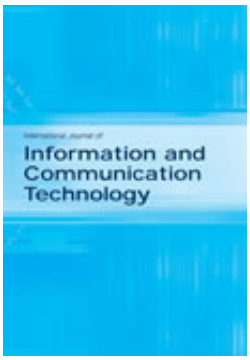


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

MA-HGCN: modality attentive hierarchical graph convolutional network for EEG-fNIRS brain-computer interface system

Jie Zhang

DOI: 10.1504/IJICT.2025.10072900

Article History:

Received:	21 May 2025
Last revised:	16 July 2025
Accepted:	16 July 2025
Published online:	26 August 2025

MA-HGCN: modality attentive hierarchical graph convolutional network for EEG-fNIRS brain-computer interface system

Jie Zhang

Guangzhou Xinhua University,
Guangzhou, Guangdong, 510000, China
Email: zhangjie@xhsysu.edu.cn

Abstract: In order to address the challenging issue of cross-modal correlation modelling in EEG-fNIRS multimodal fusion, this paper suggests a novel multi-feature graph construction approach. The graph structure data is constructed by extracting multiple time-domain and frequency-domain features from the original EEG-fNIRS signals. The modal attention hierarchical graph convolution neural network then uses the positional relationship between the scalp electrodes to extract high-level regional features. To better extract the connectivity between various modal nodes, the hierarchical graph convolution handles the adjacency between the vertical and horizontal electrodes independently, while the modal attention mechanism can assign varying weights to distinct modal nodes. Experiments show that this strategy achieves 96.13% and 98.25% in motor imagery and mental arithmetic, respectively, improves classification accuracy by 2.51% to 5.37% when compared to the current optimal methods. Furthermore, it provides a new viewpoint on multimodal brain signal fusion by validating the simplified feature combinations.

Keywords: BCIs; EEG-fNIRS; MA-HGCN; motor imagery; mental activity.

Reference to this paper should be made as follows: Zhang, J. (2025) 'MA-HGCN: modality attentive hierarchical graph convolutional network for EEG-fNIRS brain-computer interface system', *Int. J. Information and Communication Technology*, Vol. 26, No. 31, pp.40–71.

Biographical notes: Jie Zhang is a Lecturer and has a Master's degree and graduated from Jinan University in 2015. Currently, she worked in Guangzhou Xinhua University. Her research interests include computer application, brain-computer interfaces and deep learning.

1 Introduction

Brain-computer interface (BCI) technology constitutes an innovative approach to human-computer interaction, integrating multidisciplinary domains. It decodes neural activity to manipulate external devices, enabling information exchange between the brain and the environment, with applications spanning neuroscience, rehabilitation medicine, and artificial intelligence. Currently, electroencephalography (EEG) represents the most advanced BCI technology, wherein electrodes are positioned on the scalp to amplify and record spontaneous bioelectrical brain signals, generating waveforms (Lazeyras et al.,

2000; Sun et al., 2024). Functional near-infrared spectroscopy (fNIRS) is a non-invasive, real-time optical neuroimaging technique that employs near-infrared light to continuously monitor fluctuations in tissue blood oxygenation over prolonged durations (Pinti et al., 2020). Multi-modal BCI integrates diverse modalities, wherein the combination of EEG and fNIRS addresses the limited spatial resolution of EEG, while the temporal delay in fNIRS measurements is alleviated through optimised feature extraction and sophisticated algorithmic approaches.

Therefore, multi-modal BCI technology is expected to become a pivotal research focus in the future development of BCI systems and will undoubtedly expedite the practical implementation of such systems (Nguyen et al., 2017; Sargent et al., 2018; Verma et al., 2019; Uchitel et al., 2021). Pouliot selected EEG+fMRI and EEG+fNIRS to quantify nonlinear hemodynamic responses in epileptic patients, utilising the Volterra kernel to extend the acquired bimodal data to the second order, with the study revealing statistically significant nonlinearity across all modalities for all patients (Pouliot et al., 2012). Morioka et al. introduced a decoding methodology for EEG cortical currents based on fNIRS-derived information, utilising fNIRS features as prior knowledge to estimate cortical EEG signals within the EEG framework (Morioka et al., 2014). Ahn et al. (2016) introduced a feature combination approach based on feature normalisation, wherein EEG and fNIRS features were normalised within the range of 0-1, followed by the application of feature summation to achieve improved BCI performance in fatigued driving detection tasks. The study explores the augmentation of BCI performance through the integration of EEG and fNIRS, utilising deep neural networks (DNNs) for brain state classification (Chiarelli et al., 2018). Sun et al. (2020) employed SVM for classification following the direct concatenation of EEG and fNIRS features. Cicalese et al. (2020) proposed a hybrid EEG+fNIRS model employing a Pearson correlation coefficient-based feature selection approach for EEG and fNIRS feature extraction, followed by classification using a linear discriminant analysis classifier, achieving the highest accuracy through the fusion of EEG and fNIRS features. Deligani et al. (2021) introduced a mutual information-based feature selection approach to establish a framework for feature classification of fused EEG+fNIRS data, with the objective of optimising complementarity, redundancy, and correlation. Features of various modalities are usually either directly concatenated or filtered according to the statistical results of the features in the previously mentioned feature fusion steps found in the literature. These methods only concentrate on the fusion between complete multimodal feature vectors, ignoring the differences and connections between the signals of each channel in EEG and fNIRS, thereby not making full use of the useful information contained in the EEG and fNIRS signals. Using inter-channel links in the signal fusion process is still a difficult problem.

A correlation has been identified between EEG and fNIRS signal channels. When these channels are conceptualised as nodes, the features extracted from each channel are treated as node attributes, and the interconnections among associated channels are defined as edges, EEG and fNIRS can be represented as graph-structured data. Graph convolutional neural networks (GCN) constitute a class of neural networks specifically developed for analysing graph-structured data (Kipf and Welling, 2016; Yang et al., 2024). Numerous researchers have utilised GCN to represent channel features as nodes within the graph domain, aggregating adjacent node features to generate new representations, with the objective of improving classification performance for EEG and fNIRS data. This study proposes a novel multimodal data fusion framework designed to extract and interpret complementary synergistic features from multimodal brain signals

(Kumar et al., 2023). The proposed methodology is grounded in GCN, where the attributes of EEG and fNIRS multimodal data are first integrated. Node similarity is used to build the graph, and where there are no direct links between observations, the k-nearest neighbour (KNN) technique is used to create connections between nodes during graph formation, with each node being connected to only semantically comparable nodes. Qu et al. (2024) proposed a multi-source data fusion algorithm utilising GCN. A multi-source data fusion framework was subsequently developed employing GCN. Granger causality (GC) is integrated as edge feature information within the GCN model based on node feature data, with a brain region neighbourhood matrix introduced for EEG and fNIRS to facilitate multi-source data fusion. On the basis of node feature information, GC was incorporated as edge feature information into the GCN model, and the concept of brain region neighbour matrix was introduced in EEG and fNIRS to facilitate multi-source data fusion (Ren et al., 2024). However, traditional graph convolutional neural networks show some difficulties in specific graph architectures like brainwave sampling. Since EEG electrodes are interconnected via transverse and longitudinal edges, a voltage differential exists between the electrodes, and each edge encompasses distinct information; thus, it is inappropriate to denote both sides of the graph's adjacency matrix as 1 (Zeng et al., 2021; Zheng et al., 2021). Therefore, when addressing specialised graph structures, it is feasible to employ distinct graph convolution layers to separately process the longitudinal and transverse adjacencies between electrode pairs.

Moreover, the feature information in EEG and fNIRS is not uniformly significant, and in recent years, the attention mechanism has been progressively integrated into BCI recognition to extract critical feature information from EEG and fNIRS. In order to mine the significance of different electrode channels of EEG signals and different times, respectively. Kloetzer proposed an attention-based CRNN network using a channel-based attention-based convolutional neural network (CNN) model and a self-attention-based LSTM model. The model's performance in the DEAP dataset for both validity and arousal dimensional binary classification reached 93.72% and 93.38%, respectively. The model outperforms the conventional CRNN model by 30.75% and 26.26%, respectively, achieving 93.72% and 93.38% in the DEAP dataset (Kloetzer and Mahulea, 2020). This work tackles the issue of efficient feature fusion and discriminative feature learning in emotion recognition utilising EEG and fNIRS data (Chen et al., 2024). The authors suggest a novel model, the graph convolution and capsule attention network, to boost accuracy by employing a pearson correlation adjacency matrix to integrate EEG and fNIRS information into superior primary capsules. The implementation of a capsule attention module facilitates the allocation of diverse weights to the capsules. Facilitates the allocation of diverse weights to these capsules, enhancing the selection process inside the dynamic routing mechanism for superior categorisation. It is evident that the attention mechanism assigns varying weights to distinct EEG and fNIRS channels during fusion, resulting in more focused modal fusion that can significantly enhance the model's classification effect.

To enhance the utilisation of discriminative information in EEG and fNIRS signals, a multimodal attentional hierarchical graph convolutional network (MA-HGCN) model is proposed, drawing inspiration from GCNs and attention mechanisms. Firstly, a spatial modal attention mechanism is incorporated into graph feature fusion to determine the weights of diverse channels for signals from different modes; subsequently, two independent adjacency matrices are employed within the GCN framework to denote horizontal and vertical edges, thereby constructing hierarchical graph convolutional

layers. Through these layers, each channel is regarded as a node for mapping features onto the graph domain, while aggregating neighbouring nodes' features to derive new characteristics; ultimately, the model's efficacy was validated utilising publicly accessible datasets. The principal contributions of this article are summarised as follows:

1 To balance the influences of EEG and fNIRS channel features, a cross-modal attention module was developed. The multimodal attention mechanism can assign varying weights to nodes across different modalities, facilitating more comprehensive extraction of intermodal node connections.

2 By incorporating vertical and horizontal graph convolutional layers into GCN, an HGCN model tailored for EEG-fNIRS multimodal data is proposed, effectively leveraging the structural information of electrode placement positions within the brain nomenclature system. In contrast to conventional graph convolutional networks, the refined model addresses limitations in managing inconsistent adjacency relationships and enhances detection performance.

