



OPEN ACCESS

EDITED BY

Xin Huang,
Renmin Hospital of Wuhan University, China

REVIEWED BY

Miriam Welgampola,
The University of Sydney, Australia
Kemar E. Green,
Johns Hopkins Medicine, United States

*CORRESPONDENCE

Haibo Li
✉ haibo021@sues.edu.cn

RECEIVED 07 February 2023

ACCEPTED 22 May 2023

PUBLISHED 09 June 2023

CITATION

Li H and Yang Z (2023) Torsional nystagmus recognition based on deep learning for vertigo diagnosis.

Front. Neurosci. 17:1160904.
doi: 10.3389/fnins.2023.1160904

COPYRIGHT

© 2023 Li and Yang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Torsional nystagmus recognition based on deep learning for vertigo diagnosis

Haibo Li* and Zhifan Yang

College of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai, China

Introduction: Detection of torsional nystagmus can help identify the canal of origin in benign paroxysmal positional vertigo (BPPV). Most currently available pupil trackers do not detect torsional nystagmus. In view of this, a new deep learning network model was designed for the determination of torsional nystagmus.

Methods: The data set comes from the Eye, Ear, Nose and Throat (Eye&ENT) Hospital of Fudan University. In the process of data acquisition, the infrared videos were obtained from eye movement recorder. The dataset contains 24521 nystagmus videos. All torsion nystagmus videos were annotated by the ophthalmologist of the hospital. 80% of the data set was used to train the model, and 20% was used to test.

Results: Experiments indicate that the designed method can effectively identify torsional nystagmus. Compared with other methods, it has high recognition accuracy. It can realize the automatic recognition of torsional nystagmus and provides support for the posterior and anterior canal BPPV diagnosis.

Discussion: Our present work complements existing methods of 2D nystagmus analysis and could improve the diagnostic capabilities of VNG in multiple vestibular disorders. To automatically pick BPV requires detection of nystagmus in all 3 planes and identification of a paroxysm. This is the next research work to be carried out.

KEYWORDS

torsional nystagmus, deep learning, classification and identification, convolution network, benign paroxysmal positional vertigo

1. Introduction

The vestibular system informs us of three-dimensional (3D) head position in space. Vestibular asymmetry creates a hallucination of head movement and therefore generates a compensatory slow phase eye movement and a quick phase that returns the eye closer to its starting position. Nystagmus is an involuntary oscillating eye movement that accompanies vestibular disorders (Leigh and Zee, 2015). The involvement of a specific vestibular end organ can be identified by the nystagmus trajectory (Jiang et al., 2018). Nystagmus can be summarized into two types: pathological nystagmus and physiological nystagmus. A variety of diseases, such as BPPV, Meniere's disease and vestibular neuritis, are all associated with pathological nystagmus (Newman et al., 2019). Pathological nystagmus arises from asymmetries in the peripheral or central vestibular system. Physiological nystagmus can be generated by rotational or thermal stimulation of the vestibular system (Henriksson, 1956; Cohen et al., 1977). BPPV is usually

accompanied by nystagmus, which provoked by changes of the head position relative to gravity (Lim et al., 2019). To diagnose different types of BPPV, clinicians inspect the directional and velocity characteristics of positional nystagmus during provocative testing. Among them, for BPPV with the highest incidence rate in the posterior semicircular canal, the typical torsional nystagmus was regarded as an important diagnostic factor. So, nystagmus examination is very important for the diagnosis of BPPV (Wang et al., 2014).

Some nystagmus can be observed by doctors with naked eyes or Frenzel goggles that allow for better visualization of nystagmus at the bedside. However, this diagnostic method was easily affected by the subjective experience of doctors (Slama et al., 2017). Not all pathological nystagmus was visible to the naked eye, since visual fixation suppresses peripheral spontaneous nystagmus. The other method is objective. This method usually uses electronystagmography (ENG) or video nystagmography (VNG) to record eye movements. The ENG method (Costa et al., 1995; Cesarelli et al., 1998) places sensors around the eyelid. Benign positional nystagmus arising from stimulation of one or more semicircular canals produces horizontal, vertical and torsional eye movements in the plane of that canal. The potential difference measured by the sensors is related to the horizontal and vertical movement of eyes. The velocity and frequency of eye movements can be obtained through potential difference analysis. BPPV also demonstrates a crescendo decrescendo velocity profile, the identification of which could assist with separation of BPV from its mimics. However, this method is vulnerable to electromagnetic interference, in which case the measured information is not accurate enough. The VNG methods generally use infrared camera to obtain nystagmus video. The frequency and amplitude of nystagmus were obtained by analyzing the motion information of pupil in video (Eggers et al., 2019). Recognizing 3D eye movement trajectory assists in identifying the canal of origin in patients with BPPV.

Under normal test conditions, VNG system includes a series of visual and dynamic function tests (Halmágyi et al., 2001; Newman-Toker et al., 2008). At present, some researchers (Buizza et al., 1978; Van Beuzekom and Van Gisbergen, 2002; Akman et al., 2006) have done some related work on how to use technical means to detect nystagmus. Most of the proposed methods can not fully recognize nystagmus automatically. Some parts or stages of these methods need human intervention, for example, a recognition method of nystagmus proposed by Buizza et al. (1978). In this method, doctors need to calibrate the direction of phase change. Akman et al. proposed a method to detect the period of nystagmus (Akman et al., 2006). The confirmation of the end point of nystagmus still needs to be further improved with this method. Van et al. proposed a nystagmus recognition method with VNG technology (Van Beuzekom and Van Gisbergen, 2002). This method requires researchers to manually confirm the two endpoints of the phase and remove interference factors such as noise from videos.

The research work stated above can be summarized as invasive and non-invasive (Newman et al., 2019). Invasive methods, such as electromagnetic coil method, mainly embed hardware equipment into human eyes, which leads to direct contact between equipment and human eyes. This will cause direct or potential harm to human eye health. The non-invasive detection methods were mainly gaze description methods based on video image processing. These methods detect and locate the pupil based on the contours of the eyes, which were greatly improved in comfort and accuracy.

The Non-invasive inspection methods can be combined with artificial intelligence (AI) methods. At present, AI technology is developing rapidly. Deep Learning has promoted the development of Computer Vision (He et al., 2016; Mane and Mangale, 2018; Kim and Ro, 2019; Cong et al., 2020), Natural Language Processing (Karpathy and Fei-Fei, 2014; Sundermeyer et al., 2015; Young et al., 2018) and other technologies. The development of deep learning technology also provides the possibility for medical intelligent aided diagnosis. For example, CT images of thoracic nodules were analyzed to determine whether there was a tumor in the chest (Anthimopoulos et al., 2016; Setio et al., 2016). Other medical applications include automatic analysis of skin disease images (Rathod et al., 2018; Wu et al., 2019), automatic analysis of fundus disease images (Ting et al., 2017; Sertkaya et al., 2019) and automatic analysis of tumor pathological sections (Tra et al., 2016; Lavanyadevi et al., 2017), etc. A variety of algorithms based on deep learning were integrated into the innovative diagnosis and treatment system (Litjens et al., 2017). For example, Google used neural network to analyze diabetic retinopathy, and its analysis results were similar to those of human experts (Gulshan et al., 2016). In other application fields, deep learning has been applied to motion detection in videos and achieved good recognition results (Saha et al., 2016). Therefore, the recognition of nystagmus can be tried by using the method of deep learning. At present, many scholars have begun to use artificial intelligence methods to identify nystagmus (Zhang et al., 2021; Lu et al., 2022; Wagle et al., 2022). From the experimental results, the deep learning method can be used to detect nystagmus, and the recognition accuracy can be further improved.

This paper mainly focuses on a torsional nystagmus recognition method based on deep learning. With the development of deep learning technology, this paper proposed an automatic recognition method of torsional nystagmus based on deep learning technology to help doctors make rapid diagnosis.

