



A Lightning Classification Method Based on Convolutional Encoding Features

Shunxing Zhu ^{1,2}, Yang Zhang ^{1,*}, Yanfeng Fan ¹, Xiubin Sun ², Dong Zheng ¹, Yijun Zhang ³, Weitao Lyu ¹, Huiyi Zhang ¹ and Jingxuan Wang ¹

¹ State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing 100081, China; 3210301016@stu.cuit.edu.cn (S.Z.); fanyf@cma.gov.cn (Y.F.); zhengdong@cma.gov.cn (D.Z.); wtylu@cma.gov.cn (W.L.); zhanghuiyi21@mails.ucas.ac.cn (H.Z.); 22113020026@m.fudan.edu.cn (J.W.)

² College of Electronic Engineering, Chengdu University of Information Technology, Chengdu 610200, China; sunbine@cuit.edu.cn

³ Department of Atmospheric and Oceanic Sciences, Institute of Atmospheric Sciences, Fudan University, Shanghai 200433, China; zhangyijun@fudan.edu.cn

* Correspondence: zhangyang@cma.gov.cn

Abstract: At present, for business lightning positioning systems, the classification of lightning discharge types is mostly based on lightning pulse signal features, and there is still a lot of room for improvement. We propose a lightning discharge classification method based on convolutional encoding features. This method utilizes convolutional neural networks to extract encoding features, and uses random forests to classify the extracted encoding features, achieving high accuracy discrimination for various lightning discharge events. Compared with traditional multi-parameter-based methods, the new method proposed in this paper has the ability to identify multiple lightning discharge events and does not require precise detailed feature engineering to extract individual pulse parameters. The accuracy of this method for identifying lightning discharge types in intra-cloud flash (IC), cloud-to-ground flash (CG), and narrow bipolar events (NBEs) is 97%, which is higher than that of multi-parameter methods. Moreover, our method can complete the classification task of lightning signals at a faster speed. Under the same conditions, the new method only requires 28.2 μ s to identify one pulse, while deep learning-based methods require 300 μ s. This method has faster recognition speed and higher accuracy in identifying multiple discharge types, which can better meet the needs of real-time business positioning.

Keywords: lightning classification; convolutional encoder; deep learning; encoding features



Citation: Zhu, S.; Zhang, Y.; Fan, Y.; Sun, X.; Zheng, D.; Zhang, Y.; Lyu, W.; Zhang, H.; Wang, J. A Lightning Classification Method Based on Convolutional Encoding Features.

Remote Sens. **2024**, *16*, 965. <https://doi.org/10.3390/rs16060965>

Academic Editor: Pedro Melo-Pinto

Received: 14 January 2024

Revised: 1 March 2024

Accepted: 4 March 2024

Published: 10 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Lightning is one of the most common discharge phenomena in the atmosphere. According to statistics, lightning events occur approximately 30–100 times per second globally [1]. According to the spatial location of lightning, people usually classify lightning events into two types: IC and CG. According to a large amount of recorded data, cloud flash events account for about three-quarters of total lightning events [2]. In all lightning events, some lightning discharges from the cloud to the ground, which is considered a ground flash. This type of lightning poses a serious threat to human survival and may cause the death of organisms. The pulses generated by cloud flashes are different from those generated by CG. Based on the difference in waveform between IC and CG, people can simply classify discharge events based on the waveform of extremely low-frequency electromagnetic fields. Usually, CG flashes within a range of several tens of kilometers from the observation point have similar time-domain characteristics, and their initial peaks in the time-domain waveform have the characteristics of a steep rising edge and a slow falling edge [3,4]. Compared to CG, the pulses of most cloud flashes are usually narrower. In addition, due to the high diversity of discharge events, the pulse waveforms generated often exhibit

significant differences from one event to another. Therefore, the differences between the electric field waveforms generated by IC and CG are utilized by many lightning detection networks as a basis for distinguishing between different types of lightning.

For most lightning detection networks, multi-parameter methods are often used to classify lightning strikes and lightning in clouds. This method typically extracts time-domain features such as amplitude ratio, descent time, ascent time, and zero-crossing time of waveforms to characterize electromagnetic field waveforms [5,6]. According to some practical verification results, the multi-parameter method has low classification accuracy for lightning signals. For example, Flenor et al. [7] demonstrated that, in a 2005 actual inspection, approximately 54% of the ICs recorded by the National Lightning Detection Network (NLDN) were incorrectly classified as CG. Based on the inspection results, Leal et al. [8] found that when the peak current of cloud flash is greater than 50 kA, the NLDN and Earth Networks Total Lightning Network (ENTLN) wrongly classify this cloud flash as CG; Paul et al. [9] found during an inspection that 30% of the detected ground flashes were actually cloud flashes. Most lightning detection networks only provide classification results for cloud flashes and ground flashes, but there is little mention of the classification and recognition results for special cloud lightning events such as NBEs. In a few articles that provide NBE classification accuracy, their classification accuracy for NBEs is usually low, mainly due to the influence of multi-parameter classification methods. For example, when the peak current is higher than 20 kA, more than 97% of NBEs are misclassified in the NLDN, while the corresponding percentage misclassified by the ENTLN exceeds 63% [8]. Therefore, using multi-parameter methods for more precise classification of lightning discharges is a challenge.

