Thunder Signal Detection via Deep Learning

Han Zhang\textsuperscript{1,2,3*}, Biwu Yan\textsuperscript{1,2,3}, Shanqiang Gu\textsuperscript{1,2,3}, Chao An\textsuperscript{4}, Jian Li\textsuperscript{1,2,3}, Min Wu\textsuperscript{1,2,3}, Yu Wang\textsuperscript{1,2,3}, and Heng Xu\textsuperscript{1,2,3}

\textsuperscript{1}NARI Group Corporation Ltd., Nanjing, China
\textsuperscript{2}Wuhan NARI Limited Liability Company, State Grid Electric Power Research Institute, Wuhan, China
\textsuperscript{3}Hubei Key Laboratory of Power Grid Lightning Risk Prevention, Wuhan, China
\textsuperscript{4}Technical Research Center, CRRC Tangshan Co.Ltd., Tangshan, China

*E-mail: zhanghan1@sgepri.sgcc.com.cn

Abstract. In this paper, a thunder signal detection method is proposed based on the deep learning framework. The recorded thunder signal is segment-wise acquired, stored and pre-processed. In each frame, we use Mel Frequency Cepstrum Coefficient (MFCC) to extract the features of the thunder signal, which is consistent with the frequency characteristics of human perception. We then use the MFCC features derived in each frame to form a 3-channel tensor data, which is used as the further input to the designed convolutional neural network (CNN). The goal of CNN is to classify the existence of thunder for a single data frame. To improve the robustness of CNN, we included other confusing signals that are similar to thunder signals in the training and testing datasets. On the testing dataset, our proposed method outperforms the state-of-art methods in terms of accuracy, sensitivity, and specificity. Our proposed deep-learning-based thunder detection method not only increases the real-time performance of the lightning location system with thunder signals but also further improves the accuracy of other sound alarm systems.

1. Introduction
Lightning is a natural discharge of electricity with high intensity causing different kinds of disasters including the trip of transmission lines, serious forest fire, human injuries or fatalities, etc. It is well known that lightning not only produces strong electromagnetic and optical signals, but also intense acoustic signals [1]. The lightning location system with thunder signals is developed in lightning scientific research for a long time but has not applied in the engineering project [2-8]. Prompt and correct recognition of the thunder signal from ambient noise is the foundation challenge of the lightning location system with thunder signals [9]. Moreover, the thunder signal is one of the key sources of disturbance to many sound alarm systems. Therefore, there are significant impacts to distinguish thunder signals from other sound signals.

There have been several subjective terms such as clap, peal, roll and rumble to describe thunder in the literature. Peals or claps are the sudden and loud sounds which occur in a background of prolonged roll or rumble. The term roll is sometimes used to describe irregular sound variations whereas rumble is used to describe relatively weak sound of long duration [10]. However, each thunder signal has its peculiar characteristic. There is only one fact generally agreed by investigations that the power spectrum of the thunder signal is broad, in the range from a few Hz to a few hundred Hz and with a frequency peak at in the order of 100 Hz [11-12]. The quantitative characteristic parameter of the thunder waveform has not been defined yet. At present, the endpoint detection and feature extraction
of frequency spectrum have been tried in thunder recognition in which the selection of the feature parameter is subjective [13]. Recently, a branch of artificial intelligence known as deep learning has been widely adopted in environmental sound, music, and speech processing, as well as many other fields such as drug discovery, genomics, quantum chemistry, and natural language processing [14]. Deep learning techniques provide a novel way of thunder signal detection.

In this paper, a new thunder signal detection method is proposed based on the deep learning framework. The recorded thunder signal is segment-wise acquired, stored and preprocessed. In each frame, we use Mel-Frequency Cepstrum Coefficients (MFCC) to extract the features of the thunder signal, which is more consistent with the frequency characteristics of human perception. We then use the MFCC features derived in each frame to form a 3-channel tensor data, which is used as the further input to the designed convolutional neural network (CNN). The goal of CNN is to classify the existence of thunder for a single data frame. To improve the robustness of CNN, we included other confusing signals that are similar to thunder signals in the training and testing datasets. On the testing dataset, our proposed method outperforms the state-of-art methods in terms of accuracy, sensitivity, and specificity. Our proposed deep-learning-based thunder detection method not only increases the real-time performance of the lighting location system with thunder signals but also further improves the accuracy of other sound alarm systems.