3 The classification recognition performance of MA-HGCN was validated through two distinct datasets: motor imagery (MI) and mental activity (MA). Furthermore, this study demonstrates, from the perspective of varying modalities, MA-HGCN can integrate either HbO or HbR signals with EEG to sustain robust performance while minimising the number of modalities.

The remainder of the paper is structured as follows. Section 2 introduces the relevant works to the graph convolution network and the feature fusion. Section 3 elaborates on the model and theoretical underpinnings presented in this research, encompassing feature extraction, HGCN, and the attention process. Section 4 presents the experimental data and provides a comprehensive discussion of them. Ultimately, Section 5 presents the conclusions.

2 Related works

Raw signal preprocessing, feature extraction, and classification are the phases that make up a standard EEG-fNIRS multimodal detection approach. fNIRS and EEG devices first collect and amplify signals of brain activity, then filter and preprocess the signals(Nguyen et al., 2017). Following the completion of preprocessing, the system extracts particular signal features. Signal peak, slope, mean, kurtosis, skewness, and power spectral density (PSD) are frequently utilised feature characteristics. These feature data will be fed into the classifier's subsequent stage, which consists of machine learning techniques represented by SVM and LDA with rigorous mathematical derivations; CNN, which make use of the image's convolutional invariance and are useful for processing 2D image inputs; and recurrent neural network (RNN), which are primarily made up of recurrent cells. The output of this moment and its state, which is appropriate for processing temporal information, is produced by the recurrent cells after receiving the input of each moment and the state of the preceding instant. According to the recurrent convolutional neural network (RCNN), the RCNN incorporates the benefits of both recurrent and CNNs, making it a hybrid of the two. The aforementioned neural networks either ignore the structural features of the channel dimension entirely, like RNNs, or they only partially utilise them, like CNNs, which can only process a subset of the adjacent

channels. Since the electrodes on the scalp are truly a structure in non-Euclidean space, it makes sense that a standard CNN would not be able to synthesise the input of all nearby channels. Consequently, graph convolutional neural networks (GCN) an alternative method are examined in this research. The graph convolutional neural network's convolution kernel is ideal for convolution in non-Euclidean space. Its convolution object is not to locate neighbouring coordinates in Euclidean space based on current coordinates and convolve them, but rather to locate neighbouring nodes of the current node based on the graph's adjacency matrix and convolve them (Zeng et al., 2021; Zheng et al., 2021). This is ideal for the structure of the EEG location on the scalp, so we can effectively utilise the channels with the graph convolutional neural network. Structural characteristics that are full in dimensions.

2.1 GCN

The frequency domain convolution of a graph is more intuitively known as the convolution operation using the Fourier transform of a graph. The graph frequency domain convolution can be derived from the graph Fourier transform and the graph convolution operation is performed for f and g:

$$\begin{aligned} g_\theta(\Lambda) &= U^T g \\ x *_G g &= U g_\theta(\Lambda) U^T f \end{aligned} \tag{1}$$

In this section, the function $f \in R^N$ is defined as the set of all nodes in the graph, and g_θ is defined as the Fourier transform of the convolution kernel g in the graph as a function of $g_\theta(\Lambda)$ with respect to the eigenvalue Λ. The function $U^T f$ denotes the Fourier transform of f in the graph. The preceding section thus completes the derivation of the frequency domain graph convolution. However, the calculation of the feature matrix U necessitates substantial computational resources. Researchers posit that the $g_\theta(\Lambda)$ should be approximated to the k^{th} order by the truncated expansion $g_{\theta'}(\Lambda)$ of Chebyshev polynomial $T_k(f)$.

$$g_{\theta'}(\Lambda) = \sum_{k=0}^{k} \theta'_k T_k(\tilde{\Lambda}) \tag{2}$$

The primary objective $\Lambda = \frac{2}{\lambda_{\max}} \Lambda - I_N$ is to scale the diagonal matrix into the range [–1, 1]. The maximal eigenmatrix of L is denoted by $\lambda_{\max}$, $T_k(f) = cos(^{k*-1}(f))$, $\theta' \in R^k$ defined as the vector of Chebyshev coefficients, and the frequency-domain map convolution operation can be simplified as follows:

$$\begin{aligned} f *_G g &= U\left(\sum_{k=0}^{k} \theta'_k T_k(\tilde{\Lambda})\right) U^T f \\ &= \sum_{k=0}^{k} \theta'_k T_k\left(U \tilde{\Lambda} U^T\right) f \\ &= \sum_{k=0}^{k} \theta'_k T_k(\tilde{L}) f \end{aligned} \tag{3}$$

L was rescaled to obtain $\tilde{L} = \frac{2}{\lambda_{\max}} L - I_N$. Given that the eigenvalues of the symmetrically normalised Laplace matrix are [0, 2], it can be deduced that the condition $\lambda_{\max} \approx 2$ exists. Furthermore, by approximating equation (3) only up to the 1-order Chebyshev polynomials, it reduces to:

$$g_{\theta'} * f \approx \theta'_0 f + \theta'_1 (L - I_N) f = \theta'_0 f - \theta'_1 D^{-\frac{1}{2}} A D^{-\frac{1}{2}} f \tag{4}$$

The definition of condition $\theta' = \theta'_0 = -\theta'_1$, equation (4) can be reduced to the following equation, where the definition functions $S(A) = I_N + D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$ as the propagation matrix of the adjacency matrix A:

$$g_{\theta'} * f = \theta' \left(I_N + D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \right) f = \theta' S(A) f \tag{5}$$

In summary, the linear propagation formula for the frequency domain convolutional layer of the graph can be expressed as:

$$Z = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} X \Theta \tag{6}$$

It can thus be concluded that $\tilde{A} = I_N + A, \tilde{D}_{ii} = \sum_j^N \tilde{A}_{ij}$ is also applicable to $S(A) = \tilde{D}^{-\frac{1}{2}} A \tilde{D}^{-\frac{1}{2}}$. $\Theta \in R^{M \times F}$ denotes the weight matrix of the graph convolutional neural network layer, while $Z \in R^{N \times F}$ signifies the output of the graph convolutional neural network layer. In addition, the incorporation of an activation function subsequent to the implementation of the linear propagation formula accomplishes the execution of all operations associated with a graph frequency domain convolution, thereby yielding the following result:

$$H^{l+I} = \sigma \left(S(A^l) H^l W^l \right) \tag{7}$$

In this equation, l denotes the layer number, H indicates the graph convolution layer, W signifies the weight matrix (i.e., the coefficients of the Chebyshev polynomials previously referenced), and σ represents the activation function. In the context of this paper, the frequency domain based GCN is employed, with all layers utilising an identical adjacency matrix $A \in R^{N \times N}$, $\oplus$ this matrix signifies the connection between two nodes, expressed as follows:

$$A^{ij} = \begin{cases} 1, & for\ x_i \oplus x_j \\ 0, & for\ !x_i \oplus x_j \end{cases} \tag{8}$$

2.2 Feature fusion

In a class of data fusion techniques, feature fusion aims to eliminate redundant and repetitive information between multiple pieces of information while simultaneously integrating information from various access routes or times at the feature layer to compensate for the shortcomings of a single piece of information. Data fusion techniques can be categorised into three groups based on the various levels of fusion:

1 Data layer fusion is the combination of the initial data that each sensor gathered. Data layer fusion has the advantage of preserving all of the original data, which improves the representation of the correlation between the data and increases the accuracy of the classification findings. The drawback is that it is not appropriate for systems with high real-time needs and increases the computational time complexity following fusion.

2 Feature layer fusion is the process of combining feature vectors that were taken from various modal data sources. Compared to data layer fusion, feature layer fusion has the advantage of requiring less computation and improving the system's real-time performance. The drawback is that the information connection across modes cannot be described by the features that are retrieved independently from various modal information, which could lessen the system's impact.

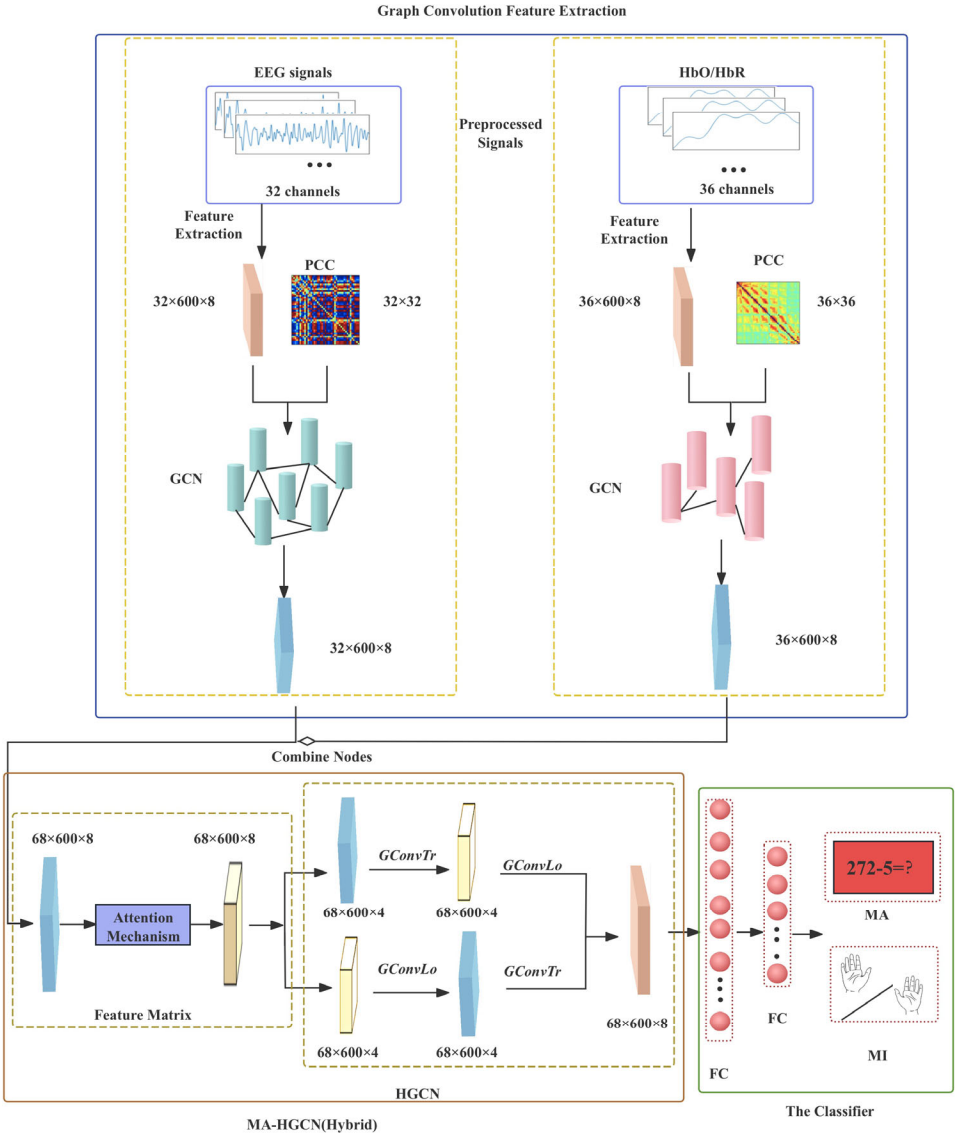
3 Decision level fusion: the target's final classification results are obtained by fusing the preliminary classification results that were obtained by processing each modal information. Independent processing of each modal information, such as feature engineering, categorisation, and raw data preparation, is necessary for decision level fusion. The benefit of this approach is that the decision layer fusion has minimal computational time cost and strong real-time performance, and the useful information modalities can be freely chosen to increase the system's fault tolerance.