Deep learning has become an increasingly important branch within the field of machine learning. In recent years, deep learning has made breakthrough progress in fields such as video recognition, audio analysis, and medical diagnosis. A study has used support vector machine (SVM) methods to classify extremely low-frequency lightning waveforms of cloud and ground flashes, with an accuracy of 97% [10]. The introduction of neural networks has significantly improved the classification ability of lightning waveforms. In a study by Wang et al. [11], they applied one-dimensional convolutional neural networks (CNNs) to lightning signal classification. In their results, 10 types of lightning signals were classified with an overall accuracy of 98%, but its recognition efficiency was very low and required high hardware equipment, making it difficult to meet the needs of real-time classification. On the other hand, although multi-parameter recognition methods have better recognition speed than CNNs and other methods, it is difficult to achieve higher accuracy recognition. Therefore, in practical business applications, a method that balances classification accuracy and classification speed is needed. Autoencoders were developed as early as 1980 [12]. People can use autoencoders to convert complex high-dimensional data into low-dimensional encoding, and apply autoencoders to various fields. For example, Mak et al. applied variational autoencoders to the field of game design [13], Kapoor et al. improved the detection accuracy of images by merging multi-layer features through bottleneck structures [14], Guo et al. achieved efficient compression of lightning signals through stacked autoencoders [15], and Ling et al. improved the accuracy of lightning prediction by leveraging the advantages of encoder decoder structures [16]. This article proposes a classification method based on convolutional encoding features, which utilizes the excellent feature extraction ability of convolutional neural networks and utilizes the special structure of encoders to extract low-dimensional features and complete the classification of extremely low-frequency lightning waveforms.

2. Methods

2.1. Data

All lightning data used in this article are from the Lightning Low Frequency Electric Field Detection Array (LFEDA). In 2014, the Chinese Academy of Meteorological Sciences established the LFEDA at the Field Experiment Base on Lightning Sciences of China

Meteorological Administration (CMA_FEELS). The LFEDA is used to detect triggered lightning experiments and natural lightning in the region. The array was completed in Conghua and surrounding areas of Guangzhou in 2014, and 10 independently operating substations have been set up in the area, forming a full flash positioning network. As shown in Figure 1, the detection station network is distributed between longitude 113.2~113.9°E and latitude 23.1~23.7°N. The central station is CHJ. Except for ZCJ station and GDJ station, the baseline length of the other 8 adjacent stations is between 6 and 42 km. ZCJ station and GDJ station are far away from the other 8 stations, with a baseline length of 30–68 km, forming a longer baseline. Each substation of the LFEDA is mainly composed of three parts: a lightning fast electric field change measuring instrument [17], a signal collector, and a GPS clock source. The lightning fast electric field change measuring instrument with a sensitivity of about 1 V/m is responsible for receiving spatial electric field change signals. The received signal is filtered and input to the signal collector. The length of each signal is 1 ms, and the pre trigger length is 0.2 ms. The final collected signal frequency range is 160 Hz to 600 kHz. The GPS clock source provides a time accuracy of about 30 ns to achieve synchronization among various substations, and the final waveform data have a time accuracy of about 100 ns. Each substation can capture discharge signals without dead time and perform high-precision time labeling, thereby achieving high-precision three-dimensional positioning of thunderstorm activity [18]. In order to meet the application requirements of real-time monitoring and early warning, the LFEDA was upgraded in 2020 to form a real-time low-frequency all flash positioning network (RT_LFEDA). Each station of the RT_LFEDA is able to perform real-time signal processing on the collected data, store the original waveform, and transmit the features in real time to the central station through the network. Each substation can independently extract discharge signal features and save the original waveform, and transmit the features in real-time to the central station through wireless networks. At the same time, we have built a cloud center station for real-time reception of substation data and real-time positioning results.

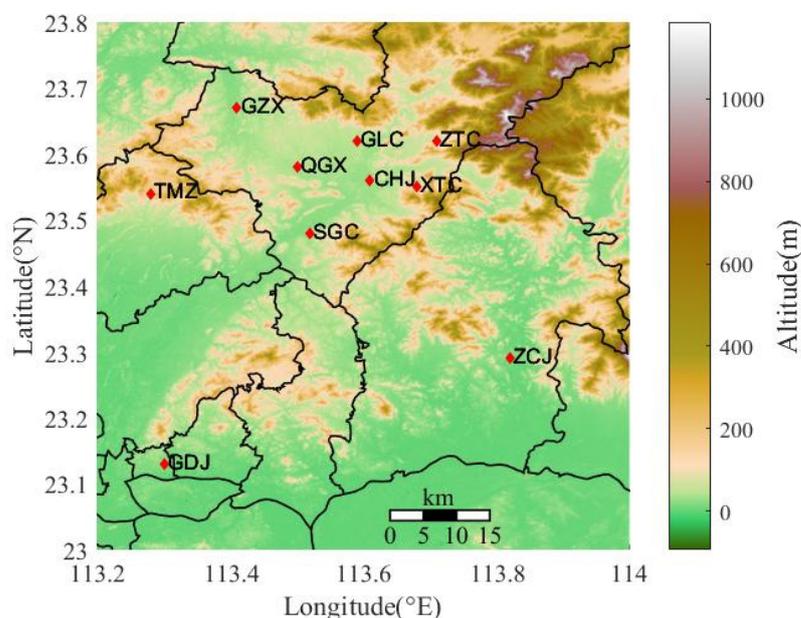


Figure 1. Layout of low-frequency electric field detection array station network.

We used data collected by RT_LFEDA for lightning discharge events during the period from June to July 2019 to establish a data set for training and testing. We selected 10,000 cloud flash, ground flash, and bipolar narrowband pulses each to form the entire training sample. The entire dataset was divided into training and testing datasets, with a segmentation ratio of 75/25. The training dataset was used during the process of adjusting model parameters and training, using quadruple cross-validation to adjust model

hyperparameters and validate model capabilities. Then, we trained the final model on the entire training dataset. During the model training phase, we did not use the test dataset. To ensure the objectivity of the evaluation, we only used the test dataset to evaluate the performance of the model in the final stage.

