Chinese Journal of Electronics
vol. 33, no. 3, pp. 793–802, May 2024
https://doi.org/10.23919/cje.2022.00.210

RESEARCH ARTICLE

Long Short-Term Memory Spiking Neural Networks for Classification of Snoring and Non-snoring Sound Events

Rulin ZHANG, Ruixue LI, Jiakai LIANG, Keqiang YUE, Wenjun LI, and Yilin LI

Electronics and Information College, Hangzhou Dianzi University, Hangzhou 310018, China

Corresponding author: Wenjun LI, Email: liwenjun@hdu.edu.cn
Manuscript Received July 10, 2022; Accepted June 21, 2023

Abstract — Snoring is a widespread occurrence that impacts human sleep quality. It is also one of the earliest symptoms of many sleep disorders. Snoring is accurately detected, making further screening and diagnosis of sleep problems easier. Snoring is frequently ignored because of its underrated and costly detection costs. As a result, this research offered an alternative method for snoring detection based on a long short-term memory based spiking neural network (LSTM-SNN) that is appropriate for large-scale home detection for snoring. We designed acquisition equipment to collect the sleep recordings of 54 subjects and constructed the sleep sound database in the home environment. And Mel frequency cepstral coefficients (MFCCs) were extracted from these sound signals and encoded into spike trains by a threshold encoding approach. They were classified automatically as non-snoring or snoring sounds by our LSTM-SNN model. We used the backpropagation algorithm based on an alternative gradient in the LSTM-SNN to complete the parameter update. The categorization percentage reached an impressive 93.4%, accompanied by a remarkable 36.9% reduction in computer power compared to the regular LSTM model.
Keywords — Snoring detect, Mel frequency cepstral coefficients, Spiking encoding, Long short-term memory spiking neural networks, Backpropagation based on alternative gradient.
Citation — Rulin ZHANG, Ruixue LI, Jiakai LIANG, *et al.*, "Long Short-Term Memory Spiking Neural Networks for Classification of Snoring and Non-snoring Sound Events," *Chinese Journal of Electronics*, vol. 33, no. 3, pp. 793–802, 2024. doi: 10.23919/cje.2022.00.210.

I. Introduction

Snoring is one of the most typical sleep problems in humans. In daily life, the likelihood of snoring during sleep ranges from 13% to 42.4% [1]. Because snoring is so prevalent and has little impact in the short term, it has yet to be taken seriously. Snoring can lower the amount of oxygen in the blood, which might mainly contribute to health issues, including heart disease, high blood pressure, stroke, diabetes, and so on [2], [3]. In addition, snoring is one of the earliest and most common symptoms of many sleep disorders, such as obstructive sleep apnea hypopnea syndrome (OSAHS) [4], which has a high prevalence and affects people of all ages. The global prevalence of OSAHS is between 2% and 10%, with an increased tendency [5]. At night, patients with OSAHS may feel shortness of breath and hypoxia. Severe instances can result in hypoxemia and hypercapnia and even lead to sudden death [6].

Polysomnography (PSG) is now the most reliable diagnostic tool for detecting sleep problems. It has various sensors which monitor multiple physiological indicators of sleep, such as respiratory airflow, electrocardiogram, electroencephalogram, and so on [7]. However, there are certain drawbacks to adopting PSG detection. On the one hand, PSG is time-consuming, costly, and requires experienced personnel. It is not popularized in each hospital [8]. As a result, some patients missed the best timing of treatment. On the other hand, when a patient wears a variety of sensors for long-term monitoring during the recording by PSG, the patient's sleep quality would undoubtedly be altered, altering the measurement data and the diagnostic results. Therefore, the development of more efficient, cost-effective, and non-contact approaches for screening and diagnosing sleep disorders is of primary scientific importance.

Associate Editor: Prof. De-Chuan ZHAN, Nanjing University.

When employing non-contact methods to identify sleep issues, an auxiliary diagnostic method based on snoring has naturally become the most straightforward and successful approach. First, the snoring sound has a simple recording procedure as an acoustic signal [9] and entirely fits the non-contact standards. Second, there is the fact that snoring is a prevalent symptom of a variety of sleep problems [10]. Such as nearly 75% of sleep snorers suffer from OSAHS [11]. Moreover, research [12] showed that acoustic parameters of snoring noises changed significantly between regular snorers and OSAHS patients. By examining the acoustic features of snoring, it is possible to detect and assess the severity of OSAHS [13]–[15]. Hence, the ability to distinguish massive snoring sound episodes from raw respiratory sounds quickly and accurately has become a sufficient and necessary requirement.

Due to the prevalence of sleep problems, snoring detection at home is an important study area. Many researchers [16]–[21] constructed neural networks to detect snoring. However, traditional neural networks have a vast number of parameters, making it challenging to directly transplant the trained model to embedded devices for completing local detection. Some studies [22] used mobile phones to collect sound to accomplish localization. Nevertheless, it was still necessary to upload the sound data to the server for calculation and then return the detection results to the mobile phone. This technique of transmitting sound data to the server for detection required a large amount of storage and computational resources, and it did not fully fulfill home detection localization. One approach to transferring the algorithm from the server to the local side was to create a network model with fewer parameters, allowing to transfer of the trained model to the integrated device immediately. This paper employed an automatic snore detection model based on a LSTM-SNN. The model inherited the accuracy of the LSTM networks in sound recognition and had the characteristics of SNN receiving spiking sequences as input for information transmission and processing and sharing spiking instead of floating point values. So the LSTM-SNN reduced the networks' computational parameters and provided a solution for porting the algorithm to integrated devices without losing accuracy.

The content of this paper is as follows: Section I introduces the research background. Section II describes the research status of snoring detection. Section III discusses data collection and model architecture. Section IV details the design of the specific parameters in the experiment and the analysis of the experimental results. Section V presents the conclusion and discussion.

II. Related Work

Many academics have presented various approaches to identify snoring occurrences automatically in recent years. For example, Bruno Arsenali *et al.* [8] worked with the Center for Sleep Medicine to gather sleep sounds from 20 people who had standard PSG reports. They used Mel frequency cepstral coefficient (MFCC) and a recurrent neural network (RNN) to identify snoring and non-snoring events in binary, which obtained an accuracy of 95%. Smartphones were utilized by Lim *et al.* [16] to capture the individuals' sleep sound data. They extracted several characteristics from the sound data and employed RNN to identify snoring. On their dataset, their model has a high accuracy rate of 98.9%. Nguyen *et al.* [17] used a microphone worn around the neck to capture nocturnal sleep sound data from 15 subjects with snoring complaints. They proposed a multilayer perceptron neural network with a correlation filter (f-MLP) and acquired an average detection rate of 96%. Jiang *et al.* [18] teamed up with a hospital to collect sleep sound data using a microphone. They extracted five spectrogram features from audio data. They also created two models, convolutional neural networks (CNNs)-deep neural networks (DNNs) and CNNs-LSTMs-DNNs, for combination experiments. Khan [19] extracted MFCC spectrograms as input features from different online public datasets and constructed CNN to categorize snoring and non-snoring occurrences. Based on the constant-q transform (CQT), Xie *et al.* [20] employed the SOMNIN [21] to create a 3-layer CNN, 1-layer LSTM for snoring sound classification. The model's categorization accuracy was more than 94%.

