier types (Table 8). Then the best classifiers were verified in ten runs then the training and testing results were showed in Table 9. The training process was a very time consuming for 5 kHz therefore the preliminary analysis was done for 500 Hz sampled lightning E-field data using z-score normalization. Further analysis showed that there was no significant improvement for 5 kHz sampled training data and the overall training time was still in average ten times greater than for 500 Hz sampling type.

Preliminary results for twelve the best ML classifiers were given in Table 8.

One can see that satisfying classification results were achieved only in case of SVM and KNN classifiers. SVM was learning slower than KNN but was about two times faster during testing and had the best accuracy of 93.3%. Therefore this type of ML classifier was selected for further more detailed analysis (Table 9). The remaining types of classifiers had accuracies much below 80%. The decision trees tested better than trained. The same trend was observed for Linear Discriminant algorithm. This suggests a relatively low performance of such ML methods in presence of small amount of data. Ensemble and Native Bayes classifiers as well as Logistic regression completely failed and were excluded from further analysis.

Results for the best performing ML classifier the Support Vector Machine were presented in Table 9.

Complete sequence of training and testing tasks was invoked 10 times, then the results were averaged to increase CG/IC lightning classification reliability. Many SVM parameters were changed and checked but each time the accuracy was not so high as for the initial guess (Table 8). For all given SVM types the averaged testing accuracy was higher than training accuracy but still below 90%. The best result was

Table 7

Detailed structure of convolutional network based on Wang et al. [19] for 500 Hz sampled lightning electric field data.

Layer Number	Layer Type	Number of Filters	Filter Size	Pooling Window Size	Padding	Stride	Activation Function
1	1D-Conv	16	32	–	causal	1	ReLU
2	Normalization	16	–	–	–	–	–
3	1D-Conv	16	32	–	causal	1	ReLU
4	Max Pooling	–	–	2	[1 1]	2	–
5	1D-Conv	32	32	–	causal	1	ReLU
6	1D-Conv	32	32	–	causal	1	ReLU
7	Max Pooling	–	–	2	[1 1]	2	–
8	1D-Conv	64	16	–	causal	1	ReLU
9	1D-Conv	64	16	–	causal	2	ReLU
10	Max Pooling	–	–	2	[1 1]	2	–
11	1D-Conv	128	8	–	causal	1	ReLU
12	Normalization	128	–	–	–	–	–
13	1D-Conv	128	8	–	-causal	1	ReLU
14	Max Pooling	–	–	3	[1 1]	2	–
15	1D-Conv	128	3	–	causal	1	ReLU
16	1D-Conv	128	3	–	causal	1	ReLU
17	Global Max Pooling	–	–	–	–	–	–
18	Dense	–	–	–	–	–	Softmax

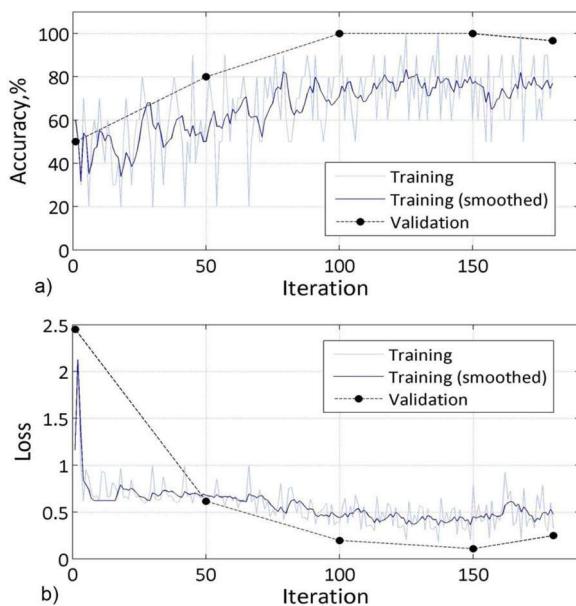


Fig. 9. Detection accuracy (a) and Loss function (b) of 1D-CNN based on Wang 500 Hz NN model [19] using the same data sampling as for Figs. 4 and 5. Plot interpretation was given in the corresponding text in the paper.

achieved for the Medium Gaussian type of SVM with kernel scale being set as auto and box constraint level of 2. It placed the SVM behind the MLP and CNN networks and before the RBF network structure. Accuracy and loss function plots were not available for this type of simulation but the confusion matrix analysis and computational performance were discussed in Section 5.

Table 8

Initial training and testing results for different types of machine learning classifiers. Sorted from the highest to the lowest testing accuracy (500 Hz sampled lightning E-field data).

Machine Learning Classifier Type	Simulation Parameters ^{a)}	Accuracy (Training), %	Accuracy (Testing), %	Training Time, s	Testing Time, ms
SVM	Preset: Fine Kernel function: Gaussian Kernel scale: Auto Box constraint level: 2 Preset: Medium	78.1	93.3	8.36	6.66
KNN	Number of k: 10 Distance metric: Euclidean Distance weight: Equal Preset: Weighted	78.1	86.7	4.74	13.15
KNN	Number of k: 10 Distance metric: Euclidean Distance weight: Squared inverse Preset: Cubic	81.2	83.3	2.84	12.19
KNN	Number of k: 10 Distance metric: Minkowski Distance weight: Equal Type: Corase	87.5	76.7	2.81	11.90
Decision tree	Number splits: 20 Split criterion: Gini's diversity index Type: Medium	68.8	76.7	4.04	14.49
Decision tree	Number splits: 20 Split criterion: Gini's diversity index	71.9	73.3	9.49	13.33
Linear Discriminant	Covariance structure: Full Type: Bagged trees	51.3	70.0	3.22	9.17
Ensemble classifier	Number of splits: 129 Number of learners: 30	68.8	66.7	4.34	5.00
Native Bayes	Model does not have parameters Type: Subspace discriminant	62.5	53.3	21.00	32.26
Ensemble classifier	Subspace dimension: 150 Number of learners: 30	59.4	53.3	6.25	11.63
Native Bayes	Kernel type: Gaussian	53.1	53.3	6.25	11.63
Logistic regression	Model does not have parameters	43.8	53.3	9.48	9.09

^{a)} Detailed description of parameters given in R2022b Matlab documentation (available online)

Table 9

Average training and testing results for SVM classifiers (10 runs). Sorted from the highest testing accuracy to the lowest (500 Hz sampled lightning E-field data).