It is typically possible to fuse the EEG and fNIRS signals at the feature layer in the study of bimodal BCI based on EEG and fNIRS because it is challenging to fuse the EEG and fNIRS signals at the data layer due to their different sampling rates and frequency bands of interest (Deligani et al., 2021). Additionally, it is challenging to make an effective classification using the fNIRS unimodal signals alone, which makes it difficult for the fNIRS signals to contribute to the final classification results when using the decision-layer fusion. At the feature layer, and fNIRS signals are combined.

3 Method

The EEG-fNIRS data recognition framework, as illustrated in Figure 1, sequentially comprises graph feature extraction, MA-HGCN-based graph feature fusion, and a classifier module. The EEG and fNIRS data are represented as graph-structured data, with the correlation between each channel feature and each channel computed and subsequently input into the MA-HGCN model as a node feature matrix and adjacency matrix, respectively, to achieve multimodal fusion recognition.

Figure 1 MA-HGCN bimodal fusion recognition model (see online version for colours)



3.1 Graph convolution feature extraction

To acquire the model inputs, which include the channel dimension, time dimension, and feature dimension collectively, the original EEG signal and fNIRS signal are first preprocessed, and the features are then recovered from the frequency and time domains, respectively. Lastly, the signal channels are described as nodes using GCN, which are mapped into the graph domain according to their topological structure information. Then, using the connectivity relationship between the channels, the information of neighbouring nodes is combined to create new channel nodes, which are fused features.

3.1.1 Pre-processing

Electrodes are applied to the scalp to record electroencephalographic (EEG) signals, which are electrical impulses produced by brain activity. Although EEG data has a relatively poor spatial resolution, it has a high temporal resolution, catching variations in the signal on the order of milliseconds. The presence of these frequency bands correlates with various brain functioning states. Each of the five common EEG signal bands has a distinct frequency and amplitude range. The δ-wave (frequency range 0.5–4 Hz), θ-wave (frequency range 4-8Hz), α-wave (frequency range 8–13 Hz), β-wave (frequency range 13-30Hz), and γ-wave (frequency range 30–100 Hz) are the five bands in question. Near-infrared light is used in fNIRS, a novel optical brain function imaging approach, to identify changes in tissue oxygenation in a non-invasive, real-time, and continuous monitoring manner. A novel form of optical brain imaging called fNIRS technology employs near-infrared light to continuously, continuously, and non-invasively measure changes in tissue blood oxygenation over an extended period of time. With its high temporal resolution, moderate spatial resolution, and robust tolerance to motion noise, fNIRS technology is inexpensive, portable, and safe. It can also measure changes in HbO, HbR, and total hemoglobin (HbT) concentrations all at once, which is much more informative and ideal for BCI.

Prior to feature extraction, pre-processing is necessary for EEG and functional NIR signals. After the EEG signals were filtered, de-baselined, and re-referenced from 0.5 to 45.0 Hz, they were downsampled to 200 Hz and their artifacts eliminated using independent component analysis (ICA). The fNIRS gadget emits near-infrared light from a source, while the detector captures the photons that are dispersed and absorbed by the cerebral medium. When the scattering and absorption coefficients are known, Lambert-Beer's law can be employed to calculate the relationship between incident and outgoing light in a non-scattering medium, thereby allowing for the determination of concentration variations of oxygenated haemoglobin (HbO) and deoxygenated hemoglobin (HbR) (Trakoolwilaiwan et al., 2017). The HbO and HbR signals were processed with a third-order Butterworth bandpass filter with a frequency range of 0.01 to 0.09 Hz. Subsequent procedures are uniformly implemented for both fNIRS and EEG inputs. Signal segments were obtained from three sessions, each Signal segments were obtained from three sessions, each comprising 20 repetitions, covering the stimulation duration (0 s to 10 s). Baseline correction was executed utilising the mean signal amplitude captured during the pre-stimulus instruction phase (–2 s to 0 s).

3.1.2 Feature extraction and selection

Following the pre-processing of EEG and fNIRS signals, feature processing can commence. For the extraction of EEG signal features, primarily from the frequency and time domains, this paper employs two methods to extract frequency domain features: PSD and differential entropy (DE).

For the frequency domain features of fNIRS, we extracted the PSD features of HbO and HbR in the range of 0.01-0.1 Hz using Fourier transform, respectively (Zhang et al., 2021). PSD features like DE, are based on the Fourier transform principle, which transforms the fluctuations of the signal surface topography into the spatial frequency distribution in the designated frequency domain range. The following is the precise

processing flow: First, the discrete Fourier transform is carried out for the finite-length discontinuous HbO and HbR signals $x(n)$, which is provided by:

$$P(\omega) = \frac{1}{N}\left|X_N\left(e^{-j\omega}\right)\right|^2 = \frac{1}{N}\left|\sum_{n=1}^{N} x(n)e^{-j\omega n}\right|^2, n = 1,...,N \tag{9}$$

In this case, $P(\omega)$ is the end result of calculating the power spectrum from the average periodogram, and $X_N(e^{j\omega})$ is the Fourier transform of the sequence $x(n)$.

For the frequency domain features of EEG signals, we extracted the DE features in the range of 0.5–45 Hz by using the short-time Fourier transform (Duan et al., 2013), which is calculated as follows:

$$DE = h(x) - \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right) dx$$
$$dx = \frac{\log\left(2\pi e\sigma^2\right)}{2} \tag{10}$$

where μ and σ represent the mean and standard deviation of the EEG signal, respectively, and e is the Euler constant.

The EEG and fNIRS characteristics derived from the time domain perspective encompass maximum, mean, variation, peak-to-peak, median, kurtosis and skewness. The maximum illustrates the attributes of the signal curve, the mean value indicates the magnitude of the signal, the rectified mean signifies the average of the absolute values of all sampling points, the standard deviation reflects the stability of the entire dataset, the peak-to-peak denotes the difference between the minimum and maximum values of the data, the value in the centre of the sorted data is called the median. The kurtosis describes the sharpness of the peaks among all sampling points, represented by the formula:

$$x_{kurt} = \frac{\frac{1}{2f}\sum_{i=1}^{2f}\left(x_i - x_{avg}\right)^4}{\left(\frac{1}{2f}\sum_{i=1}^{2f}\left(x_i - x_{avg}\right)^2\right)^2} \tag{11}$$

Skewness denotes the extent of asymmetry among all sampled points and is expressed as:

$$x_{skew} = \frac{\frac{1}{2f}\sum_{i=1}^{2f}\left(x_i - x_{avg}\right)^3}{\left(\frac{1}{2f}\sum_{i=1}^{2f}\left(x_i - x_{avg}\right)^2\right)^{\frac{3}{2}}} \tag{12}$$

User data is segmented into 10-second epochs from the commencement of the task. In this configuration, there are 20 observations every session and a total of 60 observations. The EEG and fNIRS signals are divided into 1-second intervals, resulting in 600 segments, denoted as T = 600. The quantity of signal channels is represented as C.

Subsequently, a total of $C \times T$ signal analysis segments must be extracted for feature extraction, with each segment derivable from both time and frequency domains to yield F-dimensional features. The overall input information can be delineated as follows $X \in R^{C \times T \times F}$: for EEG: $C = 32$, $T = 600$, $F = 8$; for HbO/HbR of fNIRS: $C = 36$, $T = 600$, $F = 8$.

3.1.3 EEG-fNIRS graph data construction

The EEG and fNIRS signals must be converted into graph-structured representations in order for data to be fed into a graph convolutional neural network. A set of nodes and edges, denoted as G = (V, E), make up graph-structured data, where V stands for the set of nodes and E for the set of edges. The EEG and fNIRS signals are represented as graph-structured data by treating each channel as a node and using the extracted features for each channel as node attributes to create the feature matrix. The adjacency matrix is created using the inter-channel connectivity as edges. Eight retrieved EEG features and fNIRS statistical features make up the graph-structured data's feature matrix, while the adjacency matrix is produced by computing the Pearson correlation coefficients between the two-channel signals. Bao et al. (2022) achieved notable results in EEG and fNIRS recognition tests by using the Pearson correlation coefficient as the adjacency matrix for graph convolutional networks.

$$PCC(i,k) = \frac{\frac{1}{T}\sum_{t=1}^{T}(y_{it} - \mu_i)(y_{kt} - \mu_k)}{\sigma_i \sigma_k} \tag{13}$$

The EEG-fNIRS channel adjacency matrix was created by calculating the correlation between two EEG and fNIRS signal channels using Pearson's correlation coefficient. By merging different channel EEG and fNIRS nodes, the GCN graph convolution layer facilitates feature fusion by aggregating features from numerous nearby nodes using the adjacency matrix to produce new node features. The GCN graph convolution layer receives the EEG, HbO, and HbR graph structure data separately. Features are then extracted to provide the appropriate graph convolution feature matrices for EEG, HbO, and HbR. By combining the properties of nearby nodes, the graph convolution layer creates new node characteristics. This improves the information about linked nodes in comparison to the initial features, which is useful for later modules that involve various modal nodes.