2. Materials and methods

2.1. Detail of data sources

The data set of this paper comes from Eye, Ear, Nose and Throat (Eye & ENT) Hospital of Fudan University. In the process of data acquisition, the infrared videos were obtained from eye movement recorder with the model of VertiGoggles R ZT-VNG-II, which was provided by Shanghai Zhiting Medical Technology Co., Ltd. Eye movement recorder was used to record and save the patient's eye movements video. The video format is MP4. The size of video frame is 640×480 and the frame rate is 60fps. The data set include 26,931 nystagmus videos from 1,236 patients. After removing the abnormal and disturbed data, the remaining 24,521 videos were used as the data set. The length of each nystagmus video was not required to be exactly equal. The length of video in the data set was reduced to 6–10s. The data were from patients with BPPV. The videos were monocular, including left and right eyes. All data were annotated by four ophthalmologists according to the motion characteristics of torsional nystagmus. 80% of the data were used for training and 20% for verification.

The doctors recruited eligible subjects in the otolaryngology clinic or vestibular function examination room. For patients who complained of positional vertigo, bilateral Dix-Hallpike test was performed first, and then bilateral Roll test was performed. Each body position change was rapid, but not exceeded the patient's tolerance. In case of atypical

symptoms such as hearing loss, severe headache, limb sensation or movement disturbance, consciousness disturbance, ataxia, etc., corresponding audiological or imaging examination was carried out first to eliminate other inner ear or central lesions. After judging that the conditions for enrollment were met, the subjects themselves signed the informed consent form and collected their basic information and contact information. The doctor collected nystagmus videos of patients in the whole process of Dix-Hallpike test suspension sitting position and Epley method reduction. The Epley reposition method maintained each position for at least 30s until the nystagmus disappears. After the restoration, the subjects rested for 15min, and then performed Dix-Hallpike test again. The negative person indicated that the restoration was successful. If the first reset failed, the doctor performed the reset again and collected the nystagmus video of the second reset.

2.2. Network model structure and classification process

Ethical statement. The study was conducted according to the guidelines of the Declaration of Helsinki and approved by Ethics Committee of the Eye, Ear, Nose and Throat Hospital affiliated to Fudan University (approval number: 2020518). Written informed consent was obtained from all enrolled patients.

In order to recognize nystagmus automatically by deep learning, a recognition model as shown in Figure 1 was designed in this paper. Firstly, the nystagmus video was sent to the sequence layer in the model for processing. The output video frame sequence was transmitted to the input of the sequence folding layer. Secondly, the motion characteristics of each frame in the video was extracted independently by convolution operation. Thirdly, the extracted

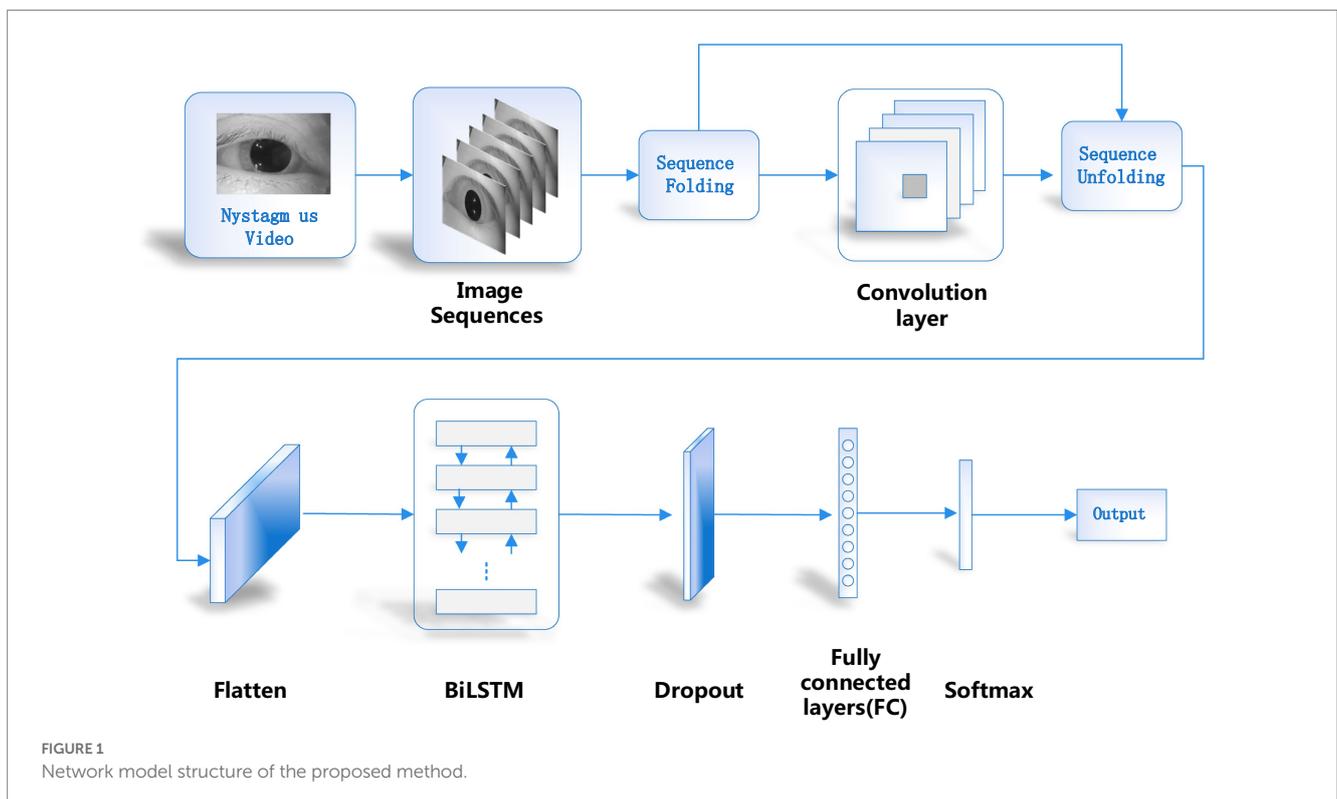
features were restored to the sequence structure after passing through the sequence unfolding layer and flattening layer. At the same time, the output was transformed into vector sequences. Finally, the obtained vector sequences were classified by using Bi-directional Long Short-Term Memory (BiLSTM) layer and output layer. The functions of each part of the network model are introduced as follows.