Machine Learning Classifier Type	Simulation Parameters ^{a)}	Accuracy (Training), %	Accuracy (Testing), %	Training Time, s	Testing Time, ms
SVM	Preset: Medium Kernel function: Gaussian Kernel scale: Auto Box constraint level: 2 Preset: Fine Kernel function: Gaussian	81.2	86.7	4.55	7.41
SVM	Kernel scale: Auto Box constraint level: 2 Preset: Medium Kernel function: Gaussian	81.2	86.7	9.24	8.13
SVM	Gaussian Kernel scale: 32 Box constraint level: 2	76.9	84.7	5.01	6.37

^{a)} Detailed description of parameters given in R2022b Matlab documentation (available online)

Table 10

Detailed architectures of multilayer LSTM neural networks used for classification of 500 Hz sampled lightning electric field data.

Layer Number	Layer Type with input parameters ^{a)}		
	11-layer LSTM	13-layer LSTM	16-layer LSTM
1	Sequence Input	Sequence Input	Sequence Input
2	BiLSTM(50, 'last')	BiLSTM(50, 'last')	BiLSTM(50, 'last')
3	ReLU	ReLU	ReLU
4	BiLSTM (100, 'last')	BiLSTM (100, 'sequence')	BiLSTM(50, 'last')
5	ReLU	Dropout(0.2)	Dropout(0.2)
6	Dropout(0.2)	ReLU	ReLU
7	Fully Connected (100)	Fully Connected(50)	BiLSTM (50, 'sequence')
8	Tanh	Tanh	ReLU
9	Fully Connected (2)	Fully Connected(7)	Fully Connected (50)
10	Softmax	Tanh	Tanh
11	Classification Output	Fully Connected(2)	Dropout(0.2)
12		Softmax	Fully Connected(7)
13		Classification Output	Tanh
14			Fully Connected(2)
15			Softmax
16			Classification Output

^{a)} Detailed description of layers and parameters given in R2022b Matlab documentation (available online)

performance of classification model whose output was the probability value between 0 and 1.

Averaged results of training and testing accuracy were presented in

Table 11.

Due to similar number of layers and the same 500 Hz input data sampling, this type of the LSTM network should be directly compared with 500 Hz CNN based on Wang (Table 4). One can see a much lower training and testing accuracy in case of the LSTM architecture. The reason could be a bad learning skills of this type of NN in presence of a relatively small dataset. This was confirmed by significant differences in results of particular training-testing sequence runs. Moreover, another problem could be a very similar shape of individual parts of CG and IC lightning E-field signatures. Proper classification of such data type needs not only remembering of individual lightning stroke shapes but also learning skills for specific arrangement of those strokes within entire lightning flash. The MLP and CNN networks were much better in such tasks. The training time was much longer than for all other simulated NN and ML classifiers. Testing time was about two times shorter than for the SVM algorithm. The accuracy and Loss function during teaching and validation process were presented in Fig. 10.

The LSTM network very fast increased its accuracy to the level of 80% but just after 5 iterations up to the end of the simulation this parameter changed slightly. The value of Loss function dropped significantly within first 20 iterations then was decreasing slightly for the remaining 80 iterations and then stopped at 0.4 for testing and 0.5 for validation. Further discussion on computational performance of the LSTM network was done in Section 5.

3. Lightning electric field signatures most difficult to learn neural networks

Among the entire preselected lightning database 19% events made the neural network learning process much harder or even impossible. Some of those representative electric field signatures were gathered in groups and presented in Fig. 11 for CG and in Fig. 12 for IC lightning events. The most problematic features of CG and IC lightning dataset for NN were close similarities of the entire or some parts of the CG and IC E-field recordings.

Six types of CG events being the most difficult to learn NN were presented in Fig. 11. A significant part of those CG lightning electric field signatures had positive polarity (Fig. 11a, b and c). Those type of electric field waveforms were similar to some kind of slow electric field IC lightning changes (see Fig. 3c and e, or Fig. 12d). Electric field recordings done in far distance from the lightning channel also made the CG/IC lightning classification difficult because of the significant noise level observed in their signatures (Fig. 11c and d). The CG/IC lightning identification has another limitation for close distance CG lightning E-field signatures. In this case (Fig. 11e) the inverse polarity of a stepped leader stage in respect to the following CG lightning stroke was similar in shape to the bipolar ELF/LF E-field change observed for many IC lightning events (Fig. 3b or Fig. 12a and b).

Four types of the most problematic IC lightning E-field signatures for NN teaching were presented in Fig. 12. Two first IC lightning cases were presented in Fig. 12 (panels a and b). A high correlation of those bipolar type of E-field signatures to close distance CG lightning events was already discussed in this section for CG lightning. Another difficulty, as in case of CG lightning events, was high noise level (Fig. 12c and d) and similarity of a stepped-like E-field sequences from distant multiple RS (Fig. 2b and k) and K-changes occurring during active IC lightning stage (Fig. 12c).