2. Method

2.1. Signal Preprocessing

It is well known that the acoustic thunder signal is non-stationary and time-varying. The waveform of thunder signal is rich in information that can be utilized in automatic detection of thunder occurrence. To improve the efficiency of data utilization and eliminate interference, it is necessary to preprocess the recorded thunder signal before further analysis.

Firstly, to make the proposed method suitable for real-time application, segment-wise data acquisition is required. Each recording, under suspicion of including thunder signal, is divided into data segments with equivalent lengths, and adjacent data segments are overlapped with a given percentage. In this paper, the length of each segment is three seconds, and 2/3 overlapping is selected. Secondly, the data segments are down-sampled with a resampling frequency of $f_s=4000$ Hz. Finally, the resampled data segments are passed through a Butterworth bandpass filter, with the lower and upper cut-off frequencies being 100 Hz and 1000 Hz respectively. Then the target is converted to judge whether the processed data segment includes thunder signal. In this paper, as long as over 10% ratio of the data segment contains thunder, it is determined that there exists the thunder signal in the current time window.

2.2. Feature Extraction of the Thunder Signal

MFCC feature is a very popular choice for performing speaker recognition. Therefore, we choose MFCC for representing the perceptual thunder features for our work. The procedure of MFCC extraction method is given as follows.

Firstly, each data segment will be emphasized by a high-pass filter $H(z)=1-\mu z^{-1}$, in which $\mu=0.9357$ is used in this paper. The high-frequency components of the data segment will be highlighted by the pre-emphasis Filter. Therefore, the flatten frequency spectrum can facilitate the next step of feature extraction.

Secondly, the data segment is divided into multiple frames by a sliding window, from which the extracted MFCCs are used to form a feature map for each data segment. An $N_{\text{FFT}}$-point window is used and 50% overlapping is employed. In this paper, we adopt $N_{\text{FFT}}=128$. Meanwhile, each frame is weighted by a Hamming window, which is given by

$$w(n) = 0.5 - 0.5 \times \cos \left( \frac{2\pi n}{N_{\text{FFT}} - 1} \right), \quad 0 \leq n \leq N_{\text{FFT}} - 1$$

(1)
The 50% overlapping between adjacent frames is necessary because only the intermediate portion of each frame is emphasized by windowing, as observed from equation (1). While a data segment is finally divided into \( M \) frames, the weighted data in the \( m \)-th frame is arranged as 
\[
x_m=[x_m(0), x_m(1), \ldots, x_m(N_{FFT}-1)]^T,
\]
where \( (\ast)^T \) denotes transpose operation.

Thirdly, the frequency spectra of \( x_m \), \( m=1,2,\ldots,M \) will be obtained by the fast Fourier transform (FFT), represented by
\[
Y_m(k) = \sum_{n=0}^{N_{FFT}-1} x_m(n) e^{-j2\pi kn/N}, \quad k=0,1,\ldots,N_{FFT}/2-1; \quad m=1,2,\ldots,M
\]
(2)

Then the frequency power of each frame is passed through a \( Q \)-order Mel filter bank \( \Psi_{Q,q=1,2,\ldots,Q} \), leading to
\[
y_m(q) = \sum_{k=0}^{N_{FFT}/2-1} |Y_m(k)|^2 \psi_q(k), \quad q=1,2,\ldots,Q
\]
(3)
where the employed Mel filter bank includes \( Q \) triangular filters, uniformly covering the whole Mel frequency range, with 50% overlapping. The order of the Mel filter bank is set as \( Q=24 \) in this paper. The Mel frequency is defined as
\[
f_{\text{Mel}}(f) = 2959 \times \log_{10} \left(1 + f / 700 \right). \quad f \sim [0, f_s/2]
\]
(4)
Finally, for the \( m \)-th frame the MFCCs are calculated by the discrete cosine transform (DCT) of the logarithm of \( y_m=[y_m(1), y_m(2), \ldots, y_m(Q)]^T \), i.e.
\[
e_m(p) = \sqrt{\frac{2}{Q}} \sum_{l=0}^{Q-1} \log \left[ y_m(l+1) \right] \cos \left[ p \frac{2l+1}{2} \frac{\pi}{Q} \right], \quad p=1,2,\ldots,P
\]
(5)

Since the thunder components are generally concentrated in the relatively low frequency region, in this paper we choose \( P=12 \), i.e. deriving the first 12-order MFCCs for each frame.

