IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 1, 2020, accepted August 12, 2020, date of publication August 18, 2020, date of current version August 28, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3017680

Disease Classification Based on Synthesis of Multiple Long Short-Term Memory Classifiers Corresponding to Eye Movement Features

YUXING MAO, YINGHONG HE, LUMEI LIU, AND XUESHUO CHEN
State Key Laboratory of Power Transmission Equipment and System Security and New Technology, Chongqing University, Chongqing 400044, China

Corresponding author: Yuxing Mao (myx@cqu.edu.cn)

This work was supported in part by the Key Project of Key Research and Development Program of Ningxia Hui Autonomous Region under Grant 2019BFG02031, and in part by the National Key Research and Development Program of China under Grant 2018YFB2100100.

ABSTRACT Medical research confirms that eye movement abnormalities are related to a variety of psychological activities, mental disorders and physical diseases. However, as the specific manifestations of various diseases in terms of eye movement disorders remain unclear, the accurate diagnosis of diseases according to eye movement is difficult. In this paper, a deep neural network (DNN) method is employed to establish a disease discrimination model according to eye movement. First, multiple eye-tracking experiments are designed to obtain eye images. Second, pupil characteristics, including position and size, are extracted, and the feature vectors of eye movement are obtained from the normalized pupil information. Based on a long short-term memory (LSTM) network, a classifier that corresponds to each feature, which is referred to as a weak classifier, is built. The experimental samples are preclassified, and the classification ability of each weak classifier for different diseases is also calculated. Last, a strong classifier is achieved for disease discrimination by synthesizing all the weak classifiers and their classification abilities. By classification testing for three categories of healthy controls, brain injury patients and vertigo patients, the experimental results demonstrated the efficiency of this method. With the deep learning method, more medical information can be excavated from eye movement to improve the values in disease diagnosis.

INDEX TERMS Eye movement, disease discrimination, deep learning, long short-term memory network, classifier.

I. INTRODUCTION

Human eyes are extensions of the brain, and eye movements are dominated by abundant nerves. The measurement and precise analysis of various parameters of eye movements can be an effective method for the diagnosis of various psychological activities, mental disorders and physical diseases. Eye movement deficits have been established as biomarker across brain injure, schizophrenia, autism, Parkinson's disease, Alzheimer's disease and multiple sclerosis [1], [6], [7], [15], [18], [20], [24], [25], [30]. In terms of existing research results throughout the world, research on congenital nystagmus is the most in-depth research [2]–[5]; it focuses on the treatment of nystagmus and its related diseases. In paper [6]–[10], eye-tracking technologies were applied to research in children with autism spectrum disorder (ASD). Natalia I. Vargas-Cuentas *et al.* [8] developed an eye-tracking algorithm as a potential tool for early diagnosis of ASD in children. Since changes in people's psychological and physiological statuses are likely to be reflected in pupil size, Antoinette *et al.* [9] applied the pupil adaptation ability as a quantitative test method for children with autism. They captured their pupils' adaptation modes in continuous stages of darkness and light and explored the quantitative measuring method of autism. Typical applications of eye movement information in the diagnosis of depression were described in paper [11]–[14]. Shengfu Lu *et al.* had patients with major depressive disorder (MDD) and nondepressed controls complete eye-tracking tasks and analyzed their attention preference to positive, negative, and neutral expressions [11]. According to the results of their study, age can also affect eye performance in free observation tasks.

The associate editor coordinating the review of this manuscript and approving it for publication was Longzhi Yang.

Some scholars conducted research on brain injury and mental disorders via eye movement analysis [15]–[18]. Chao Liu *et al.* [16] applied eye-tracking methods and their tracking parameters to assess the conditions of patients with mental disorders caused by brain injury. Jagla [17] discussed that saccadic eye movements could be utilized as a marker of mental disorders. Research on the diagnosis of schizophrenia also involved eye movement analysis [19]–[23]. Kentaro Morita *et al.* [20] considered eye movements as a biomarker of schizophrenia. They recruited 85 schizophrenia patients and 252 healthy controls to perform free fixation, stable fixation and smooth tracking tasks and employed an integrated eye movement score to distinguish patients with schizophrenia from healthy controls.

Some scholars have applied eye movement information to Alzheimer's disease (AD) research [24]–[26]. Jeremiah K. H. Lim *et al.* [24] analyzed the defects of cerebrospinal fluid analysis, brain imaging and postmortem, which are commonly applied methods for detecting AD pathological biomarkers, presented the evidence of ocular biomarkers in AD and explored potential future research approaches of eye movement analysis for AD diagnosis. Eye tracking technology can also be useful in the diagnosis of Parkinson's disease (PD) [27]–[30]. Lemos *et al.* [27] used eye-tracking technology to explore the functional differences between PD patients and healthy people in horizontal and vertical saccades and provided evidence of the distinction in functional cortical asymmetries between vertical saccades and horizontal saccades in PD patients and healthy controls for the first time.

The literature [31]–[34] showed how eye movement and nystagmus videos can be utilized in dizziness diagnosis. Theekapun Charoenpong *et al.* [32] proposed a method to diagnose vertigo by measuring the eye movement velocity of nystagmus. Videos of eye performance were recorded by infrared cameras, and diagnostic information was obtained in three steps: pupil extraction, eye movement velocity calculation and nystagmus detection. Wang Haowei *et al.* [33] investigated the performance of patients with posterior circulation ischemia vertigo (PCIV), who complained of vertigo and imbalance with PCI in videonystagmography (VNG). VNG tests were performed on 50 patients. The results showed that a vestibular central system and peripheral system could be involved in PCIV, and VNG tests had clinical significance in differential diagnosis and localization of the lesions.

A scientific basis has been established for the application of eye movements in the diagnosis and assessment of mental and physical diseases. Although some progress has been made in relevant research, it remains relatively weak in general. Current research is mainly aimed at a single disease with a unitary test method, and comprehensive analysis is lacking. On the other hand, traditional methods are employed in information processing, which is not conducive to the deep mining of the hidden features of eye movement information. Thus, reliability and versatility are limited. Two key problems need to be solved in the application of eye movement information in medical research. The first problem is to design target-tracking experiments based on the research of medical theory to induce and stimulate the appearance of eye movement disorders to ensure that the acquired eye movement information has greater diagnostic value. The second problem is to extract and quantify eye information and medical characteristics to obtain digitized data, and then explore the inner relationship between sys data and disease types. In recent years, the tremendous progress in artificial intelligence (AI) and deep learning technology can have great benefit to medical feature extraction and disease discrimination of eye images.

