Research Article

Eye Movement Prediction Based on Adaptive BP Neural Network

Yushou Tang[1] and Jianhuan Su [2]

[1]*Teacher Education College, Hechi University, Yizhou 546300, Guangxi, China*
[2]*School of Artificial Intelligence and Smart Manufacturing, Hechi University, Yizhou 546300, Guangxi, China*

Correspondence should be addressed to Jianhuan Su; 05003@hcnu.edu.cn

This paper uses adaptive BP neural networks to conduct an in-depth examination of eye movements during reading and to predict reading effects. An important component for the implementation of visual tracking systems is the correct detection of eye movement using the actual data or real-world datasets. We propose the identification of three typical types of eye movements, namely, gaze, leap, and smooth navigation, using an adaptive BP neural network-based recognition algorithm for eye movement. This study assesses the BP neural network algorithm using the eye movement tracking sensors. For the experimental environment, four types of eye movement signals were acquired from 10 subjects to perform preliminary processing of the acquired signals. The experimental results demonstrate that the recognition rate of the algorithm provided in this paper can reach up to 97%, which is superior to the commonly used CNN algorithm.

1. Introduction

With the rapid advancement of artificial intelligence, individuals began to utilize machines to detect users' emotional states, and the machines were required to provide feedback based on human emotions, a mechanism known as the human-computer interaction. Detecting eye movement is an important research topic that has gained significant interest in recent years. Electrical impulses generated around the eyes by eye movements are known as oculomotor nerve signals [1]. Physically disabled people and the elderly with limited mobility may not be able to express their wishes through their bodies but can now use their eyes for emotional communication. Hence, if we can extract useful signals from each other's eyes and design a human-computer interaction system, we can assist human beings in realising the good wishes of human-computer [2]. Eye movement signals have the benefits of large amplitude, easy waveform identification, and easy processing when compared to other bioelectric signals, providing more reliable and convenient circumstances for collecting eye movement information [3]. This work focuses on how to gather eye movement data, as well as extracting, classifying, and identifying that data [4].

Compared with other methods, this paper proposes an adaptive BP algorithm, which mainly solves the problem of different lengths of eye movement information and substantially improves the recognition rate of eye movement signals, laying a good foundation for future human-computer interaction systems [5].

This study focuses on a combination of single and multiframe human eye tracking enhanced algorithms with radial blurring. The shading component, optimization, and characteristics of the radial rendering methodology are explored, the comparison with classic blurring effects is made, and the combination of single and multiframe human eye tracking improved algorithm with radial blurring is studied. The shading component, optimization, and characteristics of the radial rendering technique are explored, as well as the comparison with standard blurring effects and the integration of single and multiframe human eye tracking enhanced algorithms with radial blurring. Traditional blurring and Unity3D [6] are used to compare the outcomes with our work. Unity3D is used to demonstrate and deploy a combination of improved single and multiframe human eye tracking techniques with radial blurring. During eye tracking research, deep learning-based eye tracking

techniques are also an important study direction. Human eye tracking is divided into single-frame image identification tasks and video frame tracking tasks by deep learning. Researchers have introduced numerous unique algorithms through research on human eye tracking [7], despite the fact that there are still challenges in deep learning-based human eye tracking. The proposed algorithms may be able to forecast position information when the eyes are obstructed.

The radial blurred scene rendering function based on human eye tracking is proposed in the virtual reality scene to improve the user's immersion, with the user immersion index equal to the scene (fps reaching 60) and the human eye tracking accuracy reaching 60 by preventing the interframe interference and inaccurate positioning.

The rest of this paper is organized as follows. The literature review is discussed in Section 2. Our proposed adaptive BP neural network for predicting eye movement is discussed in Section 3. In Section 4, we provide the experimental results. Finally, the paper is concluded and future research directions are provided in Section 5.

2. Literature Review

Smooth tracking eye movement is a kind of slow eye movement [8]. Eye movement study began very recently and was primarily theoretical at the time [9]. It is now rapidly transitioning to applied research. Professor Yao and his human-computer interaction (HCI) product based on eye movement EagleEyes is a pioneer in the field of eye movement system research, having been one of the first to build accessible HCI technology based on eye movement control [10]. The system also includes a variety of add-on software that allows users to send emails and search for information on their PCs. Professor Gipps has also collaborated with TECCE at the Department of Psychology at Boston University to perform a number of cognitive experiments employing eye movement technologies [11].

Soleymani et al. divided emotions into three groups based on validity and arousal retrieved statistical variables related to pupil diameter and gaze distance and frequency domain features from eye movement signals in a study on participant emotion recognition based on eye movement data [12]. To cross-validate and forecast emotions, the participants utilized a support vector machine (SVM). With proper recognition rates of 68.8 percent and 63.5 percent, the SVM was utilized to predict emotions. The authors proposed not only an eye movement signal feature set for emotion recognition but also a multimodal emotion recognition database MAHNOB-HCI, which includes EEG and eye movement signals [13]. These experimental results show that eye movements can express emotions and eye movement signals can also be used for the study of emotion recognition, which lays the foundation for subsequent studies of emotion recognition based on eye movement signals [14]. The research group also proposed another emotion dataset containing both eye movement and EEG signals, SEED-V, which contains eye movement data from many subjects and contains five human emotions, but the amount of data may still be relatively small and currently inaccessible.