3.2 MA-HGCN

For the MA-HGCN framework, where the EEG and fNIRS features output from the graph convolution feature extraction layer are first subjected to the spatial attention weight fusion. To accomplish modal fusion more precisely, different EEG and fNIRS channels are assigned distinct weights during the fusion process. Then, because each edge contains different information and there is a voltage difference between the EEG and fNIRS electrodes, the hierarchical graph convolution network model is proposed. This model uses the vertical and horizontal graph convolution layers, respectively, to solve the issue of the edges' differing information and to further increase the recognition rate.

3.2.1 Attention mechanism

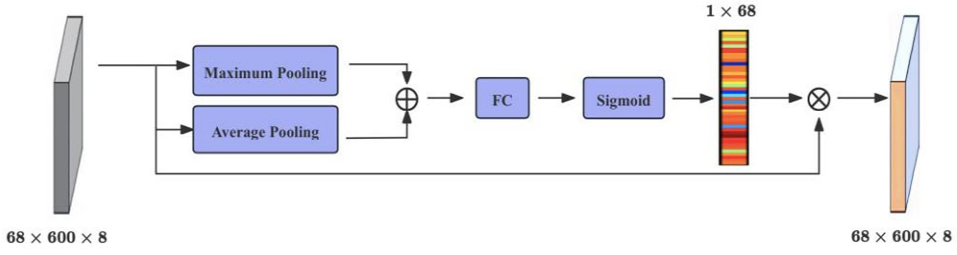
The attention mechanism, derived from the human visual system, allows neural networks to concentrate on certain subsets of inputs or characteristics, which is crucial in computer vision, natural language processing, and neural machine translation, among other domains. Bahdanau et al. (2014) developed the attention mechanism, which eliminates a significant proportion of extraneous information through a top-down information selection process, where importance is quantified by assigning weights to salient elements. The attention mechanism addresses the issue of information overload by selectively filtering a limited quantity of pertinent information while disregarding a substantial volume of irrelevant data during information processing. The fundamental concept of the attention mechanism is to allocate attention weights. The weight magnitude can be modified during the learning process, thereby enhancing the initial data weighting procedure, which in turn improves the neural network's performance. Attention processes comprise two types: spatial attention and channel attention. Hu et al. introduced the SENet (squeeze-and-excitation networks) module in 2017, which fundamentally aims to assess the significance of each feature channel for the classification objective using feature modelling (Hu et al., 2018). A reweighting algorithm is subsequently employed. The SE module's operational method has three stages for weight assignment: compression, activation, and recomputation. Park et al. implemented the attention mechanism within graph convolutional networks, allowing the model to autonomously identify and prioritise significant features (Kim et al., 2019). This facilitates the efficient allocation of computational resources towards feature representations that substantially influence classification outcomes, while diminishing the impact of less relevant features. Consequently, the integration of the attention mechanism often yields considerable enhancements in performance across various tasks.

In EEG-fNIRS signals comprising multiple channels, there exist distinct variations between the signals of each EEG and fNIRS channel. Consequently, certain electrode channels may be irrelevant to MI and mental arithmetic tasks, as well as lacking significant representation. In this research, we present a spatial attention mechanism within the graph convolutional layer fusion of EEG and fNIRS signals across channels to facilitate the adaptive identification of significant brain regions. Figure 2 illustrates that maximum pooling and average pooling operations are executed on the EEG and fNIRS direct fusion features, respectively. The resultant features, obtained by concatenating the two, are subsequently processed through a 2D convolutional layer and a sigmoid layer to produce spatial attention maps. Ultimately, the derived spatial attention weight coefficients are applied to the original input features, which can be articulated as follows:

$$
\begin{aligned}
S_{ave,(h,w)} &= \frac{1}{CB}\sum_{c=1}^{C}\sum_{b=1}^{B}\left(X'_{h,w}(c,b)\right) \\
S_{\max,(h,w)} &= \max_{c}\left(\max_{b}\left(X'_{h,w}\right)\right) \\
S &= cat\left(S_{ave,(h,w)}, S_{\max,(h,w)}\right) \\
F_s(X') &= Sigmoid\left(conv(S)\right)
\end{aligned}
\tag{14}
$$

where $X'_{h,w} \in R^{C\times B}$ indicates the value of all channels corresponding to the input signal.

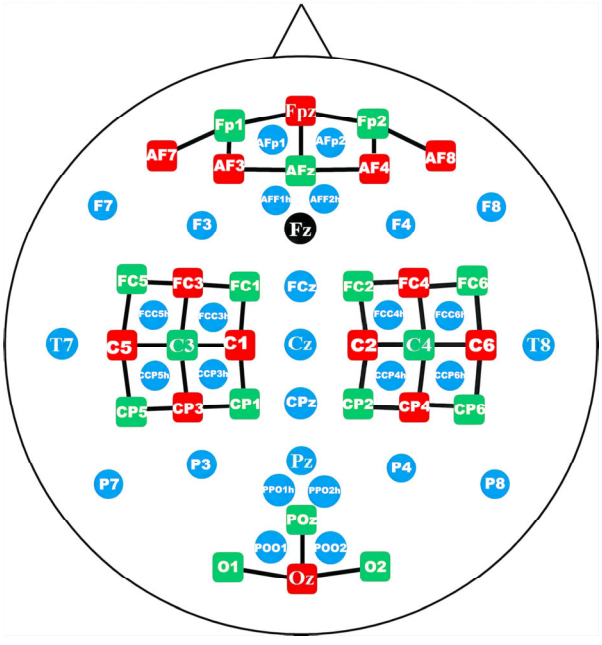
Figure 2 Modality attention (see online version for colours)



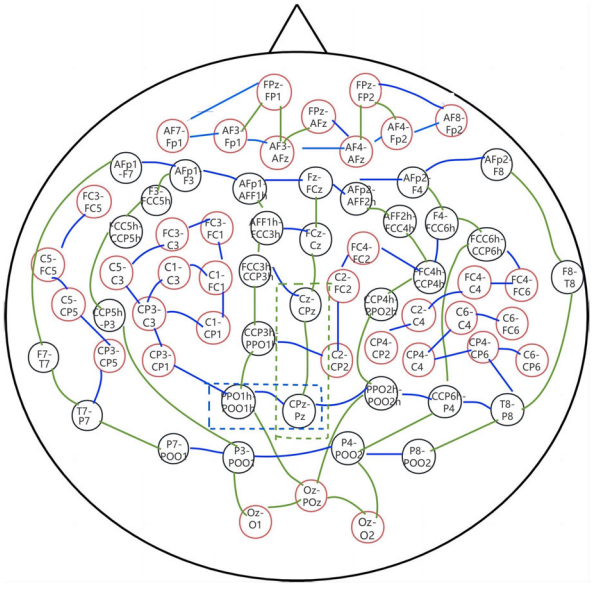
3.2.2 HGCN

In the acquisition of EEG and fNIRS signals, the absolute potential obtained from a single electrode channel lacks a definitive physical interpretation and is typically represented as the potential difference between two longitudinally adjacent electrodes. The dataset employed the International 10-5 system, as illustrated in Figure 3. Figure 3(a) illustrates the positioning and nomenclature of each electrode, while Figure 3 depicts the channels acquired during the sampling process, indicating the two longitudinally adjacent electrodes from which the potential difference of that channel is derived, with black denoting the EEG channel and red the fNIRS channel. In this specific graph structure of brainwave sampling, conventional graph convolutional neural networks will encounter significant limitations. The two laterally adjacent nodes PPO1h-POO1h and CpZ-PZ, delineated by blue dashed lines in Figure 3(b), consist of four blue nodes: PPO1h, POO1h, CpZ, and PZ. The two vertically adjacent nodes Cz-CPz and CPz-Pz, delineated by the green dashed lines in Figure 3(b), consist of three nodes: Cz, CPz, and Pz. This example illustrates that the information conveyed by the edges in Figure 3(b) lacks consistency. When an edge in Figure 3(b) connects two horizontally adjacent nodes, it actually represents the structural relationship among four electrodes. Conversely, when an edge connects two longitudinally adjacent nodes, it pertains solely to the structural relationship among three electrodes. In a standard graph convolutional neural network, all neighbour relations in the adjacency matrix are denoted by 1, implying that the horizontal and vertical neighbour relations among the nodes in Figure 3(b) are treated as equivalent types of neighbour relations. Consequently, the graph convolutional network model was enhanced by allocating the vertical and horizontal neighbour interactions among electrode pairs to two distinct graph convolutional layers: the longitudinal graph convolutional layer and the horizontal graph convolutional layer (Zeng et al., 2021; Zheng et al., 2021).

Figure 3 The score errors of the model with the change different parameters, (a) electrodes placement: EEG electrodes (blue), fNIRS sources (red) and detectors (green) (b) the electrode pairs leading in longitudinal montage coupling (see online version for colours)



(a)



(b)

Note: Adjacencies between nodes are shown by the blue and green lines, respectively, indicating horizontal and vertical adjacencies.

In this paper, the longitudinal and transversal adjacencies existing between electrode pairs are left to two graph convolutional layers, viz. Longitudinal map convolutional layer and horizontal map convolutional layer. A longitudinal map convolutional layer whose adjacency matrix has only longitudinal neighbours is denoted as *GConvLo*. A_{Lo} and W_{Lo} represent the adjacency matrix and weight matrix of $GConv_{Lo}$, respectively. The adjacency matrix of the longitudinal map convolution layer A_{Lo} can be expressed as.

$$A_{Lo}^{ij} = \begin{cases} 1, & for\ x_i \swarrow\nearrow x_j \\ 0, & for\ !x_i \swarrow\nearrow x_j \end{cases} \tag{15}$$

where $\swarrow\nearrow$ represents that the nodes are vertically neighbouring each other. The horizontal graph convolutional layer whose adjacency matrix has only horizontal neighbours is denoted as $GConv_{Tr}$. A_{Tr} and W_{Tr} represent the adjacency matrix and weight matrix of GConvTr respectively. The adjacency matrix of the horizontal graph convolutional layer A_{Tr} can be expressed as follows.

$$A_{Tr}^{ij} = \begin{cases} 1, & for\ x_i \leftrightarrow x_j \\ 0, & for\ !x_i \leftrightarrow x_j \end{cases} \tag{16}$$

where $\leftrightarrow$ represents that the nodes are horizontally neighbouring each other.