In this paper, experiments are designed to obtain eye videos, and then multiple features are extracted for analysis, which involves deep neural networks and machine learning algorithms. Based on the detected pupil area, we calculate 9 parameters as optional features. With the eye movement features, weak classifiers are established using a long short-term memory (LSTM) network [35], [36], and their weights are associated with their classifying performance on different diseases. Thus, a unified classification model for the diagnosis of multiple diseases and evaluation of curative effects are established. The original intention of this paper is to use advanced AI technology to extract more valuable eye movement features by supervised learning method. This study shows that the application of advanced AI technology in the pathological analysis of eye movement has obvious advantages and good prospects.

II. MATERIALS AND METHODS

A. EYE TRACKER AND EXPERIMENTAL DESIGN

To employ eye activities for scientific research, the first step is to record and extract eye movement information. Previously, electronystagmography (ENG) was performed. The physiological basis of ENG was to record the potential changes between the cornea and the retina using electronic instruments. The potential changes caused by eye movement can be recorded by attaching electrodes around the eyeball. This bioelectric signal is collected and amplified and then displayed as graph, which is referred to as ENG. The accuracy of ENG is susceptible to many external factors, including recent medications taken by the patient, the state of arousal during the test, interference with other biological signals and the experience of the operators *et al.* ENG is also complicated and expensive. With the development of computer video technology in recent years, VNG, where camera video is used to record the entire process of eye movement, has become the main way to obtain eye movement information.

Our laboratory has developed an infrared video eye tracker according to the application requirements of eye movement research, as shown in FIGURE 1. Its core components include CMOS cameras with infrared LED lighting, a wireless video transmission module based on WIFI, a battery, and a helmet device. The software that runs on the host computer provides the user interface for the control of the eye tracker and data reception and saving. On the interface menu, users can choose different experiments according to the experimental scheme.

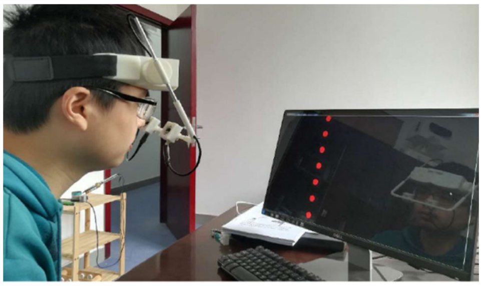

FIGURE 1. **The eye tracker developed by our team.**

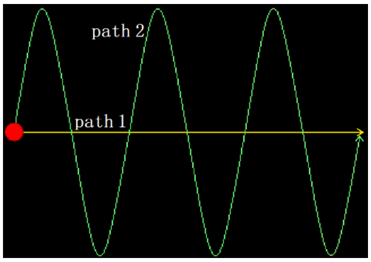

FIGURE 2. **Moving trajectory of spots in these two experiments.**

Thus, corresponding images can be represented successively on display devices, which enables subjects to track specific targets or browse interesting areas on the image. Eye videos can be obtained, and pupil information can be extracted for further study in this paper. The technical parameters of the eye tracker are described as follows:

Image resolution: $320 * 240 \sim 640 * 480$

Frame rate: 30 to 60 fps (adjustable)

Adjustable recording time

8 kinds of target-tracking experiments (expandable)

Extracted pupil features: position, area, shape, etc.

Maximum continuous battery supply time: 3 hours

A variety of experimental schemes are designed to generate and display static or dynamic images on the screen. These experimental schemes include the gaze test, saccade test, ocular pursuit test, optokinetic test, positional test, and positioning test.

In this study, we employ two experimental schemes: the optokinetic test and the ocular pursuit test, as shown in FIGURE 2. In these two experiments, a red light spot moves along the set trajectory, and subjects are told to gaze at the point and then follow it as it changes its position. In the first experiment, the light spot circularly moves at uniform speed from left to right along the horizontal median line of the screen. In the second experiment, the light spot repeatedly moves along a sinusoidal trajectory from left to right at uniform speed.

The subjects are required to sit directly in front of the screen while wearing an eye tracker. The experiment begins, and the infrared camera records videos while the subjects perform experiments. Eye movement images are concurrently transmitted to the host computer via the WIFI module of the eye tracker. After recording a set number of frames, the experiment and video recording stop.

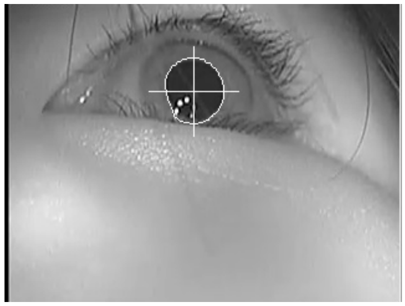

FIGURE 3. **Pupil extraction results.**

B. EYE MOVEMENT FEATURE EXTRACTION

The use of infrared LED lighting in eye image acquisition not only enhances the ability to resist light interference but also makes the pupils more visible and facilitates extraction, as shown in FIGURE 3. The pixels that correspond to the pupil area in each video frame can be obtained by conventional image analysis methods. Therefore, the pupil parameters, including position, range, size of its circumscribed rectangle, circumscribed rectangle aspect ratio, angle of the smallest circumscribed rectangle, symmetry, shape, etc., can be calculated. These parameters are the feature parameters in this paper.

Because subjects' pupils may vary in size, initial position, and FOV, the feature parameters need to be normalized. The normalized features are given by Equation (1).

$$g_i = \left(f_i - \min_{i=1\sim M}(f_i)\right) / \left(\max_{i=1\sim M}(f_i) - \min_{i=1\sim M}(f_i)\right) \quad (1)$$

where M is the total frame number in a video, f_i and g_i are certain feature parameters obtained from the i-th frame before normalization and after normalization, respectively.

After the processing, all feature parameters will be in the range $0 \sim 1$, and $\boldsymbol{g} = [g_1, g_2, \ldots, g_M]$ forms a feature vector, which corresponds to a certain parameter of a subject's pupil in a specific experimental scheme. For a subject to participate in p experimental schemes with q parameters extracted from each scheme, $p * q$ feature vectors exist for subsequent classification.