Summarizing, there is still a large field of analysis regarding to the improvement of CG/IC lightning classification with application of neural networks. A real challenge is to teach artificial neural networks taking into account similarities between positive CG and IC lightning events, bipolar IC and close CG lightning events, or distant multiple RS sequences and K-change IC signatures. Some improvement in this area can be obtained using auxiliary information from LLS databases. Neural network learning based on lightning electric field signatures involving the peak value and polarity of lightning current as well as the distance to

Table 11

Average training and testing results for LSTM classifier after 10 runs (500 Hz sampled lightning E-field data).

Neural network type	Simulation Parameters ^{a)}	Accuracy (Training), %	Accuracy (Testing), %	Training Time, s	Testing Time, ms
11-layer LSTM	Given in 2nd column of Table 10	0.80	0.80	902.50	5.43
13-layer LSTM	Given in 3rd column of Table 10	0.76	0.76	2291.90	2.74
16-layer LSTM	Given in 4th column of Table 10	0.77	0.77	1578.80	4.04

^{a)} Detailed description of parameters given in R2022b Matlab documentation (available online)

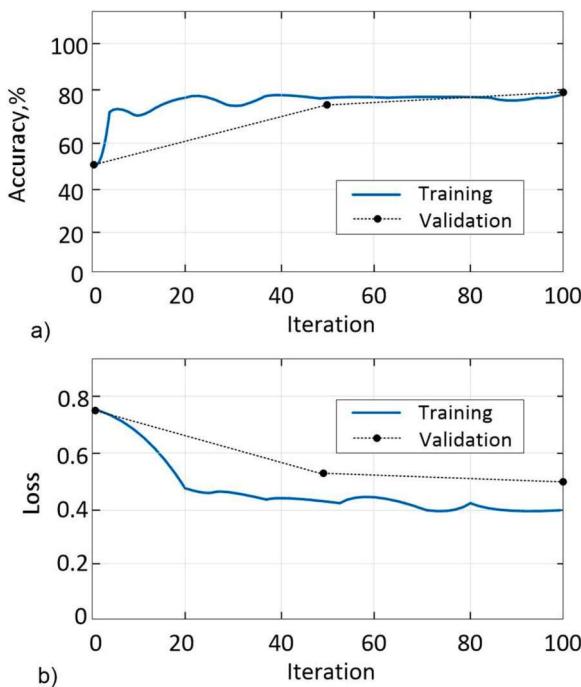


Fig. 10. Detection accuracy (a) and Loss function (b) of 11-layer LSTM network for 500 Hz sampled lightning.

the lightning channel reported for each particular E-field registration can make the NN teaching process more efficient.

4. BIG-OH analysis and computational performance of tested neural network architectures

4.1. Detailed analysis of obtained results

This section presents a detailed qualitative analysis of the obtained lightning discrimination results. The effectiveness of our classification methods can be judged by their ability to correctly recognize the CG and IC lightning discharge type and assign them to one of the two classes: positive and negative one. To this end, several measures of classification quality were introduced. The quantities relevant to the binary classifier, are the four entries in the confusion matrix (CM) given as:

$$CM = \begin{pmatrix} TP & FN \\ FP & TN \end{pmatrix} \quad (4)$$

where: TP (True Positive) means the number of correctly classified positive samples; TN (True Negative) means the number of correctly classified negative samples; FP (False Positive) represents the number of samples misclassified as positive; FN (False Negative) represent the number of samples misclassified as negative.

Based on the values of the confusion matrix, several important classification quality measures can be defined, such as *Precision*, *Sensitivity (Recall)* or *F1-score*.

Precision measure is the ratio of correctly recognized samples of a

given class to all recognized samples of this class, this metric is calculated with the following formula:

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

Recall (Sensitivity) represents the number of correctly recognized samples belonging to a given class compared to all samples belonging to that class, which is defined as follows:

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

F1-score is a balanced average of precision and recall, which means that extreme values are penalized. This metric is not symmetrical with respect to classes; its value depends on which class is defined as positive and negative. This metric is calculated as:

$$F1-score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (7)$$

The calculated classification quality metrics for the 2-layer MLP neural network were given in Table 12, for the 3-layer MLP in Table 13, for the 2-layer RBF in Table 14. Metrics for the convolutional neural networks, the best ML classifiers and LSTM network were listed in Tables 15–17, respectively.

Each one of the first rows given in tables from Table 12, 17 contain values of the confusion matrix, which details the classification results of all samples, the (TP/FN/FP/TN) values according to formula (4) contain information on the number of correct and incorrect classifications of each class of (i.e., CG or IC) lightning. The “Precision (CG/IC)” row contains the precision values of CG and IC lightning classifications. The “Recall (CG/IC)” row gives the sensitivity values of individual models in the classification of both types of lightning discharge. Finally, the “F1-score (CG/IC)” row gives the values of the balance between the precision and sensitivity of each method in classifying CG and IC lightning.

The quality metrics given in Tables 12–17 complement and confirm the obtained classification accuracies of the developed models, which were given in Table 2, Table 4, Table 9 and Table 11. 100% classification accuracy level in the 3-layer MLP models (see Table 2) represents the highest possible precision (1.0) and classification sensitivity (1.0) of CG and IC lightning events given in Table 13. RBF networks do not recognize IC-type discharges (zero precision). In 2-layer MLP networks, the highest precision (0.933) for classifying CG lightning is in a model with 25 hidden neurons. The IC lightning type are best classified by networks with 10 hidden neurons (precision of 1.0). The CG lightning type, in 1D-CNN networks, are classified with the highest (1.0) precision by the CNN based on Peng, 5 kHz, and CNN based on Wang, 5 kHz models, and the IC lightning type is incorrectly classified by the CNN based on Wang, 500 Hz model. Medium and fine Gaussian SVM models with auto Kernel function have similar quality metrics. A little bit lower classification correctness was observed when the Kernel scale was set by user. Comparing the remaining parameters of Table 16 one can see the best CG/IC Precision of fine Gaussian architecture and the highest Recall value for SVM with Kernel size set to 32. The LSTM network (Table 17) was not as good as the MLP, CNN and even SVM structures. It was confirmed by a very high level of FP classifications in the corresponding confusion matrix. The Precision and F1-score were also below the corresponding ratings given for other neural network types.