These researchers produced positive findings in their experiments but did not consider snoring data in the home environment. Sleep disorders are a common occurrence, as evidenced by several statistics. Therefore, snoring detection at home is critical for screening and diagnosing sleep problems and may become a future trend. Sound data gathered in a home environment can contain more complicated background noise than data collected in a professional sleep laboratory or quiet hospital. It necessitates a model with more generality and stability. In addition, designing a hardware-friendly model for the home environment is critical, as the model could be readily translated to the edge device and applied to the home. Using CNNs and RNNs with high parameters is not an ideal option.

For the two problems mentioned above, first, our team designed a convenient collection device for proper storage and uploading for sleep sound recording in a home environment and constructed a database of snoring clips in daily life. Second, spiking neural network (SNN) offers the advantage of having fewer parameters and consuming less energy. Data-driven learning is used in current DNN models, which demand many processing resources. SNN relies on event-based spiking computational units for learning and computation. It is more energy and resource-efficient because of its temporal sparsity. As a result, we created a low-energy LSTM-based spiking neural network to detect snoring and non-snoring events automatically. In the end, we achieved a recognition effect comparable to traditional neural networks, providing an alternative solution for sleep sound acquisition and snoring detection in a home environment.

III. Dataset and Methods

This section describes the specifics of the whole architecture of the LSTM-SNN system used in our study. The general system framework is shown in Figure 1. It consists of four parts: dataset generation, feature extraction, spike encoding, and model building. The model-building part includes spiking neural units and backpropagation methods based on alternative gradients. Details of all these four parts are described in the following subsections.

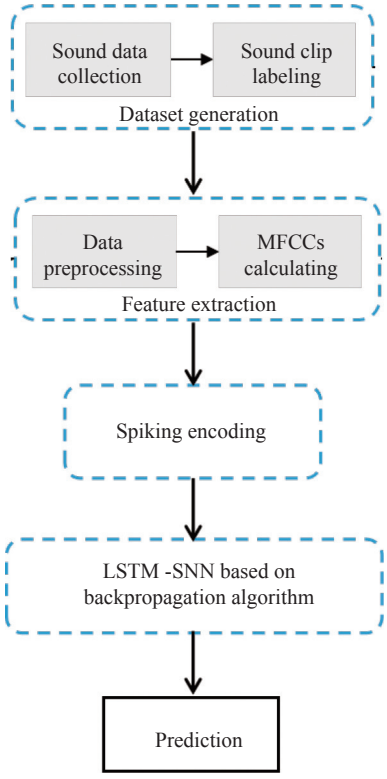


Figure 1 Framework of the LSTM-SNN algorithm in this work.

1. Dataset acquisition and annotation

In our experiment, we cooperated with Hangzhou Normal University Affiliated Hospital to record subjects' overnight sleep sound with their consent. In order to adapt to the home environment, we designed acquisition equipment to record sound data. The primary main control module of this acquisition equipment was NXP's Cortex-A7 architecture I.MX6ULL. The I.MX6ULL microcontroller unit worked at 800 MHz and easily supported the embedded Linux system. Functional modules included audio acquisition, wireless communication, storage, and power supply modules. The audio acquisition module was responsible for sleep audio recording and quantization coding. A high-resolution microphone (NIS-80V, FengHuo Electronic Technology Co., Ltd, Guangdong, China; 20–2000 Hz frequency range, −45 dB sensitivity) was used to record sleep sounds. It was connected to the audio code chip to encode the sound signal digitally. The audio data was sent online to the server through 4G wireless transmission or saved locally on an SD card. Furthermore, the equipment used the Type C interface for the wired power supply. A voltage regulator circuit was included inside the power supply module to provide the needed working voltage for each module.

In our experiment, we used a portable PSG and our acquisition equipment to collect physiological data synchronously during the nocturnal sleep of subjects. PSG was used to diagnose subjects with sleep disturbances, ensuring the diversity of sleep sounds in our dataset. Our acquisition equipment was installed at the bedside to collect respiratory sounds while sleeping. The distance from the device to the subject should be kept within 3 m, with 1.5 m being an ideal limit. The device used a 16 kHz sampling frequency, 16-bit sample bits, and monophonic sound to record sleep sounds. We formatted each piece of audio data in a fixed-size format of 100 MBytes. (each audio duration was about 50 minutes). We collected audio data from 54 subjects aged 16 to 58, comprising 41 males and 13 females, and included 21 usual snorers and 33 individuals with various degrees of OSAHS. Each individual was monitored for 7 hours each night on average. Table 1 includes all of these individuals' information.

Table 1 Subject demographic information

Features	Normal	OSAHS
No. of subjects	21	33
Age (year)	35.10±12.51	45.91±14.01
BMI (kg/m^2)	24.43±3.52	26.90±3.13
AHI (times/hour)	2.00±1.05	25.47±15.77

Note: BMI: body mass index, calculated by the weight and hight; AHI: apnea hypopnea index, which clinically is used for OSAHS diagnosis.

This study aimed to automatically categorize snoring and non-snoring events in recorded sleep audio. Thus, we needed to create a dataset that included both snoring and non-snoring incidents. Figure 2 depicts how we processed raw sleeping sound data. First, we applied the Wiener filter algorithm [23] to denoise the raw data. We can see the reduction of burrs on the audio waveform after denoise processing. Then, we utilized a dual-threshold endpoint detection algorithm based on short-term energy and short-term zero-crossing rate [24] to recognize voiced segments in raw sleep sounds. We marked the sound segments that will be intercepted with vertical lines of the same color (red or green) in Figure 2. Finally, experienced doctors marked snoring and non-snoring events. Our experiment generated a dataset containing 6052 snoring clips, including basic snorers and OSAHS sufferers of various severity, and 6052 non-snoring clips, containing various ambient noises such as speaking, phone ringing, coughing, and other sounds.