Now for each data segment we have a \( P \times M \) MFCC matrix \( \mathbf{F} \), by which we will further build a 3-channel tensor as the final feature map instead of directly using it. The 3-channel tensor consists of the MFCC itself, the 1st-order differences and the 2nd-order differences of MFCC. Therefore, the 1st, 2nd, and 3rd channel of the tensor feature map are given by \( \Delta_1=\mathbf{F}[:,1:M-2] \), \( \Delta_2=\mathbf{F}[:,2:M-1]-\mathbf{F}[:,1:M-2] \) and \( \Delta_3=\mathbf{F}[:,3:M]-\mathbf{F}[:,2:M-1] \) respectively. The purpose of calculating differences in the tensor is to exploit the time-variant information of thunder signals. To satisfy generality, the tensor is normalized per-column and per-channel.

2.3. Training Method of CNN

In practical application, the acoustic thunder signals have great diversity, and the recordings contain a variety of environmental interferences inevitably. To overcome the challenge of recognition of thunder signals and improve the detection probability of thunder signals, CNN with temporal/spatial translation invariance property is used in this paper. To achieve fast thunder detection, a four-layer CNN is established, in which the parameters will be trained by a built thunder database. Table 1 shows the structure of the applied CNN.

The tensor feature map, derived from each data segment, with the size of \( 12 \times 184 \times 3 \), is used as the input to the established CNN. Some terminologies in table 1 is illustrated as follows: (1) “pad” tells whether the input current layer is padded with zeros to make the size of input and that of output be the same; (2) “SAME” and “VALID” denote padding and no padding respectively; (3) “stride” gives the sliding length of the convolution kernel when performing convolution between the kernel and input data; (4) “Drop-Out” shows the probability of keeping each element. The size of features will be \( 3 \times 46 \times 8 \) through all the convolutional layers and pooling layers, and then a dense layer stretches the tensor to a vector. After a fully connection layer connecting the vector output of dense layer to a 2-
entry output vector evaluated by Softmax, the probability of thunder detection will be obtained. For an input vector \( s \), the Softmax function is given by 
\[
\sigma(s) = \frac{\exp(s)}{\sum_l \exp(s)}
\]
where \( l \) is an all-one vector.

The ultimate output vector is displayed as 
\[
p = [p_E, p_I]^T
\]

**Table 1.** The structure of CNN.

| Layer Type                  | Input Tensor Size |
|-----------------------------|-------------------|
| 5x5 Conv(pad-SAME,stride-1) | 12x184x3          |
| 5x5 Conv(pad-SAME,stride-1) |                   |
| 2x2 Max-Pooling(VALID,stride-2) |              |
| 5x5 Conv(pad-SAME,stride-1) |                   |
| 5x5 Conv(pad-SAME,stride-1) |                   |
| 2x2 Max-Pooling(VALID,stride-2) |            |
| Dense(1104) + Drop-Out(0.5) |                  |
| Fully Connection + 2-way Softmax |                |

The CNN described above is trained under the TensorFlow framework. We select Adam as the optimizer, set the learning rate to be 0.001, and use cross-entropy as loss function. The one-hot code is used in customizing the label corresponding to each data segment. \([1,0]^T\) and \([0,1]^T\) denote the labels corresponding to existence and inexistence of thunder in the current data segment respectively.

The detection performance will be evaluated using three indexes: accuracy, sensitivity, and specificity, defined by the formulas as followed:

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}
\]

\[
\text{Sensitivity} = \frac{TP}{TP + FP}
\]

\[
\text{Specificity} = \frac{TN}{FP + TN}
\]

where \( TP, TN, FP, \) and \( FN \) denote the number of true positive, true negative, false positive, and false negative cases in detection results respectively.

3. Experimental Results

3.1. Instrumentation and Database

From 2011 to 2020, we conducted lightning field observation experiments in Wuhan of Hubei Province. Two microphone arrays are used to observe lightning flashes. The microphone array is a three-dimensional structure which is made up of 19 microphones with a frequency response from 20 Hz to 100 kHz. As shown in figure 1, 7 microphones are installed in the vertical bar of the steel structure while 12 microphones are distributed on the horizontal bar symmetrically. The microphones are connected to the central logger through the shielded cable, sampling at 50 kHz.
Abundant of thunder data produced by the intra-cloud lightning and cloud-to-ground lightning has been collected since initial installation. We have obtained over 1800 thunder signals including different kinds of thunder signal such as clap, peal, roll and rumble. Meanwhile, other sound signals such as the sound of rain, whistle, music, people and crash are observed by our microphone array. To improve the robustness of CNN, we included other confusing signals such as the firecracker and gunshot that are similar to thunder signals and some environmental noises such as the car-honking and the sound of rainy and windy in the training and testing datasets.