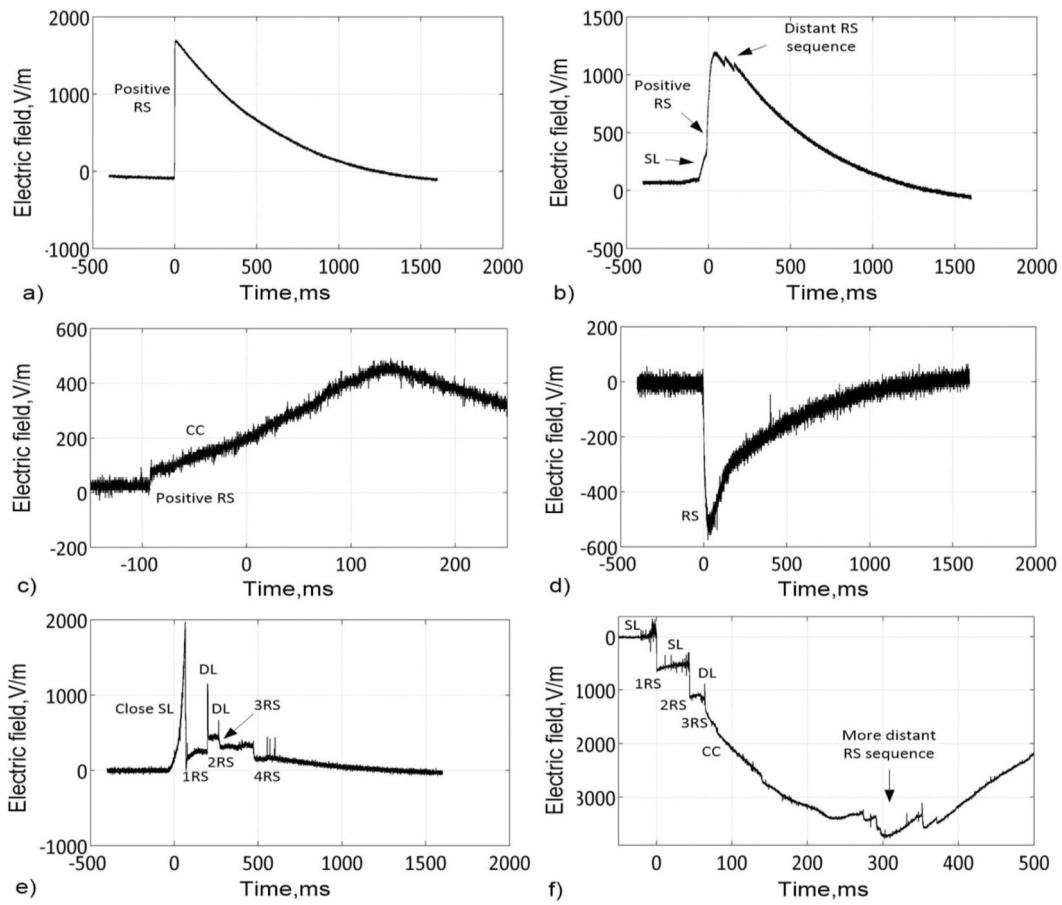


Fig. 11. The most problematic CG lightning E-field signatures during neural network learning process. Panel (a): a single positive RS with some distant CG strokes superimposed (panel b). Distant positive CG stroke with a well pronounced CC stage and high noise level. Panel (d): distant negative RS with high noise level. Panels (e-f): close and middle distance multiple RS CG lightning with opposite stepped leader stage and intensive preliminary breakdown sequences.

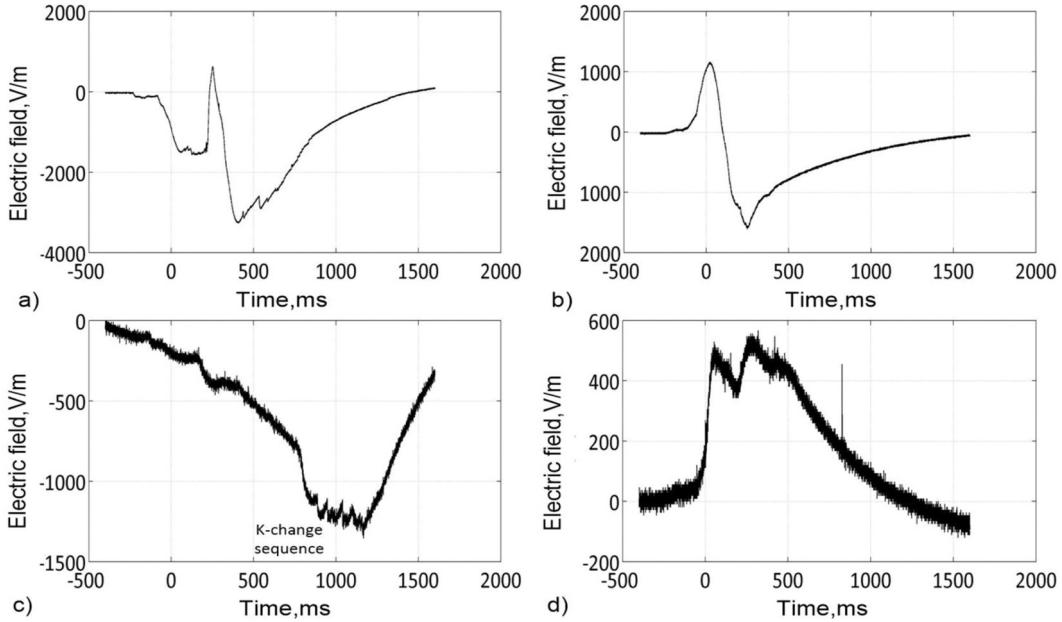


Fig. 12. The most problematic IC lightning E-field signatures for neural network learning. Panels (a-b): bipolar IC lightning sequences similar to close CG lightning with a well pronounced SL stage. Panel (c): K-change IC lightning signature similar to sequence of far multiple CG strokes. Panel (d): positive IC event close in shape to the positive CG lightning.