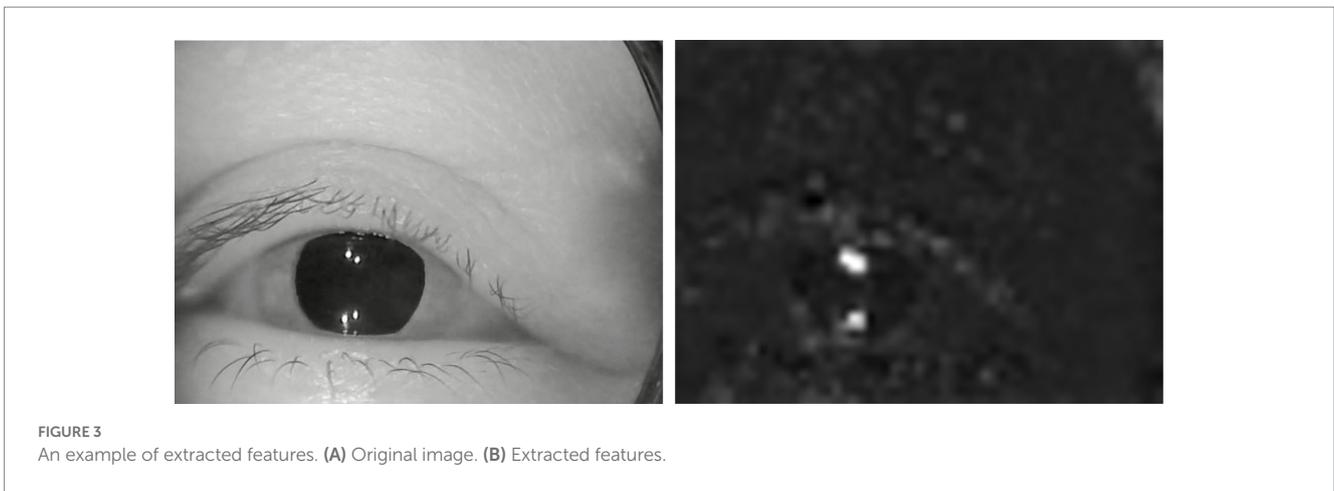
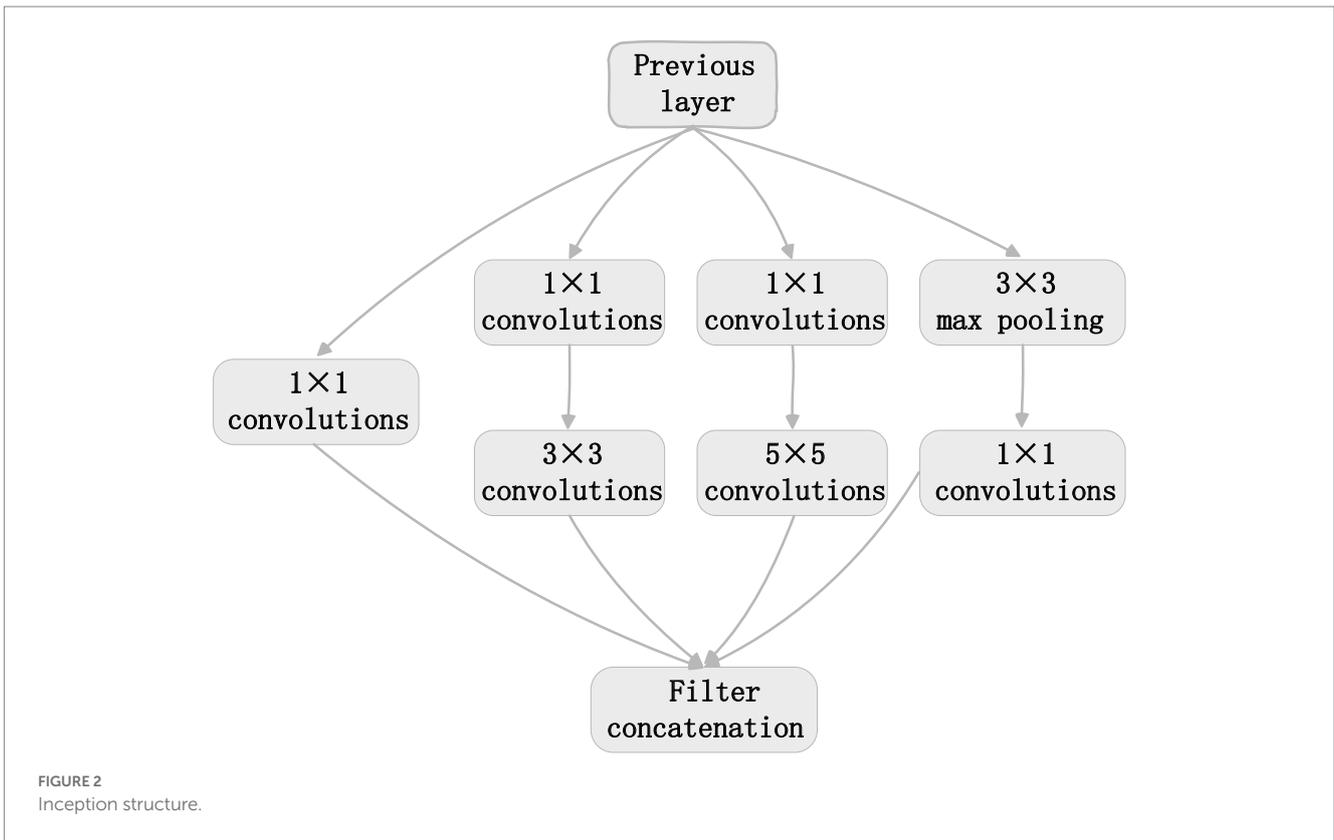
2.3. Converting video into video sequence and sequence folding

Firstly, a single video was processed to obtain the relevant parameters of the video, such as the height, width, number of channels and frames of the video. Then the video was cropped. This paper adopted the longest edge of the cropped video and adjusted its size to obtain a 224×224 fixed size. In order to enable the feature extraction network to obtain the features of single frame, a sequence folding layer was constructed to convert sequences into images. The sequence folding layer converts a batch of image sequences into a batch of images. The sequence unfolding layer restores the sequence structure of the input data after the sequence was folded.

2.4. Feature extraction

Feature extraction was mainly completed by five modules. The first module includes convolution layers and the maximum pooling layer. Convolution layer: the kernel size is 7×7 ; the step of sliding window is 2; the number of output channels is 64. Pooling layer: the window size is 3×3 ; the step of sliding window is 2; the output channel number is 64.





The second module has two convolution layers and a maximum pool layer. Convolution layer: the kernel size is 3×3 ; the step of sliding window is 1; the output channel number is 192. Pooling layer: the window size is 3×3 ; the step of sliding window is 2; the output channel number is 192. The third module has two Inception modules in series, followed by a maximum pool layer. Figure 2 illustrates the structure of the Inception module. The Inception module adopts the idea of network in network (NIN). It extracts the local features of the image by using multiple convolution kernels with different scales. Each branch in the Inception module adopts 1×1 convolution kernel. It can effectively improve the receptive field of convolution kernel and reduce the dimension to accelerate the network calculation and strengthen the

real-time performance. As can be seen from Figure 2, the Inception module has four main components: 1×1 , 3×3 , 5×5 convolution and 3×3 pooling. An example of extracted features in the four components of the inception module was shown in Figure 3. The main purpose of this structure is to extract the multi-scale information through a variety of convolution kernels of different sizes, and then fuse them, so as to have better image representation ability. In practice, using 3×3 and 5×5 convolution directly will lead to too much calculation. So, 1×1 convolution layer should be concatenated in front. The nonlinearity of the network can be increased at the same time.

The numbers of channels output by 4 lines of the first Inception are 64, 128, 32 and 32. The total number of output channels is the

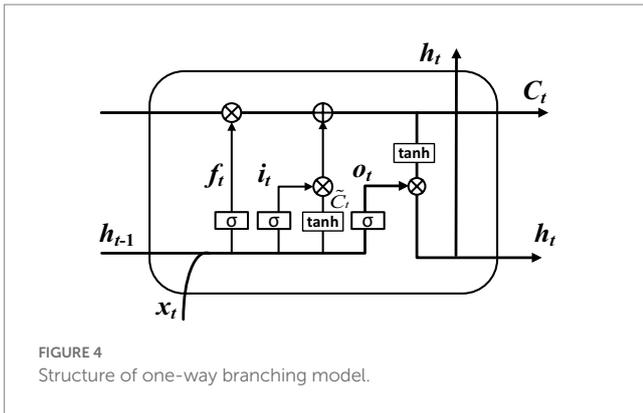


FIGURE 4 Structure of one-way branching model.

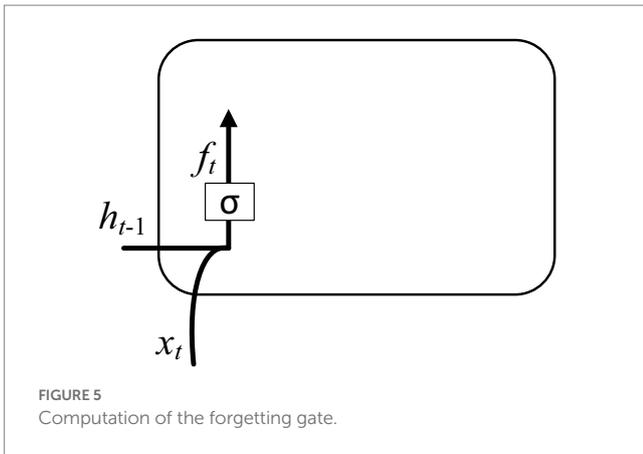


FIGURE 5 Computation of the forgetting gate.

accumulation of the four lines, which is 256. The numbers of channels output by 4 lines of the second Inception are 128, 192, 96 and 64 respectively, and the total number of output channels is 480. Pooling layer: the window size is 3 × 3; the step of sliding window is 2; the output channel number is 480.

The fourth module has five Inception blocks in series, followed by a maximum pool layer. The numbers of channels output by 4 lines of the first Inception are 192, 208, 48 and 64 respectively, and the total number of output channels is 512. The numbers of channels output by 4 lines of the second Inception are 160, 224, 64 and 64 respectively, and the total number of output channels is 512. The numbers of channels output by 4 lines of the third Inception are 128, 256, 64 and 64 respectively, and the total number of output channels is 512. The numbers of channels output by 4 lines of the fourth Inception are 112, 288, 64 and 64 respectively, and the total number of output channels is 528. The numbers of channels output by 4 lines of the fifth Inception are 256, 320, 128 and 128 respectively, and the total number of output channels is 832. Pooling layer: the window size is 3 × 3; the step of sliding window is 2; the output channel number is 832.

