Research Article

Liang Huawei* and Bi Xiumei

Study on the influence of biomechanical factors on emotional resonance of participants in red cultural experience activities

https://doi.org/10.1515/pjbr-2025-0011
received May 5, 2025; accepted June 21, 2025

Abstract: As an important part of Chinese culture, red culture has a profound historical heritage and rich spiritual connotations. The red culture experience aims to inspire participants' patriotic feelings and national pride by recreating historic and revolutionary scenes and sharing stories of the revolution. To further enhance the impact of red cultural experience activities, this paper employs a minimum spanning tree (MST) analysis to investigate the emotional effects of eye and brain movement features on participants in these activities. A multimodal physiological signal emotion recognition model was constructed using EEG and eye movement data, and the accuracy of emotion classification for happy, sad, fearful, and neutral emotions was investigated. The research results show that the emotion recognition effect is optimal when eye movement, differential entropy feature, and MST attribute are integrated simultaneously. The three can achieve effective integration between EEG signals and eye movement signals across modes to obtain more emotion-related information and fully utilise the complementarity between this multimodal information.

Keywords: red cultural experience activities, biomechanics, eye movement, brain electricity, emotion recognition, biomechanics, generative AI, multimodal fusion

1 Introduction

Emotion is a physiological state that integrates all the feelings, thoughts, and activities of the human body, as well as a psychological and physiological response produced by various external stimuli. Red culture is a vital part of Chinese heritage that commemorates the revolutionary history of the Communist Party, promotes collectivism, fosters national pride, and cultivates patriotism. It preserves historical memory through stories, media, and experiential activities, particularly targeting young people. By adapting to modern formats like digital media, red culture remains relevant, inspiring civic responsibility and emotional connection to shared national values. Electroencephalography is considered a physiological measure in which the electrical activity of nerve cells is detected in the human cerebral cortex, while emotion-related brain activity can also be detected through electroencephalography. The proposed method selects the emotions of happy, sad, fear, and neutral for their broad representation of emotional states across the valence-arousal spectrum, aligning with established theories such as Ekman's and Russell's models. These emotions show distinct patterns in EEG and eye-tracking data, enhancing classification accuracy. They are also relevant to the red cultural film stimuli used, which are designed to evoke these specific emotional responses. The SEED-IV dataset supports this choice by providing consistent, pre-rated clips for reliable emotional elicitation and labelling. In the study of EEG signals, Thammasan et al. utilised power spectral density to extract information features from the original EEG signal and then applied an emotion classification algorithm to classify continuous music with binary emotions. The classification accuracy reached 72.9% [1]. Wu et al. employed correlation and coherence analysis to estimate EEG connectivity between electrodes, aiming to identify subjects' emotional levels in arousal and valence dimensions. The results revealed that connectivity features in the gamma band contained information conducive to valence classification [2].

*** Corresponding author: Liang Huawei,** School of Marxism, Fuyang Normal University, Fuyang, Anhui, 236037, China,
e-mail: liang_huawei90@outlook.com
Bi Xiumei: High School Biology Group of the High School Affiliated to Fuyang Normal University, Fuyang, Anhui, 236037, China

In recent years, eye-tracking technology has also become a crucial research tool in various fields, including psychology and cognitive science. Eye movement signals have become more convenient to capture, and participants' instinctive cognitive behaviours can be known by interpreting their eye focus. In the study of emotion recognition related to eye movement signals, Oliva and Anikin found that pupil responses are closely tied to the time process of emotion processing and provide rich information sources. The intensity of perceived emotion can only enhance pupil dilation [3]. Moreover, affective feedback significantly affects gaze behaviour, and its performance in arousal is better than that of titer. Lu et al. applied features related to eye movements to identify positive, negative and neutral emotions, with an average classification accuracy of 77.80% [4]. Zheng et al. utilised features such as fixation time, saccade amplitude and duration, and blink frequency to classify the degree of four emotions, achieving an accuracy of 67.82% [5]. Generative AI applications can enhance user engagement in cultural experiences by utilising multimodal emotion recognition, which leverages EEG and eye movement signals. These applications can predict or simulate user emotional responses, create personalised narratives or audio guides, and offer supportive interventions during moments of sadness or fear. A GAN-based synthetic data generator can produce realistic multimodal emotional signals, address data imbalance, and enhance model training. A personalised cultural feedback agent can analyse emotional patterns and generate customised exhibit suggestions, thereby enriching the post-experience journey. These generative AI extensions are scalable and emotionally intelligent.