The hierarchical graph convolutional neural network model comprises a hierarchical graph convolution module that includes both graph convolution modules. Feature maps are initially propagated to the longitudinal graph convolution layer, followed by the transversal graph convolution layer. Conversely, feature maps may first be directed to the transversal graph convolution layer before proceeding to the longitudinal graph convolution layer. Ultimately, the feature maps from all pathways are concatenated and integrated at the conclusion of the hierarchical map convolution module to produce the comprehensive feature map. In summary, the propagation formula of a hierarchical graph convolution module is expressed as:

$$\begin{aligned} h^{l+1,1} &= \sigma\left(S\left(A_{Lo}^{l,1}\right)H^{l}W_{Lo}^{l,1}\right) \\ H^{l+1,1} &= \sigma\left(S\left(A_{Tr}^{l,1}\right)h^{l+1,1}W_{Tr}^{l,1}\right) \\ h^{l+1,2} &= \sigma\left(S\left(A_{Tr}^{l,2}\right)H^{l}W_{Tr}^{l,2}\right) \\ H^{l+1,2} &= \sigma\left(S\left(A_{Lo}^{l,2}\right)h^{l+1,2}W_{Lo}^{l,2}\right) \\ H^{l+1} &= \sigma\left(\left[H^{l+1,1};H^{l+1,2}\right]W^{l+1}\right) \end{aligned} \tag{17}$$

In the context of the hierarchical map CNN, H_l and H_{l+1}denote the input and output of the l^{th} layer, respectively. When $l = 0$, $H^0 = X \in R^{N\times M}$ represents the feature input generated by the electrode pairs in the EEG position naming system after preprocessing, where N represents the number of electrode pairs and M represents the number of extracted features. In particular, the activation function $\sigma(x) = \tanh(x)$ is utilised. L_2 normalisation is incorporated into the weights W to mitigate the risk of overfitting, while L_1 normalisation is included in the adjacency matrix A to impose constraints on its variation during adaptive adjustment.

Subsequent to the integration of all features present within all nodes, a pair of fully connected layers are utilised for the purpose of summarising the characteristics of all electrode pairs. The application of a softmax layer enables the classification of time periods. Thereafter, the classification results for each time period are obtained. Subsequently, the classifications obtained from multiple time periods are aggregated to obtain the classification for the entire time period.

4 Experiments

4.1 Dataset

We employed an open-access EEG-fNIRS dataset (Shin et al., 2017) comprising MI and MA tasks, involving 30 healthy participants (14 males and 15 females, 29 right-handed and 1 left-handed). The EEG data were acquired at a sampling rate of 1,000 Hz utilising a multi-channel EEG amplifier equipped with 30 active electrodes arranged in accordance with the worldwide 10-5 standard. The NIRS data were collected at a sampling rate of 12.5 Hz, utilising 36 physiological channels equipped with 14 sources and 16 detectors, allocated throughout the frontal region (9 sources) and left-handed orientation. A total of 36 physiological channels were established utilising 14 sources and 16 detectors, allocated among the frontal (9 channels near Fp1, Fp2, and Fpz), motor (12 channels adjacent to C3 and C4), and visual (3 channels surrounding Oz) cortical regions. Figure 1 illustrates the arrangement of the EEG electrodes and NIRS optodes.

- Dataset A (MI): In the MI task, participants were directed to visualise the act of opening and closing their hands as though grasping a ball, following visual prompts. A black arrow, oriented left or right, was displayed in the centre of the screen for 2 seconds, succeeded by a fixation cross during the task duration. Participants executed the hand-gripping exercise at a frequency of 1 Hz, with each trial concluding with a beep and a 'STOP' notification. The task was executed 20 times in one instance. The test was conducted 20 times within a single session, comprising 10 trials for each hand condition, and sessions included 10 randomised blocks of left or right hand MI.
- Dataset B (MA): In the MA task, participants viewed an initial subtraction presented centrally on the screen for 2 seconds. They were directed to commit the numbers to memory throughout this period. Subsequently, the subtraction vanished, and a fixation cross emerged while participants were instructed to maintain focus on the screen for 2 seconds. Subsequently, the subtraction vanished, and a fixation cross emerged while participants were instructed to constantly subtract the one-digit number from the outcome of the preceding computation. The job persisted for the specified duration.

4.2 Evaluation metrics

This study presents a classification model evaluated using four test metrics: accuracy (*ACC*), sensitivity (Sens), classification error confidence interval (*Error*), and their representations on receiver operating characteristic (ROC) curves, with the area under curve (*AUC*) computed. In a binary classification task, the count of accurate predictions for positive instances is denoted as *TP*, the count of erroneous predictions for negative

instances is *FN*, the count of erroneous predictions for positive instances is *FP*, and the count of accurate predictions for negative instances is *TN*. Accuracy is the likelihood of correctly forecasting outcomes over all samples, represented by the formula:

$$ACC = \frac{TP + FP}{TP + TN + FP + FN} \tag{18}$$

Sensitivity is the probability of predicting correctly in an actual positive case and is given by the formula:

$$Sens = \frac{TP}{TP + FN} \tag{19}$$

A confidence interval for classification error is an estimated range for a population parameter derived from a sample statistic, utilised to measure the uncertainty of the estimate. The confidence interval indicates the extent to which the true value of the parameter is likely to reside within the measurement result, hence reflecting the level of confidence in the measured parameter value. Confidence intervals for classification error are often computed using equation (3), whereby error is the classification error rate (95% confidence interval), const represents a constant that specifies the selected probability, and n indicates the sample size.

$$\begin{aligned} error &= \frac{FP + FN}{TP + TN + FP + FN} \\ ClassError &= error + const\sqrt{\frac{error(1 - error)}{n}} \end{aligned} \tag{20}$$

The *ROC* curves is a graphical representation of the true positive rate (*TPR*) and false positive rate (*FPR*), where *TPR* denotes sensitivity, and *FPR* indicates the likelihood of an erroneous prediction in a positive scenario, *FPR = FP / (TN + FP)*. In the binary classification task, the *ROC* curve is generated by systematically varying the classification thresholds, resulting in various sets of *TPR* and *FPR*. Upon acquiring the *ROC* curves, the area under the *ROC* curves (*ROC-AUC*) can be calculated to assess the overall performance of the model in terms of *TPR* and *FPR*.

4.3 Results

4.3.1 Comparison of different modes

To compare the recognition outcomes of uni-modal and multi-modal fusion, distinct modality recognition experiments were conducted on EEG signals and the statistical features of HbO and HbR, utilising the direct fusion method as the baseline classifier for classification in this study. Table 1 displays the average recognition rates of 29 individuals across various modalities. Even utilising the most basic fusion technique, multi-modality demonstrates a notable enhancement in recognition rates compared to uni-modality, with an average accuracy increase of around 3.99%–11.04% (MI) and 3.65%–9.86% (MA).

Table 1 Average recognition results of different modality

Modality	*Average Acc (%)*	
	MI	*MA*
EEG	76.67	75.27
HBO	71.66	79.36
HBR	71.59	79.48
EEG+HbO	80.66	82.48
EEG+HbR	79.67	82.6
EEG+HbR+HbO	82.63	84.13

4.3.2 MA-HGCN experimental results

To assess the efficacy of the proposed hierarchical graph-based convolutional fusion method, the statistical features of EEG signals, along with HbO and HbR signals, are employed to examine the fusion recognition outcomes across various feature combinations, with results presented in Tables 2 and 3.The mean recognition rates for the MI datasets: EEG+HbO, EEG+HbR, and EEG+HbO+HbR are 95.65%, 95.13%, and 96.13%, respectively, with corresponding mean sensitivities of 92.24%, 90.85%, and 92.85%. The error rate confidence intervals are 5.12%±4.725, 5.61%±6.01%, and 4.98%±5.35%. For the MA dataset, the mean recognition rates for EEG+HbO, EEG+HbR, and EEG+HbO+HbR were 97.89%, 97.23%, and 98.13%, respectively. The mean sensitivities were 95.63%, 94.08%, and 95.13%. The confidence intervals for the error rates were2.89%±3.01%, 3.12% ± 3.43%, and 3.14% ± 4.76%. *ROC* curves, illustrated in Figure 4, were employed to evaluate the model's performance based on the TPR and FPR. For MI tasks, the *ROC-AUCs* for EEG+HbO, EEG+HbR and EEG+HbO+HbR are 0.962, 0.963 and 0.962; for MA, the ROC-AUCs for EEG+HbO, EEG+HbR and EEG+HbO+HbR are 0.983, 0.982 and 0.987. The MA-HGCN model demonstrates a significant enhancement in recognition accuracy compared to fusion approaches that solely integrate distinct modal eigenvectors as a collective entity. The model employs a hierarchical graph convolution module that comprehensively addresses the disparities and interconnections between the signals of each EEG and fNIRS channel, managing two types of adjacencies between graph nodes – vertical and horizontal – thereby effectively compensating for the limitations of singular information and enhancing the classification efficacy of multimodal data. Specifically, in comparison to the fusion of the three signals – EEG, HbO, and HbR – the disparity in recognition rate and *ROC-AUCs* value when utilising only two of these signals is negligible; nonetheless, it still attains a greater recognition rate than unimodal approaches. Consequently, the ensuing discussion relies exclusively on EEG+HbO and EEG+HbR, which provide a reduction in the number of modalities while preserving satisfactory outcomes.