2. Feature extraction

Mel frequency cepstral coefficient (MFCC) [25] is a nonlinear characteristic based on human hearing. It has

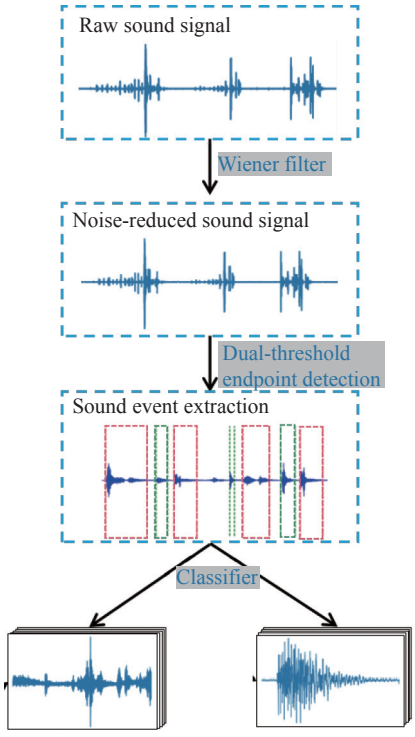


Figure 2 Process of labeling for raw audio data.

been widely used in audio recognition studies, including snoring recognition [22], and it also performed well for audio signals with a lower signal-to-noise ratio [26]. In our experiment, we calculated MFCCs as input characteristics for each sample. For a sound signal $x(t)$, the MFCCs calculation process is as follows:

Step 1 Data preprocessing including pre-emphasis, framing, and windowing. Pre-emphasis was passing the speech signal through a high-pass filter. Then the signal was separated into 2040 ms per frame, with overlapping between two consecutive frames to minimize significant changes between them. Per frame passed a window function, such as a Hamming window, reduced spectral leakage.

Step 2 Calculated the fast Fourier transform (FFT) of each frame to obtain the spectrum.

$$X(k)=\sum_{n=0}^{N-1}x(n)e^{-\mathrm{j}2\pi nk/N},k=0,1,\ldots,N-1 \quad (1)$$

where k is the spectral line (frequency component), n is the sampling point, and N is the total number of sampling points.

Step 3 For each frame, calculated the power spectrum and passed through a set of Mel-scale triangular filters, which could obtain the energy value of the signal. It was achieved by

$$s(m)=\ln\left(\sum_{k=0}^{N-1}|X(k)^2H_m(k)|\right),0\le m\le M \quad (2)$$

where $H(k)$ is the frequency response of triangle filters, m denotes the mth filter, and M is the number of Mel-scale triangular filters.

Step 4 The discrete cosine transform (DCT) was applied to decorrelate the filter bank coefficients.

$$C(l)=\sum_{m=0}^{M}s(m)\cos\left(\frac{\pi l(m-0.5)}{M}\right),l=1,2,\ldots,L \quad (3)$$

where L is the order of MFCCs.

In this work, we divided the sound sample into 32 ms per frame, set the number of FFT points to 512, and chose the number of Mel-scale triangular filters in the filter bank to be 26. The number of MFCCs was set to 39, including 13-dimensional MFCCs, 13-dimensional first-order difference coefficients, and 13-dimensional second-order difference coefficients. The first dimension coefficients were replaced with logarithmic energy values. Further, we normalized the feature matrix. Then the feature matrix size of all samples was unified 39×280. We called the audio processing library librosa to calculate MFCCs. Figure 3 plots a non-snoring class sample and a snoring class sample in the time domain and its MFCC representation.

3. Model

1) Spike encoding

A spiking neural network is a type of neural network replicating biological neuron cells fed a series of spikes as input. It is required to encode the input features into discrete pulse sequences to create an effective spiking neural network [27]. The collected MFCC features were encoded in this paper using a threshold coding approach based on Gaussian distribution [28]. We created a MFCCs spectrum for all samples throughout the feature extraction method in section III-2. Then we generated a threshold matrix comprising values to cover the MFCCs spectrum. The Gaussian distribution of this threshold matrix had a mean of 0 and a variance of 1. We compared each sample's MFCCs spectrogram to the threshold matrix as (4). When the sample's MFCCs spectrum value exceeded the encoding threshold, the characteristic data was encoded as a spike; otherwise, it was encoded as null.

$$S_{ij}(l)=\begin{cases}0, & C_{ij}(l)\le\mu_{ij}\\ 1, & C_{ij}(l)>\mu_{ij}\end{cases} \quad (4)$$

where μ_{ij} denotes the threshold at the (i,j) position of the threshold matrix. As a result, the MFCC spectrogram was converted into the 0-1 matrix. Figure 4 depicts the threshold encoding procedure.

2) LSTM spiking neural networks

Long short-term memory network (LSTM) [29] is a particular case of RNN, which has the outer loop of RNN and an intracellular self-loop. It works well with time series signals [30]. The classification model in this paper adopted the LSTM-SNN [28] that leverage the LSTM capability of learning temporal dependencies and the SNN advantages of energy saving. As shown in Figure 5, the

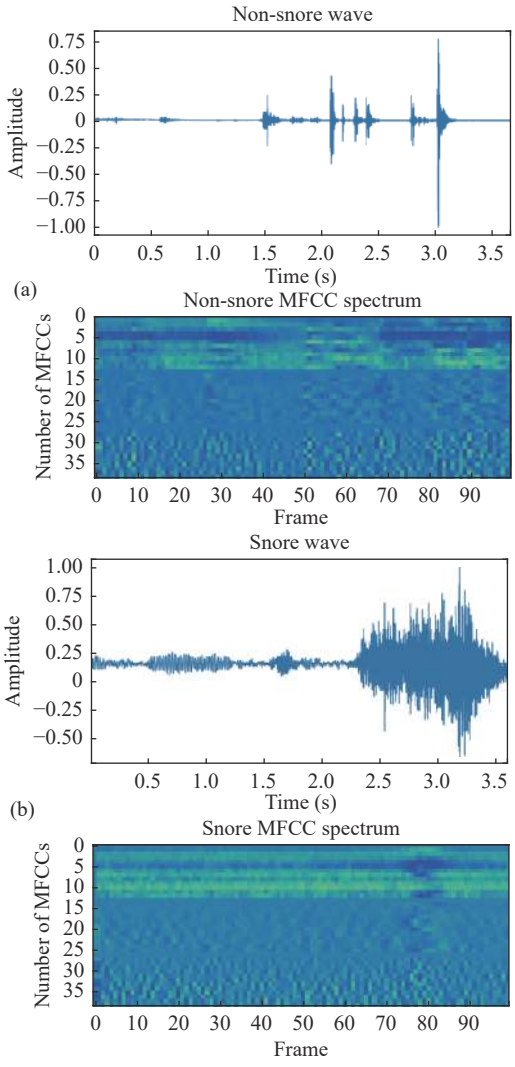


Figure 3 Sound signals and their corresponding MFCC spectrum. (a) The non-snoring sound; (b) The snoring sound.