The proposed multimodal approach, which integrates EEG differential entropy (DE), minimum spanning tree (MST)-based brain connectivity, and eye movement features, enhances the precision of emotion recognition. These insights enable user interfaces to adapt dynamically to users' emotional states, improving engagement and usability. Generative AI systems can utilise emotional feedback to personalise content in real time, thereby deepening user resonance. In educational settings, recognising emotions can help tailor learning experiences to enhance motivation and retention. Biomechanical cues such as gaze patterns and connectivity metrics inform more intuitive interaction designs. These findings support the development of emotionally aware, human-centred technologies across various application domains.

Regarding the recognition of emotions in multimodal physiological signals, Islam et al. combines EEG signals with facial expressions in a multimodal manner. He finds that the emotion recognition rate of multimodal fusion is 61.88%, while the emotion recognition rates of single modes are 60.2 and 52.43%, respectively [6]. Kwon et al. combined electroencephalogram and electrodermal response and used CNN to merge the EEG spectrogram and electrodermal response features to complete multimodal emotion recognition [7]. Abtahi et al. employed audio/video, electromyography, and EEG to investigate four different modes of emotion [8].

Biomechanical signals, such as EEG and eye movement, provide direct insight into participants' embodied emotional responses during red cultural experience activities. These responses are not merely physiological but are deeply influenced by the symbolic and historical significance of revolutionary content, which evokes culturally anchored emotions such as patriotism and collective memory. This makes emotion recognition in such a setting uniquely valuable, as it captures how sociocultural meaning shapes sensorimotor expression. The integration of these biomechanical factors thus offers a novel lens for understanding emotion, extending beyond generic stimuli to include culturally embedded affective resonance.

It can be seen from the above collation and induction of relevant literature that in the studies on emotion recognition based on eye movement signals, this article finds that most features, such as pupil response and pupil diameter, are used for classification. Still, there is no clear indication of which features or combinations of features are most beneficial to emotion recognition tasks [9]. Ting et al. propose the ethnokinesiology framework, which links sensorimotor and cultural factors in shaping movement. It highlights motor accents as culturally influenced variations, supporting analysis of biomechanics in Red Cultural activities and their role in emotional engagement [10]. In addition, the number of studies on emotion recognition using eye movement signals is still relatively limited, and single-mode EEG data cannot fully and accurately represent a person's emotional state. Therefore, utilising the complementarity of multimodal signals to represent emotion states can enhance the accuracy of the emotion recognition model and facilitate valuable exploration in selecting emotion classification features. Neural oscillations and oculomotor behaviour reflect emotional processing by capturing both internal brain activity and external behavioural responses. EEG features, such as DE and connectivity patterns extracted through MST analysis, reveal changes in neural complexity and functional network organisation during emotional states. Simultaneously, eye movement metrics, such as fixation patterns, pupil diameter, and saccades, indicate variations in arousal and attention. When combined, these two modalities enhance the accuracy and depth of emotion recognition by integrating both observable behaviour and physiological

signals, thereby offering a more comprehensive understanding of emotional states.

This study aims to investigate how biomechanical factors such as EEG and eye movement influence participants' emotional resonance in red cultural experience activities. It focuses on developing a multimodal emotion recognition model by combining EEG and eye movement signals to improve classification accuracy. The research examines the effectiveness of features such as DE and MST attributes. It compares unimodal and multimodal approaches across four emotional states: happiness, sadness, fear, and neutrality. The goal is to enhance emotional analysis by leveraging the complementarity of multimodal signals in immersive cultural settings.

To systematically investigate the emotional impact of biomechanical features during red cultural experience activities, this article is structured as follows: Section 2 introduces the methodology, including EEG and eye movement feature extraction, the construction of a multimodal emotion recognition model, and the use of MST analysis. Section 3 presents the experimental setup, dataset, signal preprocessing techniques, and the emotion classification framework, along with detailed performance results. Section 4 concludes the paper by summarising key findings and outlining directions for future research in multimodal emotion recognition within cultural contexts.

2 Feature extraction model based on multiple modes

2.1 EEG feature extraction model

2.1.1 DE features

Given that its probability density distribution function is $P(x)$ for EEG time series X, the complexity of the EEG series can be expressed by DE [9]. The DE feature of the fixed subband in the EEG time series is calculated as follows:

$$\mathrm{DE} = \frac{1}{2}\ln(2\pi e\sigma^2). \tag{1}$$

In preprocessing, bandpass filtering removes the DC component of the EEG signal, resulting in a mean value of 0. Therefore, the variance formula of the EEG signal is:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2. \tag{2}$$