C. WEAK CLASSIFIER BUILDING BASED ON LSTM

1) LSTM NETWORK FOR CLASSIFICATION

The information contained in feature vectors is time-varying, and some symptoms may be hidden at special time points. Therefore, the use of recurrent neural network (RNN), which is a class of artificial neural network, is suitable. In an RNN, the connections among units form a directed cycle to create an internal state of the network, which enables it to exhibit dynamic temporal behavior. RNNs can utilize their internal memory to process arbitrary sequences of inputs. RNNs have repeating neuron-like blocks, and each unit has a simple structure, such as a tanh activation layer, so can neither handle long-term memory problems nor evaluate the importance of memory content for selection.

Long short-term memory (LSTM) is a special deep learning RNN, and an LSTM network is well-suited for

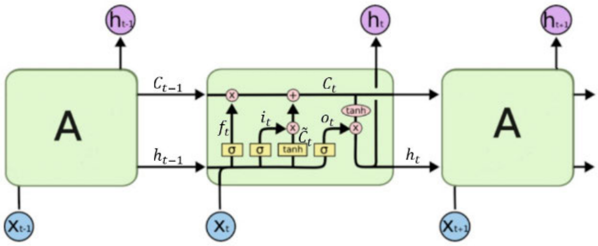

FIGURE 4. **LSTM blocks with input, output, and forget gates.**

learning from experience to classify, process and predict time series when very long time lags of unknown size exist between important events.

LSTM blocks contain three or four "gates", which are used to control the flow of information into or out of their memory; refer to FIGURE 4. These gates are implemented using a logistic function to compute a value between 0 and 1. Multiplication is applied with this value to partially allow or deny information to flow into or out of the memory. Specifically, an "input gate" controls the extent to which a new value flows into the memory; a "forget gate" controls the extent to which a value remains in memory; and an "output gate" controls the extent to which the value in memory is used to compute the output activation of the block.

In FIGURE 4, C_t is the cell state vector for carrying historical memory and then adding to new output. The specific processing algorithm of the cell is described in Equation (2).

$$\begin{cases} f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \\ i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \\ \tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \\ C_t = f_t^* C_{t-1} + i_t^* \tilde{C}_t \\ o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \\ h_t = o_t^* \tanh(C_t) \end{cases} \tag{2}$$

where σ is a sigmoid function. When using the LSTM network for classification, the first step is to determine *h_num*, which is the number of neurons in the hidden layer of the cell. The input of each cell is the feature data of a time slice; assume that it is a one-dimensional vector that contains *x_num* elements. The dimensions of the weight matrices W_f, W_i, W_c, and W_o are *h_num* rows and *h_num* + *x_num* columns. The forget gate vector f_t, which represents the weight of remembering old information, has the dimensions of *h_num* row and 1 column.

2) WEAK CLASSIFIERS BUILDING BASED ON SINGLE FEATURE

An LSTM network is used to classify the eye movement features in this paper. Based on the eye-tracking experimental scheme, the time-series feature vectors obtained from an eye video are divided into equal-length time slices, and then input to each cell. After the input is completed, the output information that is obtained will be input into a fully connected network, which will output the classification results, as shown in FIGURE 5.

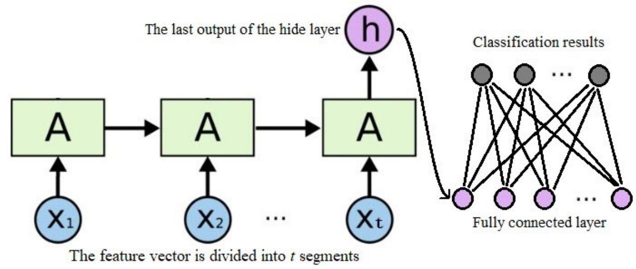

FIGURE 5. **Technical solution of LSTM network for time-varying eye feature classification.**

Dividing the video involves dividing M frames into t time slices. For each feature, the input of a cell is a one-dimensional vector that contains $x\_num = M/t$ elements. After inputting all of the vectors, the output classification result will be calculated via the fully connected network. The training is performed on the labeled samples to obtain the weights and bias parameters of this LSTM-based weak classifier.

D. STRONG CLASSIFIER CONSTRUCTION

The classifiers established using the LSTM network are referred to as weak classifiers since they are aimed at a single feature vector. The number of feature vectors is given by $m = p^*q$, where p is the number of experiments, and q is the number of features that need to be extracted. In the process of multiclassification based on eye features, some features are helpless in diseases classification. Thus, an evaluation of the classification ability of weak classifiers for better fusion is necessary to obtain a strong classifier.

1) CLASSIFICATION ABILITY EVALUATION OF WEAK CLASSIFIERS

A classifier with better classification ability not only has higher accuracy but can also distinguish positive samples more clearly.

Set the number of labeled sample categories to k. The output of a LSTM weak classifier is $\mathrm{c} = [c_1, c_2, \ldots, c_k]$, where c_i represents the probability that the input sample belongs to the i-th category calculated by this classifier.

Consider N as the total number of samples and N_i as the number of samples in the i-th category. Thus,

$$N = N_1 + N_2 + \ldots + N_k \tag{3}$$

Subsequently, m weak classifiers are constructed for m feature vectors. $P_i^{j,l}$ is defined as the probability that the l-th sample belongs to the j-th category calculated by the i-th weak classifier, as shown by Equation (4).

$$P_i^{j,l} = \mathrm{c}_j \left| \begin{array}{l} \text{sample } l \;\; l = 1, 2, \cdots, N \\ \text{classifier } i \;\; i = 1, 2, \ldots, m \end{array} \right. \quad j = 1, 2, \ldots, k \tag{4}$$

$s_i^{j,l}$ is defined as the risk of obtaining inaccurate classification results when classifying the l-th sample, which effectively belongs to the j-th category using the i-th classifier. The definition of $s_i^{j,l}$ is the relative value of the difference between the maximum probability value of the correct class and the incorrect class, as given by Equation (5).

$$s_i^{j,l} = \left(P_i^{j,l} - \max_{\substack{r=1\sim k \\ r\neq j}} P_i^{r,l}\right) / \left(\max_{r=1\sim k} P_i^{r,l} - \min_{r=1\sim k} P_i^{r,l}\right) \quad (5)$$

The classification result is correct when $s_i^{j,l} \geq 0$, and the larger is the value of $s_i^{j,l}$, the lower is the risk of obtaining inaccurate classification results. The mean value of risk $s_i^{j,l}$ among the examples in the j-th category can be obtained by Equation (6).

$$C_i^j = \sum_{l=1}^{N_j} s_i^{j,l} / N_j \quad (6)$$

C_i^j is used to calculate the i-th classifier's classification ability for the j-th class.