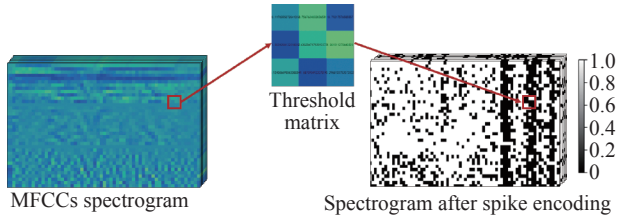


Figure 4 Process of the threshold coding.

LSTM spiking cell we used, like the typical LSTM cell, has three crucial gate structures: the input gate checks the information entering the cell, the forget gate helps to discard superfluous information, and the output gate produces the cell's result. The difference is that the LSTM spiking cell used two threshold activation functions $\sigma_1(\cdot)$ and $\sigma_1(\cdot)$ instead of sigmoid and tanh in the conventional LSTM. These activation functions determine the output of spikes or nulls at each time step in their respective gates. The key of such an LSTM spiking unit, like ordinary LSTMs, is the unit state which acts as

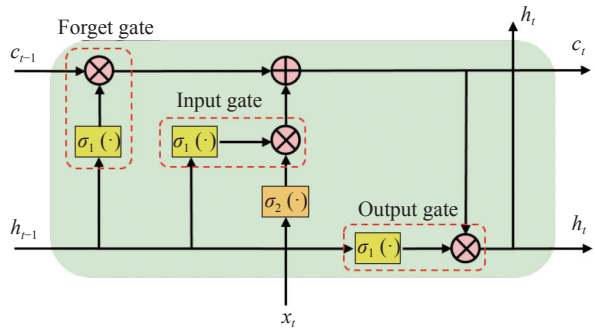


Figure 5 The LSTM spiking unit.

a conduit and manager of information flow between units [28]. More specifically, given a set of spiking inputs $\{x_1, x_2, \ldots, x_T\}$, the gates and states are characterized as follows:

$$\begin{aligned}
\boldsymbol{f}_t &= \sigma_1(\omega_{f,h}\boldsymbol{h}_{t-1} + \omega_{f,x}\boldsymbol{x}_{t-1} + \boldsymbol{b}_{f,h} + \boldsymbol{b}_{f,x}) \\
\boldsymbol{i}_t &= \sigma_1(\omega_{i,h}\boldsymbol{h}_{t-1} + \omega_{i,x}\boldsymbol{x}_{t-1} + \boldsymbol{b}_{i,h} + \boldsymbol{b}_{i,x}) \\
\boldsymbol{g}_t &= \sigma_2(\omega_{g,h}\boldsymbol{h}_{t-1} + \omega_{g,x}\boldsymbol{x}_{t-1} + \boldsymbol{b}_{g,h} + \boldsymbol{b}_{g,x}) \\
\boldsymbol{c}_t &= \boldsymbol{f}_t \odot \boldsymbol{c}_{t-1} + \boldsymbol{i}_t \odot \boldsymbol{g}_t \\
\boldsymbol{o}_t &= \sigma_1(\omega_{o,h}\boldsymbol{h}_{t-1} + \omega_{o,x}\boldsymbol{x}_t + \boldsymbol{b}_{o,h} + \boldsymbol{b}_{o,x}) \\
\boldsymbol{h}_t &= \boldsymbol{o}_t \odot \boldsymbol{c}_t
\end{aligned} \tag{5}$$

where $\boldsymbol{f}_t$ is forget gate layer, $\boldsymbol{i}_t$ is input gate layer, $\boldsymbol{g}_t$ is modulated input, $\boldsymbol{c}_t$ is unit state, $\boldsymbol{o}_t$ is output gate layer, $\boldsymbol{g}_t$ is hidden state, $\odot$ represents the Hadamard product, $\sigma_1(\cdot)$ and $\sigma_2(\cdot)$ are threshold activation functions that can map their input to a spike if it exceeds the threshold value θ_1 and θ_2, respectively. Both thresholds were set to 0.1 in our experiments, as shown in (6). Notice that the unit state ($\boldsymbol{c}_t = \boldsymbol{f}_t \odot \boldsymbol{c}_{t-1} + \boldsymbol{i}_t \odot \boldsymbol{g}_t$) can take the values 0, 1, or 2. Since the gradients around 2 are not as informative, we threshold this output to output 1 when it is 1 or 2.

$$\sigma_1(u) = \sigma_2(u) = \begin{cases} 0, & u \le 0.1 \\ 1, & u > 0.1 \end{cases} \tag{6}$$

3) Backpropagation based on alternative gradient

Most models employ the gradient descent approach to update parameters during backpropagation, either directly or indirectly. Gradient descent is used to minimize the loss function. Equation (7) approximately expresses the relationship between the model loss value L and the weight $\boldsymbol{\omega}$ in the LSTM unit:

$$\frac{\partial L}{\partial \boldsymbol{\omega}} = \frac{\partial L}{\partial \sigma} \cdot \frac{\partial \sigma}{\partial u} \cdot \frac{\partial u}{\partial \boldsymbol{\omega}} \tag{7}$$

The activation process of the LSTM spiking unit described above was represented as a step function with an infinite derivative at the threshold and zero at all other points. That is to say, $\frac{\partial \sigma}{\partial u}$ cannot be directly calculated in the LSTM spiking unit. As a result, we could not use gradient descent methods to optimize the LSTM-SNN directly. Recently, alternative gradient methods [31]–[33]

were proposed to deal with non-differentiable pulse sequences. They tried to update the gradient information based on various approximation functions of the step function using the derivative of the approximate function as a replacement gradient without modifying the spiking neuron's activation process. The study by Lotfi Rezaabad *et al.* [28] found that the choice of the alternative derivative function is preferable to a Gaussian distribution with suitable variance. Based on the derivatives of the sigmoid and tanh functions employed in standard LSTM cells, we looked for other alternative functions, including the Gaussian function, Gaussian error function, and fast sigmoid function, as shown in Figure 6. In this work, we concluded the best result was achieved by replacing $\frac{\partial \sigma_1}{\partial u}$ with a Gaussian distribution $G_1(u)$ with mean 0 and variance 4, and $\frac{\partial \sigma_2}{\partial u}$ with a Gaussian distribution $G_2(u)$ with mean 0 and variance 0.3. Therefore, the relationship between the model loss value L and the weight ω during the back-propagation process of the LSTM-SNN model can be expressed as

$$\frac{\partial L}{\partial \omega} = \frac{\partial L}{\partial \sigma} \cdot \frac{\partial \sigma}{\partial u} \cdot \frac{\partial u}{\partial \boldsymbol{\omega}} \approx \frac{\partial L}{\partial \sigma} \cdot G(\boldsymbol{u}) \cdot \frac{\partial u}{\partial \boldsymbol{\omega}} \tag{8}$$

4) Evaluation metrics

In this paper, the performance of the model was evaluated with multiple metrics, including accuracy, sensitivity, specificity, precision, and the F1 score. The calculation formulas are shown as follows:

$$\begin{aligned}
\text{Accuracy} &= \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \\
\text{Sensitivity} &= \frac{\text{TP}}{\text{TP}+\text{FN}} \\
\text{Specificity} &= \frac{\text{TN}}{\text{TN}+\text{FP}} \\
\text{Precision} &= \frac{\text{TP}}{\text{TP}+\text{FP}} \\
\text{F1}_{\text{score}} &= 2\frac{\text{Precision}\cdot\text{Sensitivity}}{\text{Precision}+\text{Sensitivity}}
\end{aligned} \tag{9}$$

where TP, TN, FP, and FN represent the number of the true positives, true negatives, false positives, and false negatives, respectively.