2.2. Classification Methods

At present, there are two main methods for distinguishing lightning types: one is based on the lightning waveform itself, directly classifying lightning waveform through different methods, such as using convolutional neural networks to directly classify lightning waveform data; another approach is to extract appropriate features from lightning waveform data and use different methods to classify these features. The classic method is to perform time-domain feature analysis on the waveform of lightning electric field changes, and achieve lightning classification by statistically analyzing the characteristic parameters of different types of lightning waveforms. The features used in this method were extracted based on artificially defined standards, which may introduce artificial errors and affect the classification results. At the same time, there is still doubt as to whether these features can fully represent waveform information. Therefore, this article proposes a lightning classification method based on convolutional encoding features; Figure 2 shows a flowchart of this method. Firstly, we filtered the lightning waveform data collected by the collector to reduce the noise contained in the data. Due to the signal length collected by the collector being 1 ms, we extracted a waveform near the signal peak as the original waveform for constructing the dataset. We also standardized the data, which can accelerate the convergence speed of the classifier. Next, we used the processed data to train the convolutional encoder. We removed the decoder part of the trained convolutional encoder, and the remaining part was called the feature extraction model. The feature extraction model was used to extract convolutional encoding features of lightning signals. At the same time, we manually selected three types of lightning signals, including IC, CG, and NBE, each containing a dataset of 10,000 samples. We used feature extraction models to extract their convolutional encoding features, forming a dataset of convolutional encoding characteristics including the three types of lightning mentioned above. Finally, we selected the appropriate classifier to train the classification model. We connected the output of the feature extraction model with the input for the lightning coding feature classification model to form our final model. Through the above process, our lightning classification method based on convolutional encoding features was ultimately formed.

2.2.1. Waveform Preprocessing

Data preprocessing is mainly divided into three parts:

- (1) Digital filters can be used to filter out high-frequency noise in lightning signals and preserve the composite lightning signals that appear in the original signal;
- (2) Peak search and interception: finding the discharge pulse by peak-searching the original data, and intercepting a certain length of waveform signal before and after the peak;
- (3) Normalization processing: Data normalization is the process of using a certain algorithm to map data to a specified range, remove unit constraints, and convert it into dimensionless pure values. Under normal circumstances, standardization allows for numerical comparability of features between different dimensions, which helps improve the accuracy and convergence speed of classifiers. This article uses the minimum maximum normalization method to uniformly map data to [0,1].

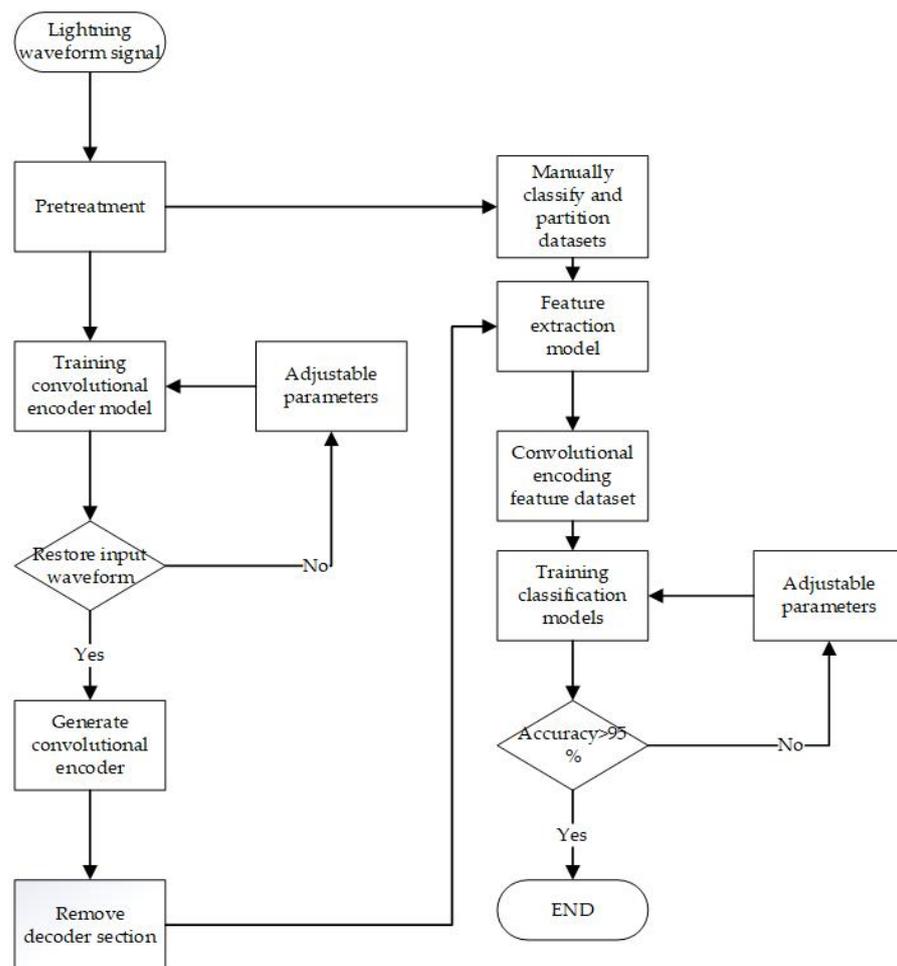


Figure 2. Flowchart of lightning classification based on convolutional encoding features.