The variance of the EEG signal is the mean of its energy. According to Parseval's theorem that the mean energy is proportional to the energy spectrum, it can be obtained:

$$\sum_{n=0}^{N-1}|x_i^2| = \frac{1}{N}\sum_{k=0}^{N-1}|X_k^2| = P_i. \tag{3}$$

Sequence $\{x_i\}$ is obtained in the formula by a short-time Fourier transform from sequence $\{X_k\}$. Therefore, the DE value of the EEG signal is related to its average energy. In the calculation process, the logarithm of the mean energy is taken to obtain the DE feature [11]. Compared to the frequency band energy, the difference between different sub-bands is much smaller, which can help balance the data gap, reduce errors, and improve algorithm accuracy. In the actual calculation of the DE feature, the logarithmic coefficient between it and the average energy of the EEG signal is ignored, and the formula for calculating the DE feature of each sub-band is obtained:

$$h_i(X) \approx \log(P_i). \tag{4}$$

2.1.2 Function connection specifications

In this article, the phase lag index is introduced, which is a measure of the asymmetric phase difference distribution between two signals, which can not only reflect the statistical interdependence and coupling strength between time series but also is expected to be less sensitive to the influence of common source and amplitude effects. The instantaneous phase expression of the original time series $x(t)$ is as follows:

$$\phi_x(t) = \arctan\frac{\tilde{x}(t)}{x(t)}. \tag{5}$$

The formula $\tilde{x}(t)$ is the Hilbert transform of the original time series $x(t)$. As the original time series, the EEG signal is the projection of the analytic signal on the real number domain. After restoring the analytic signal formula, the phase differences between the two signals can be calculated simultaneously [12]. Therefore, the calculation formula of the phase lag index of the two signals is as follows:

$$\mathrm{PLI} = |\langle \mathrm{sign}(\Delta\phi(t))\rangle| = \left|\frac{1}{N}\sum_{n=1}^{N}\mathrm{sign}(\Delta\phi(t_k))\right|. \tag{6}$$

In the formula, t_k represents the same time of two signals, and sign is a symbolic function whose phase lag index ranges from [0, 1].

2.1.3 MST

MST is a graph-theoretical approach used to model functional brain connectivity. It simplifies complex brain networks by connecting all nodes (EEG channels) with the minimum total

edge weight, helping to quantify information flow and network efficiency during emotional processing. The MST is a valuable method for EEG analysis in emotion recognition, as it provides a threshold-free, sparse, and interpretable representation of brain connectivity. By using the inverse of the Phase Lag Index as edge weights, MST highlights key functional relationships while minimising noise. It enables the extraction of meaningful topological features that reflect changes in emotional state. Compared to traditional connectivity methods, MST ensures consistency across subjects and is less sensitive to noise. When combined with other features, such as DE and eye movement data, MST enhances emotion classification accuracy, providing complementary insights into brain function during emotional processing.

The MST is used to simplify and represent functional brain networks by connecting EEG channels with minimal redundancy while preserving key connections. It utilises phase lag index-based weights to ensure robustness against noise and computes graph metrics, such as degree, diameter, and tree hierarchy, to quantify network properties. MST features enhance emotion recognition when combined with EEG and eye movement data by providing topological and spatial insights into brain activity.

In this article, the MST is obtained using the greedy idea. First, M nodes need to be regarded as M separate spanning trees, and N edges are sorted in ascending order according to the weight value. Next, each of the N edges is extracted in sequence. If both nodes are found to be on two different trees, they are merged into one tree; otherwise, they are discarded, and the process continues traversing [13]. When all edges are traversed and all spanning trees are merged into one tree, the filtered edges and all nodes form the smallest spanning tree in this functional brain network. In this paper, the weight of an edge is defined as the reciprocal of the phase lag index. The constructed MST contains 62 nodes and 61 edges, that is, $M = 62$ and $N = 61$.

Six commonly used metrics are calculated to better measure the structure of the MST. They are Degree, Diameter (Dia), Eccentricity (Ecc), Leaf fraction (Lf), Betweenness Centrality (BC), and Tree hierarchy (Th).

The degree is the most basic and widely used metric in network research. It can describe the higher-order topological relationships in brain networks. The diameter is the maximum path length between any two nodes. It measures the global transmission efficiency of the network and can describe the size of the brain network. In small-diameter networks, information can be efficiently processed between distant brain regions [14]. However, it should be noted that the diameter of the MST is related to the number of leaf nodes.

Eccentricity is the maximum path length between a given node and any other node in the network. A higher eccentricity means that the node of the MST is closer to the central node. Average eccentricity E_i represents the average of all nodes deviating from the centre node:

$$E_i = \frac{\sum_{j=1}^{N}(\max_{i\in N} d_{ij})}{N}. \tag{7}$$

The leaf fraction is the ratio of leaf nodes to the number of nodes in the minimum-spanning tree. Intermediate centrality is an index to judge the importance of nodes in a network.