The fifth module has two Inception blocks in series, followed by a pooling layer. The output channel number of 4 lines are 256, 320, 128 and 128, respectively, in the first Inception, and the total number of output channels is 832. The numbers of channels output by 4 lines of the second Inception are 384, 384, 128 and 128 respectively, and the total number of output channels is 1,024. The pooling layer adopts global average pooling and the convolution layer with height and width of 1 is obtained. The number of output channels is 1,024.

2.5. Recovering sequence structure

The sequence structure was deleted by the sequence folding layer. So, the sequence structure should be restored after feature extraction. The recovery task of sequence structure was completed by sequence unfolding layer. The sequence unfolding layer takes the minibatchsize output information of the sequence folding layer as the minibatchsize input information of the sequence unfolding layer. The output of the sequence unfolding layer was reconstructed into vector sequences. The spatial dimension of the tensor was flattened to channel dimension. Flatten layer flattens input spatial dimension into a single channel. This layer retains the observation dimension (N) and sequence dimension (S) after flattening.

2.6. Sequence classification

Long Short-Term Memory (LSTM) model can record the relationship between elements in a spatial distance. This memory function can be realized by training LSTM model. But one disadvantage of LSTM model is that the order of memorizing information can only be from front to back. In order to better classify the types of nystagmus, this paper uses BiLSTM to solve this problem. BiLSTM is composed of two LSTMs with opposite directions. Figure 4 shows the structure of the one-way branching model in BiLSTM. In the figure, x_t , o_t , C_t , f_t , h_t , \tilde{C}_t and i_t represent input vector, output gate, cell state, forgetting gate, hidden layer state, temporary cell state and memory gate, respectively.

The classification calculation process was completed by the following steps. Step 1: The discarded information was determined by calculating the forgetting gate. The input is the hidden layer state h_{t-1} at time $t-1$ and the input vector x_t at time t . The output is the value f_t of the forgetting gate at time t . As shown in Figure 5.

The input of h_{t-1} and x_t were calculated to obtain a forgetting gate output f_t through the sigmoid function, and its expression is shown in Equation (1).

$$f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f) \tag{1}$$

Where $f_t \in [0,1]$ (0 indicates to discard the information completely, and 1 indicates to retain the information completely); σ indicates the activation function; w_f represents a learnable connection vector; x_t is input; b_f represents the offset value.

Step 2: The retained information was determined by calculating the memory gate. The input is the hidden layer state h_{t-1} at time $t-1$ and the input vector x_t at time t . The output is the value i_t of the memory gate at time t and the value \tilde{C}_t of the temporary cell state at time t . As shown in Figure 6. The value of the memory gate was obtained after that the hidden layer state value at time $t-1$ and the input vector at time t pass through the sigmoid activation function. The value of the temporary cell state was obtained after the hidden layer state value at time $t-1$ and the input vector at time t pass through the tanh activation function. The output values of two activation functions were multiplied to obtain the value of the input gate. The corresponding equation can be written as:

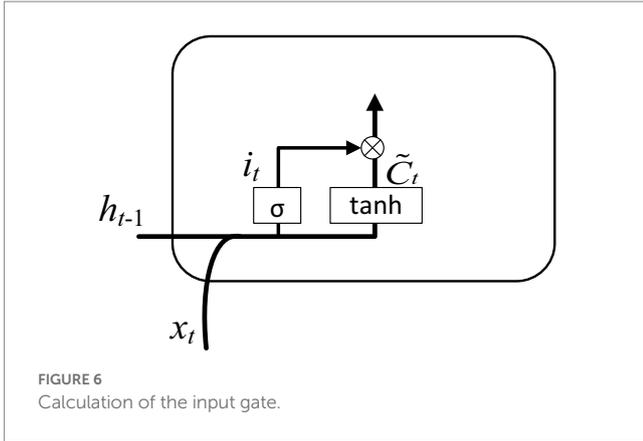


FIGURE 6
Calculation of the input gate.

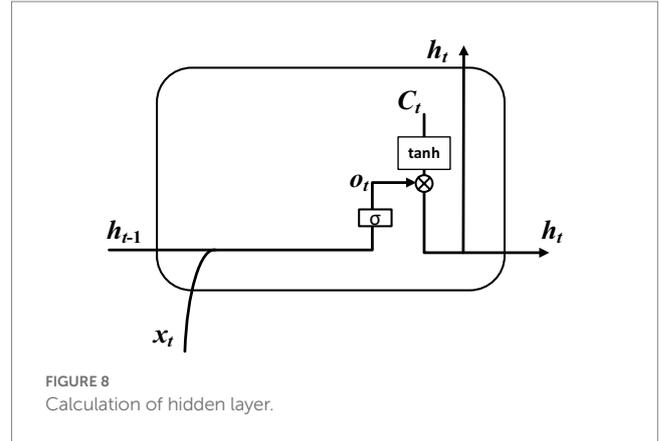


FIGURE 8
Calculation of hidden layer.

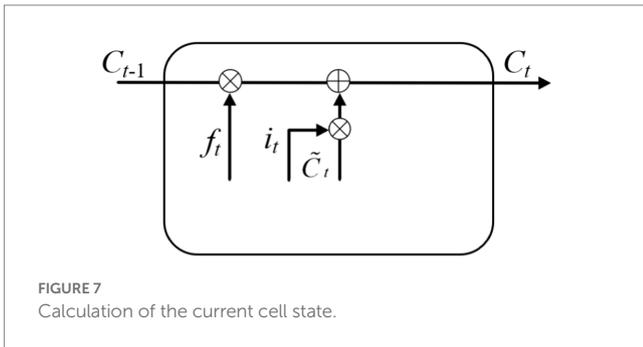


FIGURE 7
Calculation of the current cell state.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{3}$$

Where: \tanh represents the activation function; W_i and W_c represent the learnable connection vectors; b_i and b_c represent the offset values.

Step 3: The cell state C_t was obtained through the joint action of forgetting gate and input gate. The input is the memory gate i_t at time t , the forgetting gate f_t at time t , the temporary cell state \tilde{C}_t at time t and the cell state C_{t-1} at time $t-1$. The output is the cell state C_t at time t . As shown in Figure 7. The corresponding equation can be written as:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{4}$$

Step 4: The value of the output gate and the value of the hidden layer state were determined by calculation. The input is the hidden layer state h_{t-1} at time $t-1$, the input vector x_t at time t and the cell state C_t at time t . The output is the value o_t of the output gate at time t and the value h_t of the hidden layer at time t . As shown in Figure 8. The value of output gate was obtained after that the hidden layer state value at time $t-1$ and the input vector at time t pass through the sigmoid activation function. The value of hidden layer state was obtained after that the output gate value at time t and the cell state at time t pass through the tanh activation function. The corresponding expression can be written as:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{5}$$

$$h_t = o_t * \tanh(C_t) \tag{6}$$

W_o and b_o represent learnable connection vectors and offset values, respectively.

Through the above steps, we can get the corresponding sequence, which is $\{h_0, h_1, \dots, h_{n-1}\}$. BiLSTM consists of two branches in different directions mentioned above. The parameters on each branch are independent of the other branch. One branch can only fit time-related data from one direction. BiLSTM has two branches in opposite directions so that it can capture patterns that one branch may ignore. The structure can be seen in Figure 9.

If the hidden layer state sequence calculated by one branch of BiLSTM was represented by h^r , the hidden layer state sequence of the other branch in the opposite direction was represented by h^l . The final output result is as follow:

$$h_t = \alpha h^r + \beta h^l \tag{7}$$

$$y_t = \sigma(h_t) \tag{8}$$

Where α , β are constants and $\alpha + \beta = 1$. σ is the activation function.

After that the output results pass through the classification layer, the type results of nystagmus recognition can be obtained. The output layer includes dropout layer, full connection layer, softmax layer and classification layer.