2.2.2. Lightning Waveform Convolutional Encoding Feature Extraction Method

The key to correctly identifying discharge types using feature-based recognition methods is whether more effective features can be extracted. In previous studies, time-domain features such as the rise and fall time of primary and secondary pulses, pulse interval, and pulse to peak ratio were often used to describe the corresponding lightning waveform. However, limited time-domain features cannot fully describe the discharge waveform information, and traditional manual feature extraction methods may introduce errors, making it difficult to improve the recognition accuracy. Therefore, in order to extract more effective lightning waveform features, it is necessary to study better feature extraction methods. Artificial intelligence technology provides an effective way to accurately extract lightning features. Convolutional neural networks have excellent feature extraction capabilities. This paper applies convolutional neural network technology to lightning feature extraction, trains and establishes a lightning autoencoder, and extracts convolutional encoding features for high-precision classification of lightning signals.

The convolutional autoencoder network is a commonly used feature extraction technique, which is divided into encoder subnetworks for feature extraction and decoder subnetworks for input restoration based on different network functions. The simple encoder structure is shown in Figure 3. The part between the input layer and the hidden layer is the encoder subnetwork, and the part between the hidden layer and the output layer is the decoder subnetwork. The encoder subnetwork maps high-dimensional data raw data to a low-dimensional data space to obtain representation features with higher information density, while the decoder subnetwork is completely the opposite, restoring low dimensional features to the original data. Autoencoders, also known as bottleneck

structures based on network structure, have a feature dimension smaller than the input and output in the middle layer, which forces the model to retain important information for reconstructing data samples in the encoding features of the middle layer. In order to minimize the reconstruction error of the restored data, it is required to retain the same information as most samples. In order to obtain richer and more informative feature representations, many studies have made appropriate adjustments in autoencoder networks. Convolutional autoencoders use convolutional kernels in the middle layer, which have efficient high-dimensional feature extraction and compression capabilities and exhibit excellent performance on various types of data.

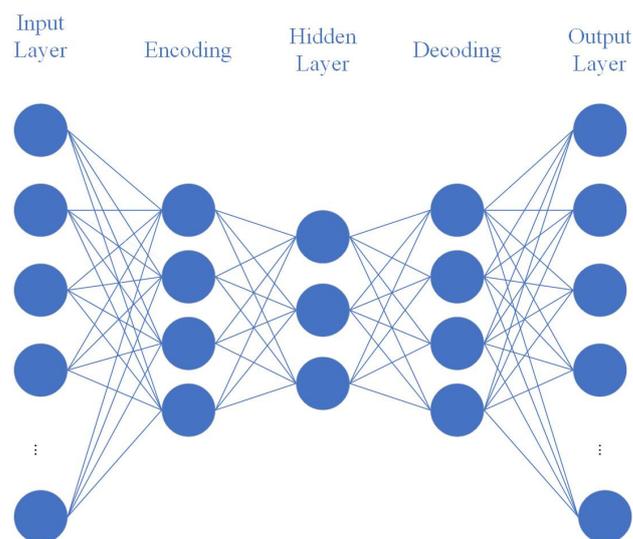


Figure 3. Autoencoder structure.

This paper uses a convolutional autoencoder to obtain the encoding features of lightning pulse signals. Each layer of the convolutional encoder has convolution kernels of different sizes, which have a good recognition effect on waveform edge trends. In waveform feature extraction, waveform edge recognition is particularly important for waveform description. The convolutional encoder framework mainly consists of three parts: encoder, decoder, and feature output module. The convolutional encoder is composed of alternating convolutional layers and pooling layers. The decoder has a completely opposite structure to the encoder, expanding the encoding features into a reconstructed waveform that is basically consistent with the input waveform through deconvolution and pooling layers. There is a feature output module between the encoder and decoder, which can convert effective encoding into convolutional encoding features for output. By minimizing the error at both ends of the convolutional encoder, we can obtain more effective convolutional encoding features. The training process extracts pulse waveform encoding features through an encoder, and the decoder decodes the encoding features into waveforms. In neural networks, loss functions are commonly used to measure the quality of training results. In this experiment, we used mean square error (MSE) as the loss function. The smaller the loss value, the smaller the difference between the predicted value and the true value. During the training process, we judged the quality of the model by observing the changes in the loss function. In this task, we can also judge the quality of the model by comparing the consistency of the original waveform before and after entering the convolutional encoder. We drew the waveforms of the input and output of the convolutional encoder and by observing whether there is a significant difference between the two we can intuitively judge the quality of the model. Figure 4 compares the waveforms of the input and output of the convolutional encoder after decoding. We can see that the decoded output waveform is basically consistent with the input waveform, indicating that the convolutional autoencoder can restore the original waveform well. We removed the decoder part of the

convolutional encoder, and the remaining part was called the feature extraction model. In practical applications, we only need to input the lightning pulse waveform into the feature extraction model to extract the convolutional encoding features of lightning pulses.

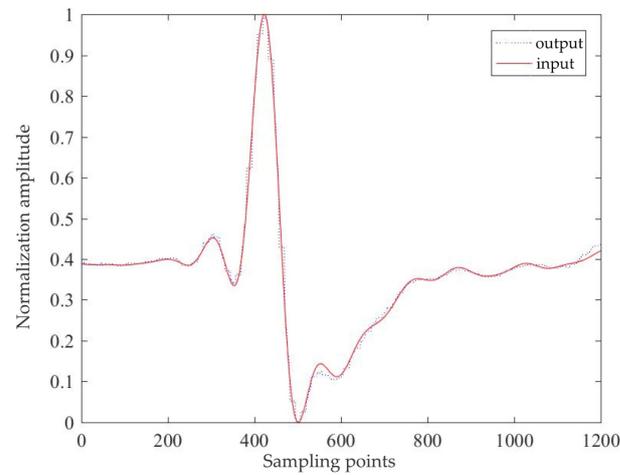


Figure 4. Comparison between input waveform and reconstructed waveform. The red curve represents the input waveform, and the blue dashed line represents the reconstructed waveform.