The tree level quantifies the balanced relationship between the integration capability of the MST and node overload, and its calculation formula is as follows:

$$\mathrm{Th} = \frac{L}{2\mathrm{MBC}_{\max}}. \tag{8}$$

In the formula, the coefficient of molecular position 2 ensures that the value of the tree level is between 0 and 1 [15]. In the MST, if $L = M$, the tree has a star structure, as shown in Figure 1, when Th = 0.5; if $L = 2$, the tree has a linear structure, as shown in Figure 2, where Th = 0.

2.2 Eye movement feature extraction model

For the eye movement information collected by SMI eye tracking glasses, after preprocessing, the mean and standard deviation are calculated in the X and Y directions, respectively. The short-time Fourier transform is then used to perform time-frequency conversion and extract DE features [16]. DE is a statistical feature derived from EEG signals that capture the complexity or unpredictability of brain activity within a specific frequency band. It is particularly effective in emotion recognition due to its sensitivity to subtle neural dynamics. The mean and standard deviation of Fixation Dispersion, Saccade Duration, Fixation Duration and Saccade AMPLituDE commonly used in eye movement

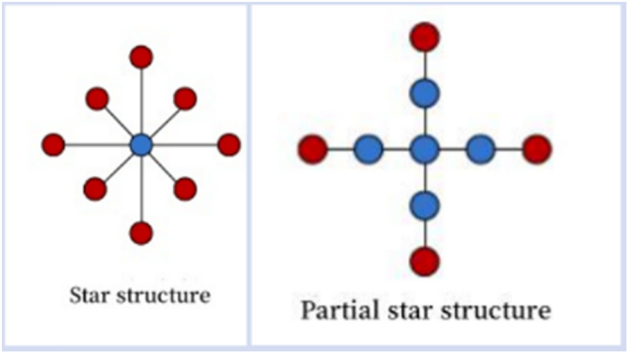

Figure 1: Star MST topology network structure.

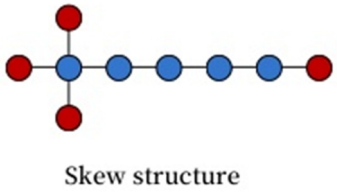

Figure 2: Partial MST topology network structure.

studies were extracted. At the same time, nine statistical events characteristics, such as Blink Frequency and Fixation Frequency, were recorded. The specific eye movement characteristics and dimensions are shown in Table 1.

2.3 Feature fusion

In this article, the feature layer fusion strategy is employed to construct a new feature vector by serial concatenation fusion of all extracted feature matrices. The new feature vector is then sent to the classifier for training and testing, and finally, the classification result is obtained. Feature-level fusion using serial concatenation is adopted to effectively integrate the complementary aspects of DE, MST attributes, and eye movement features. This approach enables early interaction among modalities, allowing the model to learn richer and more informative joint representations for emotion recognition. The improved classification performance supports this strategy over alternative fusion methods such as decision-level or intermediate-level fusion. The fusion feature calculation formula is as follows:

$$\begin{aligned} D_i &= (d_1,\ d_2,\ \ldots,\ d_n)\ M_i = (m_1,\ m_2,\ \ldots,\ m_n) \\ E_i &= (e_1,\ e_2,\ \ldots,\ e_n)\ Y_i = [D_i,\ M_i,\ E_i]. \end{aligned} \tag{9}$$

In the formula, D_i is the DE feature of the i sample, M_i is the MST attribute of the i sample, E_i is the eye movement feature of the i sample, and Y_i is the fusion feature.

The proposal demonstrates that the combination of multimodal features, including EEG DE, MST attributes, and eye movement metrics, increases the dimensionality of the feature space, increasing the risk of overfitting and reducing classification accuracy. This is particularly problematic when working with small sample sizes, such as the SEED-IV dataset. To mitigate these risks, the study uses a two-sample T-test for feature selection, identifying and retaining only the most discriminative features based on statistical significance. This dimensionality reduction enhances the model's robustness and efficiency, thereby improving classification performance.

To improve feature sensitivity and classify each emotion more accurately, this paper employs the two-sample *T*-test method to select relevant and important features and then classify them after reducing the feature dimension [17]. The feature weight is based on the absolute value and combined variance estimation of the two-sample *t*-test, and the calculation formula is as follows:

$$W(i) = \left| \frac{m_k - m_n}{\sqrt{\frac{\sigma_k^2}{N_k} + \frac{\sigma_n^2}{N_n}}} \right|. \tag{10}$$

In the formula, $W(i)$ is the weight of feature i, m is the mean value of the feature, N_x is the number of samples in the emotion category, m_x is the mean value of the emotion classification in the expected feature, and σ_x^2 is the variance of the emotion classification in the expected feature [18]. In this article, the critical weight value is set to 0.05; if it is less than 0.05, it is considered significant, and *vice versa*.