IV. Experiment and Results

1. Experience setups

We separated the annotated sound clips into three sets in our experiment: a training set, a validation set, and a test set, which accounted for 60%, 20%, and 20% of the total, respectively. The training and validation sets data came from 44 subjects, while the test set came from another ten subjects. Table 2 shows the comprehensive details of the sound data. We used the validation set in fine-tuning the parameters to enhance the evaluation results. Once we obtained the best parameters from the training and validation set, we used the test set to validate the model's performance on the snoring detection task.

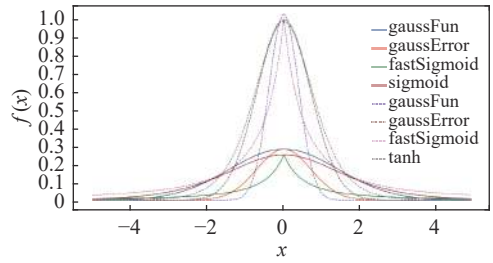


Figure 6 Derivatives for multiple alternative activation functions.

Table 2 The partition of our dataset

Name of dateset	Training set	Validation set	Test set
Snoring	3632	1210	1210
Non-snoring	3632	1210	1210
Number of subjects	34	10	10

The experiment constructed a one-layer LSTM spiking neural network with 70 spiking units. Each unit had an input size of 39×4 with one hidden layer of size 1000. Activation function $\sigma_1(\cdot)$ and $\sigma_2(\cdot)$ activation thresholds were set to 0.1. The output layer used softmax to connect to the last spike unit, $y_t = \text{softmax}(w_y \times h_t + b_y)$. Figure 7 shows the network structure. We illustrated the spike raster plots shown as Figure 8. In detail, we adopted the random threshold method to encode MFCCs as the input of the model. The output was the hidden layer of the last LSTM spiking unit.

In the experiment, we adopted the cross-entropy loss function as the loss function, set the initial value of the learning rate to 0.005, and adopted the piecewise constant decay. We used the Adam optimization algorithm for parameter update, where the exponential decay rate β_1 was set to 0.9, β_2 to 0.999, and ω to 10^{-8}. All weights in (5) were initialized by np.random.randn, and all biases were initialized to 0.

Experiments were performed on a server equipped with NVIDIA GEFORCE RTX-2080Ti GPU and Intel® Core™ i7-8700K CPU. The computer operating system is Ubuntu 18.10.

2. Results

This section records results of the model to recognize different sounds. We evaluated the model's performance through the training process, as shown in Figure 9. As the training progressed, the training set's accuracy and the test set's accuracy continued to improve. After about 800 epochs, the model started to converge, and the accuracy curve and loss curve of the training set and validation set were quite close and remained stable. For a more detailed analysis, we applied the trained model to the test set for prediction, and the accuracy rate reached 93.4%. We calculated the confusion matrix of the test results, as shown in Figure 10. Each row of the matrix rep-

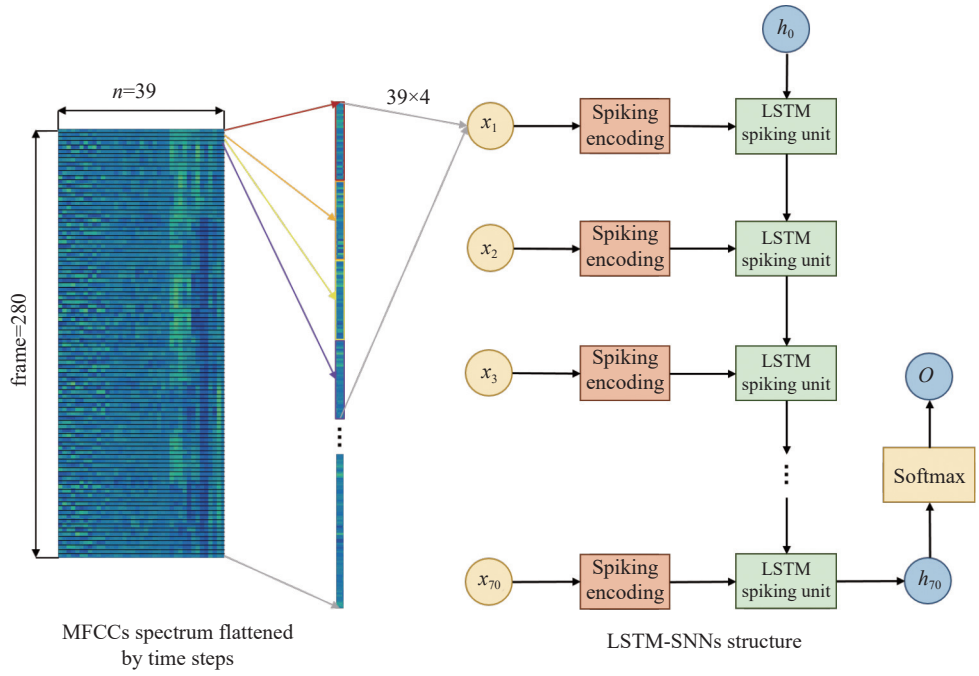


Figure 7 The architecture of the LSTM-SNN model in this work.

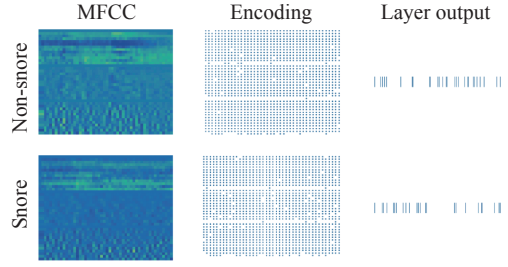


Figure 8 The spike raster plots of the LSTM-SNN model.

resented the true value of the sample, and each column represented the predicted value of the sample. According to the confusion matrix, the sensitivity was 93.1%, the specificity was 93.6%, the precision was 93.6%, and the F1 score was 93.3%. As a result, the LSTM-SNN model is effective in the snore detection task. We conducted a comparative experiment using traditional LSTM. The traditional LSTM input and hidden layer sizes were consistent with the LSTM-SNN. We compared two model's consumptions. The accuracy of traditional LSTM model on the test set was 94.6%, and the parameter size was 34.1 MBytes. The accuracy of the LSTM-SNN model was 93.4%, and the parameter size was 24.9 MBytes. The comparison results are shown in Figure 11. Compared with LSTM, the accuracy of the LSTM-SNN decreased by 1.2%, but the number of parameters decreased by 36.9%. We also compared our results with other studies, and the details are documented in Table 3. The experiments for comparison used multilayer CNNs or RNNs for classification, which inevitably led to a large amount of computation. Our model significantly reduced the model's size but retained the high recognition accuracy of mainstream networks, which was very meaningful for the terminal algorithm.