The convolutional encoder we designed mainly consists of an encoder and a decoder. The encoder section is composed of alternating stacking of convolutional layers and pooling layers. The decoder section is composed of alternating convolutional layers and upsampling layers. The specific parameters of the model are shown in Table 1. After inputting 1200×1 waveform data through the input layer, the encoder part passes through a stacked structure of 4 Conv1D layers and MaxPooling1D layers to obtain 8×16 feature data. The parameters and structure of the decoder and encoder parts are completely symmetrical, and the 8×16 feature data obtained in the encoder part is reconstructed into waveform data after passing through the decoder part. We have designed a feature output module between the encoder and decoder sections, which consists of a Conv1D layer with special parameters. This module will not be used during the training process and will only function when feature extraction is required.

Table 1. Detailed parameters of convolutional encoder structure.

No.	Module	Layer	Filter Number	Kernel Size	Pooling Window Size	Activation Function	Output Shape
1		Input	-	-	-	-	(1200,1)
2		1D-Conv	128	3	-	ReLU	(1200,128)
3		MaxPooling1D	-	-	5	-	(240,128)
4		1D-Conv	64	3	-	ReLU	(240,64)
5	Encoder	MaxPooling1D	-	-	5	-	(48,64)
6		1D-Conv	32	3	-	ReLU	(48,32)
7		MaxPooling1D	-	-	3	-	(16,32)
8		1D-Conv	16	3	-	ReLU	(16,16)
9		MaxPooling1D	-	-	2	-	(8,16)
10	Feature output module	1D-Conv	1	1	-	ReLU	(8,1)
11		1D-Conv	16	3	-	ReLU	(8,16)
12		UpSampling1D	-	-	2	-	(16,16)
13		1D-Conv	32	3	-	ReLU	(16,32)
14		UpSampling1D	-	-	3	-	(48,32)
15	Decoder	1D-Conv	64	3	-	ReLU	(48,64)
16		UpSampling1D	-	-	5	-	(240,64)
17		1D-Conv	128	3	-	ReLU	(240,128)
18		UpSampling1D	-	-	5	-	(1200,128)
19		Output	1	1	-	-	(1200,1)

1D-Conv: 1D convolutional layer. ReLU: Rectified Linear Unit.

2.2.3. Classifier

There are quite a few classification algorithms in machine learning, and different classification algorithms will produce different results in different applications. Therefore, choosing the appropriate classification algorithm is also very important. Random forest was proposed by Leo Breiman (2001) [19], which is an ensemble learning algorithm. Firstly, it randomly selects N samples from the original sample set with replacement as the training set. Next, it randomly selects m features from each sample for training. Finally, it uses all trained decision trees to classify or predict the data to be validated, and uses methods such as voting or averaging to obtain the final result. This approach of collecting results from multiple decision trees can help to improve the accuracy of classification. In a random forest, each decision tree is trained based on different samples and features, so they can complement each other, thereby improving the accuracy and robustness of the entire model.

SVM is a common binary machine learning model that is often used for linear or nonlinear classification. SVM performs well when the classified object is linearly separable. The mathematics behind SVM are beyond the scope of this article and are explained in detail by Hastie et al. [20], so it will not be elaborated in detail in this article. The ultimate goal of SVM is to find the most suitable classification hyperplane and apply it effectively to complete classification tasks. Although SVM is a binary classifier, it can also perform multi-classification tasks through some methods.

We input the dataset formed by convolutional encoding features into two classifiers for training. By comparing the classification results and recognition speed on the test set, we selected the classifier with better performance. Here, we only compared random forest and SVM classifiers, which are used to classify lightning waveforms using the convolutional encoding features described in this article. Table 2 shows the performance of two classification models when the number of encoding features is 8. We compared the classification results of the two classifiers using the same training set and found that the classification performance of the random forest model was slightly higher than that of SVM. The classification accuracy of the two classifiers only differs by 2%. However, the random forest classifier recognizes a single waveform in less than 30 microseconds, while the SVM classifier recognizes a single pulse in approximately 87.2 microseconds. Therefore, the random forest model is more suitable as a classifier for classifying lightning waveforms using convolutional encoding features. For multi-class classification models, we used Equation (1) to calculate the accuracy of each category.

$$Accuracy = \frac{Correct\ Predictions}{Total\ Samples} \quad (1)$$

Table 2. The results of convolutional coding features on SVM and random forest classifiers.

Classifier	Random Forest	SVM
Single waveform recognition Time (μ s)	28.2	87.2
Accuracy	97%	95%

3. Results

3.1. Determination of Parameters

During the research process, we found that different pulse lengths can affect the classification results. Therefore, we studied the changes in accuracy corresponding to lightning waveforms with different lengths. Based on the duration distribution of lightning discharge pulses, we selected pulse waveforms with total lengths of 80 μ s, 100 μ s, 120 μ s, and 140 μ s. We extracted eight convolutional encoding features from four different lengths of lightning pulses, and presented their classification results in Table 3. As the pulse length gradually increases, we found that the corresponding classification accuracy also improves. However, when the pulse length exceeds 120 μ s, there was no significant change in accuracy. Therefore, we believe that when the pulse length reaches 120 μ s our method can accurately distinguish the three types of lightning signals.

Table 3. Comparison of accuracy of four pulse lengths.

	Pulse Length			
	80 μ s	100 μ s	120 μ s	140 μ s
Accuracy	93%	95%	97%	97%

At the same time, we also studied the impact of the number of convolutional encoding features on accuracy. In Table 4, we provide the classification accuracy for each lightning type when the pulse length is 120 μ s and the number of convolutional encoded features is set to 12, 8, and 4. When the number of extracted features exceeds eight, the overall event accuracy of the test dataset remains stable at around 97%. The accuracy of CG in the test dataset was slightly higher than that of IC, which is similar to the accuracy of NBE. The overall accuracy of event classification in the test dataset was basically consistent with the accuracy of the training dataset, with little difference in accuracy for each category, indicating that our model can be well applied to “unprecedented” test data.