2.4 Feature classification

The core idea of support vector machine (SVM) is to find the classification hyperplane that separates samples reasonably and maximizes the margin [19]. The classification

Table 1: The details of the features extracted from eye movement data

Eye movement signal indicators	Extracted feature	Characteristic dimension
Pupil diameter	Mean value, standard deviation and 0–0.2 Hz, 0.2–0.4 Hz, DE features at 0.4–0.6 Hz, 0.6–1 Hz	12
Fixation bias	Mean, standard deviation	4
Fixation duration	Mean, standard deviation	2
Saccade duration	Mean, standard deviation	2
Saccade amplitude	Mean, standard deviation	2
Statistical event	Blink frequency, gaze frequency, maximum gaze time, total gaze deviation, maximum gaze deviation, saccade frequency	9

hyperplane model, interval maximisation strategy, and convex quadratic programming algorithm are the three key elements of SVM.

For binary classification problems, the model is illustrated in Figure 3, assuming the data is linearly separable.

In this case, the hyperplane data satisfy:

$$\omega^T x + b = 1, \quad y_i = 1\omega^T + b = -1, \quad y_i = -1. \tag{11}$$

If the data space is linearly indivisible, the lower-dimensional data can be mapped to a higher-dimensional space by a kernel function to become linearly divisible. In this article, LinearKernel is used as the kernel function of SVM. The formula is as follows:

$$k(x, y) = x^t y. \tag{12}$$

Finally, the objective function of the dual problem is obtained as follows:

$$L(\omega, b, a) = \frac{1}{2} \|\omega\|_2 + \sum_{i=1}^{m} a_i(1 - y_i(\omega^T x_i + b)). \tag{13}$$

3 Research on emotion recognition based on biomechanical factors

3.1 Data set

The data set used in this study was SEED-IV (SJTU Emotion EEG Dataset for Four Emotions), which included 62-channel EEG and eye movement signals from 15 subjects collected using the ESI neural scanning system and SMI eye tracking glasses. SEED-IV dataset uses 168 red culture movie clips as stimulus materials, recruits 44 volunteers to watch movie clips and score them from two dimensions of arousal and titer, and selects 72 movie clips with the highest matching degree and easy-to-induce target emotions for subsequent experiments [20]. Titer refers to the emotional intensity or strength perceived by participants, often in conjunction with arousal, to map emotions on a two-dimensional plane. Each participant conducted three experiments at different times, each approximately a week apart, watching 24 red culture movie clips that contained four emotions: happiness, sadness, fear, and neutral, with six movie clips for each emotion. The emotional experiment paradigm of the SEED-IV dataset is shown in Figure 4.

It can be seen that there is a 5 s prompt time at the beginning of each movie segment. The duration of the movie segment is approximately 2 min, and the PANAS scale self-assessment time is 45 s after the end [21]. Based on the feedback, the data were discarded if the subject failed to trigger the right emotion or if the emotion was not awakened strongly enough. According to the electrode distribution diagram of the 62-channel international 10–20 system shown in Figure 5, the ESI neural scanning system was utilised to collect stable EEG signals. At the same time, SMI eye-tracking glasses were employed to capture eye movement information.

3.2 Signal preprocessing

3.2.1 EEG preprocessing

The EEG signal is extremely weak, typically only about 50 μV, and exhibits high time-varying sensitivity, which makes it easily susceptible to interference from other signals. Therefore, the original EEG signal collected contains a lot of noise and artefacts and is usually preprocessed to some extent. During the experiment, the original EEG signal was downsampled to 200 Hz, and a 1–75 Hz bandpass filter was used to filter noise and remove artefacts [22].

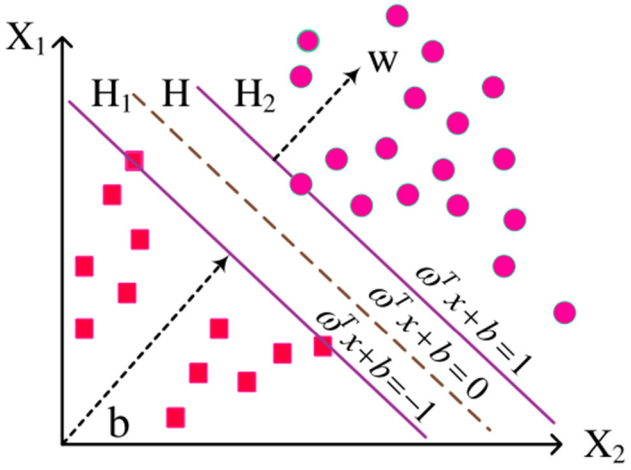

Figure 3: Linear separable model of SVM.