To facilitate subsequent processing, the sigmoid function is used to transform C_i^j into W_i^j, as shown in Equation (7).

$$W_i^j = 1/\left(1 + e^{-C_i^j}\right) \quad (7)$$

where the value of W_i^j is between 0 and 1 to represent the normalized value of the classification ability of the i-th classifier for the j-th category.

2) STRONG CLASSIFIER CONSTRUCTION METHOD

The weak classifiers are combined to obtain a more efficient multiclassifier. After each weak classifier classifies the input samples, the final classification result can be acquired by the combination of results obtained from the weak classifiers. The probability of a test sample to be classified into the j-th category is given by Equation (8).

$$P^j = \left(\sum_{i=1}^{m} W_i^j P_i^j\right) / \left(\sum_{i=1}^{m} W_i^j\right) \quad (8)$$

The maximum value in $P^j, j = 1, \ldots, k$ corresponds to the classification result of the input test sample.

III. RESULTS

A. PARTICIPANT RECRUITING AND GROUPING

The optokinetic test and ocular pursuit test are carried out in this research. Subjects with eye trackers are instructed to gaze at a dynamic point on a screen and track its moving trajectory, while the camera on the eye tracker records 250 frames of images at a frame rate of 30 fps.

The pupil information is extracted for each frame of an image, and the abscissa x, ordinate y, and area s of the pupil are employed as eye movement features. Set f_i, $i = 1, 2, \ldots, 6$ as one-dimensional feature vectors with 250 elements, which correspond to 250 frames in a video. The normalized feature vector $g_i (g_i \in R^{250})$ can be obtained according to formula (1). Six kinds of features exist, as listed in TABLE 1.

We cooperate with medical institutions and invite 98 subjects to participate in the experiment, including 34 healthy volunteers, 34 patients with brain injury (including cerebral infarction), and 30 patients with vertigo, which are marked as category 1, category 2 and category 3, respectively. All 98 samples were divided stochastically into 3 groups, of which 2 groups (32 subjects in each group) are for training the weak classifiers and computing the classification capabilities and 1 group (34 subjects) is for testing the strong classifier. The distribution of the three category samples in the three groups is shown in TABLE 2.

TABLE 1. Definition of 6 eye movement feature vectors g_i.

	x	y	s
Optokinetic test	g_1	g_2	g_3
Ocular pursuit test	g_4	g_5	g_6

TABLE 2. Constitution of three sample sets.

	Category 1	Category 2	Category 3
Training group 1	11	11	10
Training group 2	11	11	10
Test group	12	12	10

We performed eye-tracking experiments on all 98 participants, and for each subject, 6 feature vectors exist. One sample is randomly selected from each of the three categories of samples, and the same feature of the three samples are described in a diagram. The 6 diagrams are shown in FIGURE 6.

According to the diagrams, intuitively distinguishing the differences among different categories of samples or to know which feature is the most valuable for each category of samples is challenging. Thus, the construction of a classifier is necessary to effectively perform the classification.

B. CLASSIFICATION ABILITY OF WEAK CLASSIFIERS

Six 250-element feature vectors exist for six features of each sample in all 3 types of 98 samples. For each feature, an LSTM classifier is constructed using the same structure. Divide the 250 elements in each feature vector into 10 time slices in order, with 25 elements for each slice. Thus, the number of input nodes is $x\_num = M/t = 250/10 = 25$. The number of hidden layer neurons, h_num is 64, and the number of output nodes is 3.

Thirty-two samples from training group *1* are used to train an LSTM classifier for each of the 6 features, and Python + TensorFlow is applied for the network construction. FIGURE 7 shows the change in classification accuracy of the six classifiers with an increase in the number of iterations during the training process.

As shown in FIGURE 7, the convergence rates of the six classifiers vary. However, the classifiers can correctly classify all training samples in less than 60 iterations. Thus, 6 weak classifiers are obtained based on the samples in training group 1.

We utilize these 6 weak classifiers to classify the samples in training group 2 and evaluate the classifiers' classification ability. The feature vectors of the samples in training group