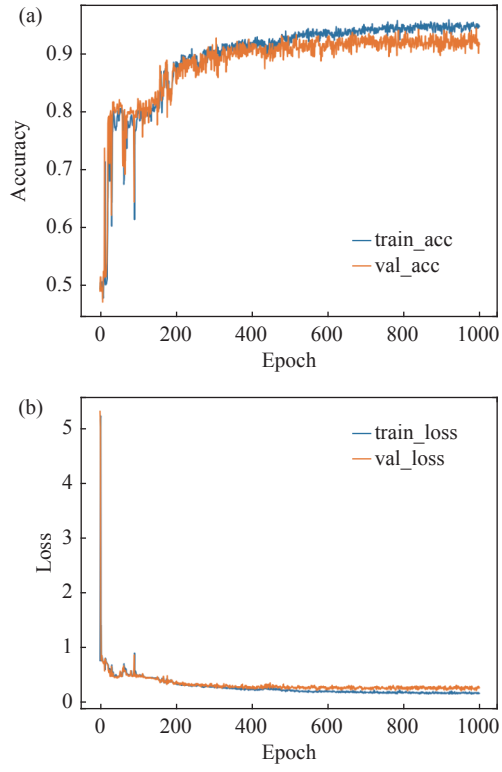


Figure 9 Experimental results. (a) Accuracy curves of training and validation set; (b) Loss function curves of training and validation set.

V. Conclusion and Discussion

In this work, we used a self-developed acquisition apparatus to capture sleep sound signals and built a self-made sound event dataset to establish an algorithm suit-

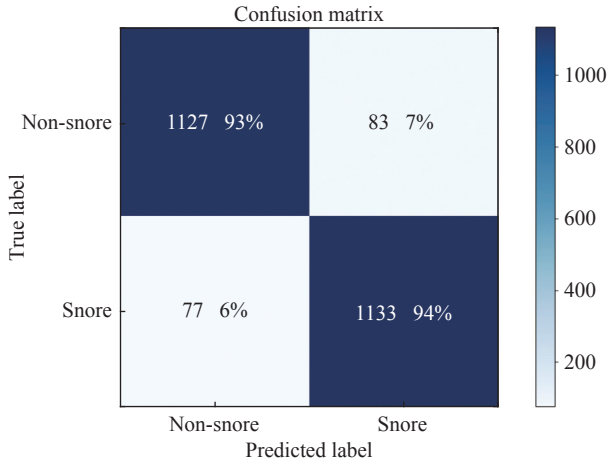


Figure 10 Confusion matrix of test set.

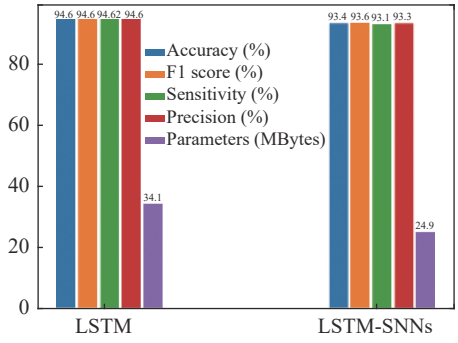


Figure 11 Performance of the regular LSTM model and our LSTM-SNN model in our test dataset.

ed for large-scale home snoring identification. Based on this, we employed the long short-term memory spiking neural networks model. We spiked the features with encoding and utilized a threshold activation function in the LSTM spike units to minimize the computational complexity of the model after extracting MFCCs features for sound events. We used a gradient substitution approach to complete parameter updating to address the problem of updating SNN parameters. Finally, we used a set of assessment measures to verify our model's performance on the snoring detection job.

We employed the LSTM-SNN model instead of typical convolutional neural networks or recurrent neural networks, which were used in much recent research on snoring detection [16]–[20]. We spiked the feature data before feeding it into the LSTM spiking neural networks. So, the feature fed into the network was matrices of 0s and 1s. Further, we adjusted the activation function in the LSTM spiking unit to two threshold activation functions. When calculating the input features, the calculation results exceeding the threshold were 1, and the calculation results not exceeding the threshold were 0. The information transmitted in the LSTM spiking units was matrices of 0s and 1s. When the calculation core operated on the matrix, the value 0 did not participate. Thus, the LSTM spiking neural networks exhibited recognition performance equivalent to classic neural networks on the snoring detection test while using less energy and resources. In the daily environment, the collection equipment we designed was ideal for collecting, storing, and transferring sleep sound data. It prepared the ground for us to complete mass home detection. We also labeled the sound data ourselves, which had a low signal-to-noise ratio and was more in line with the sound characteristics of a home environment. Based on the dataset, we adopted an automatic classification model that has improved generalization performance in the home detection scenario.

We also exploited key points coding [34] of MFCC spectrums to find effective spike coding algorithms. We detected key points on the spectrum by looking for a local maximum. Figure 12 shows the MFCC spectrum of snoring and non-snoring after key points coding. In the comparison experiments, we used the feature matrix after key points encoding as the input of the LSTM-SNN. The parameter settings were kept the same as Section IV.1. The results are shown on Figure 13. It can be seen that the validation loss starts to rise after 25 epochs, which indicates that the model is overfitting. The accuracy can only reach about 85%. By contrast, the threshold-based encoding is better than the key points for snoring detection.

Table 3 Comparison with other works

Authors	Number of subjects	Features	Models	Results (%)
Jiang *et al.* [18]	15	Mel-spectrogram	CNN-LSTM-DNN	Accuracy: 95.07 Sensitivity: 95.42 Soecificity: 95.82 F1 Score: 95.02 Precision: 94.62
Khan *et al.* [19]	–	MFCC image	CNN	Accuracy: 96
Xie *et al.* [20]	38	CQT-spectrogram	CNN-LSTM	Accuracy: 95.3 Sensitivity: 92.2 Specificity: 97.7
This work	54	MFCCs	LSTM-SNN	Accuracy: 93.4 Sensitivity: 93.1 Specificity: 93.6 F1 Score: 93.6 Precision: 93.3

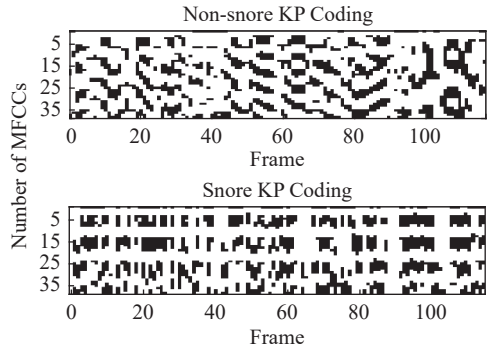


Figure 12 The key point coding of non-snore and snore.