3. Results

3.1. Model training and verification process

The data set of this paper comes from Eye, Ear, Nose and Throat (Eye & ENT) Hospital of Fudan University. In the process of data acquisition, the infrared videos were obtained from eye movement recorder with the model of VertiGoggles R ZT-VNG-II, which was provided by Shanghai Zhiting Medical Technology Co., Ltd. Eye movement recorder was used to record and save the patient's eye movements video. The video format is MP4. The size of video frame is 640×480 and the frame rate is 60fps. The

data set include 26,931 nystagmus videos from 1,236 patients. After removing the abnormal and disturbed data, the remaining 24,521 videos were used as the data set. All data were annotated by four ophthalmologists according to the motion characteristics of torsional nystagmus. 80% of the data were used for training and 20% for verification. The model training and verification process are shown in Figure 10.

Figure 10 shows that the classification accuracy of the training and verification process tends to be stable with the increase of iterations. The average accuracy after stabilization is shown in Table 1.

The loss during training and verification is shown in Figure 11.

As can be seen from Figure 11, with the increase of the number of iterations, the loss in the training and verification process has decreased to a stable state. In order to further evaluate the designed method. Figure 12 shows Area Under Curve (AUC).

3.2. Results of feature extraction by different methods

In addition, we also study the impact of different methods to extract video frame features on the classification effect. We use eight

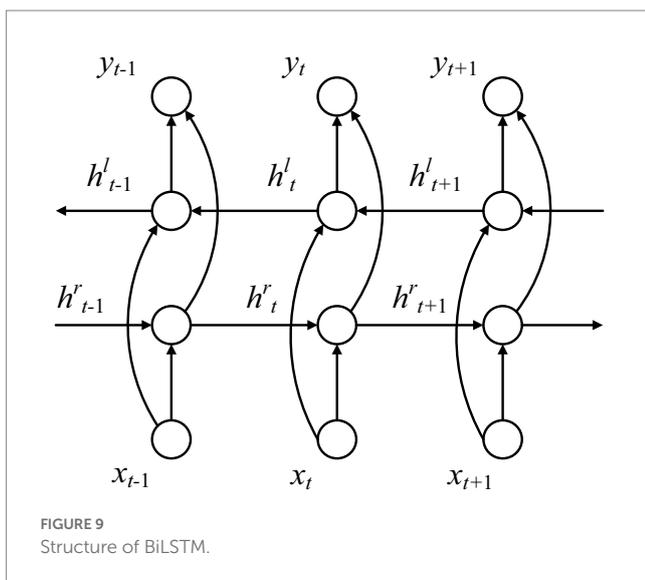


FIGURE 9 Structure of BiLSTM.

Fire modules to extract the video frame features, and the structure of other parts remains unchanged. This method is named method 2. The structure of Fire module (Kim and Kim, 2020) is shown in Figure 13.

It can be seen from Figure 12 that the proposed method can identify torsional nystagmus more accurately. In addition, sensitivity@0.912 and specificity@0.946.

The same data set was used for training and verification. Figure 14 shows the classification accuracy of training and verification by method 2.

As can be seen from Figure 14, the classification accuracy tends to be stable with the increase of iterations, whether in the training process or verification process. Figure 15 shows the Loss during the training and verification process by method 2 using the same data set.

As can be seen from Figure 15, with the increase of iterations, the Loss of method 2 decreased to a stable state, whether in the training process or verification process. Method 2 was compared with the method proposed in this paper. The comparison results of classification accuracy in training set are shown in Figure 16.

As can be seen from Figure 16, the recognition accuracy of two methods tends to be stable with the increase of iterations. The average accuracy after stabilization is shown in Table 2.

It can be seen from Table 2 that the proposed method has high recognition accuracy in training process. The recognition accuracy of two methods in verification set is shown in Figure 17.

Figure 17 shows that the recognition accuracy of two methods in verification set tends to be stable with the increase of iterations. The average recognition accuracy of two methods after stabilization is shown in Table 3.

It can be seen from Table 3 that the proposed method has high recognition accuracy in verification set.

4. Discussion

The proposed method was compared with Zhang's method (Zhang et al., 2021) and Zhou's method (Zhou et al., 2022). The same data set was used for training and verification, respectively. The recognition accuracy of different methods in training set is shown in Figure 18.

As can be seen from Figure 18, the recognition accuracy of all methods in training set tends to be stable with the increase of

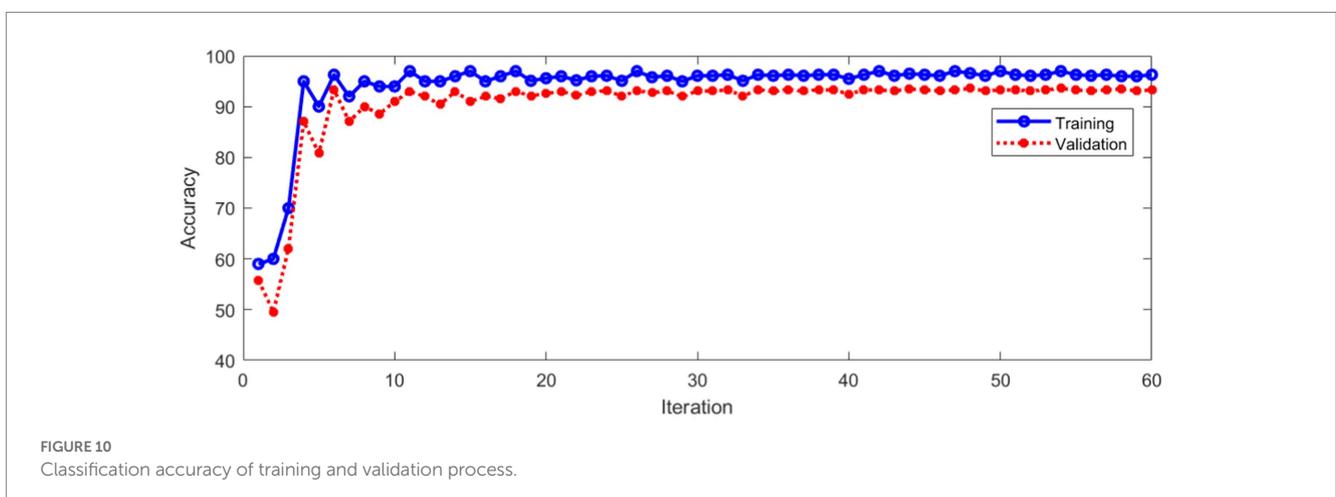


FIGURE 10 Classification accuracy of training and validation process.

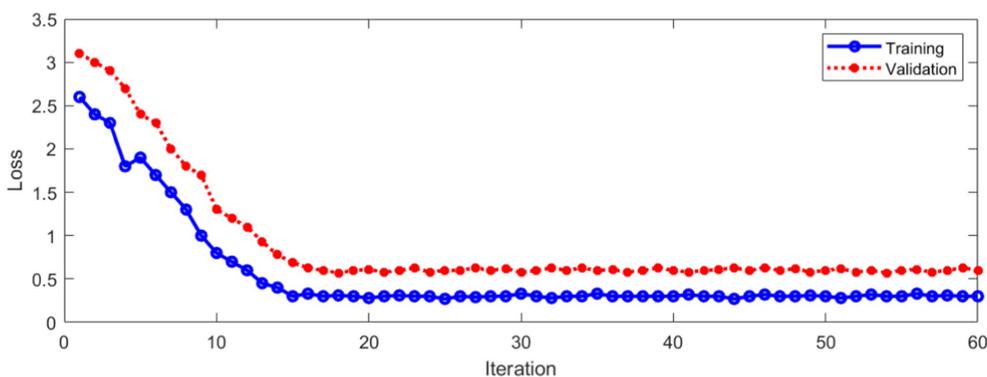


FIGURE 11 Loss during training and verification.

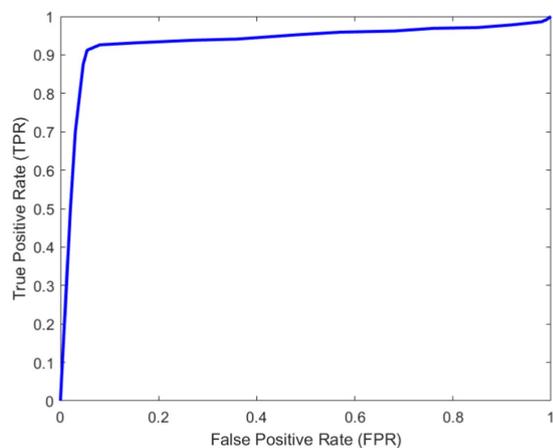


FIGURE 12 AUC curve.