Table 4. The characteristic bits of convolutional coding (4, 8 and 12 respectively) and the accuracy of three types of discharge events.

Number of Feature	Classification Accuracy				Total
	IC	CG	NBE		
12	97%	98%	97%		97%
8	97%	97%	97%		97%
4	93%	96%	93%		94%

3.2. Model Testing

On 7 July 2019, multiple thunderstorms occurred in Conghua District, Guangzhou City, Guangdong Province, China. We selected the waveform data recorded by the substation of the RT_LFEDA located at the Conghua Meteorological Bureau in Guangzhou from 15:48 to 16:48 on that day. A total of 9241 pieces of data within an hour were collected to test the actual accuracy of the model proposed in this article. Based on the results of manual inspection, we sorted the recognition results of the model. Based on the inspection results, we placed the actual identification results of each lightning type in the discharge process of the model in Table 5. Our method achieved classification accuracies of 96.83%, 97.01%, and 97.37% for IC, CG, and NBE in this test. We found that this was very similar to the results in the test set, indicating that our method has good generalization ability. Therefore, we believe that our method could also achieve good results in practical applications.

Table 5. The classification accuracy of the model for each type of lightning signal.

	Classification Lightning			
	IC	CG	NBE	Total
True Classification	8071	583	297	8951
Error Classification	264	18	8	290
Accuracy	96.83%	97.01%	97.37%	96.86%

4. Discussion

4.1. Comparison of Classification Methods Based on Different Features

Based on the test set data, this section compares the classification performance of classification methods based on waveform time-domain parameter features and convolutional encoder features. As for the identification method of waveform time-domain parameter characteristics, this paper refers to the method proposed by Cai et al. [21] to extract time-domain characteristic parameters such as rise time, fall time, pulse width, pulse

interval, pulse amplitude ratio, the ratio of the maximum waveform fluctuation amplitude before pulse to its peak to peak value, the ratio of the difference between the maximum and minimum values in the data, and the ratio of the negative and positive amplitudes of the pulse; we then use a random forest classifier for multi-classification testing. Table 6 shows the accuracy of two methods for each type of lightning waveform. In terms of accuracy, the convolutional encoding features of the three types of lightning waveforms exhibit higher precision compared to their waveform time-domain features. The reason for this difference may be that some cloud flashes and NBEs have similar waveforms, and the extracted time-domain waveform cannot describe the differences between the two in some cases, resulting in the inability of waveform time-domain feature-based recognition methods to accurately distinguish the two waveforms. This article selects one cloud flash pulse and one NBE pulse as shown in Figure 5. The waveforms of the two pulses have high similarity near the peak. When using a multi-parameter method for recognition, both pulses are recognized as NBEs, and using convolutional encoding features can distinguish between the two. In addition, the multi-parameter method also requires detailed feature engineering to extract the feature vectors of waveform time-domain features. When facing different datasets, it requires multiple fine-tuning of feature vector selection to maintain high performance, which has certain limitations and lower efficiency than convolutional encoding features.

Table 6. Comparison of classification accuracy between time domain features and convolutional coding features.

Method	A Random Forest Classification Model Based on Waveform Time-Domain Features			A Random Forest Classification Model Based on Convolutional Encoding Features			
	Type	IC	CG	NBE	IC	CG	NBE
Accuracy		89%	90%	88%	97%	97%	97%

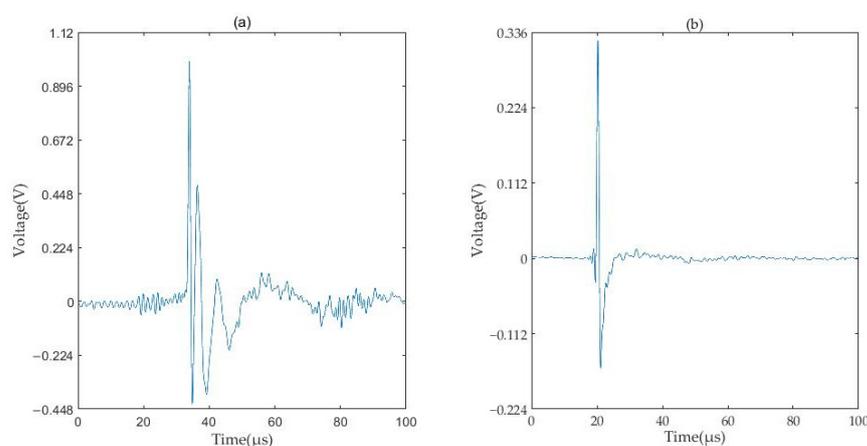


Figure 5. One cloud flash pulse waveform (a) and one NBE pulse waveform (b).

4.2. Comparing Classification Methods Based on Waveform and Encoding Features

Wang et al. [11] applied deep learning methods to the classification of lightning signals. The model structure of this method mainly consists of stacked convolutional layers and pooling layers, and the classification results are ultimately obtained through fully connected layers and output layers. The convolutional layer extracts different features of lightning waveforms through convolutional kernels of different sizes, the pooling layer reduces the dimensionality of the extracted features, and finally the fully connected layer integrates all features to calculate the probability of each type included in the training set. In order to compare with the effectiveness of this study, we tested the method proposed by Wang et al. [11] using our dataset. We obtained the classification accuracy of two models for three types of lightning and the time required for each model to recognize one pulse, as shown in Table 7. For three types of lightning discharges, both classification methods

maintained high recognition accuracy, with values of 99% and 97%, respectively. However, the method based on convolutional encoding features had significantly better identification speed than the CNN method, with recognition times for individual pulses of 300 μs and 28.5 μs , respectively. Considering the rapid progress of lightning positioning technology and the increasing number of lightning discharge events captured within a unit time, higher requirements are put forward for real-time identification speed. Compared to CNN methods, the method proposed in this article can better meet the requirements of higher performance real-time positioning services.