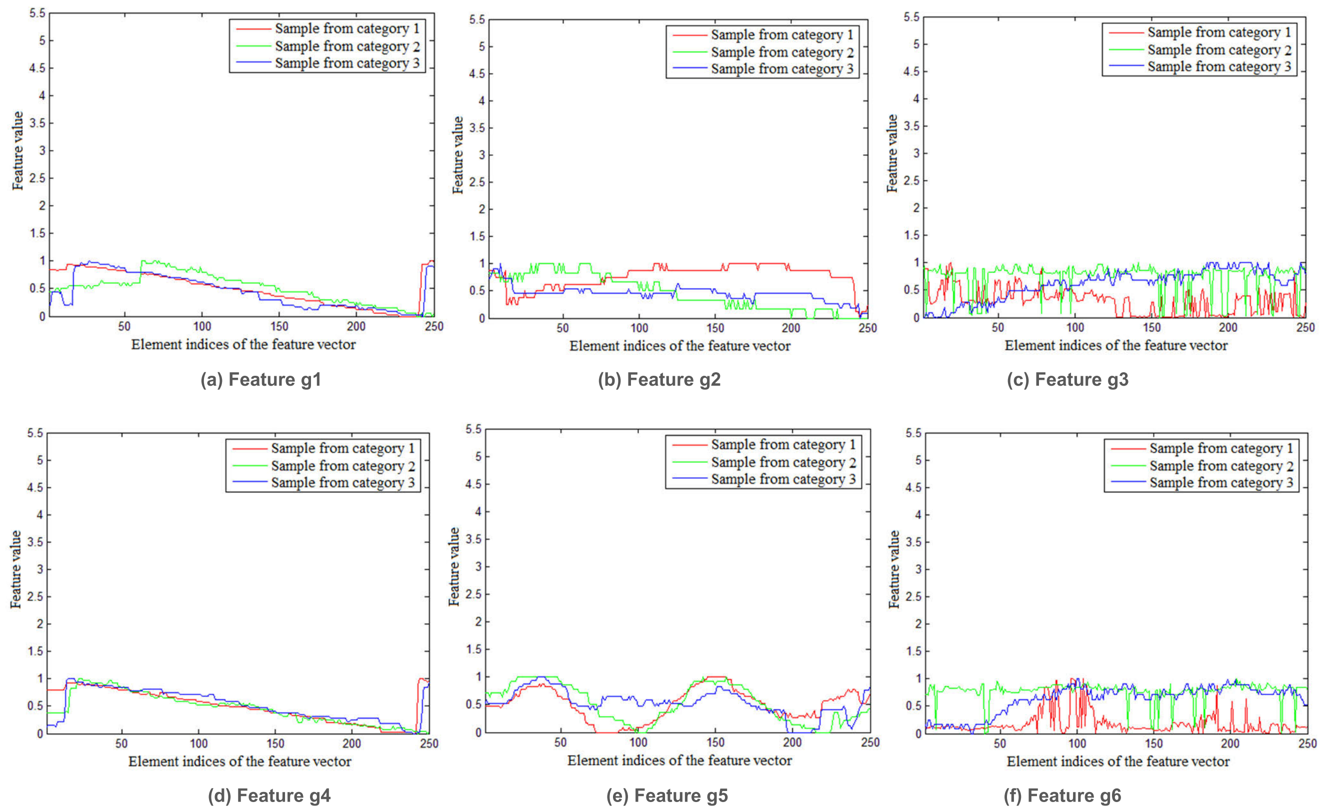

FIGURE 6. Feature diagrams of subjects from 3 classes.

TABLE 3. Recall of the six weak classifiers for each category of samples in training group 2 and the total accuracy.

	Classifier 1	Classifier 2	Classifier 3	Classifier 4	Classifier 5	Classifier 6
Category 1	0.727	0.636	0.818	0.727	0.727	0.909
Category 2	0.545	0.727	0.727	0.636	0.636	0.727
Category 3	0.700	0.800	0.800	0.700	0.800	0.800
Total	0.656	0.719	0.781	0.688	0.719	0.813

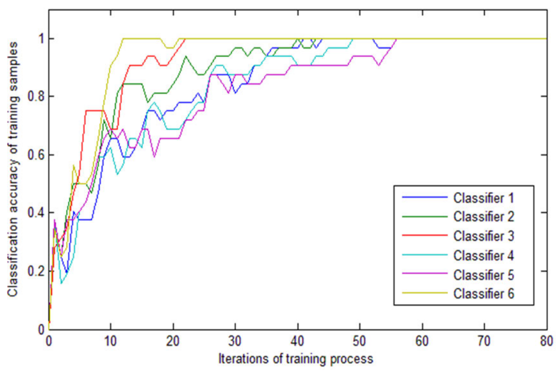

FIGURE 7. Change in the classification accuracy during the process of training.

2 are input into their corresponding weak classifiers. The recall of the six weak classifiers for the 3 categories of samples in training group 2 and the total accuracy for each classifier to training group 2 are shown in TABLE 3.

To construct a strong classifier using weak classifiers, their classification ability needs to be calculated. Although the recall of each weak classifier can reflect the classification ability in a certain category, the reflection is not detailed enough and has only the judgment results. No specific probability value exists for a sample that belongs to a class.

According to the definition of W_i^j in Equation (7), the output results of the fully connected layers after the LSTM networks can be used to calculate W_i^j, which represents the normalized value of the classification ability of the i-th classifier for the samples in category j. The classification ability calculated by W_i^j is shown in FIGURE 8, which is not the same as the previously calculated recall. Use of the classification ability W_i^j to construct of the strong classifier is helpful in solving the overfitting problem caused by a small sample size.

C. CLASSIFICATION RESULTS OF THE STRONG CLASSIFIER

After calculating the classification ability of each weak classifier, the strong classifier is constructed according to Equation (8) to calculate P^j, $j = 1, 2, 3$, which is the probability that a test sample will be classified into a specific class and the class with the maximum probability value is the ultimate classification result.

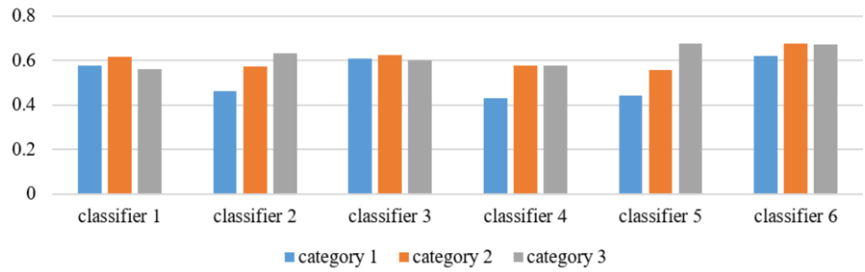

FIGURE 8. Classification ability of 6 weak classifiers for the three categories of samples.

TABLE 4. The key indicators of six weak classifiers and strong classifier testing with the samples in test group.

		Weak classifier 1	2	3	4	5	6	Strong classifier
Category1	Precision	0.7778	0.6667	**0.9091**	**0.9091**	0.5000	0.8462	**0.9167**
	Recall	0.5833	0.8333	0.8333	0.8333	0.6667	**0.9167**	**0.9167**
	Fscore	0.6667	0.7407	**0.8696**	**0.8696**	0.5714	0.8800	**0.9167**
Category2	Precision	0.5238	**0.8182**	0.6875	0.6667	0.7000	0.7500	**0.9231**
	Recall	**0.9167**	0.7500	**0.9167**	0.8333	0.5833	0.7500	**1.0000**
	Fscore	0.6667	**0.7826**	**0.7857**	0.7407	0.6364	0.7500	**0.9600**
Category3	Precision	0.5000	**1.0000**	**1.0000**	0.8750	0.7500	0.7778	**1.0000**
	Recall	0.2000	**0.8000**	0.7000	0.7000	0.6000	0.7000	**0.9000**
	Fscore	0.2857	**0.8889**	0.8235	0.7778	0.6667	0.7368	**0.9474**
Total Accuracy		0.5882	0.7941	**0.8235**	0.7941	0.6176	0.7941	**0.9412**

In order to verify the effectiveness of our method, we compute four key indicators of precision, recall, Fscore, and total accuracy to evaluate the performances of the classifiers. By verifying the six weak classifiers and one strong classifier with thirty-four samples of the test group, the precision, recall and Fscore indicators corresponding to each category and the total accuracy are obtained and presented in TABLE 4.

As shown in TABLE 4, the six weak classifiers have different classification performances to the three sample categories, which indicates that the correlation strength of eye movement features to specific diseases is different. For example, weak classifier 2 has a good classification effect for category 3, weak classifier 4 for category 1, and weak classifier 3 for both category 1 and 2. In terms of recall, weak classifier 1 performs best on category 2, while weak classifier 6 on category 1. Overall, among the six weak classifiers, weak classifier 3 has the highest total accuracy, reaching 82.35%. Due to the classification abilities of each weak classifier for different categories being considered, the strong classifier gains the best performances among all indicators. Of the 34 test samples, 32 samples are classified correctly, and the total accuracy is 94.12%. This result verified the effectiveness of our classification method.