3. Adaptive BP Neural Network for Reading Eye Movement Prediction Analysis

3.1. Adaptive BP Neural Network Design. In this study, we mainly focus on the eye movement signals based on two aspects. First, we judge the subjects' emotions based on their eye movement signals, because it is often said that the eyes are the windows of the human mind, so the expression of human emotions can be seen from the eye movement signals. Although certain results have been achieved, they are not yet ideal. The research on the emotional expression of eye movement is relatively new and there are still certain difficulties, especially while extracting the eye movement information. There are numerous challenges in the extraction and recognition of eye movement signals at present. The extraction of temporal features of eye movement signals has been studied, and the temporal features with the ability of emotion representation are extracted by combining the features of eye movement signals. This paper extracts the abstract features of eye movement information at a high level and uses the adaptive BP neural network to automatically learn the features of eye movement signals for the maximum extraction of signals, which can improve the accuracy of recognition. The algorithm structure is shown in Figure 1.

Earlier, the detection of eye signals was mainly based on human observation, but this was very ineffective and the analysis results were highly inaccurate. With the rapid development of computing, we can use computers to perform waveform analysis. Since the eye movement signals are different and can vary when humans observe different objects or people in different scenes, it is feasible to analyze the eye movement signals using waveform features. When using a computer to observe the eye movement signal, different eye movement information corresponds to different waveforms, so that the amplitude, wavelength, and so on of the signal can be observed for further analysis of the eye movement signal at a later stage.

The analysis of waveforms, as small as the oculomotor signal, is mainly performed using the frequency domain method, which is based on the Fourier transform. The two main types include parametric and nonparametric estimation. In this paper, we mainly use the parametric estimation method to establish the corresponding power spectral density using the following equation:

$$P(w) = \sum_{k=-\infty}^{+\infty} r(k^e) e^{jwk}. \tag{1}$$

An alternative definition of power spectral density is given in the following equation:

$$P(w) = \lim E\left[\frac{1}{N} | \sum_{n=1}^{N} x(n^2) e^{jwk} \right]^2. \tag{2}$$

Eye movements are mainly divided into eye skipping and blinking. Eye hopping is the process of eye movement from one point to another, which mainly reflects the attention of the eye, learning efficiency, and learning difficulty. It is examined by the speed, in the context of time, and so on. For
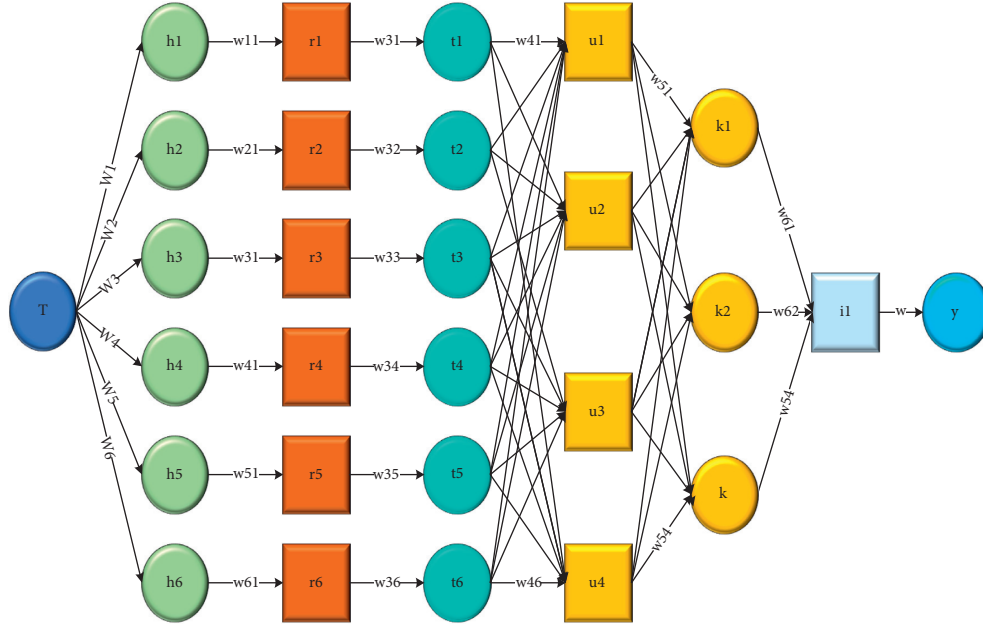
FIGURE 1: Structure of adaptive BP neural network.

example, when you are reading a book, if you are interested in point A and point B, then you see the distance between the two from point A to point B. We call this the distance of eye darting, if the distance from point A to point B is longer, then the time will also be longer and vice versa. This can also reflect the fatigue of the eyes.

Blinking is a process by which the eyes open and close. In a general sense, blinking reflects a person's fatigue program and is related to its frequency. If a person blinks more frequently and for a longer period, the size of the pupils of the eyes is inconsistent, which in a certain way reflects the person's psychological changes and this change is an important indicator for analyzing the person's mental activity. For example, if a person is more interested in something, the pupils of this person will be dilated, which indicates that the subject is very excited, reflected in the study that the subject wants to learn more knowledge.