Table 7. Comparison of classification accuracy and recognition time between the CNN model and convolutional coding features.

Method	Waveform Based Convolutional Neural Network Classification Model			A Random Forest Classification Model Based on Convolutional Encoding Features		
Single pulse recognition Time (μs)	300			28.2		
Type	IC	CG	NBE	IC	CG	NBE
Accuracy	99%	99%	99%	97%	97%	97%

It is worth noting that the above comparison is a comparison of the effectiveness of different identification methods. During the comparison process, the identification accuracy obtained by each method is true. This is because for each method, we use the results of manually identifying waveforms as the testing criteria. Using the results of manual waveform identification as a standard is also a commonly used method in identifying other lightning discharge signals, and its reliability has been widely proven. Due to the unique waveform characteristics of discharge signals such as CG and NBEs involved in this article, the results of manually identifying waveforms themselves are relatively reliable.

5. Conclusions

This article proposes a lightning classification method based on convolutional encoding features, which can recognize and classify various types of lightning such as IC, CG, and NBEs. The main results are as follows:

(1) This paper proposes a multi-type lightning discharge recognition method based on encoding features and a random forest classifier. The method utilizes the excellent feature extraction ability of convolutional neural networks to effectively extract the convolutional encoding features of lightning waveform signals. After comparative testing, the random forest classifier outperformed SVM in accuracy and recognition speed for the same dataset classification. Therefore, we chose a random forest classifier as the final classifier for this method. Comparative tests were conducted on two factors that may affect the classification results; namely, the length of lightning pulse signals and the number of bits extracted from convolutional encoding features. After verification, it was shown that when the pulse length is 120 μs and 8-bit convolutional encoding features are extracted, this method can achieve a high recognition accuracy of 97% for the IC, CG, and NBE lightning types;

(2) Compared with multi-parameter classification methods, the new method solves the problem that multi-parameter classification methods cannot accurately identify similar cloud discharge signals (such as NBEs and cloud flash pulses), and also improves the classification accuracy of IC and CG, which to some extent reduces the complexity of lightning signal classification tasks. It was found that for the same set of data, the new method has a higher classification accuracy for the three types of lightning than the multi-parameter classification method, with an accuracy rate of about 97%. In addition, the new method does not require detailed feature engineering to extract pulse time-domain features, such as rising edge time, falling edge time, pulse width, and peak to peak ratio before and after the pulse, reducing the possibility of low accuracy caused by human error;

(3) Compared with waveform classification methods based on convolutional neural networks, our method can quickly identify lightning signals. From the recognition results, both methods have high recognition accuracy, with an accuracy difference of only about

2%. The time required for the recognition of single pulse signals by the new method is about one-tenth of that of waveform classification methods based on convolutional neural networks. The recognition speed has been greatly improved. The method can also quickly classify lightning signals while maintaining high classification accuracy, providing important support for real-time positioning business applications.

The type of lightning discharge is a key parameter for lightning monitoring and research. High-precision lightning classification information can enhance the applicability of lightning positioning data and can also be used for positioning, reducing the error of lightning positioning to a certain extent. In addition, this method is not only applicable to the classification of CG, IC, and NBEs, but can also be extended to identify more types of lightning discharges. Currently, various total lightning positioning technologies have been applied to business positioning systems. However, due to the complexity of cloud flash signals, there are many errors in using traditional signal features to identify lightning discharge types. The method proposed in this article can be applied to business positioning systems in two ways. The first way is to apply the feature extraction model to each substation of the lightning positioning system, extract the encoded features in real time, and send them to the central station. At the central station, the lightning type identification results are obtained through a random forest classification model. The second approach is to directly use the feature extraction model and random forest classification model proposed in this article in the substations to obtain the lightning discharge type results, and then send the lightning signal classification results to the central station.

Author Contributions: Conceptualization, Y.Z. (Yang Zhang); methodology, S.Z. and Y.Z. (Yang Zhang); software, S.Z.; validation, S.Z., Y.Z. (Yang Zhang), D.Z., and Y.Z. (Yijun Zhang); investigation, S.Z.; resources, Y.Z. (Yang Zhang) and Y.F.; data curation, S.Z., Y.Z. (Yang Zhang), D.Z., and Y.F.; writing—original draft preparation, S.Z. and Y.Z. (Yang Zhang); writing—review and editing, S.Z. and Y.Z. (Yang Zhang); visualization, S.Z.; supervision, Y.Z. (Yang Zhang), X.S., H.Z., J.W., and W.L.; project administration, Y.Z. (Yang Zhang) and W.L.; funding acquisition, Y.Z. (Yang Zhang) and W.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the 2022 China Electric Power Research Institute Laboratory Open Fund Project (Project Title: Research on Comprehensive Lightning Observation and Discharge Characteristics in High Altitude Areas of Tibet), the S&T Development Fund of CAMS (2023KJ050), the Basic Research Fund of the Chinese Academy of Meteorological Sciences (Grant 2021Z011, 2023Z008), and the National Key Research and Development Program of China (2019YFC1510103).

Data Availability Statement: All lightning data from real-time low-frequency electric-field detection array (RT_LFEDA) were collected by State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, and the data used in this paper can be obtained from the website (<https://zenodo.org/records/10782763>, accessed on 13 January 2024).