In addition, we draw the ROC curves and calculate the AUC values for each classifier to compare and analyze the performance in detail, as shown in FIGURE 9. Because the experiment includes three categories, but these performance indicators are only defined for binary classification, we convert the 3 categories into a 3 binary classification problems by selecting one category and evaluating it against the remaining two categories to calculate the parameters. Figures 9(a) $\sim$ (f) respectively correspond to six weak classifiers, each showing 5 curves. Three of them are ROC curves relating to the above 3 binary classification problems, and the other two are the average values obtained by adopting macro and micro methods from the three curves. Meanwhile, the AUC values were calculated for each of the 5 curves, which also represent the classification performance.

From these figures, we can see that the performance of the weak classifier is not stable. The performance of different classifiers is very different, and the classification performance of the same classifier for different categories is also very different. Because the weak classifier comes from different features, and some features are not related to some diseases, it has no classification value. As can be seen from the FIGURE 9, weak classifier 3 has the best performance and weak classifier 5 has the worst performance, which are consistent with the parameters listed in the TABLE 4.

In order to further evaluate the disease classification ability of the strong classifier, we analyze the details of the wrongly classified samples. As shown in TABLE 5, there is one sample in category 1 is classified into category 3 and another in category 3 is classified into category 2. Although the two samples are wrongly classified, the differences between the probabilities corresponding to the real category and the classification category is low ($\approx$8%). Similarly, we compute and present ROC curves and AUC values in FIGURE 10. All the AUC values towards the three categories are close to 1, showing better performance than that of other classifiers. It is verified that the strong classifier has high effectiveness and reliability in classifying these diseases.

IV. DISCUSSION

Medical research has confirmed that eye movement is related to a variety of psychological and physical diseases. However, due to an unclear mechanism, obtaining a direct correlation

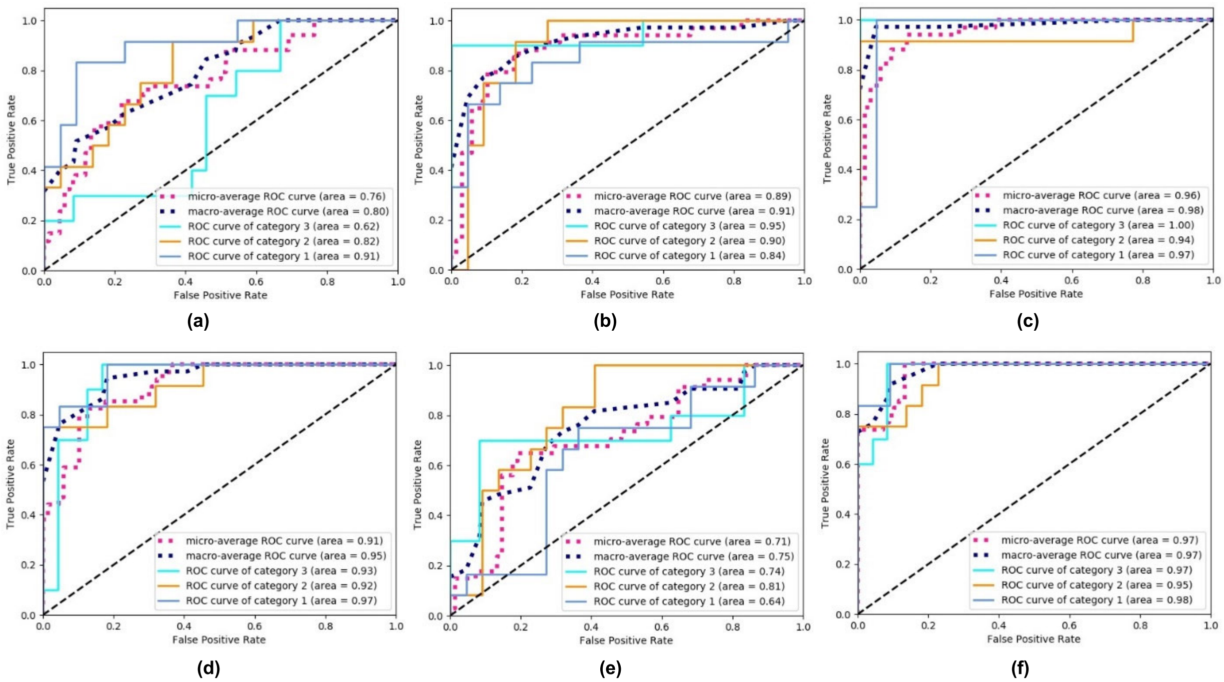

FIGURE 9. **The ROC curves and AUC values of six LSTM classifiers (a) Weak classifier 1 (b) Weak classifier 2 (c) Weak classifier 3 (d) Weak classifier 4 (e) Weak classifier 5 (f) Weak classifier 6.**

TABLE 5. **The details of the two samples wrongly classified by strong classifier.**

Real category	Three classification probabilities Category 1	Category 2	Category 3	Classification category
1	0.383599	0.146901	0.469500	3
3	0.096607	0.490268	0.413125	2

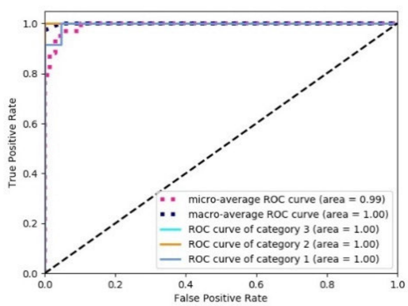

FIGURE 10. **The ROC curves and AUC values of strong classifier.**

between eye movement characteristics and disease discrimination is difficult. Thus, medical diagnosis research usually focuses on a single disease because distinguishing among multiple diseases is especially difficult.

Our research intends to take advantage of AI technology and apply deep neural networks to classify diseases according to eye movement information. During training, the effectiveness of each classifier is evaluated and calculated to automatically design the weights of these weak classifiers, and then a strong classifier can be obtained by their weighted combination. The strong classifier is able to classify multiple diseases with high accuracy and can be more practical than diagnosis methods for a single disease.

The key point of research in this paper is to design multiple experimental schemes in the absence of prior pathological knowledge and extract multiple features to build a classifier. By training with two groups of training samples, the evaluation and selection of weak classifiers are realized, which is helpful for the application of multiple eye information in medical diagnosis.