Table 12

Quality metrics values for 2-layer MLP neural networks used for CG/IC lightning classification (test set).

Neural Network Type	Number of Hidden Neurons				
<i>2-layer MLP</i>					
Hidden Neuron Number	10	25	100	150	225
Confusion matrix for CG/IC classification (TP/FN/FP/TN)	13/2/ 0/15	14/1/ 1/14	13/2/ 1/14	14/1/ 3/12	13/2/ 2/13
Precision (CG/IC)	1.0/ 0.882	0.933/ 0.933	0.929/ 0.875	0.834/ 0.923	0.867/ 0.867
Recall (CG/IC)	0.867/ 1.0	0.933/ 0.933	0.867/ 0.933	0.933/ 0.800	0.867/ 0.867
F1-score (CG/IC)	0.929/ 0.938	0.933/ 0.933	0.897/ 0.903	0.875/ 0.857	0.867/ 0.867

Table 13

Quality metrics values for 3-layer MLP neural networks used for CG/IC lightning classification (test set).

Neural Network Type	Number of Hidden Neurons				
<i>3-layer MLP</i>					
Hidden Neuron Number	9–3	25–5	100–10	150–13	225–15
Confusion matrix for CG/IC classification (TP/FN/FP/TN)	15/0/ 0/15	15/0/ 0/15	15/0/ 0/15	15/0/ 0/15	15/0/ 0/15
Precision (CG/IC)	1.0/ 1.0	1.0/ 1.0	1.0/1.0	1.0/1.0	1.0/1.0
Recall (CG/IC)	1.0/ 1.0	1.0/ 1.0	1.0/1.0	1.0/1.0	1.0/1.0
F1-score (CG/IC)	1.0/ 1.0	1.0/ 1.0	1.0/1.0	1.0/1.0	1.0/1.0

Table 14

Quality metrics values for 2-layer RBF neural networks used for CG/IC lightning classification (test set).

Neural Network Type	Number of Hidden Neurons				
<i>2-layer RBF</i>					
Hidden Neuron Number	62	62	62	62	62
Spread	10	20	30	40	50
Confusion matrix for CG/IC classification (TP/FN/FP/TN)	15/0/ 15/0	15/0/ 15/0	15/0/ 15/0	15/0/ 15/0	15/0/ 15/0
Precision (CG/IC)	0.5/0	0.5/0	0.5/0	0.5/0	0.5/0
Recall (CG/IC)	1.0/0	1.0/0	1.0/0	1.0/0	1.0/0
F1-score (CG/IC)	0.667/ 0	0.667/ 0	0.667/ 0	0.667/ 0	0.667/ 0

Table 15

Quality metrics values for convolution neural networks used for CG/IC lightning classification (test set).

Neural Network Type	CNN based on Peng, 5 kHz	CNN based on Wang, 5 kHz	CNN based on Wang, 500 Hz
Confusion matrix for CG/IC classification (TP/FN/FP/TN)	14/1/0/15	14/1/0/15	15/0/1/14
Precision (CG/IC)	1.0/0.938	1.0/0.938	0.938/1.0
Recall (CG/IC)	0.933/1.0	0.933/1.0	1.0/0.933
F1-score (CG/IC)	0.968/0.968	0.968/0.968	0.968/0.968

4.2. Computational time performance of tested neural networks

This section presents the computational complexity of the methods used in the form of computation times in the model learning and model testing phases.

The computational complexity of the neural networks (MLP, RBF,

Table 16

Quality metrics values for the best ML classifiers used for CG/IC lightning classification (test set).

Neural Network Type	Medium Gaussian SVM with auto Kernel function	Fine Gaussian SVM with auto Kernel function	Medium Gaussian SVM with 32 Kernel scale
Confusion matrix for CG/IC classification (TP/FN/FP/TN)	14/1/2/13	14/1/1/14	15/0/5/10
Precision (CG/IC)	0.875/0.875	0.933/0.933	0.750/1.00
Recall (CG/IC)	0.933/0.933	0.933/0.933	1.00/0.670
F1-score (CG/IC)	0.903/0.903	0.933/0.933	0.857/0.80

Table 17

Quality metrics values for LSTM network used for CG/IC lightning classification (test set).

Neural Network Type	11-layer LSTM	13-layer LSTM	16-layer LSTM
Confusion matrix for CG/IC classification (TP/FN/FP/TN)	13/2/4/ 11	12/3/4/ 11	14/1/6/9
Precision (CG/IC)	0.804/ 0.819	0.759/ 0.838	0.712/ 0.938
Recall (CG/IC)	0.800/ 0.788	0.813/ 0.707	0.947/ 0.600
F1-score (CG/IC)	0.786/ 0.783	0.767/ 0.736	0.807/ 0.719

CNN, LSTM) and ML classifiers was reported at the learning and testing stages. The unit learning time was determined by dividing the total time necessary to process the entire training dataset by the number of learning epochs and events. On the other hand, the unit testing time was defined as the time to process one lightning event by the pretrained NN. To determine the computational complexity of the NN classification models, a 10-fold learning and testing process was carried out for each NN model and architecture. The average values of learning time and testing time were calculated for 30 lightning events. Next, the unit testing time was inversed which gave another parameter called the NN speed being the number of lightning events possible to classify in time interval of one second. This parameter has important practical meaning for the real-time operating LLS.