TABLE 1 Recognition accuracy in training and verification stage.

Recognition accuracy	Stage
96.1%	Training
92.9%	Validation

TABLE 2 Recognition accuracy of two methods in training set.

Recognition accuracy	Method
96.1%	The proposed method
89.9%	Method 2

TABLE 3 Recognition accuracy of two methods in verification set.

Recognition accuracy	Method
92.9%	The proposed method
87.4%	Method 2

TABLE 4 Recognition accuracy of different methods in training set.

Recognition accuracy	Method
96.1%	The proposed method
91.2%	Zhang's method
93.9%	Zhou's method

TABLE 5 Recognition accuracy of different methods in verification set.

Recognition accuracy	Method
92.9%	The proposed method
89.4%	Zhang's method
90.7%	Zhou's method

TABLE 6 Statistical comparisons of the variable performance accuracy.

Method	Precision	Recall
The proposed method	94.3%	91.2%
Zhang's method	90.1%	87.6%
Zhou's method	91.9%	88.4%

iterations. The average recognition accuracy of different methods after stabilization is shown in Table 4.

It can be seen from Table 4 that the proposed method has high recognition accuracy in training set. The recognition accuracy of different methods in verification set is shown in Figure 19.

Figure 19 shows that the recognition accuracy of different methods in verification set tends to be stable with the increase of iterations. The average recognition accuracy of different methods after stabilization is shown in Table 5.

It can be seen from Table 5 that the proposed method in this paper has high recognition accuracy in verification set. This shows that the proposed method has a good effect in torsional nystagmus recognition. In addition, other statistical comparisons of the variable performance accuracy across models are shown in Table 6.

It can be seen from Table 6 that the proposed method has high precision and recall rate, which indicates that the recognition

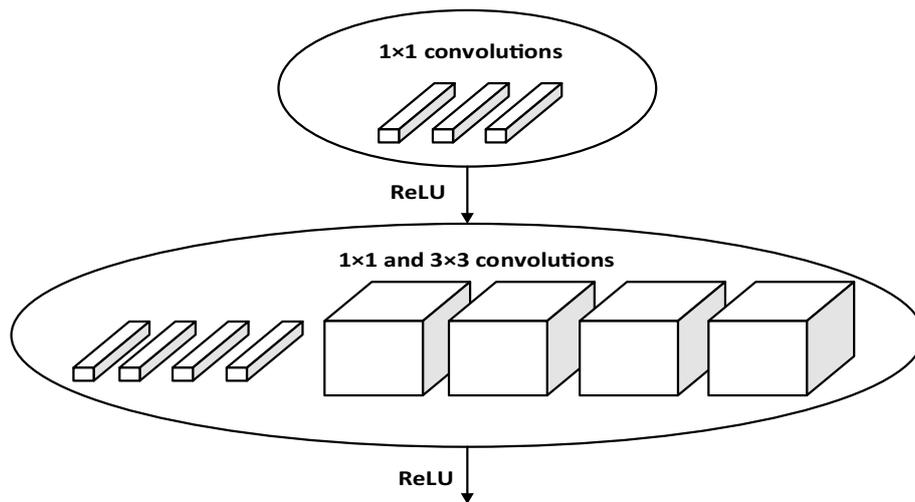


FIGURE 13
Fire module structure.

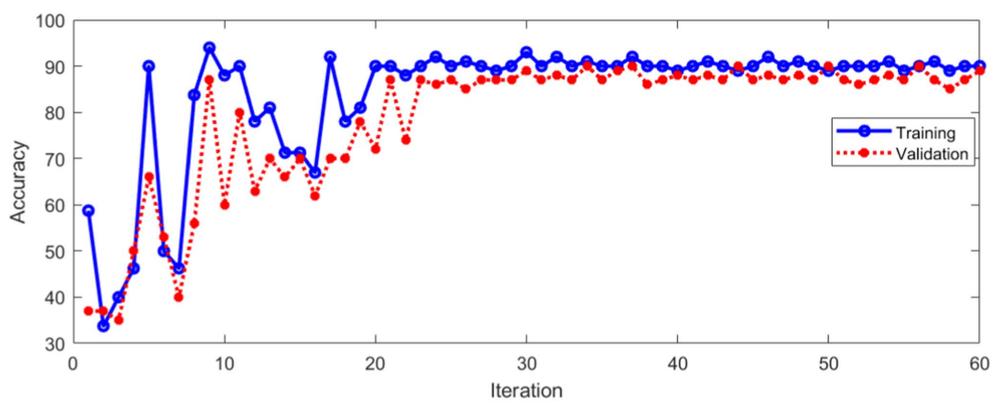


FIGURE 14
Classification accuracy of method 2.

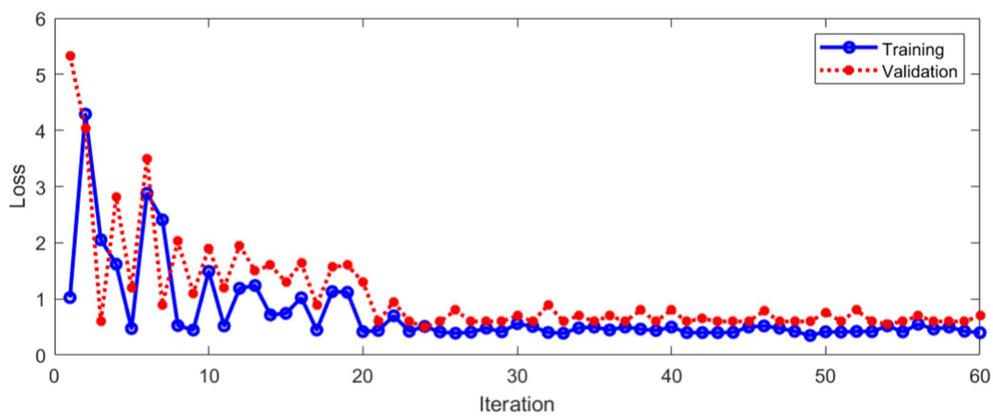


FIGURE 15
Loss during training and verification with method 2.

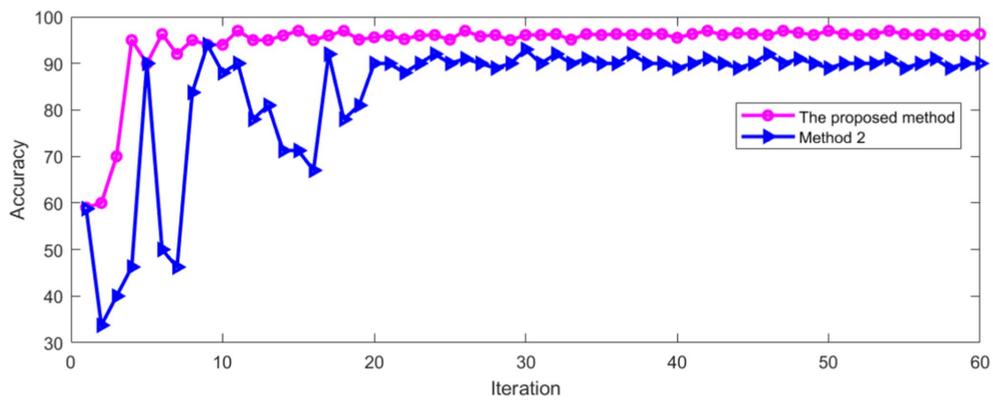


FIGURE 16
Comparison of accuracy in training set.