Pearson's correlation coefficient (PCC) is a statistical method used to analyze the correlation between variables. For two variables X and Y, the data values obtained experimentally, are expressed as X ($X1$, $X2$, Xi) ($i = 1, 2, 3, \ldots, n$) and Y ($Y1$, $Y2$, Yi) ($i = 1, 2, 3, \ldots, n$). The equation for the average of the two sets of data is shown as follows:

$$E(X) = \frac{\sum_i^n X_i^2}{n},$$
$$E(Y) = \frac{\sum_i^n Y_i^2}{n}, \tag{3}$$

where the covariance is computed using the following equation:

$$\mathrm{Cov}(X, Y) = \frac{\sum_i^n \left(X_i^2 - E(X)\right)\left(Y_i^2 - E(Y)\right)}{n}. \tag{4}$$

The Pearson correlation coefficient is computed using the following equation:

$$p_{x,y} = \frac{\sum_i^n \left(X_i^2 + E(X)\right)\left(Y_i^2 + E(Y)\right)}{n}. \tag{5}$$

Since the MAHNOB-HCI dataset has some noise, we preprocess it before using this dataset. For preprocessing, we use the following equation:

$$X_{\mathrm{data}} = \left[X_{pdl}, X_{pdr}\right]. \tag{6}$$

To test the generalization ability of the experimental model in this paper, we must divide 80% of the sample size as the training set and 20% as the validation set. Then, the above algorithm is used to train the eye movement information data to obtain the corresponding sentiment classification model. The accuracy rate is defined as the ratio of the number of correctly classified samples to the number of all samples classified as positive among the positive samples, as shown in the following equation:

$$P = \frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}} - F_{\mathrm{FP}}}. \tag{7}$$

The recall is defined as the ratio of the number of correctly classified samples among positive samples to the number of all samples classified as positive, and it measures the ability of the classification to correctly classify positive samples as follows:

$$R = \frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}} + F_{\mathrm{FP}}}. \tag{8}$$

$F1$ score is a statistical measure of the accuracy of a binary classification model. It combines both the accuracy and recall of a classification model. $F1$ Score can be considered as a kind of summed average of the accuracy and

recall of a model, whose maximum value is 1 and minimum value is 0. $F1$ Score is twice the average of the sum of accuracy and recall, and the $F1$ Score combines the accuracy and recall of a classifier, as follows:

$$F1 = \frac{RP}{R + P}. \tag{9}$$

Sequential data are processed using the memory capability of recurrent neural networks. In traditional neural networks, each layer is fully connected from the input layer to the output layer, but the nodes are not fully connected, and traditional neural networks have poor predictive power for sequential data, where the preceding and following inputs and outputs are not correlated [15]. In recurrent neural networks, the previous information is remembered and preserved and the current output is applied. In the hidden layer, the nodes are connected so that the input of the hidden layer not only is the input of the current input layer but also includes the output of the previously hidden layer, which works well when processing multiple image frames, as shown in Figure 2.

3.2. Experimental Design of Reading Eye Movement. In this paper, we propose a classification method based on eye movements, that is, the eyes can be classified into the following categories during transit: jumping, gaze, and smooth tracking (shown in Figure 3). The main steps are as follows:

(i) Firstly, the inaccurate eye movement data removed upon preprocessing

(ii) Using the speed of eye movements to distinguish between the above-mentioned categories and make them into several segments

(iii) Classification of the obtained data using wavelet transform and vector machines

(iv) Automatic recognition of individual segments using their eigenvalues and clustering algorithms

The above steps mainly get what is more interesting to the user and what is not interesting and what is most interesting or maybe the upper and lower objects are cross-validated with the environment and some additional information obtained. Because the pupil changes when a person is interested in some objects, it is more practical and accurate to use the pupil change to determine the user.

Regarding the dataset we used, this paper uses the cross-missing method for validation. By interpreting the dataset, we can know that, for each subject, we can have the first person perform the eye movement test first and then as a sample and the rest of the people as a model training. This way the accuracy of the emotions of the other subjects can be tested by comparison.

In each eye movement signal, we can extract a sequence of features from it, which then makes up a complete sample. At the same time, we can slide each window, so that each window can also be considered as a smaller sample for easy analysis. Since the whole extracted sample is different, the length of each window is also different, as well as, of course,

the number of truncations. These subsamples are mostly from smaller samples and are identical. Using these subsamples, we can have an idea about the complete network input and the output value of the BP neural network and subsamples can be obtained as well. This is shown in Figure 4. Note that it is not desirable if intuitive extraction samples are used.

For a complete waveform in longitudinal pupil coordinates, an upward trend in the waveform indicates upward eye movements and a downward trend in the waveform indicates downward eye movements. When the waveform is segmented, each segment represents the pattern of upward and downward eye movements per unit of time. This time segment includes the amplitude of eye movements and the order of eye movements [16]. For a combination of two segments, the correlation coefficient between the two segments indicates the similarity between the two segments. If correlation coefficients are found between all segments, the sum of these coefficients reflects the sum of similarities between all segments in the longitudinal coordinate waveform of the pupil, that is, the complexity of the combination of all pupil up-and-down movement patterns. In this paper, we use an adaptive BP neural network algorithm to select the segmentation time length and correlation coefficients for each pupil position coordinate waveform when using the combined waveform complexity as the pupil position coordinate feature value.