3.2. Analysis of Typical Case
The acoustic thunder signals can be classified into clap thunder and rumble thunder from the perspective of whether the sound being crisp or not. The analysis will be given by typical cases of thunder waveform, where short time Fourier transform (STFT) spectrum and MFCC spectrum will also be provided.

3.2.1. Clap Thunder. The typical clap thunder is shown in figure 1, in which the waveform, the STFT spectrogram and the MFCC spectrogram are shown in figure 2(a), figure 2(b), and figure 2(c) respectively. It can be observed from the STFT spectrogram that the clap thunder signal consists of some high-frequency components almost over 200 Hz, and these components are more focused on some frames. It can be manifested from the normalized MFCC spectrogram that the energy of clap thunder is most concentrated on the first 3 orders of MFCCs.

3.2.2. Rumble Thunder. Another data segment example contains typical rumble thunder. The waveform, STFT spectrogram and MFCC spectrogram are displayed in figure 3(a), figure 3(b) and figure 3(c) respectively. It can be discovered from the STFT and MFCC spectrograms that the relatively higher frequency components corresponding to thunder are more diffuse in the time domain than those in clap thunder. The phenomenon can be more easily displayed through MFCC than STFT, which also shows the benefits in using MFCC features for thunder detection.
3.3. Evaluation of Detection Performance

Restricted by the size of the data set, in this paper we concern on thunder detection, i.e. existence or in-existence of thunder in a data segment to be examined. Specifically, to improve the robustness of thunder detection by the CNN, some recordings of artificial percussion noises that imitate thunder signal are fused into the data set.

The data set consists of 1409 thunder segments and 2606 ambient noise segments, where 196 artificial percussion noise segments are included in the negative sample set. In the network training procedure, the input data is randomly shuffled before being entered into the network, to create a more scattered distribution of the thunder signal data set. 80% of the data set is used as the training set and the rest 20% is used as the test set. The number of network iterations is set to be 5000. As the 2-entry output vector \( p = [p_E, p_I] \) contains the estimated probability of thunder existence \( p_E \) and the estimated probability of thunder in-existence \( p_I \), we consider the thunder signal being detected when \( p_E > 0.95 \).

Figure 4 shows the curve trend graphs of accuracy and loss with the number of iterations during the training process. The final indexes evaluating the thunder detection performance on test set are shown in table 2. It can be found that the specificity can still reach 0.91 even while there are various types of thunder signals as well as other interferential signals.

|                | Accuracy | Sensitivity | Specificity |
|----------------|----------|-------------|-------------|
| Test results   | 0.87     | 0.79        | 0.91        |

After training, a feed-forward network depicted in table 1 can be formed using the learning parameters. Applying this established CNN can realize segment-wise thunder detection in real scenarios. Figure 5 displays an example of thunder detection based on the proposed method. There exists a rumbling thunder in the front portion of the recording with a duration of 26 s while a clapping thunder happened in the intermediate portion, as shown in figure 5(a). From the STFT spectrogram in

---

**Figure 3.** A typical rumble thunder recording: (a) the waveform; (b) the STFT spectrogram; (c) the MFCC spectrogram.

**Figure 4.** The curve trend graphs in training: (a) accuracy; (b) loss.
figure 5 (b), it can be observed that the energy of clapping thunder is less diffuse in time that that of rumbling thunder. The same recording is identified by two different methods. One is the current general method based on the endpoint detection and feature extraction of frequency spectrum, the other is the proposed method in our paper. It is observed that the clapping thunder is recognized while the rumbling thunder is not recognized by the current method as shown in figure 5(c). Nevertheless, the proposed method can detect not only the clapping thunder but also the rumbling thunder as shown in figure 5(d). There may be two reasons: (1) the waveform of the rumbling thunder in the time domain is more divergent which results in hardly any distinguishing feature in the frequency spectrum; (2) the rumbling thunder mostly generated by intra-cloud lightning or remote lightning is more interfered by other environmental noises. As can be seen from the comparison between the two methods, the proposed CNN-based thunder signal detection method performs better especially for the rumbling thunder and in the environment with more ambient interference.