The drawbacks of our study lie in several aspects: First, during the test, the subjects are required to stay awake, participate in the test as required, and try to keep their heads still, which limits the scope of use. Second,there are too few samples, resulting in the classifier not being sufficiently robust. Third, the algorithm should be optimized to reduce the calculation complexity and shorten the time consumption. Future work focuses on designing more effective experimental schemes for specific diseases based on the investigation of a medical mechanism to extract more valuable features and improve the ability of classifiers to distinguish specific diseases. Additional experimental samples should be collected to ensure the reliability of data collection and improve the robustness of the classifiers. Currently, we are cooperating with relevant medical institutions to design experimental schemes and conduct relevant research on Alzheimer's disease, Parkinson's disease, depression, etc.

V. CONCLUSION

To deep mine eye information for maximum utilization in medical diagnosis, AI technology is applied for self-learning and diagnosis. A variety of eye-tracking experiments are designed to obtain eye images, and after image processing, pupil information is excavated and normalized feature vectors are formed. For each feature, a weak classifier is constructed using an LSTM network. The ability of each weak classifier is evaluated by a self-learning method to obtain their weights, and then a strong classifier for multiple diseases is obtained by synthesizing the weak classifiers. The effectiveness of this method, which is validated by experiments, can be helpful for future diagnosis of multiple diseases or other applications.

REFERENCES

[1] C. K. Sheehy, E. Bensinger, A. Romeo, L. Rani, B. Shi, J. M. Gelfand, S. B. Stevenson, and A. J. Green, "High fidelity fixational eye movements as a novel potential biomarker for disability in multiple sclerosis," *Multiple Sclerosis J.*, vol. 24, no. 2, pp. 572–573, 2018.

[2] K. Feil, N. Boettcher, O. Kremmyda, C. Muth, J. Teufel, A. Zwergal, T. Brandt, and M. Strupp, "Pharmacotherapy of vestibular disorders, nystagmus and cerebellar disorders," *Laryngo-Rhino-Otologie*, vol. 97, no. 1, pp. 14–23, 2018.

[3] J. H. Rim, S.-T. Lee, H. Y. Gee, B. J. Lee, J. R. Choi, H. W. Park, S.-H. Han, and J. Han, "Accuracy of next-generation sequencing for molecular diagnosis in patients with infantile nystagmus syndrome," *JAMA Ophthalmology*, vol. 135, no. 12, pp. 1376–1385, 2017.

[4] E.-S. Lee, K.-B. Sung, and T.-K. Lee, "Teaching video NeuroImages: Upbeat and horizontal gaze-evoked nystagmus in bilateral medial medullary infarction," *Neurology*, vol. 89, no. 19, p. e238, Nov. 2017.

[5] O. Kremmyda, S. Bardins, A. Straube, and T. Eggert, "Changes in quick phases of downbeat nystagmus during visual fixation," *J. Neurol.*, vol. 264, no. S1, pp. S93–S95, Oct. 2017.

[6] M. Murias, S. Major, K. Davlantis, L. Franz, A. Harris, B. Rardin, M. Sabatos-DeVito, and G. Dawson, "Validation of eye-tracking measures of social attention as a potential biomarker for autism clinical trials," *Autism Res.*, vol. 11, no. 1, pp. 166–174, Jan. 2018.

[7] J. Bradshaw, F. Shic, A. N. Holden, E. J. Horowitz, A. C. Barrett, T. C. German, and T. W. Vernon, "The use of eye tracking as a biomarker of treatment outcome in a pilot randomized clinical trial for young children with autism," *Autism Res.*, vol. 12, no. 5, pp. 779–793, May 2019.

[8] N. I. Vargas-Cuentas, A. Roman-Gonzalez, R. H. Gilman, F. Barrientos, J. Ting, D. Hidalgo, K. Jensen, and M. Zimic, "Developing an eye-tracking algorithm as a potential tool for early diagnosis of autism spectrum disorder in children," *PLOS ONE*, vol. 12, no. 11, pp. 1–13, 2017.

[9] A. S. DiCriscio and V. Troiani, "Pupil adaptation corresponds to quantitative measures of autism traits in children," *Sci. Rep.*, vol. 7, no. 6476, pp. 1–9, Dec. 2017.

[10] J. Wei, J. Fang, and W. Zhao, "Word learning with multiple cues in children with autistic spectrum disorder: Evidence from eye movements," *Acta Psychologica Sinica*, vol. 46, no. 3, pp. 385–395, 2014.

[11] S. Lu, J. Xu, M. Li, J. Xue, X. Lu, L. Feng, B. Fu, G. Wang, N. Zhong, and B. Hu, "Attentional bias scores in patients with depression and effects of age: A controlled, eye-tracking study," *J. Int. Med. Res.*, vol. 45, no. 5, pp. 1518–1527, Oct. 2017.

[12] M. Li, S. Lu, G. Wang, L. Feng, B. Fu, and N. Zhong, "Alleviated negative rather than positive attentional bias in patients with depression in remission: An eye-tracking study," *J. Int. Med. Res.*, vol. 44, no. 5, pp. 1072–1086, Oct. 2016.

[13] A. Duque and C. Vazquez, "A failure to show the efficacy of a dot-probe attentional training in dysphoria: Evidence from an eye-tracking study," *J. Clin. Psychol.*, vol. 74, no. 12, pp. 2145–2160, Dec. 2018.

[14] X. Xu, X. Lyu, X. Chang, and Y. Yu, "Exploration of eye movement of depressive disorder liver Qi stagnation syndrome patients," *Chin. Arch. Traditional Chin. Med.*, vol. 35, no. 4, pp. 889–891, 2017.

[15] M. Hunfalvay, "Eye tracking as a biomarker for concussion," *J. Head Trauma Rehabil.*, vol. 35, no. 2, pp. E183–E184, 2020.

[16] C. Liu, Q. Zhang, and W. Cai, "The parameters of eye-track in patients with mental disorder due to brain damage," *Forensic Sci. Int.*, vol. 277, no. 1, p. 64, 2017.

[17] F. Jagla, "Saccadic eye movements as a marker of mental disorders," *Physiological Res.*, pp. S365–S371, Oct. 2016.