The delineation of periods is particularly important when extracting the complexity of the integrated waveform. Human response times to different physiological signals of stimulus events may vary between emotional expressions, with heart rate and skin electrical response times to emotional arousal ranging from 3 to 6 seconds. The same emotional reaction times exist for pupil position coordinates, and each time segment of the waveform contains one emotional fluctuation if the length of the time segment can be matched to the smallest unit of the kinetic nerve response. The correlation between segments is calculated in this way; that is, it indicates the degree of correlation of different emotional fluctuations and the sum of the correlation coefficients of all segments within the video temporal degree indicates the combined emotional fluctuations of the subject while watching that video. In other words, it represents the combined waveform complexity.

4. Analysis of Results

4.1. Algorithm Performance Results. From Figure 5, the higher value is the blinking point, and if the fluctuation is not too large, that is, the transition is relatively smooth, then it is the eye gaze point. It is clear from Figure 5 that, by comparing the standard values with the algorithmic results, the BP-based adaptive network algorithm proposed in this paper matches the standard values to a higher standard. In this figure, we can also see that the existing schemes, that is, I-SC, IVDT, and CNN, all show some incorrect classification data and cannot correctly identify the correct trajectory of the eye movement signal. Therefore, the adaptive BP neural network algorithm outperforms the I-SC, IVDT, and CNN
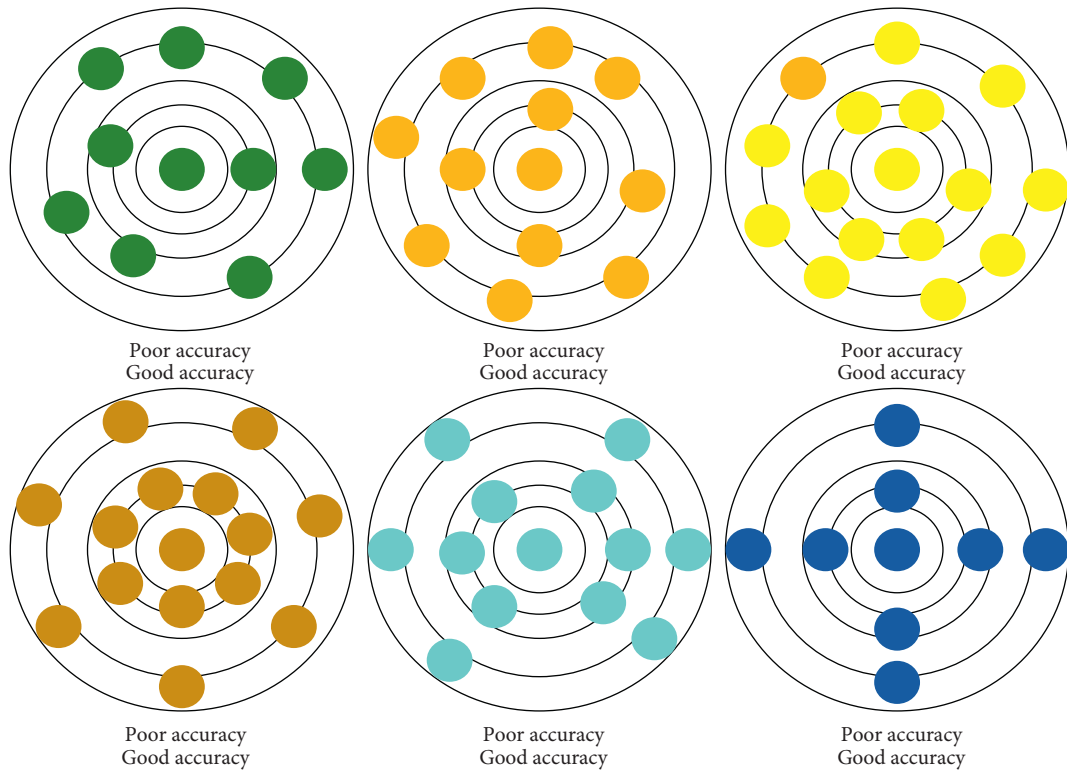
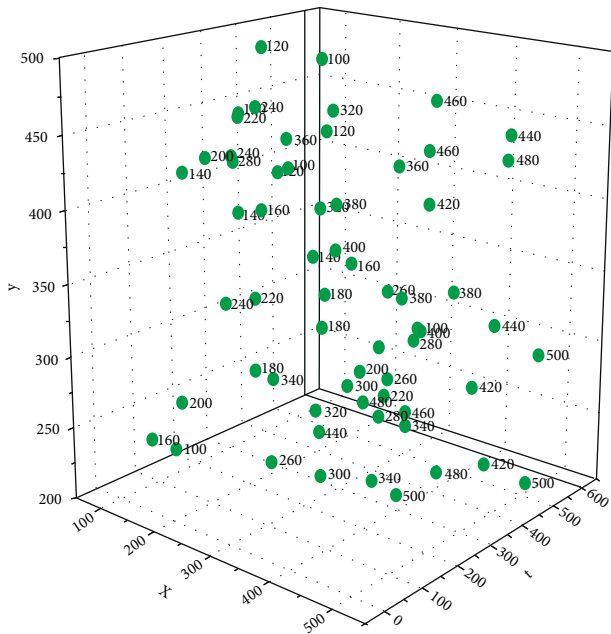FIGURE 2: Error analysis model of the line-of-sight tracking system.