**Figure 5.** An example of thunder detection based on the proposed method: (a) STFT spectrogram of the recording; (b) original labels marked by blocks; (c) detection results marked by blocks.

**4. Summary and Conclusions**

In this study, we applied the deep learning method in recognizing the thunder signal. The recorded signal is segment-wise preprocessed. In each frame, MFCC is applied to extract the features of the thunder signal. We then use the MFCC features derived in each frame to form a 3-channel tensor data, which is used as the further input to the designed CNN. The goal of CNN is to classify the existence of thunder for a single data frame. In this paper, the database was made up by 1409 thunder segments and 2606 ambient noise segments. To improve the robustness of CNN, we included other confusing signals in the training and testing datasets. On the testing dataset, our proposed method outperforms the current methods with splendid assessment result of accuracy=0.87, sensitivity=0.79, and specificity=0.91. Moreover, the recognition results by the proposed method are compared with the method based on the endpoint detection and feature extraction of frequency spectrum. The contrast result indicates that our proposed deep-learning-based thunder detection method has better robustness for the rumbling thunder. Therefore, the proposed method not only increases the real-time
performance of the lightning location system with thunder signals but also improves the accuracy of other sound alarm systems.

This work is an attempt at achieving the thunder signal detection by deep learning. It is well known that deep learning is most profitable while applied to large amounts of training datasets. To improve deep learning performance significantly in computer vision a database of over 14 million hand-labeled pictures was a major element. Therefore, future work aims to accumulate more thunder data to improve the robustness of the recognition methods. Furthermore, there are other features of thunder signals and various deep learning based classification frameworks to recognize and classify thunder signals. We will also try to compare various classification models and selectively combine them to further improve thunder signal detection performance.

Acknowledgments
This work was supported by National Natural Science Foundation of China(Grant No. 51777037), Science and Technology Program of State Grid Corporation of China(Grant No. 524625200017) and National key Research and development program(Grant No. 2016YFB1200403).

References
[1] Rakov V A, Uman M A 2003 Lightning: Physics and Effects (UK: Cambridge University Press)
[2] Few A A 1970 Lightning channel reconstruction from thunder measurement J. Geophys. Rse. 75 7515-23
[3] Arechiga R O, Johnson J B, and Edens H E 2011 Acoustic localization of triggered lightning J. Geophys. Res. 116 1-11
[4] Qiu S, Zhou B H and Shi L H 2012 Synchronized observations of cloud-to-ground lightning using VHF broadband interferometer and acoustic arrays J. Geophys. Rse. 117 1-9
[5] Bodhika J A P, Dharmaratna W G D, Fernando M and Cooray V 2013 Reconstruction of lightning channel geometry by localizing thunder sources J. Atmos. Sol-terr 102 81-90
[6] Dayeh M A et al. 2015 First images of thunder: Acoustic imaging of triggered lightning Geophys. Res. Lett. 42 6051-7
[7] Zhang H, Zhao C, Gu S Q, Feng W X, Wang H T, Lu E Z, Guo J T, Tao H T 2016 Lightning imaging with thunder using broadband direction-of-arrival estimation technique International Conference on Lightning Protection, IEEE
[8] Zhang H, Gu S Q, Chen J H, Zhao C, Wu M, and Yan B W 2019 Single-Station-Based Lightning Mapping System With Electromagnetic and Thunder Signals IEEE Transactions on Plasma Science 2019 1-8
[9] Latham D J 1964 A study of thunder from close lightning discharges M.S. Thesis (New Mexico: New Mexico Institute of Mining and Technology)
[10] Depasse P 1994 Lightning acoustic signature J. Geophys. Res. 99 25933-40
[11] Ouyang Y H and Yuan P 2012 Audible thunder characteristic and the relation between peak frequency and lightning parameters J. Earth Syst. Sci. 121 211-20
[12] Wei K J, Li H R, Sun X J and Zhang C L 2015 A Method for Real-time Recognition of Thunder Research and Exploration in Laboratory 34 27-30
[13] Hinton G et al. 2012 Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups IEEE Signal Process. Mag. 29 82-97
[14] Purwins H, Li B, Virtanen T, Schluter J, Chang S Y and Sainath T N 2019 Deep Learning for Audio Signal Processing IEEE Journal of Selected Topics in Signal Processing 2019 1-1