Motor and fronto-parietal areas were the activation regions linked to MI and MA, respectively. While the classification accuracy of fNIRS was maximum after a few seconds of delay, the classification rate of EEG was not significantly delayed. While MA-related regions were better identified, many participants had MI-related regions that were inactivable. This is because individuals are more accustomed to the MA task during MA, which makes it easier for them to concentrate on it. As a result, MA classification accuracy is typically higher than MI.Experiments conducted on a dataset of

EEG-functional near-infrared spectroscopy with 29 people in this research demonstrate that the suggested approach achieves an average accuracy of 98.25% and 96.13% for the classification of MI tasks. Our suggested approach performs better on both kinds of tasks and could offer a broad framework for multimodal brain signal fusion processing and BCIs that are based on several modes.

Table 2 MA-HGCN classification results for MI tasks

Subject	*Accuracy (%)*	*Sensitivity (%)*	*Error (%)*
	EEG+HbO / EEG+HbR / EEG+HbO+HbR		
1	96/94.33/94.13	91.67/86.33/85.34	5±5.5/8.3±7/8.1±6.2
2	98.33/95/97.32	96.67/96.67/96.67	1.7±3.2/5±5.5/3.7±1.2
3	93.67/92.67/92.62	83.33/83.33/83.33	8.3±7/8.3±7/8.3±7
4	94.33/91.67/90.71	86.67/83.33/83.33	6.7±6.3/18.3±9.8/11.7±11.3
5	93.67/92/91.23	90/86.67/86.67	8.3±7/10±7.6/7.3±7
6	96/98.33/97.15	90/96.67/96.67	5±5.5/1.7±3.2/3.2±1.5
7	92.67/90.33/91.34	83.33/76.67/79.17	8.3±7/11.7±8.1/11.3±7.2
8	100/90.67/91.67	100/79.33/83.33	0±0/13.3±8.6/10.5±7.6
9	98.33/98.33/97.37	100/96.67/95.17	1.7±3.2/1.7±3.2/1.7±3.2
10	96/96/96	93.33/93.33/93.33	5±5.5/5±5.5/5±5.5
11	97.67/96/96	100/90/93.33	3.3±4.5/5±5.5/4.3±4.3
12	93.67/100/94.35	96.67/100/100	8.3±7/0±0/8.3±7
13	100/92.67/91.97	100/86.67/86.67	0±0/8.3±7/10±10.3
14	90.33/92/91.33	100/90/90	11.7±10.4/10±7.6/11.7±10.4
15	96/96/95.56	93.33/96.67/93.67	5±5.5/15±9/5.3±5.7
16	93.67/92/93.63	83.33/80/81.45	8.3±7/10±7.6/8.3±7
17	94.33/97.67/96.37	83.33/100/90	6.7±9.4/3.3±4.5/6.7±5.4
18	92.67/93.67/92.67	86.67/87.67/86.67	8.3±7/8.3±7/8.3±7
19	96/94/93	96.67/80/80	5±5.5/10±10.1/9±5.5
20	93.67/98.33/96.15	90/96.67/96.67	8.3±7/3.7±3.2/8.3±7
21	98.33/98.33/98.53	100/100/100	1.7±3.2/1.7±3.2/1.7±3.2
22	100/100/100	100/100/100	0±0/0±0/0±0
23	94.33/97.67/94.37	86.67/96.33/91.35	6.7±6.3/3.3±4.5/6.7±6.3
24	97.67/97.67/94.33	96.67/96.67/96.67	3.3±4.5/3.3±4.5/5.3±4.5
25	91/91/91	93.33/95.33/93.33	10±7.6/10±7.6/10±7.6
26	98.33/100/100	96.67/100/96.67	1.7±3.2/0±0/0±0
27	96/98.33/96.55	93.33/100/93.33	5±5.5/1.7±3.2/5±5.5
28	96/93.67/93.97	93.33/85.33/85.35	5±5.5/8.3±7/5±5.5
29	95.33/93.67/92.67	80/79.33/81.33	6.7±4.4/8.3±5.8/6.7±5.4
Avg	95.65/95.73/96.13	91.24/90.85/92.85	5.12±4.72/5.61±6.01/4.98±5.35
Std	2.01/1.88/1.98	2.75/1.95/2.32	—/—/—

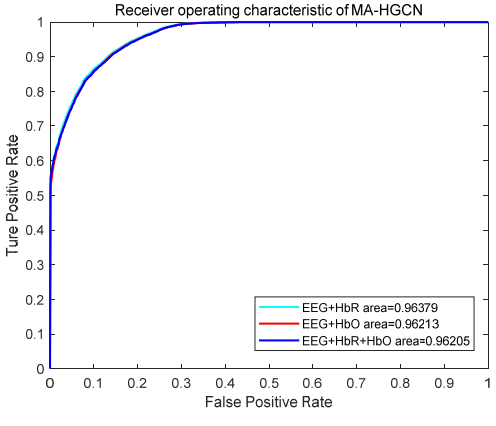
Table 3 MA-HGCN classification results for MA tasks

Subject	*Accuracy (%)*	*Sensitivity (%)*	*Error (%)*
	EEG+HbO/ EEG+HbR/ EEG+HbO+HbR		
1	100/100/100	100/100/100	0±0/0±0/0±0
2	96/100/96	90/100/90	5±5.5/0±0/4.3±5.1
3	98.33/94.67/95.33	96.67/83.33/93.33	1.7±3.2/8.3±1.7/1.3±1.7
4	100/100/100	100/100/100	0±0/0±0/0±0
5	96.67/96.67/96.67	93.33/96.67/96.67	3.3±4.5/3.3±4.5/3.3±4.5
6	100/98.33/100	100/96.67/96.67	0±0/1.7±3.2/0±0
7	95/91/89	90/90/90	5±5.5/15±9/11±7.3
8	100/100/100	100/100/100	0±0/0±0/0±0
9	93.67/100/93.67	96.67/100/100	8.3±7/0±0/9.1±10
10	98.33/100/93.33	96.67/100/100	1.7±3.2/0±0/6.3±5.7
11	100/90/92	100/90/89	0±0/5±11/11±10.2
12	93.33/98.33/93.33	86.67/100/90	6.7±6.3/1.7±3.2/6.7±6.2
13	93.33/96.33/96.35	100/86.67/86.67	6.7±6.3/6.7±6.3/6.7±6.3
14	98.33/95/95.33	96.67/90/90	1.7±3.2/5±5.5/5±5.5
15	95/94/94	93.33/96.67/93.37	5±5.5/15±9/11±9.2
16	91.67/98.33/94.67	86.67/96.67/93.37	8.3±7/1.7±3.2/7.3±3.2
17	100/100/100	100/100/100	0±0/0±0/0±0
18	100/100/100	100/100/100	0±0/0±0/0±0
19	98.33/100/98.33	100/100/100	1.7±3.2/0±0/1.3±3.1
20	91.67/98.33/94.67	90/96.67/91.17	8.3±7/1.7±3.2/7.3±1.2
21	98.33/98.33/98.33	100/100/100	1.7±3.2/1.7±3.2/1.7±3.2
22	95/95/97	93.33/90/90	5±5.5/5±5.5/5±5.5
23	96.67/98.33/97.67	93.33/100/100	3.3±4.5/1.7±3.2/3.3±3.2
24	100/95/98	100/90/90	0±0/5±5.5/4.3±5.5
25	96.67/95/93.37	93.33/96.67/96.67	3.3±4.5/5±5.5/5.7±5.5
26	98.33/98.33/98.33	96.67/96.67/96.67	1.7±3.2/1.7±3.2/1.7±3.2
27	98.33/97.5/97.33	96.67/95/95	1.7±3.2/2.5±4.8/2.5±4.8
28	96.67/98.33/94.67	96.67/96.67/91.37	3.3±4.5/1.7±3.2/7.7±3.2
29	96.33/100/94.33	86.67/100/91.37	6.7±6.3/0±0/6.3±6.7
Avg	97.89/97.23/98.13	94.63/94.08/95.13	2.89±3.01/3.12±3.43/3.14±4.76
Std	1.96/1.74/1.94	2.32/1.99/2.01	—/—/—

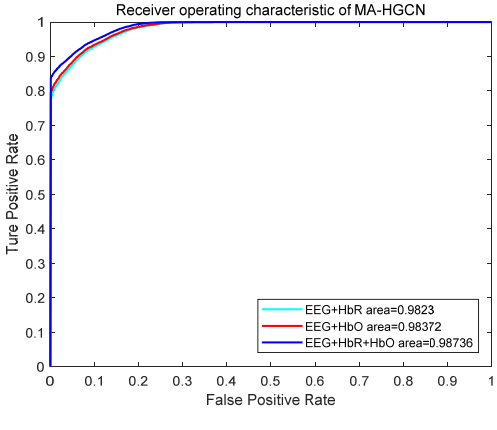
Spatial comparable EEG channel signals are more correlated because the EEG signal records the total of synchronised synaptic activity in groups of cortical pyramidal cells in the scalp following volumetric conduction to the electrodes. The degree of attenuation of near-infrared light emitted from the transmitting optic pole, after it has been scattered in brain tissue and propagated to the receiving optic pole, is what converts fNIRS signals into signals of changes in hemoglobin concentration. Signals that are not in the brain tissue and are not in the propagation path have no effect on the signal. Therefore, the

connection between the channels of the fNIRS signal is not related to the spatial distribution, and it is the functional connection between different brain regions that affects the connection between the channels. By mapping the neural signals onto the graph using the brain network as the adjacency matrix, HGCN allows for the full utilisation of the structural information between the channels, the aggregation of the information between each node and its neighbouring nodes, and the creation of new channel nodes as fused features. The influence of the noise-affected channels on the results when computed with the graph convolution layer approximated by Chebyshev polynomials will also be lessened because channels that are more affected by noise have a lower strength of functional connectivity with other channels and the values of the corresponding elements in the Laplace matrix will be smaller. HGCN prevents the classification findings from being impacted by the subpar signal in subjects when the fNIRS signal is less effective than the EEG signal. As a result, the feature fusion method employing HGCN can more reliably produce better classification results on bimodal signals as opposed to unimodal signals. It also avoids the issue of influencing bimodal signal classification because of the subpar classification performance on fNIRS signals.