FIGURE 3: Typical eye movement data with gaze, eye bounce, and smooth tailing.

FIGURE 4: Segmentation of eye movement data.

algorithms in terms of accuracy, recall, and F1 scores for both the initial and intermediate excess and final points of the eye movement signal.

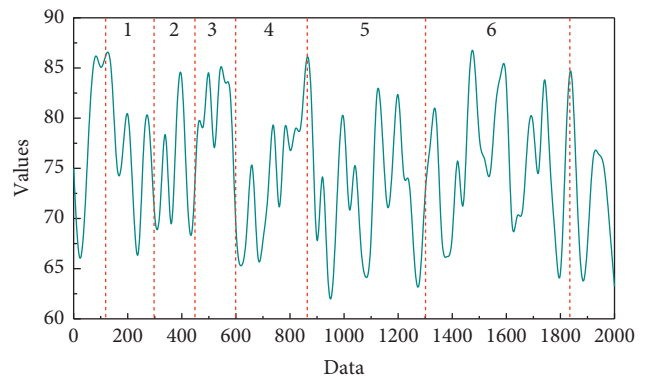To further analyze the advantages and disadvantages of these four algorithms, white noise is added to the standard values in this paper, and the comparison results are shown in Figure 6. In this figure, we can see that the adaptive BP neural network algorithm is effective in classifying eye movement data containing noise due to the consideration of the input values of the eye movement signal and the characteristics of continuity and burst. The simulation results show that the adaptive BP neural network algorithm is effective in classifying eye movement data containing noise. According to the performance metrics of each algorithm given in this figure, the gaze and smooth tailing classification performance of our BP neural network algorithm outperform the other three algorithms in terms of recall and accuracy metrics.
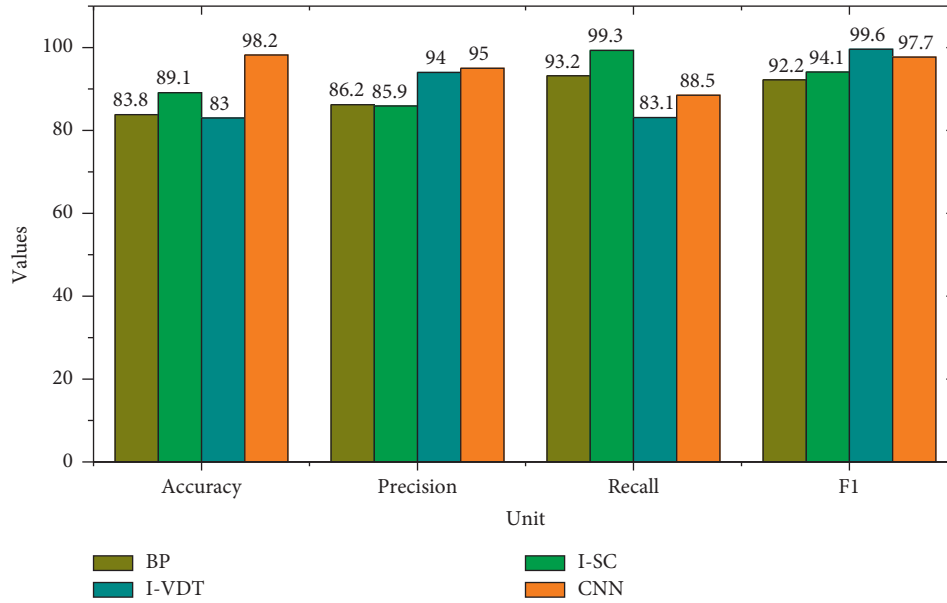
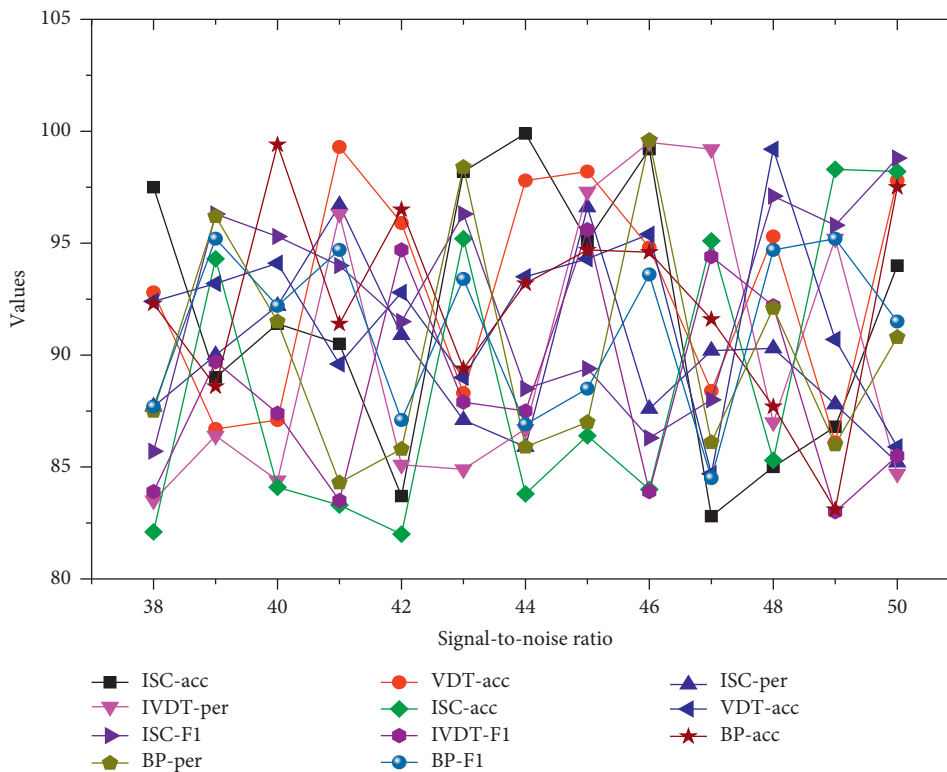FIGURE 5: Classification performance on the eye tracking dataset.