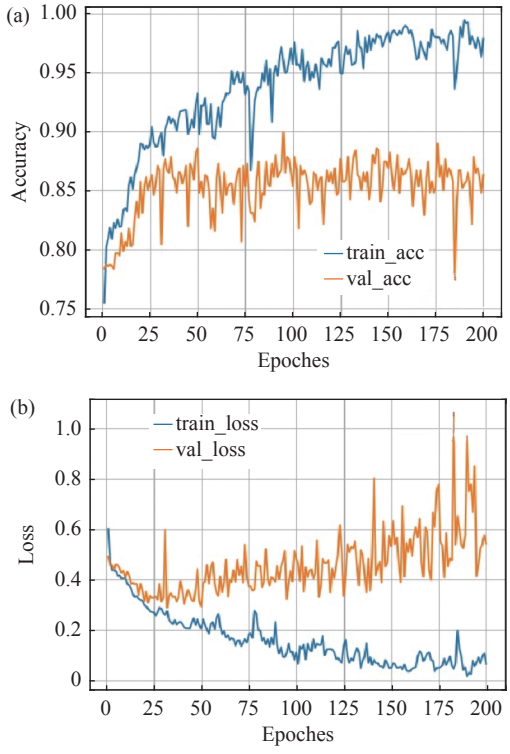


Figure 13 Results of the model based on the key point coding approach. (a) The accuracy curve of training and validation sets; (b) The loss function curve of training and validation sets.

The input size of the spike unit and the size of the hidden layer were the two most noticeable elements that determined the classification effect of the LSTM spiking neural networks. The spike unit's input size determined the model's complexity. Furthermore, the hidden layer's size specifies how much data was transferred from one unit to the next. If it was set too small, it quickly led to a decline in learning ability, and if it was too large, it needed to be more computationally complex and prone to overfitting. After many trials, we used a 39×4 spike unit input size and a 1000 hidden layer size to get the best classification results.

The model currently ignores the impact of respondents' gender, age, BMI, AHI, and other characteristics. It is also possible to consider the silent period in the automated categorization for the night's sleep sound data. In future work, on the one hand, the dataset's variety can be enhanced and more spike coding and spike calculation methods can be tried. The model structure may be improved based on the present model to improve the detection algorithm's accuracy. On the other hand, snoring detection is the first step in identifying OSAHS by snoring. Then we can look at ways to employ snoring features to classify OSAHS patients automatically.

Acknowledgements

This work was supported by the Zhejiang Key Research and Development Project (Grant No. 2022C01048) and the Zhejiang Province Public Welfare Project (Grant No. LGG22F010012).

References

[1] W. C. Ni, M. G. Cao, and Y. L. Tang, "Effect of sleep apnea syndrome with nasal continuous positive airway pressure," *Chinese Journal of Ophthalmology and Otorhinolaryn*, vol. 6, no. 4, pp. 214–215, 2009.

[2] P. E. Brockmann, O. Bruni, L. Kheirandish-Gozal, *et al.*, "Reduced sleep spindle activity in children with primary snoring," *Sleep Medicine*, vol. 65, pp. 142–146, 2020.

[3] J. Wang, C. Janson, E. Lindberg, *et al.*, "Dampness and mold at home and at work and onset of insomnia symptoms, snoring and excessive daytime sleepiness," *Environment International*, vol. 139, article no. 105691, 2020.

[4] D. Pevernagie, R. M. Aarts, and M. De Meyer, "The acoustics of snoring," *Sleep Medicine Reviews*, vol. 14, no. 2, pp. 131–144, 2010.

[5] B. Y. Chen and Q. Y. He, "Systemic damage in obstructive sleep apnea syndrome," *National Medical Journal of China*, vol. 92, no. 18 , pp. 1225–1227, 2012. (in Chinese)

[6] X. X. Chen, L. Chen, Z. M. Deng, *et al.*, "Diagnostic significance of dynamic electrocardiogram and synchronous respiration monitoring in snoring patients," *Journal of Guangdong Medical University*, vol. 39, no. 3, pp. 274–277, 2021. (in Chinese)

[7] O. Yildirim, U. B. Baloglu, and U. R. Acharya, "A deep learning model for automated sleep stages classification using PSG signals," *International Journal of Environmental Research and Public Health*, vol. 16, no. 4, article no. 599, 2019.

[8] B. Arsenali, J. van Dijk, O. Ouweltjes, *et al.*, "Recurrent neural network for classification of snoring and non-snoring sound events," in *Proceedings of 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Honolulu, HI, USA, pp. 328–331, 2018.

[9] M. Camacho, M. Robertson, J. Abdullatif, *et al.*, "Smartphone apps for snoring," *The Journal of Laryngology & Otology*, vol. 129, no. 10, pp. 974–979, 2015.

[10] Y. Yang, Y. Qin, W. N. Huang, *et al.*, "Acoustic characteristics of snoring sound in patients with obstructive sleep apnea hypopnea syndrome," *Journal of Clinical Otorhinolaryngology Head and Neck Surgery*, vol. 26, no. 8, pp. 360–363, 2012. (in Chinese)

[11] M. Dilkes and A. Adams, *Stop Snoring The Easy Way: And the Real Reasons You Need To*. The Orion Publishing Group Ltd, London, UK, pp. 11–16, 2017.

[12] H. J. Xu, L. S. Yu, W. N. Huang, *et al.*, "A preliminary study of acoustic characteristics of snoring sound in patients with obstructive sleep apnea/hypopnea syndrome (OSAHS) and with simple snoring," *Journal of Audiology and Speech Pathology*, vol. 17, no. 3, pp. 235–238, 2009. (in Chinese)

[13] K. Qian, C. Janott, V. Pandit, *et al.*, "Classification of the excitation location of snore sounds in the upper airway by acoustic multifeature analysis," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 8, pp. 1731–1741, 2017.

[14] S. M. Cervera, D. Nikolić, A. Barney, *et al.*, "A model of breathing abnormalities in sleep for development of classification and diagnosis techniques," in *Proceedings of 2010 3rd International Symposium on Applied Sciences in Biomedical and Communication Technologies*, Rome, Italy, pp. 1–2, 2010.