[18] S. Brakemeier, A. Sprenger, I. Meyhöfer, J. E. McDowell, L. H. Rubin, S. K. Hill, M. S. Keshavan, G. D. Pearlson, C. A. Tamminga, E. S. Gershon, and S. S. Keedy, "Smooth pursuit eye movement deficits as a biomarker for psychotic features in bipolar disorder—Findings from the PARDIP study," *Bipolar Disorders*, pp. 1–10, 2019. [Online]. Available: https://onlinelibrary.wiley.com/doi/pdfdirect/10.1111/bdi.12865

[19] J. E. Silberg, I. Agtzidis, M. Startsev, T. Fasshauer, K. Silling, A. Sprenger, M. Dorr, and R. Lencer, "Free visual exploration of natural movies in schizophrenia," *Eur. Arch. Psychiatry Clin. Neurosci.*, vol. 269, pp. 407–418, Jan. 2018.

[20] K. Morita, K. Miura, M. Fujimoto, H. Yamamori, Y. Yasuda, M. Iwase, K. Kasai, and R. Hashimoto, "Eye movement as a biomarker of schizophrenia: Using an integrated eye movement score," *Psychiatry Clin. Neurosci.*, vol. 71, no. 2, pp. 104–114, Feb. 2017.

[21] S. Dowiasch, B. Backasch, W. Einhäuser, D. Leube, T. Kircher, and F. Bremmer, "Eye movements of patients with schizophrenia in a natural environment," *Eur. Arch. Psychiatry Clin. Neurosci.*, vol. 266, no. 1, pp. 43–54, Feb. 2016.

[22] X. Yangyang, Z. Tianhong, W. Jijun, and X. Zeping, "Research progress of smooth pursuit eye movement in schizophrenia," *J. Int. Psychiatry*, vol. 44, no. 3, pp. 395–397, 2017.

[23] Y. Luo, P. Xitian, X. Shuai, and L. Hongying, "An eye tracker used for the detection of schizophrenia," *Beijing Biomed. Eng.*, vol. 35, no. 6, pp. 639–644, 2016.

[24] J. K. H. Lim, Q.-X. Li, Z. He, A. J. Vingrys, V. H. Wong, N. Currier, J. Mullen, B. V. Bui, and C. T. Nguyen, "The eye as a biomarker for Alzheimer's disease," *Frontiers Neurosci.*, vol. 10, no. 536, pp. 1–14, 2016.

[25] O. A. Coubard, "What do we know about eye movements in Alzheimer's disease? The past 37 years and future directions," *Biomarkers Med.*, vol. 10, no. 7, pp. 677–680, Jul. 2016.

[26] G. Fernández, F. Manes, L. E. Politi, D. Orozco, M. Schumacher, L. Castro, O. Agamennoni, and N. P. Rotstein, "Patients with mild Alzheimer's disease fail when using their working memory: Evidence from the eye tracking technique," *J. Alzheimer's Disease*, vol. 50, no. 3, pp. 827–838, Feb. 2016.

[27] J. Lemos, D. Pereira, L. Almendra, D. Rebelo, M. Patrício, J. Castelhano, G. Cunha, C. Januário, L. Cunha, A. Freire, and M. Castelo-Branco, "Distinct functional properties of the vertical and horizontal saccadic network in health and Parkinson's disease: An eye-tracking and fMRI study," *Brain Res.*, vol. 1648, pp. 469–484, Oct. 2016.

[28] S. Stuart, L. Alcock, A. Godfrey, S. Lord, L. Rochester, and B. Galna, "Accuracy and re-test reliability of mobile eye-tracking in Parkinson's disease and older adults," *Med. Eng. Phys.*, vol. 38, no. 3, pp. 308–315, Mar. 2016.

[29] E. Pretegiani and L. M. Optican, "Eye movements in Parkinson's disease and inherited parkinsonian syndromes," *Frontiers Neurol.*, vol. 8, no. 592, pp. 1–7, Nov. 2017.

[30] A. Przybyszewski, S. Szlufik, J. Dutkiewicz, P. Habela, and D. Koziorowski, "Eye movement (EM) measurements as objective and precise biomarker for symptoms development in patients with Parkinson's disease," *J. Neurological Sci.*, vol. 357, pp. e283–e284, Oct. 2015.

[31] J. L. Johnston, P. M. Daye, and G. T. D. Thomson, "Inaccurate saccades and enhanced vestibulo-ocular reflex suppression during combined eye–head movements in patients with chronic neck pain: Possible implications for cervical vertigo," *Frontiers Neurol.*, vol. 8, no. 23, pp. 1–9, Jan. 2017.

[32] T. Charoenpong, P. Pattrapisetwong, and V. Mahasitthiwat, "A new method to detect nystagmus for vertigo diagnosis system by eye movement velocity," in *Proc. 14th IAPR Int. Conf. Mach. Vis. Appl. (MVA)*, May 2015, pp. 174–177.

[33] W. Haowei, G. Ying, Z. Jing, and Z. Huifang, "The manifestation of videonystagmography in the patients with posterior circulation ischemia vertigo," *J. Clin. Otorhinolaryngology Head Neck*, vol. 32, no. 2, pp. 99–102, 2018.

[34] Z. Xiaoliang, F. Wenqing, and C. Dongmin, "Analysis of clinical diagnosis application of videonystagmograph combined with inquiry and physical examination in vertigo patients," *Chin. J. Mod. Drug Appl.*, vol. 11, no. 12, pp. 32–34, 2017.

[35] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[36] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017.

YUXING MAO received the B.S. and M.S. degrees from the Radio Electrical Department, Beijing University, China, in 1989 and 1992, respectively, and the Ph.D. degree in electrical engineering from Chongqing University, China, in 2004. He worked as a Visiting Scientist with the Center for Pattern Recognition and Machine Intelligence, CENPARMI, Concordia University, Canada, in 2005. He is currently a Ph.D. Supervisor and an Associate Professor with the School of Electrical Engineering, Chongqing University. He has published 60 articles in these fields. His research interests include image processing, computer vision, and wireless sensor networks. He is a Senior Member of the China Society of Image and Graphics.

YINGHONG HE was born in 1996. She received the B.S. degree from Chongqing University, China, where she is currently pursuing the degree with the School of Electrical Engineering. Her research interests include image processing and biomedical engineering.

LUMEI LIU was born in 1994. She received the B.S. degree from Qingdao University, China. She is currently pursuing the degree with the School of Electrical Engineering, Chongqing University. Her research interests include image processing and deep learning.

XUESHUO CHEN was born in 1997. He received the B.S. degree from Chongqing University, China, where he is currently pursuing the degree with the School of Electrical Engineering. His research interests include the Internet of Things and edge computing.