FIGURE 6: Performance comparison under different noises.

4.2. Analysis of Experimental Results. As shown in Figure 7, the $F1$ values for arousal and validity were compared for all three categories under three different input conditions, indicating that the $F1$ mean value was better than the other two inputs for both arousal and validity. Hence, it indicates that the combined features are more helpful in improving the model performance. This is because the sample size of the high arousal category is much lower than the other two categories, resulting in the model not being able to learn the sample features of this category well and thus not being able to distinguish this category properly. The $F1$ values of the medium arousal category in Figure 7 are both lower than those of the low and high arousal categories, indicating that the model is less capable of identifying medium arousal compared to the other two arousal categories.
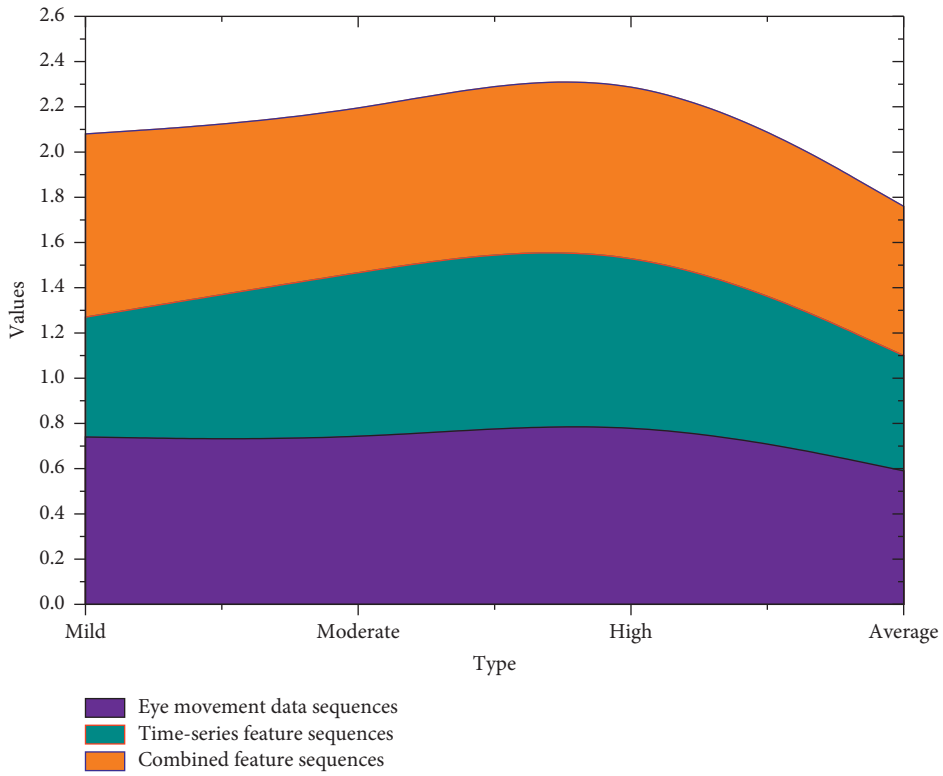
FIGURE 7: F1 values for three categories of wake-up degree with three different inputs.
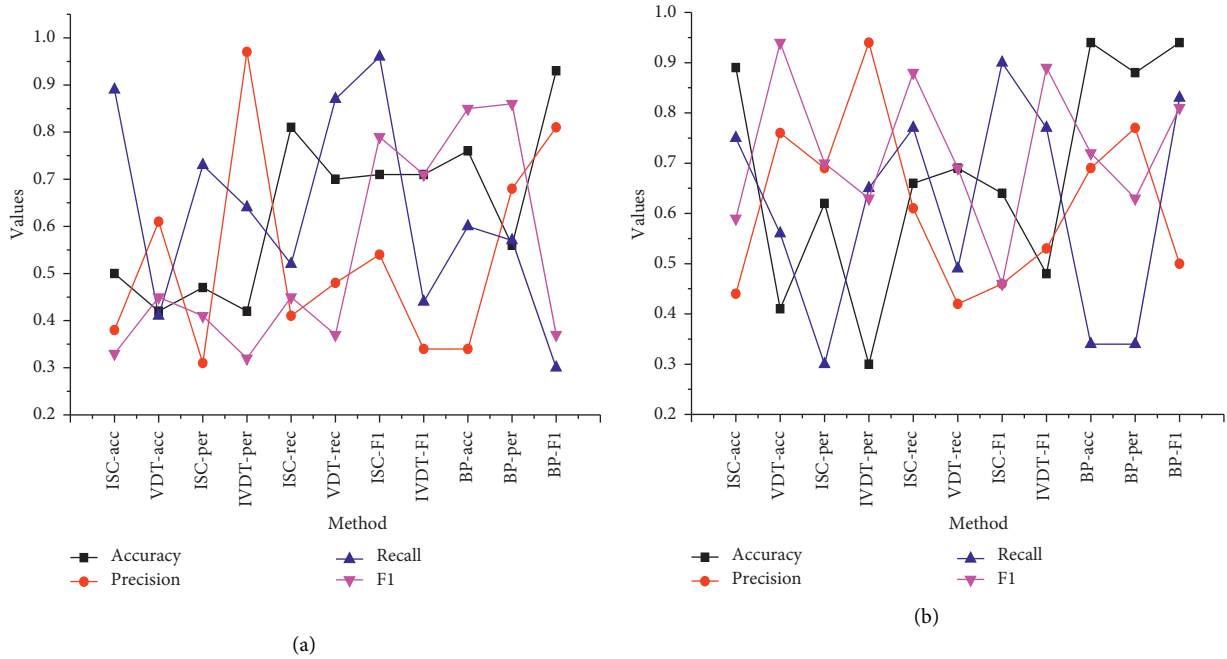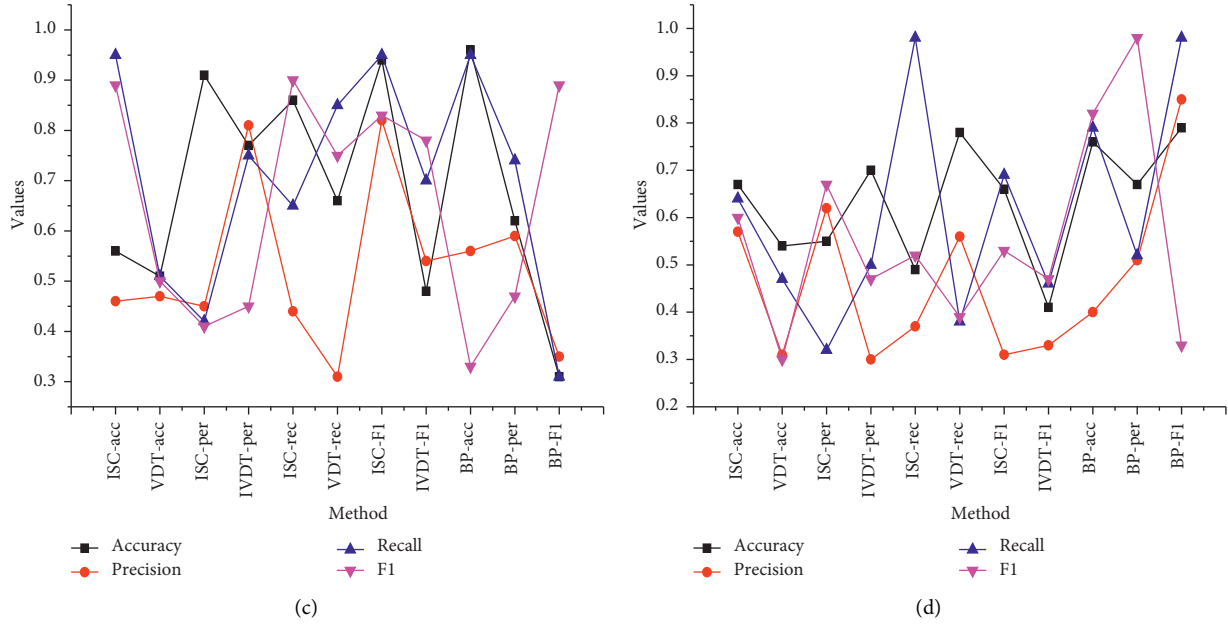