The average values of learning speed, learning time, and unit classification time for classification models as MLP and RBF neural networks were shown in Table 18. Average computation times for convolutional neural networks were listed in Table 19. Finally, the computational performance of SVM classifiers and LSTM multilayer networks was presented in Table 20 and Table 21, respectively.

The average unit learning time for most of NN networks despite LSTM increased with an extension of the network architecture (i.e., the number of neurons). However, in CNN networks it also increased with the number of neurons and convolution filters, but decreased with an increase in the number of input data samples and the number of epochs.

Average total learning times in NN and 1D-CNN networks were proportional to the number of network parameters (i.e. the number of neurons or the number of convolution filters). The average total learning time for deep networks (CNNs) was much longer than for classical networks (MLP, RBF). In neural networks, the learning time significantly depends on the architecture, the specific operation of the optimization algorithm, the drawn values of the initial parameters (i.e. weights) of the network, or the number of input data samples, and was different in each learning process. Only in RBF networks learning times being within a small range (4–5 seconds) for all variants of this network. Total learning times for SVM classifiers were at the range of MLP and RBF neural networks and were much lower than values of the corresponding parameters for LSTM architectures which were counted in tens of minutes. Similarly, the classification time of the test set was proportional to the

Table 18

Average computation times in classification models (10 runs, MLP and RBF neural networks).

Neural Network Type	Number of Hidden Neurons				
<i>2-layer MLP</i>					
Hidden Neuron Number	9	25	100	150	225
Total learning time (s)	4.171	4.976	8.413	8.676	12.37
Unit learning time (ms/epoch/ N^a)	0.331	0.395	0.668	0.689	0.982
Unit testing time (ms)	7.63	10.1	20.0	27.6	36.9
NN speed (events/s)	131	99	50	36	27
<i>3-layer MLP</i>					
Hidden Neuron Number	9–3	25–5	100–10	150–13	225–15
Total learning time (s)	3.654	6.405	8.285	9.299	12.07
Unit learning time (ms/epoch/ N^a)	0.290	0.508	0.658	0.738	0.958
Unit testing time (ms)	8.72	10.5	20.5	27.3	38.1
NN speed (events/s)	114	91	48	36	23
<i>2-layer RBF</i>					
Hidden Neuron Number	62	62	62	62	62
Spread	10	20	30	40	50
Total learning time (s)	4.728	4.229	4.175	4.281	4.406
Unit learning time (ms/epoch/ N^a)	1.173	1.049	1.035	1.062	1.093
Unit testing time (ms)	15.5	15.7	15.1	15.2	16.8
NN speed (events/s)	64	63	66	65	59

^a) N is the number of teaching events in the NN input dataset.

Table 19

Average computation times in classification models (10 runs, convolutional neural networks).

Neural Network Type	CNN based on Peng, 5 kHz	CNN based on Wang, 5 kHz	CNN based on Wang, 500 Hz
Total learning time (s)	78.07	99.64	35.93
Unit learning time (ms/epoch/ N^a)	41.31	31.63	19.01
Unit testing time (ms)	81.5	164	74.9
NN speed (events/s)	12	6	13

^a) N is the number of teaching events in the NN input dataset.

Table 20

Average computation times in classification models (10 runs, the best ML classifiers).

Neural Network Type	Medium Gaussian SVM with auto Kernel function	Fine Gaussian SVM with auto Kernel function	Medium Gaussian SVM with 32 Kernel scale
Total learning time (s)	4.55	9.25	5.02
Unit testing time (ms)	7.41	8.13	6.37
NN speed (events/s)	135	123	157

Table 21

Average computation times in classification models (10 runs, LSTM networks).

Neural Network Type	11-layer LSTM	13-layer LSTM	16-layer LSTM
Total learning time (s)	902.50	2291.90	1578.80
Unit learning time (ms/epoch/ N^a)	601.67	1527.93	1052.50
Unit testing time (ms)	5.43	2.74	4.04
NN speed (events/s)	184	365	248

^a) N is the number of teaching events in the NN input dataset.

number of neurons and filters in a given neural network, and was less than 0.5 seconds. The average values of classification unit testing times in MLP and RBF networks were less than 0.05 seconds. In CNN, the average unit test times were less than 0.2 seconds. Thus, the MLP neural networks have greater potential for practical application in real-time LLS, as they can classify from 23 to 131 lightning events in one second. Even faster were SVM classifiers and LSTM neural networks which all achieved classification speeds well above 100 events/s.

Calculated values of NN speed parameter can be referred to the corresponding values required by a typical LLS. The average number of lightning events which have to be classified in time interval of one second for one particular LLS station can be estimated on the basis of average LLS registration range and lightning occurrence statistics. Typical range of LLS antenna can be assumed as 200 km and being the same in each direction [1,2,5]. For example, the lightning density in the south-east part of Poland is 2.5 flashes/km²/year [31]. The number of thunderstorm days for the same geographical area is 25. The average duration of the particular thunderstorm based on several-season observations is about 2 hour. Taking into account all mentioned parameters the average number of lightning events occurring in Polish conditions and detected per 1 second by particular LLS station is estimated as 1.75 flashes/s. This value can be treated as relatively small comparing with hotspot areas of the world where the lightning density can reach 100 flashes/km²/year [32]. However, such high concentrations of lightning are occurring only in a very small areas of Earth. For most world regions values of lightning density being about 4 flashes/km²/year are typical. Therefore, the value of required NN speed reported for Polish conditions should be increased by the factor 4/2.5. It gives 2.78 flashes/s necessary to register for typical real-time LLS station. Such value of registration speed is one range of values lower than obtained for even most complex MLP and RBF NN structures and all SVM classifiers, CNN and LSTM NN. However, taking into account the reported CG/IC classification accuracy, only some of the proposed MLP and CNN neural network architectures could be applied in practice to analyze of lightning discharges in a real-time LLS. In fact, there is still some margin that would allow to use the proposed algorithms even in areas with much more lightning activity than the world average.