Figure 4 The ROC curves and ROC-AUCs of (a) MI tasks (b) MA tasks of MA-HGCN (see online version for colours)



(a)



(b)

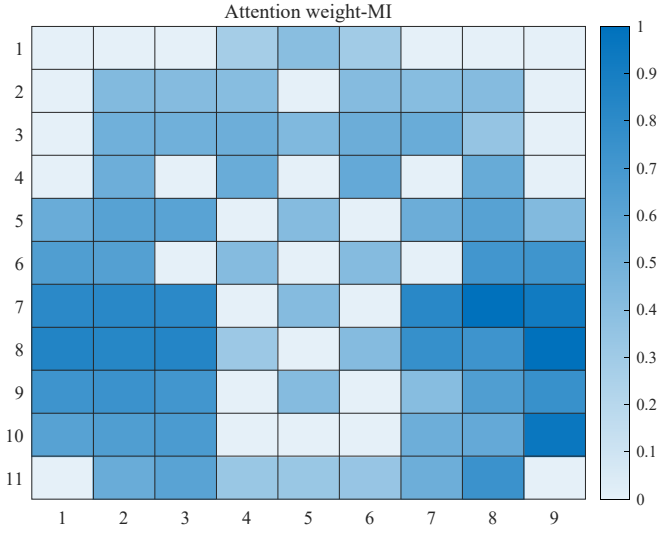
The graph feature fusion module employs a modal attention mechanism, which assigns varying weights to different channels across modalities during the training process, thereby more effectively capturing each channel's contribution to emotion recognition. As illustrated in Figure 5, the mean attention weights of each channel are depicted for all 29 participants. The channel positions are consistent with the EEG-fNIRS spatial position mapping matrix, indicating that varying weights were assigned to different channels. This activation of occipital lobe channels in a MI taskwas also observed in a study by Sammer (2007) and was interpreted as being elicited by subjects imagining images of movement. As can be seen in Figure 5(b), the various channels were assigned different weights, with the right motor cortex having a higher weight than the left motor cortex. Both the characterisation of the left motor cortex channels for the fNIRS signals and the characterisation of the right motor cortex channels for the EEG signals varied more significantly in MI. The experimental findings further support the complementary nature of the brain network features of EEG and fNIRS when interpreted from the standpoint of physiological laws of electrophysiological signals and hemodynamic signals. This finding is largely in line with well-known neurophysiological behaviours (Cui et al., 2010). Figure 5(c) shows that MA studies have focused mainly on the prefrontal cortex, while parietal and occipital channels play a lesser role, and the importance of prefrontal regions in MA recognition is reflected in the findings of the literature (Barch et al., 1997). In addition, under the same type of task, EEG signal and fNIRS signal play a larger role in spatially located overlapping channels with relatively high weighting coefficients.

Figure 5 (a) EEG-fNIRS spatial position mapping matrix (b) Distribution map of attention weight for MI Tasks (c) Distribution map of attention weight for MA tasks (see online version for colours)

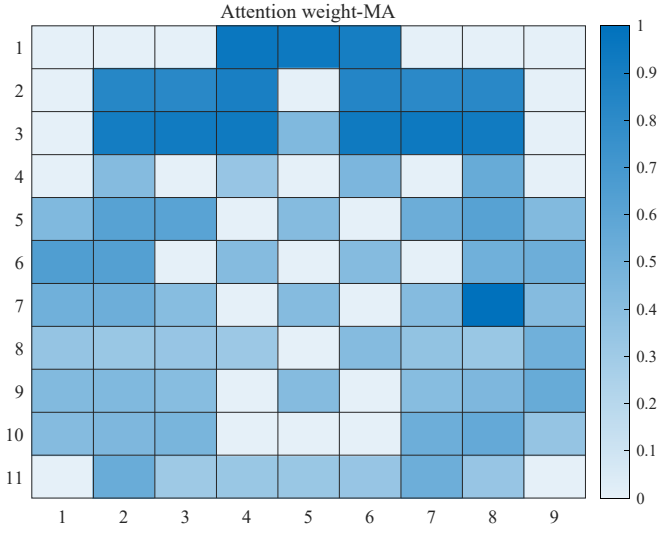
			FPz-FP1	FPz-AFz	FPz-FP2			
	AF7-Fp1	AF3-Fp1	AF3-AFz		AF4-AFz	AF4-Fp2	AF8-Fp2	
	AFp1-F7 F3-FCC5h	AFp1-F3	AFp1-AFF1h AFF1h-FCC3h	Fz-FCz	AFp2-AFF2h AFF2h-FCC4h	AFp2-F4	AFp2-F8 F4-FCC6h	
FC3-FC5	FC3-C3	FC3-FC1		FCz-Cz		FC2-FC4	FC2-C2	FC4-FC6
F7-T7	FCC5h-CCP5h		FCC3h-CCP3h		FCC4h-CCP4h		FCC6h-CCP6h	F8-T8
C5-FC5	C5-C3	C1-FC1		Cz-CPz		C2-FC2	C2-C4	C6-FC6
C5-CP5	CP3-C3	C1-C3	CCP3h-PPO1h		CCP4h-PPO2h	C2-CP2	CP4-C3	C6-C4
T7-P7	CCP5h-P3	PPO1h-POO1h		CPz-Pz		PPO2h-POO2h	CCP6h-P4	T8-P8
CP3-CP5	CP3-CP1	C1-CP1				CP4-CP2	CP4-CP6	C6-CP6
	P7-POO1	P3-POO1	Oz-O1	Oz-POz	Oz-O2	P4-POO2	P8-POO2	

(a)

Figure 5 (a) EEG-fNIRS spatial position mapping matrix (b) Distribution map of attention weight for MI Tasks (c) Distribution map of attention weight for MA tasks (continued) (see online version for colours)



(b)



(c)

4.3.3 Ablation experiments

To validate the functional contributions of each component within the proposed model, five distinct models were employed for comparative experimental analysis.

1 Direct splicing fusion CF (features from different modalities are directly concatenated following graph convolution).

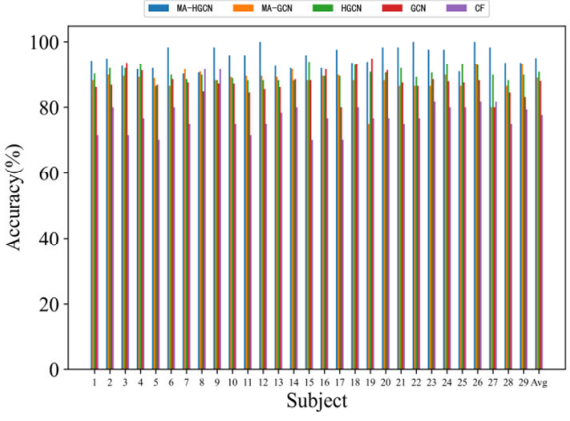
2 Graph convolution splicing fusion GCN (integration of diverse feature modalities post-graph convolution across the entire graph convolution framework)

3 Hierarchical graph convolution fusion model HGCN (the features of different modalities are processed through hierarchical graph convolution, followed by the convolution of the overall hierarchical graph).

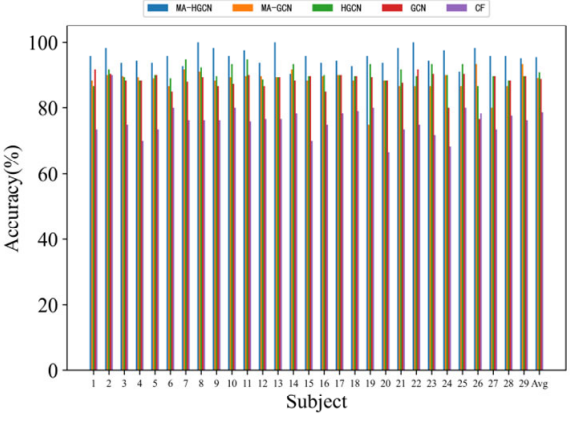
4 MA-GCN with modal attention (different modal features are convolved with a map and then the overall attention mechanism is convolved with a map).

5 The MA-HGCN model incorporates both modal attention and hierarchical map convolution, and subsequently, the fused feature vectors are processed through a fully connected network to generate the final output. Subsequently, modal fusion comparison ablation experiments are performed on EEG+HbO and EEG+HbR feature combinations to validate the efficacy of graph convolution fusion, modal attention mechanisms, and hierarchical graph convolution structures in classification recognition.

Figure 6 The comparison of MI tasks for different models, where modality, (a) EEG+HbR (b) EEG+HbO (see online version for colours)

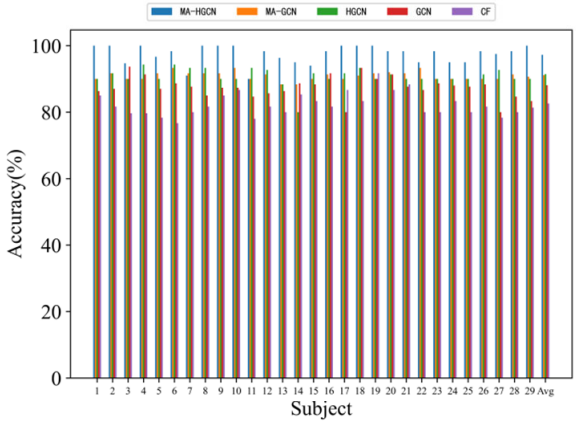


(a)

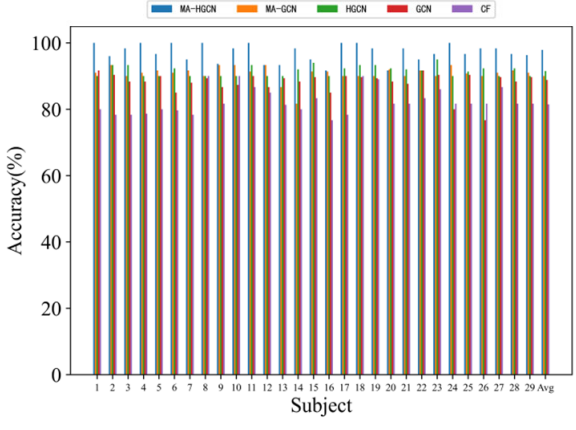


(b)

Figure 7 The comparison of MA tasks for different models, where modality, (a) EEG+HbR (b) EEG+HbO (see online version for colours)