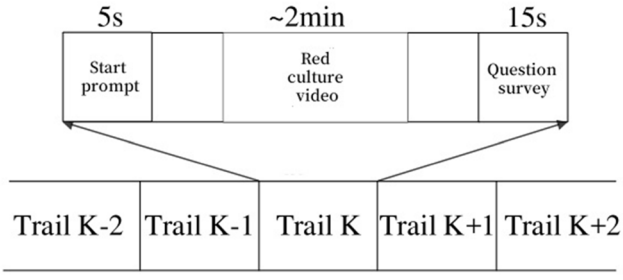

Figure 4: SEED-IV dataset emotion experiment paradigm.

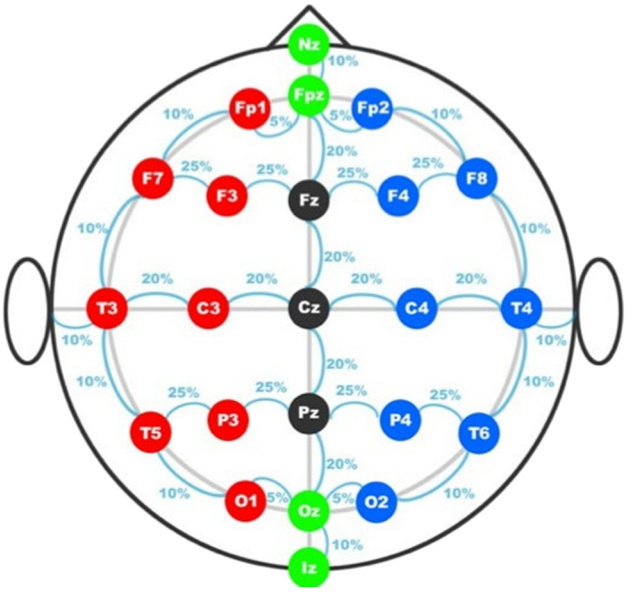

Figure 5: The 10–20 electrode placement system.

Lag analysis and cross-correlation are used to ensure temporal alignment between up-sampled eye movement signals and EEG data. Lag analysis detects time shifts between the signals, while cross-correlation identifies the optimal alignment point by measuring similarity at different time lags. Aligning both signals ensures that features represent the same time window, which is essential for accurate multimodal emotion recognition. The processed EEG signals were mapped to the theta, alpha, beta, and gamma bands, and 4 s (800 sampling points) was used as a non-overlapping segment to divide time Windows to generate a sufficient number of samples, and the number of times Windows was regarded as the number of samples, which was convenient for subsequent feature extraction and emotion recognition research.

3.2.2 Eye movement signal preprocessing

The eye tracking system captures eye movement signals triggered by events, and the sampling frequency is significantly lower than that of EEG acquisition equipment, at only 20 Hz. In this article, the eye movement signal is upsampled to 200 Hz to ensure consistency with the EEG signal, facilitating subsequent multimodal feature fusion [22]. The multimodal emotion recognition model, which combines eye movement and EEG signals, offers improved accuracy, robustness, and interpretability compared to single-mode systems. By integrating neural and behavioural data, it captures a wider range of emotion-related information. The fusion of EEG DE, MST connectivity, and eye movement features enhances classification performance, achieving a higher average accuracy (91.4%) than individual modalities. This approach also reduces vulnerability to noise and increases generalisability across users and contexts. The changes in Pupil Diameter in eye movement signals are highly affected by the ambient brightness, and only by removing the light reflection component can pupil information related only to emotion be obtained. Based on the assumption that different subjects exhibit similar pupil changes in response to the same stimulus material, a light reflection model was established, and the effect of brightness was approximately removed using the PCA (Principal Component Analysis) algorithm.

The proposed method applied a preprocessing pipeline to ensure clean and synchronised multimodal data. EEG signals were downsampled to 200 Hz, bandpass filtered (1–75 Hz) to remove noise and artefacts, and segmented into 4 s windows. Eye movement data, originally at 20 Hz, were upsampled to 200 Hz for alignment. PCA was used to remove brightness-related artefacts from pupil diameter measurements. This ensured accurate, denoised, and time-aligned signals for reliable emotion recognition.

3.3 Analysis of experimental results

Taking EEG and eye movement emotion data as the research object, the DE features, MST attributes, and eye movement characteristics of the full band were fused by serial splicing, respectively, and the four emotions of happy, sad, fear and neutral were classified. The proposed used 168 red culture movie clips, rated by 44 volunteers for emotional arousal and valence. Based on these ratings, 72 clips that best matched the target emotions (happiness, sadness, fear, neutral) were selected. Each participant viewed 24 clips across three sessions, and emotional responses were self-assessed. Clips that failed to evoke clear emotions were excluded to ensure the quality of the data. The 5-fold cross-validation strategy was employed when the LIBSVM toolkit was used for training; the samples were divided into 5 parts on average, with 4 parts randomly selected as training data and 1 part used as test data. The average value of the test results was obtained as the classification accuracy rate of the model, and the experimental results are shown in Table 2.