5. Conclusions

Discrimination between CG/IC lightning events is very important issue for the lightning location systems which play an important role in early warning of human beings against hazards of dangerous lightning impact and in lightning protection of buildings, power systems, airports and other structures. For this purpose, the most common neural networks such as MLP, RBF, convolutional and LSTM structures as well as machine learning methods dedicated to classification tasks were verified in this paper. Intentionally, a relatively small dataset of 200 CG/IC lightning events but with whole 2-seconds long EF waveforms recorded by ELF/MF sensor was analyzed to determine which artificial neural network will learn the best under given circumstances. Length of each record was considerably longer than in case of other research reported in literature, and therefore, it increased significantly the number of data samples to analyze but the long-time components of lightning such as stepped leader and continuing current can be also recognized. As a result, an appropriate compromise was obtained between the number of events and the length of one observation, appropriate for the effective real-time operation of the discrimination algorithm for different types of lightning discharges that can be implemented in working lightning location systems.

The proposed approach significantly improved the neural network learning process. Some part of initially preselected CG/IC lightning event dataset made the neural network learning process harder. Thus, a special care should be taken into account during analysis of positive CG, bipolar IC lightning changes, distant multiple RS sequences and K-change IC lightning signatures which made most of difficulties during

neural network teaching and testing.

Direct comparison of MLP and RBF network structures showed a better performance of the multilayer perceptron network. RBF networks were not able to learn effectively without input data normalization. This problem did not occur in case of more complex MLP NN structures. The one-dimensional convolutional NN verified in this paper were based on structures already applied to CG/IC lightning classification for similar lightning identification tasks. Their detection accuracy and performance were at a very good level and were comparable to results given in the literature for CG/IC lightning CCN classification. Despite the high speed of classification for SVM and LSTM architectures their accuracy was too low for CG/IC lightning classifier applications. The best overall neural network became 3-layer MLP which achieved 100% detection accuracy using a relatively small lightning event dataset. The time efficiency and CG/IC lightning identification effectiveness of this network seems to be sufficient to consider a practical implementation of the reported MLP NN solution in detection procedures of the real-time lightning location systems. Further research on CG/IC classification is planned for the same lightning event dataset using 2D input data obtained from the power spectrum density analysis of the presented lightning E-field signatures.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors would like to thank the Nowcast GmbH for providing the lightning detection data from the LINET database.

The participation of all Authors in this work was partially supported within statutory activities No. PB23.ET.24.001 of the Ministry of Science and Higher Education of Poland.