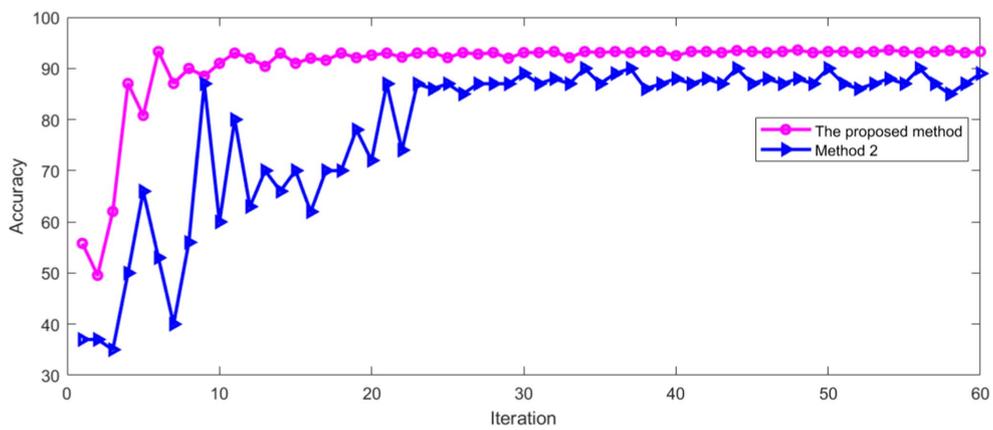


FIGURE 17
Comparison of accuracy in verification set.

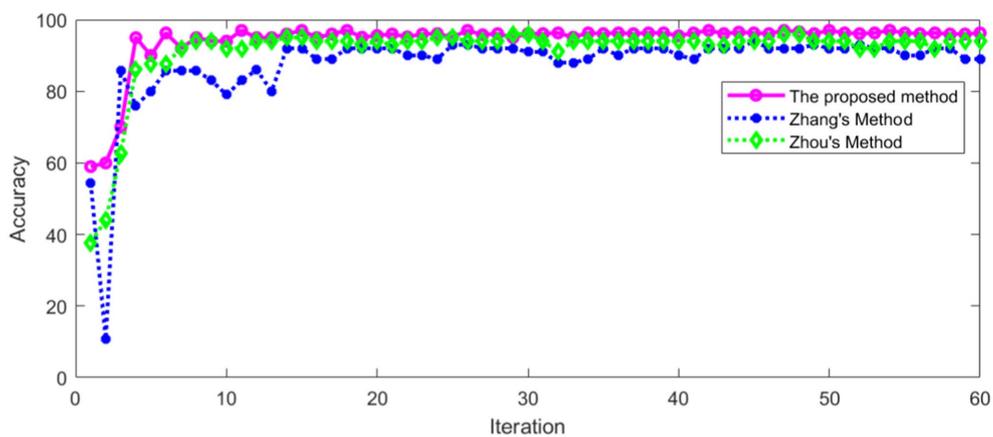


FIGURE 18
Comparison of accuracy in training set.

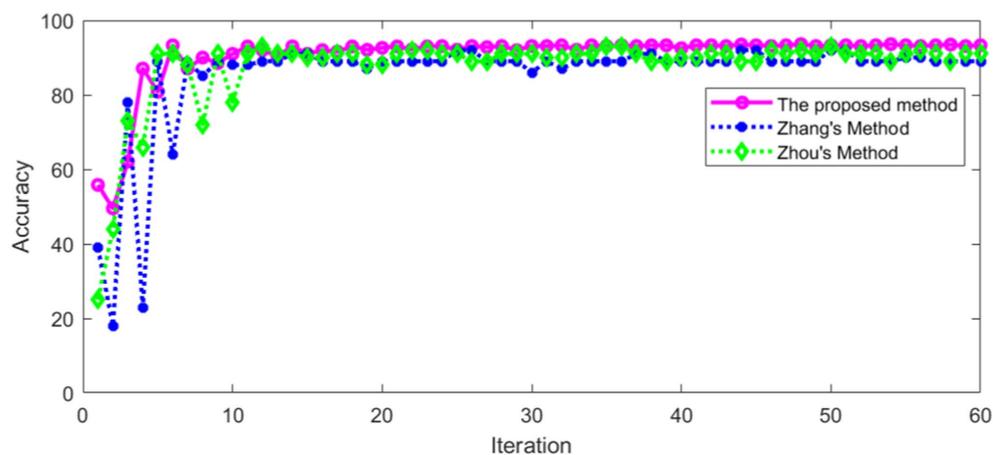


FIGURE 19
Comparison of accuracy in verification set.

performance of the algorithm is better than other methods. In addition, we also compared with the method proposed by [Wagle et al. \(2022\)](#), and the recognition accuracy of their method is 82.7%, which is lower than the proposed method.

Based on the real data of a large number of clinical patients, the characteristics and types of torsional nystagmus were intelligently recognized through the deep learning algorithm. The diagnosis of BPPV can be accurately predicted by combining the body position information, so as to realize the intelligent diagnosis and treatment of BPPV, improve the diagnosis efficiency and reduce the pain of patients. It is expected to comprehensively improve the diagnosis and treatment capacity of medical institutions at all levels for typical BPPV patients.

5. Conclusion

In this paper, a recognition model of torsional nystagmus was proposed based on deep learning network. From the experimental results, the nystagmus recognition model used convolution neural network to extract the frame features of the video sequence, and classified the obtained vector sequence, which can effectively identify torsional nystagmus. This shows that the recognition of torsional nystagmus can be accomplished by using deep learning network models with different structures. Although these changes in nystagmus are very complex for clinicians, they are indeed extractable features for deep learning. Once these specific nystagmus classification features are obtained, computer-aided clinical screening and classification of typical diseases can widely benefit patients with vertigo disease and help improve the diagnosis efficiency of vertigo disease. Compared with the existing methods, the proposed method further improved the recognition accuracy. In the future, we will label the slow phase velocity (SPV) of the nystagmus, so that we can analyze the performance of the model according to the SPV of the nystagmus. The development of an accurate torsion detection method has implications for correct interpretation of nystagmus overall. BPV is not the only disorder producing torsional nystagmus: stroke, vestibular migraine can present with torsional nystagmus; vestibular neuritis and Menieres disease can also generate horizontal torsional nystagmus. Our

present work complements existing methods of 2D nystagmus analysis and could improve the diagnostic capabilities of VNG in multiple vestibular disorders. To automatically pick BPV requires detection of nystagmus in all 3 planes and identification of a paroxysm. This is the next research work to be carried out.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Ethics statement

The study was conducted according to the guidelines of the Declaration of Helsinki and approved by Ethics Committee of the Eye, Ear, Nose and Throat Hospital affiliated to Fudan University (Approval number: 2020518). Written informed consent was obtained from all enrolled patients.

Author contributions

HL wrote the main manuscript. ZY prepared the [Figures 16, 17](#). All authors contributed to the article and approved the submitted version.