[15] Osborne J E, Osman E Z, Hill P D, *et al.*, "A new acoustic method of differentiating palatal from non-palatal snoring," *Clinical Otolaryngology & Allied Sciences*, vol. 24, no. 2, pp. 130–133, 1999.

[16] S. J. Lim, S. J. Jang, J. Y. Lim, *et al.*, "Classification of snoring sound based on a recurrent neural network," *Expert Systems with Applications*, vol. 123, pp. 237–245, 2019.

[17] T. L. Nguyen and Y. Won, "Sleep snoring detection using multi-layer neural networks," *Bio-Medical Materials and Engineering*, vol. 26, no. S1, pp. S1749–S1755, 2015.

[18] Y. M. Jiang, J. X. Peng, and X. W. Zhang, "Automatic snoring sounds detection from sleep sounds based on deep learning," *Physical and Engineering Sciences in Medicine*, vol. 43, no. 2, pp. 679–689, 2020.

[19] T. Khan, "A deep learning model for snoring detection and vibration notification using a smart wearable gadget," *Electronics*, vol. 8, no. 9, article no. 987, 2019.

[20] J. L. Xie, X. Aubert, X. Long, *et al.*, "Audio-based snore detection using deep neural networks," *Computer Methods and Programs in Biomedicine*, vol. 200, article no. 105917, 2021.

[21] M. M. van Gilst, J. P. van Dijk, R. Krijn, *et al.*, "Protocol of the SOMNIA project: An observational study to create a neurophysiological database for advanced clinical sleep monitoring," *BMJ Open*, vol. 9, no. 11, article no. e030996, 2019.

[22] J. P. Sun, X. Y. Hu, Y. Y. Zhao, *et al.*, "SnoreNet: Detecting snore events from raw sound recordings," in *Proceedings of 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Berlin, Germany, pp. 4977–4981, 2019.

[23] L. Zhang and W. G. Gong, "Improved wiener filtering speech enhancement algorithm," *Computer Engineering and Applications*, vol. 46, no. 26, pp. 129–131, 2010. (in Chinese)

[24] M. H. Wang, E. H. Zhang, and M. H. Wang, "Research and improvement on endpoint detection based on dual-threshold algorithm," *Computer & Digital Engineering*, vol. 45, no. 11, pp. 2223–2228, 2017. (in Chinese)

[25] A. Gupta and H. Gupta, "Applications of MFCC and vector quantization in speaker recognition," in *Proceedings of 2013 International Conference on Intelligent Systems and Signal Processing*, Vallabh Vidyanagar, India, pp. 170–173, 2013.

[26] S. Y. Cheng, C. Wang, K. Q. Yue, *et al.*, "Automated sleep apnea detection in snoring signal using long short-term memory neural networks," *Biomedical Signal Processing and Control*, vol. 71, article no. 103238, 2022.

[27] Q. H. Wang, L. N. Wang, and S. Xu, "Research and application of spiking neural network model based on LSTM structure," *Application Research of Computers*, vol. 38, no. 5, pp. 1381–1386, 2021. (in Chinese)

[28] A. Lotfi Rezaabad and S. Vishwanath, "Long short-term memory spiking networks and their applications," in *Proceedings of the International Conference on Neuromorphic Systems 2020*, Oak Ridge, TN, USA, article no. 3, 2020.

[29] A. Graves and A. Graves, "Long short-term memory," in *Supervised Sequence Labelling with Recurrent Neural Networks*, A. Graves, Ed. Springer, Berlin, Heidelberg, Germany, pp. 37–45, 2012.

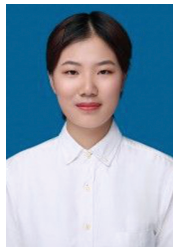
[30] G. Van Houdt, C. Mosquera and G. Nápoles, "A review on the long short-term memory model," *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5929–5955, 2020.

[31] Y. J. Wu, L. Deng, G. Q. Li, *et al.*, "Spatio-temporal backpropagation for training high-performance spiking neural networks," *Frontiers in Neuroscience*, vol. 12, article no. 331, 2018.

[32] E. O. Neftci, H. Mostafa, and F. Zenke, "Surrogate gradient learning in spiking neural networks: Bringing the power of gradient-based optimization to spiking neural networks," *IEEE Signal Processing Magazine*, vol. 36, no. 6, pp. 51–63, 2019.

[33] I. Sporea and A. Grüning, "Supervised learning in multilayer spiking neural networks," *Neural Computation*, vol. 25, no. 2, pp. 473–509, 2013.

[34] Q. Yu, Y. L. Yao, L. B. Wang, *et al.*, "A multi-spike approach for robust sound recognition," in *Proceedings of 2019 IEEE International Conference on Acoustics, Speech and Signal Processing*, Brighton, UK, pp. 890–894, 2019.



Rulin ZHANG received the B.E. degree from College of Electronic and Information Engineering, Shandong University of Science and Technology, Qingdao, China, in 2019. She is currently pursuing the M.S. degree in Electronics and Information College, Hangzhou Dianzi University, Hangzhou, China. Her research interests include spike neural networks and voice recognition.
(Email: rlzhang_0609@163.com)

Ruixue LI graduated from the School of Electrical and Information Engineering, Tianjin University and received the Ph.D. degree in 2020. She is now a Post-doctor in Electronics and Information College, Hangzhou Dianzi University. Her current research interests include computational neuroscience, pattern recognition, audio analysis and image processing.
(Email: rxli2012@163.com)

Jiakai LIANG received the B.E. degree in electronic information engineering from the Wenzhou University, Wenzhou, China, in 2021. He is currently pursuing the M.S. degree in Electronics and Information College, Hangzhou Dianzi University, Hangzhou, China. His research interests include spiking neural networks and image processing.
(Email: ljk211040090@hdu.edu.cn)

Keqiang YUE received the Ph.D. degree from College of Electrical Engineering Zhejiang University, Hangzhou, China, in 2014. He is currently an M.S. Supervisor with Electronics and Information College, Hangzhou Dianzi University, Hangzhou, China. His research interest include internet of things system development and intelligent communication.
(Email: kqyue@hdu.edu.cn)

Wenjun LI received the Ph. D. from Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghia, Chian, in 2004. He is currently a Professor of Hangzhou Dianzi University, Hangzhou, China. His research interest include integrated circuit design and intelligent computing and hardware.
(Email: liwenjun@hdu.edu.cn)

Yilin LI received the Ph. D. from School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an, China, in 2018. He is currently teaching in Electronics and Information College, Hangzhou Dianzi University, Hangzhou, China. His research interests include multidisciplinary design optimization, weak signal detection and AI model compression.
(Email: ericlee@hdu.edu.cn)