(a)

(b)

FIGURE 8: Continued.

FIGURE 8: Model metrics.

This paper analyzes the problem of the relationship between the temporal performance and the size of the time window of the adaptive BP neural network algorithm proposed in this paper. Generally, we classify the emotional problems generated by eye movements into three broad categories: the emotional initial point, the emotional peak point, and the emotional endpoint. In these three types of emotion points, if we can reasonably allocate the size of each time window, the integrity of emotion can be maintained. The data thus obtained will be more reasonable and the classification recognition will be better. The emotional brain signals we extracted were divided into a total of eight-time windows, their features were analyzed using wavelet methods, and then the vector machines were used to classify each window of emotion. In this experiment, we used 6–15 seconds to obtain 65% accuracy. Thus, the information about emotions in the EEG signal can be properly localized around 6 to 15 seconds. In our experiments, this classifier works better for the 10 s time window dataset, especially in the adaptive BP neural network-based algorithm proposed in this paper; the model gives the best results in the 10 s time window, as shown in Figure 8.

With the BP adaptive neural network algorithm, the distance between the leftward signal and the leftward template is the shortest, and the difference between the templates in the other algorithms is still relatively large. It can also be seen from Figures 8(b) and 8(c), respectively. In Figure 8(d), there is a crossover of signals when blinking twice and blinking once, which means that there is 1 mismatch in 50 datasets with a 98% match rate. The algorithm is not too difficult to implement and is an effective recognition method. The adaptive BP neural network algorithm proposed in this paper is a fuzzy input matching algorithm, which mainly solves the problem of when the input eye movement signal is long or short.