DE features from EEG signals significantly enhance emotion classification accuracy by effectively capturing frequency-domain brain activity. Eye movement features provide essential behavioural insights, while MST attributes contribute spatial connectivity information. The

Table 2: Classification results of multimodal physiological signal feature fusion

Classification object	Eye movement + DE (%)	Eye movement + MST (%)	Eye movement + MST + DE (%)
Fear & neutral	81.3	70.2	83.2
Fear & happy	97.7	97.3	98.3
Fear & sad	79.6	78.9	84.8
Happy & neutral	92.5	85.6	93.2
Happy & sad	91.6	78.9	91.7
Neutral & sad	96.6	88.8	97.2
Average classification accuracy	89.9	83.3	91.4

combination of DE, MST, and eye movement features achieves an accuracy of 91.4%, highlighting the advantages of multimodal fusion. Integrating neural and behavioural signals through MST enhances emotion recognition, providing a comprehensive and culturally sensitive approach to understanding emotional responses with superior classification performance compared to single or dual modalities.

The average classification accuracy of the fusion of eye movement and EEG features is better than that of a single mode. Single modalities, such as EEG and eye movement signals, provide only partial views of emotional resonance, thereby limiting their effectiveness. EEG provides internal neurophysiological responses but lacks contextual and behavioural information, while eye movement reflects external cognitive responses but cannot access the brain's intrinsic states. The multimodal approach, which combines EEG features and eye movement features, enhances recognition accuracy by fusing spatial, frequency-domain, and behavioural information, leading to more comprehensive and reliable emotional assessments. The emotion recognition effect of the fusion of DE features is better than that of the fusion of MST attributes, and the emotion recognition effect of the fusion of DE features and MST attributes is the best. This may be because the extracted DE features consider the frequency domain information of EEG signals. At the same time, MST attributes retain the spatial information between channels from the aspect of functional connection, and eye movement features compensate for the limitations of external behavioural features. The three can effectively integrate EEG signals and eye movement signals across modes to obtain more emotion-related information. And make full use of the complementarity between this multimodal information to have greater advantages than single-modal emotion recognition.

4 Conclusion

Emotion is a crucial factor that influences learners' daily learning and life, and it accompanies learners throughout their entire learning process. However, the lack of emotion can easily cause learners to experience low learning enthusiasm, burnout, and poor emotional engagement, as well as other problems that restrict the long-term development of online learning. In this paper, the influence of biomechanical factors on the emotional resonance of participants in the red cultural experience was investigated. A multimodal physiological signal emotion recognition model based on EEG and eye movement was constructed, and the accuracy of participants' classification of happy, sad, fearful, and neutral emotions was evaluated.

Overall, this study has achieved certain milestones and met the established research objectives. The research results presented in this paper may serve as a reference for subsequent studies in this field, particularly in terms of feature extraction and fusion strategies. However, since only two EEG and eye movement modes are considered in this paper, the need for emotion recognition in many applications will still be limited. Then, it will not be limited to certain physiological signals. Instead, more emotion-related modal data are used for fusion, such as ECG, myoelectricity, skin electricity, and voice, as well as expression, gesture, and other data to improve emotion recognition.

Funding information: Anhui Provincial Construction of "Three Comprehensive Education" Pilot Provincial Construction and Ideological and Political Work Ability Improvement Project of Colleges and Universities (SZTSJH-2022-7-9); Anhui Provincial Graduate Education and Teaching Reform Research Project 2022jyjxggyj342; Anhui Provincial Social Science Innovation and Development Research Project 2024CX033; Anhui Provincial University Outstanding Young Backbone Teachers Domestic Visiting Training Program, Project No. GXGNFX2022028; Fuyang Social Science Planning Project, Project No. FSK2024023.

Author contributions: Liang Huawei is responsible for designing the framework, analysing the performance, validating the results, and writing the article. Bi Xiumei is responsible for collecting the information required for the framework, providing software, conducting critical reviews, and administering the process. All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Conflict of interest: The authors state no conflict of interest.

Data availability statement: All data generated or analysed during this study are included in this published article.

References

[1] N. Thammasan, K. Moriyama, K. I. Fukui, and M. Numao, "Continuous music-emotion recognition based on electroencephalogram," *IEICE Trans. Inf. Syst.*, vol. 99, no. 4, pp. 1234–1241, 2016.

[2] X. Li, Y. Wu, M. Wei, Y. Guo, Z. Yu, H. Wang, et al., "A novel index of functional connectivity: phase lag based on Wilcoxon signed rank test," *Cognit. Neurodyn.*, vol. 15, no. 4, pp. 621–636, 2021.