Acknowledgments: We are thankful to all members of the lightning group of the State Key Laboratory of Severe Weather, as well as Shaodong Chen, Xu Yan for the assistance in experimental maintenance.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Rakov, V.; Uman, M.; Raizer, Y. Lightning: Physics and Effects. *Phys. Today* **2004**, *57*, 63–64.
2. Rakov, V.A. Lightning phenology and parameters are important for lightning protection. In Proceedings of the IX International Symposium on Lightning Protection, Foz do Iguaçu, Brazil, 26–30 November 2007.
3. Haddad; Michael, A.; Rakov, V.A.; Cummer, S.A. New measurements of lightning electric fields in Florida: Waveform characteristics, interaction with the ionosphere, and peak current estimates. *J. Geophys. Res. Atmos.* **2012**, *117*, D10. [[CrossRef](#)]
4. Lin, Y.T.; Uman, M.A.; Tiller, J.A.; Brantley, R.D.; Beasley, W.H.; Krider, E.P.; Weidman, C.D. Characterization of lightning return stroke electric and magnetic fields from simultaneous two-station measurements. *J. Geophys. Res. Ocean.* **1979**, *84*, 6307–6314. [[CrossRef](#)]
5. Murphy, M.J.; Cramer, J.A.; Said, R.K. Recent history of upgrades to the U.S. National lightning detection network. *J. Atmos. Ocean. Technol.* **2021**, *38*, 573–585. [[CrossRef](#)]
6. Wooi, C.-L.; Abdul-Malek, Z.; Salimi, B.; Ahmad, N.A.; Mehrazamir, K.; Vahabi-Mashak, S. A comparative study on the positive lightning return stroke electric fields in different meteorological conditions. *Adv. Meteorol.* **2015**, *2015*, 307424. [[CrossRef](#)]

7. Fleenor, S.A.; Biagi, C.J.; Cummins, K.L.; Krider, E.P.; Shao, X.-M. Characteristics of cloud-to-ground lightning in warm-season thunderstorms in the Central Great Plains. *Atmos. Res.* **2009**, *91*, 333–352. [[CrossRef](#)]
8. Leal, A.F.; Rakov, V.A.; Rocha, B.R. Compact intracloud discharges: New classification of field waveforms and identification by lightning locating systems. *Electr. Power Syst. Res.* **2019**, *173*, 251–262. [[CrossRef](#)]
9. Paul, C.; Heidler, F.H.; Schulz, W. Performance of the European Lightning Detection Network EUCLID in Case of Various Types of Current Pulses from Upward Lightning Measured at the Peissenberg Tower. *IEEE Trans. Electromagn. Compat.* **2020**, *62*, 116–123. [[CrossRef](#)]
10. Zhu, Y.; Bitzer, P.; Rakov, V.; Ding, Z. A machine-learning approach to classify cloud-to-ground and intracloud lightning. *Geophys. Res. Lett.* **2021**, *48*, e2020GL091148. [[CrossRef](#)]
11. Wang, J.; Huang, Q.; Ma, Q.; Chang, S.; He, J.; Wang, H.; Zhou, X.; Xiao, F.; Gao, C. Classification of VLF/LF Lightning Signals Using Sensors and Deep Learning Methods. *Sensors* **2020**, *20*, 1030. [[CrossRef](#)] [[PubMed](#)]
12. Hinton, G.E.; Salakhutdinov, R.R. Reducing the dimensionality of data with neural networks. *Science* **2006**, *313*, 504–507. [[CrossRef](#)] [[PubMed](#)]
13. Mak, H.W.L.; Han, R.; Yin, H.H.F. Application of Variational AutoEncoder (VAE) Model and Image Processing Approaches in Game Design. *Sensors* **2023**, *23*, 3457. [[CrossRef](#)] [[PubMed](#)]
14. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation in CVPR. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; IEEE: New York, NY, USA, 2014; pp. 580–587.
15. Guo, J.; Wang, J.; Xiao, F.; Zhou, X.; Liu, Y.; Ma, Q. An Efficient Compression Method for Lightning Electromagnetic Pulse Signal Based on Convolutional Neural Network and Autoencoder. *Sensors* **2023**, *23*, 3908. [[CrossRef](#)] [[PubMed](#)]
16. Lin, T.; Li, Q.; Geng, Y.-A.; Jiang, L.; Xu, L.; Zheng, D.; Yao, W.; Lyu, W.; Zhang, Y. Attention-Based Dual-Source Spatiotemporal Neural Network for Lightning Forecast. *IEEE Access* **2019**, *7*, 158296–158307. [[CrossRef](#)]
17. Krehbiel, P.R.; Brook, M.; McCrory, R.A. An analysis of the charge structure of lightning discharges to ground. *J. Geophys. Res.* **1979**, *84*, 2432–2456. [[CrossRef](#)]
18. Chen, Z.; Zhang, Y.; Zheng, D.; Zhang, Y.; Fan, X.; Fan, Y.; Xu, L.; Lyu, W. A Method of Three-Dimensional Location for LFEDA Combining the Time of Arrival Method and the Time Reversal Technique. *J. Geophys. Res. Atmos.* **2019**, *124*, 6484–6500. [[CrossRef](#)]
19. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
20. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*; Springer: New York, NY, USA, 2009. [[CrossRef](#)]
21. Cai, L.; Liu, W.; Zhou, M.; Wang, J.; Yan, R.; Tian, R.; Fan, Y. Differences of Electric Field Parameters for Lightning Strikes on Tall Towers and Nonelevated Objects. *IEEE Trans. Electromagn. Compat.* **2022**, *64*, 2113–2121. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.