Funding

This work was funded by the Shanghai Hospital Development Center, grant number SHDC2020CR3050B.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Akman, O. E., Broomhead, D. S., Clement, R. A., and Abadi, R. V. (2006). Nonlinear time series analysis of jerk congenital nystagmus. *J. Comput. Neurosci.* 21, 153–170. doi: 10.1007/s10827-006-7816-4
- Anthimopoulos, M., Christodoulidis, S., Ebner, L., Christe, A., and Mougiakakou, S. G. (2016). Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE Trans. Med. Imaging* 35, 1207–1216. doi: 10.1109/TMI.2016.2535865
- Buizza, A., Schmid, R., Zanibelli, A., Mira, E., and Semplici, P. (1978). Quantification of vestibular nystagmus by an interactive computer program. *ORL J. Otorhinolaryngol. Relat. Spec.* 40, 147–159. doi: 10.1159/000275399
- Cesarelli, M., Bifulco, P., and Loffredo, L. (1998). "EOG baseline oscillation in congenital nystagmus" in *VIII Mediterranean conference on medical biological engineering and computing-MEDICON* (Lemesos, Berlin: Springer), 14–17.
- Cohen, B., Matsuo, V., and Raphan, T. (1977). Quantitative analysis of the velocity characteristics of optokinetic nystagmus and optokinetic after-nystagmus. *J. Physiol.* 270, 321–344. doi: 10.1113/jphysiol.1977.sp011955
- Cong, W., Zhang, J., Niu, L., Liu, L., Ling, Z., Li, W., et al. (2020). "Dovenet: deep image harmonization via domain verification" in *2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR)* (Seattle, WA, NJ: IEEE), 8394–8403.
- Costa, M. H., Tavares, M. C., Richter, C. M., and Castagno, L. A. (1995). "Automatic analysis of electronystagmographic signals" in *38th Midwest symposium on circuits and systems. Proceedings* (Amsterdam, NJ: IEEE), 1349–1352.
- Eggers, S. D., Bisdorff, A. R., Von, B. M., Zee, D. S., Kim, J., Pérez-Fernández, N., et al. (2019). Classification of vestibular signs and examination techniques: nystagmus and nystagmus-like movements. *J. Vestib. Res.* 29, 57–87. doi: 10.3233/VES-190658
- Gulshan, V., Peng, L. H., Coram, M., Stumpe, M. C., Wu, D. J., Narayanaswamy, A., et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 316, 2402–2410. doi: 10.1001/jama.2016.17216
- Halmágyi, G. M., Aw, S. T., Cremer, P. D., Curthoys, I. S., and Todd, M. J. (2001). Impulsive testing of individual semicircular canal function. *Ann. N. Y. Acad. Sci.* 942, 192–200. doi: 10.1111/j.1749-6632.2001.tb03745.x
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition" in *2016 IEEE conference on computer vision and pattern recognition (CVPR)* (Las Vegas, NV, NJ: IEEE), 770–778.
- Henriksson, N. G. (1956). Speed of slow component and duration in caloric nystagmus. *Acta Otolaryngol.* 46, 3–29. doi: 10.3109/00016485609120817
- Jiang, Y., He, J., Xiao, Y., Li, D., and Shen, Y. (2018). "Nystagmus signal feature extraction and tracking for diagnosis of the vestibular system" in *14th IEEE international conference on signal processing* (Beijing, NJ: IEEE), 308–311.
- Karpathy, A., and Fei-Fei, L. (2014). Deep visual-semantic alignments for generating image descriptions. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 664–676. doi: 10.1109/TPAMI.2016.2598339
- Kim, H., and Kim, G. (2020). Single image super-resolution using fire modules with asymmetric configuration. *IEEE Signal Process. Lett.* 27, 516–519. doi: 10.1109/LSP.2020.2980172
- Kim, J. U., and Ro, Y. M. (2019). "Attentive layer separation for object classification and object localization in object detection" in *2019 IEEE international conference on image processing (ICIP)* (Taipei, NJ: IEEE), 3995–3999.
- Lavanyadevi, R., Machakowsalya, M., Nivethitha, J., and Kumar, A. N. (2017). "Brain tumor classification and segmentation in MRI images using PNN" in *2017 IEEE international conference on electrical, instrumentation and communication engineering (ICEICE)* (Karur, NJ: IEEE), 1–6.
- Leigh, R. J., and Zee, D. S. *The neurology of eye movements. 5th Edn.* New York: Oxford University Press (2015).
- Lim, E., Park, J. H., Jeon, H. J., Kim, H., Lee, H., Song, C., et al. (2019). Developing a diagnostic decision support system for benign paroxysmal positional vertigo using a deep-learning model. *J. Clin. Med.* 8:633. doi: 10.3390/jcm8050633
- Litjens, G. J., Kooi, T., Bejnordi, B. E., Setio, A. A., Ciompi, F., Ghafoorian, M., et al. (2017). A survey on deep learning in medical image analysis. *Med. Image Anal.* 42, 60–88. doi: 10.1016/j.media.2017.07.005
- Lu, W., Li, Z., Li, Y., Li, J., Chen, Z., Feng, Y., et al. (2022). A deep learning model for three-dimensional nystagmus detection and its preliminary application. *Front. Neurosci.* 16:930028. doi: 10.3389/fnins
- Mane, S., and Mangale, S. (2018). "Moving object detection and tracking using convolutional neural networks" in *2018 second international conference on intelligent computing and control systems (ICICCS)* (Madurai, India, NJ: IEEE), 1809–1813.
- Newman, J. L., Phillips, J. S., Cox, S. J., Fitzgerald, J. E., and Bath, A. (2019). Automatic nystagmus detection and quantification in long-term continuous eye-movement data. *Comput. Biol. Med.* 114:103448. doi: 10.1016/j.combiomed.2019.103448
- Newman-Toker, D. E., Kattah, J. C., Alvernia, J. E., and Wang, D. Z. (2008). Normal head impulse test differentiates acute cerebellar strokes from vestibular neuritis. *Neurology* 70, 2378–2385. doi: 10.1212/01.wnl.0000314685.01433.0d
- Rathod, J., Wazhmode, V., Sodha, A., and Bhavathankar, P. (2018). "Diagnosis of skin diseases using convolutional neural networks" in *2018 second international conference on electronics, communication and aerospace technology (ICECA)* (Coimbatore, NJ: IEEE), 1048–1051.
- Saha, S., Singh, G., Sapienza, M., Torr, P. H., and Cuzzolin, F. (2016). Deep learning for detecting multiple space-time action tubes in videos. *arXiv preprint arXiv:1608.01529*. doi: 10.5244/C.30.58
- Sertkaya, M. E., Ergen, B., and Togacar, M. (2019, 2019). "Diagnosis of eye retinal diseases based on convolutional neural networks using optical coherence images" in *23rd international conference electronics* (Palanga, NJ: IEEE), 1–5.
- Setio, A. A., Ciompi, F., Litjens, G. J., Gerke, P. K., Jacobs, C., Riel, S. J., et al. (2016). Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. *IEEE Trans. Med. Imaging* 35, 1160–1169. doi: 10.1109/TMI.2016.2536809
- Slama, A. B., Mouelhi, A., Sahli, H., Manoubi, S., Mbarek, C., Trabelsi, H., et al. (2017). A new preprocessing parameter estimation based on geodesic active contour model for automatic vestibular neuritis diagnosis. *Artif. Intell. Med.* 80, 48–62. doi: 10.1016/j.artmed.2017.07.005
- Sundermeyer, M., Ney, H., and Schlüter, R. (2015). From feedforward to recurrent LSTM neural networks for language modeling. *IEEE/ACM Trans. Audio, Speech, Language Process.* 23, 517–529. doi: 10.1109/TASLP.2015.2400218
- Ting, D. S., Cheung, C. Y., Lim, G., Tan, G. S., Quang, N. D., Gan, A. T., et al. (2017). Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA* 318, 2211–2223. doi: 10.1001/jama.2017.18152
- Tra, P. N. H., Hai, N. T., and Mai, T. T. (2016). "Image segmentation for detection of benign and malignant tumors" in *2016 international conference on biomedical engineering (BME-HUST)* (Hanoi, NJ: IEEE), 51–54.
- Van Beuzekom, A. D., and Van Gisbergen, J. A. M. (2002). Interaction between visual and vestibular signals for the control of rapid eye movements. *J. Neurophysiol.* 88, 306–322. doi: 10.1152/jn.2002.88.1.306
- Wagle, N., Morkos, J., Liu, J., Reith, H., Greenstein, J., Gong, K., et al. (2022). aEYE: a deep learning system for video nystagmus detection. *Front. Neurol.* 13:963968. doi: 10.3389/fneur.2022.963968
- Wang, H., Yu, D., Song, N., Su, K., and Yin, S. (2014). Delayed diagnosis and treatment of benign paroxysmal positional vertigo associated with current practice. *Eur. Arch. Otorhinolaryngol.* 271, 261–264. doi: 10.1007/s00405-012-2333-8
- Wu, Z., Zhao, S., Peng, Y., He, X., Zhao, X., Huang, K., et al. (2019). Studies on different CNN algorithms for face skin disease classification based on clinical images. *IEEE Access* 7, 66505–66511. doi: 10.1109/ACCESS.2019.2918221
- Young, T., Hazarika, D., Poria, S., and Cambria, E. (2018). Recent trends in deep learning based natural language processing. *IEEE Comput. Intell. Mag.* 13, 55–75. doi: 10.1109/MCI.2018.2840738
- Zhang, W., Wu, H., Liu, Y., Zheng, S., Liu, Z., Li, Y., et al. (2021). Deep learning based torsional nystagmus detection for dizziness and vertigo diagnosis. *Biomed. Signal Process. Control* 68:102616. doi: 10.1016/j.bspc.2021.102616
- Zhou, H., Zhao, X., Gao, Y., Li, H., and Cheng, R. (2022). Research on nystagmus video classification algorithm based on attention mechanism. *Laser Optoelectron. Prog.* 59:1617001. doi: 10.3788/lop202259.1617001