5. Conclusion

This paper's major focus is on the classification and detection of eye movement signals, with the goal of creating a new form of human-computer interaction. To achieve the goal of human-computer interaction, the majority of contemporary human-computer interaction approaches rely on human motion. Human-computer interaction technology based on eye movement signals can effectively tackle the problem of interaction through the eyes, allowing for new technological advancements that will benefit a broader spectrum of individuals. In this paper, the software and hardware environment for acquiring eye movement signals, as well as the acquisition method and electrode pad positioning, were all designed. Four types of eye movement data were collected from ten subjects in the experimental area, and preliminary signal processing was performed. None of the existing studies can address the fact that eye movement signals differ in length and from person to person. As a result, the idea of eye movement signal categorization based on an adaptive BP algorithm is proposed, which not only handles this problem well but also shows in simulation results that it may significantly enhance recognition rate.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this study.

References

[1] X. Jiang and Y.-D. Zhang, "Chinese sign language finger-spelling via six-layer convolutional neural network with leaky rectified linear units for therapy and rehabilitation," *Journal of Medical Imaging and Health Informatics*, vol. 9, no. 9, pp. 2031–2090, 2019.

[2] A. Wolf, K. Ueda, and Y. Hirano, "Recent updates of eye movement abnormalities in patients with schizophrenia: a scoping review," *Psychiatry and Clinical Neurosciences*, vol. 75, no. 3, pp. 82–100, 2021.

[3] L. Si, Z. Wang, X. Liu, C. Tan, and R. Anon, "Assessment of rib spalling hazard degree in mining face based on background subtraction algorithm and support vector machine," *Current Science*, vol. 116, no. 12, pp. 2001–2012, 2019.

[4] M. Yu, T. Quan, Q. Peng, X. Yu, and L. Liu, "A model-based collaborate filtering algorithm based on stacked AutoEncoder," *Neural Computing and Applications*, vol. 5, 2021.

[5] P. K. Upadhyay and C. Nagpal, "Time-frequency analysis and fuzzy-based detection of heat-stressed sleep EEG spectra," *Medical, & Biological Engineering & Computing*, vol. 59, no. 1, pp. 23–39, 2021.

[6] R. M. D. Hegde and H. H. Kenchannavar, "A survey on predicting resident intentions using contextual modalities in smart home," *International Journal of Advanced Pervasive and Ubiquitous Computing*, vol. 11, no. 4, pp. 44–59, 2019.

[7] M. A. Rahman, M. S. Uddin, and M. Ahmad, "Modeling and classification of voluntary and imagery movements for brain-computer interface from fNIR and EEG signals through convolutional neural network," *Health Information Science and Systems*, vol. 7, no. 1, p. 22, 2019.

[8] T. L. Alvarez, M. Scheiman, E. M. Santos et al., "The convergence insufficiency neuro-mechanism in adult population study (CINAPS) randomized clinical trial: design, methods, and clinical data," *Ophthalmic Epidemiology*, vol. 27, no. 1, pp. 52–72, 2020.

[9] J. Kim, H. Kim, and T. Hong, "Automated classification of indoor environmental quality control using stacked ensembles based on electroencephalograms," *Computer-Aided Civil and Infrastructure Engineering*, vol. 35, no. 5, pp. 448–464, 2020.

[10] L. Yao, T. Li, Y. Li, W. Long, and J. Yi, "An improved feedforward neural network based on UKF and strong tracking filtering to establish energy consumption model for aluminum electrolysis process," *Neural Computing & Applications*, vol. 31, no. 8, pp. 4271–4285, 2019.

[11] I. Mackrous, J. Carriot, M. Jamali, and K. E. Cullen, "Cerebellar prediction of the dynamic sensory consequences of gravity," *Current Biology*, vol. 29, no. 16, pp. 2698–2710, 2019, e4.

[12] C. I. De Zeeuw, S. G. Lisberger, and J. L. Raymond, "Diversity and dynamism in the cerebellum," *Nature Neuroscience*, vol. 24, no. 2, pp. 160–167, 2021.

[13] Y. Fu, C. Li, T. H. Luan, Y. Zhang, and F. R. Yu, "Graded warning for rear-end collision: an artificial intelligence-aided algorithm," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 2, pp. 565–579, 2019.

[14] A. Khosla, P. Khandnor, and T. Chand, "A comparative analysis of signal processing and classification methods for different applications based on EEG signals," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 2, pp. 649–690, 2020.

[15] G. Petit, T. Savi, M. Consolini, T Anfodillo, and A Nardini, "Interplay of growth rate and xylem plasticity for optimal coordination of carbon and hydraulic economies in Fraxinus ornus trees," *Tree Physiology*, vol. 36, no. 11, pp. 1310–1319, 2016.

[16] S. Kar, H. Sanderson, K. Roy, E. Benfenati, and J. Leszczynski, "Ecotoxicological assessment of pharmaceuticals and personal care products using predictive toxicology approaches," *Green Chemistry*, vol. 22, no. 5, pp. 1458–1516, 2020.