[3] M. Oliva and A. Anikin, "Pupil dilation reflects the time course of emotion recognition in human vocalizations," *Sci. Rep.*, vol. 8, no. 1, pp. 4871–4881, 2018.

[4] Y. Lu, W. L. Zheng, B. Li, and B. L. Lu, "Combining eye movements and EEG to enhance emotion recognition," *International Joint Conference on Artificial Intelligence (IJCAI)*, 2015, pp. 1170–1176.

[5] W. L. Zheng, W. Liu, Y. Lu, B. L. Lu, and A. Cichocki, "EmotionMeter: A multimodal framework for recognizing human emotions," *IEEE Trans. Cybern.*, vol. 49, no. 3, pp. 1110–1122, 2019.

[6] M. A. Islam, A. Hamza, M. H. Rahaman, J. Bhattacharjee, and M. M. Rahman, *Mind reader: A facial expression and EEG based emotion recognizer*, India: IEEE; 2018.

[7] Y. H. Kwon, S. B. Shin, and S. D. Kim, "Electroencephalography based fusion two-dimensional (2D)-convolution neural networks (CNN) model for emotion recognition system," *Sensors (Basel)*, vol. 18, no. 5, pp. 1383–1396, 2018.

[8] F. Abtahi, T. Ro, W. Li, and Z. Zhu, "Emotion analysis using audio/video, EMG and EEG: A dataset and comparison study," *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2018, pp. 10–20.

[9] Y. Ye, T. Pan, Q. Meng, J. Li, and L. Lu, "Online ECG emotion recognition for unknown subjects via hypergraph-based transfer learning," *IJCAI*, 2022, pp. 3666–3672.

[10] L. H. Ting, B. Gick, T. M. Kesar, and J. Xu, "Ethnokinesiology: towards a neuromechanical understanding of cultural differences in movement," *Philos. Trans. B*, vol. 379, no. 1911, p. 20230485, 2024.

[11] S. K. Khare, V. Blanes-Vidal, E. S. Nadimi, and U. R. Acharya, "Emotion recognition and artificial intelligence: A systematic review (2014–2023) and research recommendations," *Inf. Fusion.*, vol. 102, p. 102019, 2024.

[12] S. N. M. S. Ismail, N. A. A. Aziz, and S. Z. Ibrahim, "A comparison of emotion recognition system using electrocardiogram (ECG) and photoplethysmogram (PPG)," *J. King Saud. Univ.-Comput. Inf. Sci.*, vol. 34, no. 6, pp. 3539–3558, 2022.

[13] M. A. Hasnul, N. A. Ab Aziz, and A. Abd Aziz, "Evaluation of TEAP and AuBT as ECG's feature extraction toolbox for emotion recognition system," *2021 IEEE 9th Conference on Systems, Process and Control (ICSPC 2021)*, IEEE, 2021, pp. 52–57.

[14] F. Tokmak, A. Subasi, and S. M. Qaisar, "Artificial intelligence-based emotion recognition using ECG signals," *Applications of Artificial Intelligence Healthcare and Biomedicine*, Amsterdam, Netherlands: Academic Press, 2024, pp. 37–67.

[15] E. M. Younis, S. Mohsen, E. H. Houssein, and O. A. Ibrahim, "Machine learning for human emotion recognition: a comprehensive review," *Neural Comput. Appl.*, vol. 36, no. 16, pp. 8901–8947, 2024.

[16] A. F. Claret, K. R. Casali, T. S. Cunha, and M. C. Moraes, "Automatic classification of emotions based on cardiac signals: a systematic literature review," *Ann. Biomed. Eng.*, vol. 51, no. 11, pp. 2393–2414, 2023.

[17] L. Billeci, C. Sanmartin, A. Tonacci, I. Taglieri, G. Ferroni, R. Marangoni, et al., "Wearable sensors to measure the influence of sonic seasoning on wine consumers in a live context: a preliminary proof-of-concept study," *J. Sci. Food Agric.*, vol. 105, no. 3, pp. 1484–1495, 2025.

[18] R. Zhou, C. Wang, P. Zhang, X. Chen, L. Du, P. Wang, et al., "ECG-based biometric under different psychological stress states," *Comput. Methods Prog. Biomed.*, vol. 202, p. 106005, 2021.

[19] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked autoencoders are scalable vision learners," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 16000–16009.

[20] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, et al., "Swin transformer: Hierarchical vision transformer using shifted windows," *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 10012–10022.

[21] Y. Tian, G. Chen, and Y. Song, "Enhancing aspect-level sentiment analysis with word dependencies," *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, 2021, pp. 3726–3739.

[22] M. H. Phan and P. O. Ogunbona, "Modelling context and syntactical features for aspect- based sentiment analysis," *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 3211–3220.