References

- [1] V.A. Rakov, M.A. Uman, *Lightning: Physics and Effects*, Cambridge Univ. Press, New York, 2003.
- [2] W. Schulz, G. Diendorfer, S. Pedeboy, D.R. Poelman, The European lightning location system EUCLID—Part 1: Performance analysis and validation, in: Natural Hazards and Earth System Sciences, 16, 2016, pp. 595–605, <https://doi.org/10.5194/nhess-16-595-2016>.
- [3] A.F. Leal, V.A. Rakov, B.R. Rocha, Compact intracloud discharges: New classification of field waveforms and identification by lightning locating systems, Electr. Power Syst. Res. 173 (2019) 251–262, <https://doi.org/10.1016/j.epsr.2019.04.016>.
- [4] K.L. Cummins, M.J. Murphy, An overview of lightning locating systems: History, techniques, and data uses, with an in-depth look at the US NLDN, IEEE Trans. Electromagn. Compat. 51 (3) (2009) 499–518. (<https://doi.org/10.1109/TEMC.2009.2023450>).
- [5] Betz H.D., Schumann U., Laroche P., eds., 2008, Lightning: principles, instruments and applications: review of modern lightning research.
- [6] M. Stolzenburg, T.C. Marshall, S. Bandara, B. Hurley, R. Siedlecki, Ultra-high speed video observations of intracloud lightning flash initiation, Meteorol. Atmos. Phys. (2021) 1–26, <https://doi.org/10.1007/s00703-021-00803-3>.
- [7] H.D. Betz, B. Meneux, LINET systems—10 years experience, In 2014. International Conference on Lightning Protection (ICLP), IEEE, 2014, pp. 1553–1557, <https://doi.org/10.1109/ICLP.2014.6973377>.
- [8] Hugo K.M., 1998, A Comparative Analysis of Total Lightning Observations and Cloud-to-Cloud Lightning Observations in the Southeastern United States Region, Air Force Inst Of Tech Wright-Pattersonabf Oh.
- [9] C.D. Elvidge, F.C. Hsu, K.E. Baugh, T. Ghosh, National trends in satellite-observed lightning, Glob. Urban Monit. Assess. Earth Obs. 23 (2014) 97–118.
- [10] F. Erdmann, E. Defer, O. Caumont, R.J. Blakeslee, S. Pédeboy, S. Coquillat, Concurrent satellite and ground-based lightning observations from the Optical Lightning Imaging Sensor (ISS-LIS), the low-frequency network Meteorage and the SAETTA Lightning Mapping Array (LMA) in the northwestern Mediterranean region, Atmos. Meas. Tech. 13 (2) (2020) 853–875, <https://doi.org/10.5194/amt-13-853-2020>.
- [11] O.I. Abiodun, A. Jantan, A.E. Omolara, K.V. Dada, N.A. Mohamed, H. Arshad, State-of-the-art in artificial neural network applications: A survey, Heliyon 4 (11) (2018) e00938, <https://doi.org/10.1016/j.heliyon.2018.e00938>.
- [12] K. Gregor, I. Danihelka, A. Graves, D. Rezende, D. Wierstra, Draw: A recurrent neural network for image generation, Int. Conf. Mach. Learn., PMLR (2015) 1462–1471.
- [13] S. Kiranyaz, T. Ince, M. Gabbouj, Real-time patient-specific ECG classification by 1-D convolutional neural networks, IEEE Trans. Biomed. Eng. 63 (3) (2015) 664–675, <https://doi.org/10.1109/TBME.2015.2468589>.
- [14] D. Johari, T.K.A. Rahman, I. Musirin, Artificial neural network based technique for lightning prediction, In 2007. 5th Student Conference on Research and Development, IEEE, 2007, pp. 1–5, <https://doi.org/10.1109/SCORED.2007.4451448>.
- [15] K. Abhishek, M.P. Singh, S. Ghosh, A. Anand, Weather forecasting model using artificial neural network, Procedia Technol. 4 (2012) 311–318, <https://doi.org/10.1016/j.protcy.2012.05.047>.
- [16] M. Khan, H. Wang, A. Riaz, A. Elfatihany, S. Karim, Bidirectional LSTM-RNN-based hybrid deep learning frameworks for univariate time series classification, J. Supercomput. 77 (7) (2021) 7021–7045, <https://doi.org/10.1007/s11227-020-03560-z>.
- [17] R.S. Alkhawaldeh, B. Al-Ahmad, A. Ksibi, N. Ghatareh, E.M. Abu-Taieh, G. Aldehim, M. Ayadi, S.M. Alkhawaldeh, Convolution Neural Network Bidirectional Long Short-Term Memory for heartbeat Arrhythmia classification, Int. J. Comput. Intell. Syst. 16 (1) (2023), <https://doi.org/10.1007/s44196-023-00374-8>.
- [18] C. Peng, F. Liu, B. Zhu, W. Wang, A convolutional neural network for classification of lightning LF/VLF waveform. 2019 11th Asia-Pacific International Conference on Lightning (APL), IEEE, 2019, pp. 1–4. (https://doi.org/10.1007/978-981-16-344-0_6_11).
- [19] J. Wang, Q. Huang, Q. Ma, S. Chang, J. He, H. Wang, X. Zhou, F. Xiao, C. Gao, Classification of VLF/LF lightning signals using sensors and deep learning methods, Sensors 20 (4) (2020) 1030, <https://doi.org/10.3390/s20041030>.
- [20] Y. Zhu, P. Bitzer, V.A. Rakov, Z. Ding, A machine-learning approach to classify cloud-to-ground and intracloud lightning, Geophys. Res. Lett. 48 (1) (2021), <https://doi.org/10.1029/2021GL096714>.
- [21] M. Botsch, *Machine learning techniques for time series classification*, Cuvillier Verlag, 2023.
- [22] A. Gupta, H.P. Gupta, B. Biswas, T. Dutta, Approaches and applications of early classification of time series: A review, IEEE Trans. Artif. Intell. 1 (1) (2020) 47–61, <https://doi.org/10.1109/tai.2020.3027279>.
- [23] Y. Pu, S.A. Cummer, F. Lyu, Y. Zheng, M.S. Briggs, S. Lesage, O.J. Roberts, Unsupervised Clustering and Supervised Machine Learning for Lightning Classification: Application to Identifying EIPs for Ground-based TGF Detection, J. Geophys. Res.: Atmospheres (2023) e2022JD038369, <https://doi.org/10.1029/2022jd038369>.
- [24] V. Cooray, The Lightning Flash, Inst. Eng. Technol. (2014), <https://doi.org/10.1029/2022jd038369>.
- [25] Karnaś G., Masłowski G., Baranski P., Berlinski J., Pankaniń G., 2013, Instrumentation and data analysis process at the new lightning research station in Poland, Przegląd Elektrotechniczny, ISSN, 33(2097), pp.217-220.
- [26] G. Karnaś, Time synchronization of electric field measurement and high-speed video registration at the lightning observation station in Rzeszow, Poland, Przegląd Elektrotech. 91 (11) (2015) 235–238, <https://doi.org/10.15199/48.2015.11.55>.
- [27] G. Karnaś, G. Masłowski, P. Baranski, Power spectrum density analysis of intra-cloud lightning discharge components from electric field recordings in Poland, 2016 33rd Int. Conf. Light. Prot. (ICLP) (2016) 1–6, <https://doi.org/10.1109/ICLP.2016.7791383>.
- [28] G. Karnaś, G. Masłowski, P. Baranski, A novel algorithm for determining lightning leader time onset from electric field records and its application for lightning channel height calculations, Electr. Power Syst. Res. 178 (2020) 106021, <https://doi.org/10.1016/j.epsr.2019.106021>.
- [29] V.A. Rakov, The physics of lightning, Surv. Geophys. 34 (6) (2013) 701–729, <https://doi.org/10.1007/s10712-013-9230-6>.
- [30] L. Oneto, Model selection and error estimation in a nutshell, Springer International Publishing, 2020, https://doi.org/10.1007/978-3-030-24359-3_3.
- [31] S.-E. Enno, J. Sugier, R. Alber, M. Seltzer, Lightning flash density in Europe based on 10 years of ATDnet data, Atmos. Res. 235 (2020) 104769, <https://doi.org/10.1016/j.atmosres.2019.104769>.
- [32] Network, V.L.D., 2021, Total Lightning Statistics 2021: Vaisala Annual Lightning Report.