(a)



(b)

The outcomes of the ablation trials are illustrated in Figures 6 and 7. The MI dataset reveals average recognition rates for MA-HGCN, MA-GCN, HGCN, GCN, and CF as follows: (95.65%, 90.82%, 89.94%, 88.82%, and 80.66%) for EEG+HbO and (95.73%, 90.08%, 90.87%, 88.08%, and 79.67%) for EEG+HbR. In the MA dataset, the recognition rates are (97.89%, 90%, 91.51%, 88.82%, and 82.48%) for EEG+HbO and (97.23%, 91.12%, 91.44%, 88.08%, and 82.6%) for EEG+HbR. The recognition rate of the fusion method, which simultaneously incorporates modal attention and hierarchical graph convolution structures, is markedly superior to that of other models. Furthermore, the recognition rate is enhanced when either the modal attention mechanism or the hierarchical graph convolution structure is applied independently, thereby demonstrating the efficacy of both the modal attention mechanism and the hierarchical graph convolution in augmenting classification recognition rates. The detailed analyses are as follows:

- GCN versus CF: When examined alongside the unimodal results, it is evident that direct splicing of the feature vectors in the second layer of the model does not enhance the classification accuracy of bimodal signals for certain subjects compared to unimodal signals. This occurrence is particularly prevalent among subjects whose unimodal signal classification accuracy exhibit a disparity exceeding 5% between EEG and fNIRS. Conversely, the implementation of GCN in the second layer resulted in an enhancement of the participants' average bimodal signal accuracy by approximately 10.16% (MI) and 7.34% (MA). This outcome may be attributed to the following factors: Multichannel neural signals, including EEG and fNIRS signals, are non-Euclidean data because the topological relationships between channels are lost when the signals are represented as matrix data. Consequently, the topological relationship between any two channels in the collected brain signals is independent of their respective row and column spacing in the storage matrix. Utilising GCN, the brain network serves as the adjacency matrix to represent neural signals on the graph, thereby optimising the structural information among channels, consolidating data between each node and its adjacent nodes, and generating new channel nodes as integrated features. The experimental research indicates that feature fusion employing GCN is more appropriate for bimodal neural signals compared to feature fusion utilising completely connected layers.
- HGCN versus GCN: Comparative analysis with the CF results demonstrates that the GCN model has achieved superior outcomes compared to prior models, underscoring the significance of leveraging the adjacency of channels in the detection job. The HGCN presented in this paper further enhances the results, achieving an accuracy increase of 2.79% (MI) and 3.36% (MA) relative to the GCN. This demonstrates that employing distinct graph convolutional layers to address varying adjacency yields superior outcomes when managing specialised graph structures. The HGCN model comprehensively accounts for the relationships between neighbouring and asymmetric features, so acquiring more extensive spatial feature information and significantly enhancing classification accuracy. The findings indicate that the HGCN model, utilising horizontal and vertical branch propagation, is capable of acquiring novel discriminative characteristics, hence enhancing classification performance.
- MA-HGCN versus HGCN: In comparison to HGCN, the classification accuracy of MA-HGCN is enhanced by around 2% (MI) and 3.04% (MA). The application of the attention mechanism to graph convolutional networks enables the model to autonomously identify and prioritise significant features, effectively optimising computational resources to concentrate on feature representations that substantially influence classification outcomes while diminishing the impact of less relevant features, thereby enhancing classification performance.

4.3.4 Cross-subject experiments

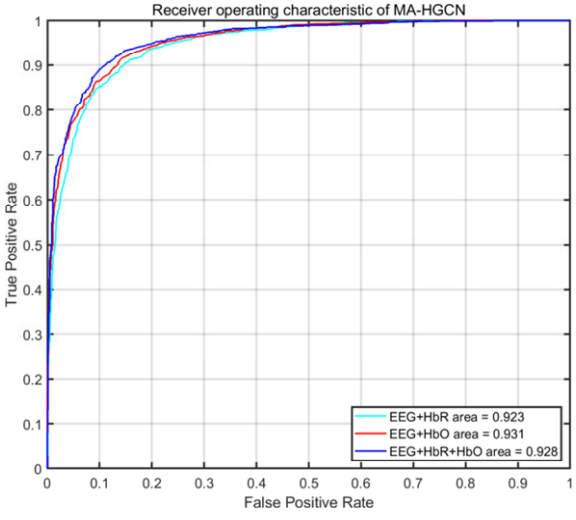
This study uses cross-subject experiments (cross-validation) to fully validate the model's effectiveness. All of the data from all subjects are merged and then split into a training set and a test set in a 9:1 ratio, with 90% of the data coming from the software training set and 10% from the test set. The system automatically and randomly selects the data from the two categories. From the start of the task, each user's data was separated into 1-second sessions. 200 observations were made for each of the three sessions with this

configuration, yielding a total of 600 observations. A total of 600 × 29 = 17,400 observations are acquired by combining the data of 29 people. Each time, 1,740 of the 17,400 observations (the number of floors fluctuates) are chosen as the training set for a particular implementation, with the remaining data being used for testing. In the meanwhile, this work employs 5-fold cross-validation, which involves randomly, creating the training set and the test set five times, to guarantee the dataset's unpredictability when divided by a 9:1 ratio. Both sets are then tested to provide average findings. The feature combinations included EEG+HbR+HbO, EEG+HbO, and EEG+HbR, with the results presented in Table 4 and Figure 8. The results are inferior to the subject-dependent experiments owing to the significant disparity in the distribution of training and test data in the cross-subject trials. Figure 9 illustrates the cross-subject categorisation recognition accuracy of the five fusion procedures. Despite generally yielding lower results than subject dependent trials, the fusion approaches employing hierarchical graph convolution and modal attention processes nonetheless achieve reasonably high accuracy, consistent with the findings of the subject-dependent experiments.

Table 4 Cross-subject results of MA-HGCN

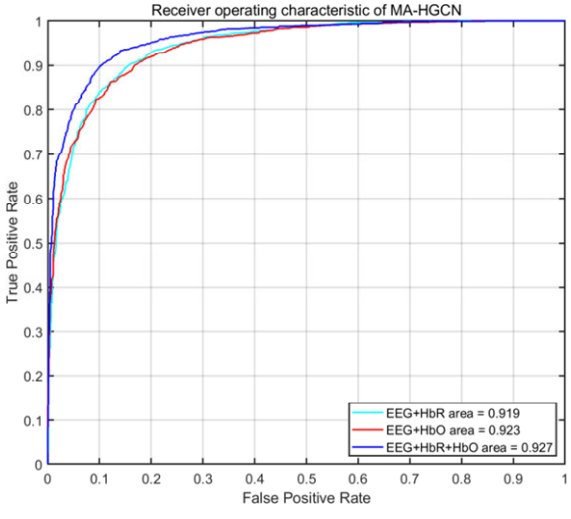
Modality	*MI*			*MA*		
	ACC	*Sens*	*Error*	*ACC*	*Sens*	*Error*
EEG+HbR	91.79	92.20	8.2±8.5	92.77	89.54	7.2±10
EEG+HbO	91.43	92.66	9.6±8.4	92.41	84.94	8.6±9.2
EEG+HbR+HbO	91.64	92.91	9.4±7.4	92.83	88.73	8.8±9.1

Figure 8 Cross-subject results of the ROC curves (a) MI tasks (b) MA tasks (see online version for colours)



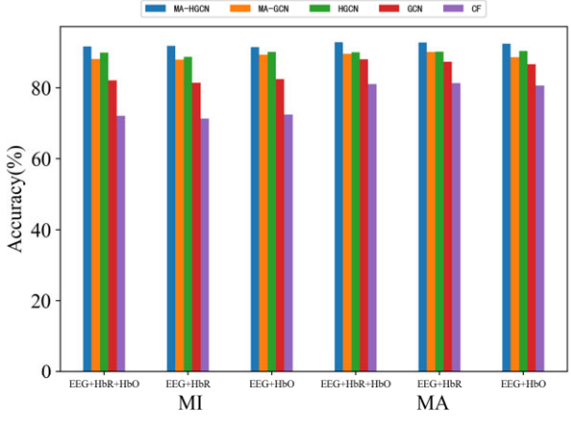
(a)

Figure 8 Cross-subject results of the ROC curves (a) MI tasks (b) MA tasks (continued) (see online version for colours)



(b)

Figure 9 Cross-subject results of different fusion methods (see online version for colours)



4.3.5 Comparison with other research methods

Table 5 enumerates the studies in the literature that utilise the identical dataset. This study, in contrast to existing literature, employs a simplified model utilising only two signals for fusion, thereby reducing the number of modalities while achieving superior classification accuracy compared to the fusion methods referenced in the cited works.

Table 5 Results of different fusion methods

Reference	*Year*	*Task*	*Type Method*	*Multi-modality*	*Acc(%)*
Shin et al. (2017)	2017	MA	Feature-level	EEG+HbR+HbO	86.2
		MI	Feature-leve	EEG+HbR+HbO	75.9
Rabbani and Islam (2021)	2021	MA	Feature-level + Decision-level	EEG+HbR+HbO	92.52
		MI	Feature-level + Decision-level	EEG+HbR+HbO	78.56
Jiang et al. (Qiu et al., 2022)	2022	MA	Logistic regression+Decision-level	EEG+HbR+HbO	94.88
		MI	Logistic regression+Decision-level	EEG+HbR+HbO	92.19
Ali et al. (2023)	2023	MA			
		MI	Temporal features and binary E-WOA	EEG+HbR+HbO	93.22
Present study	2024	MA	MA-HGCN	EEG+HbO	97.89
		MI	MA-HGCN	EEG+HbR	95.73

5 Conclusions

This study suggests a modal attention hierarchical graph convolution (MA-HGCN)-based fusion recognition approach for EEG and fNIRS, taking into account the distinctions and relationships between the signals of EEG and fNIRS channels. A modal attention mechanism module is designed to extract the multimodal signals of EEG and fNIRS frequency bands and brain regions that are more reflective of changes in MI and mental arithmetic tasks. This is done by first taking into account the relationship between neighbouring electrodes, computing the information between lateral and vertical electrodes separately, and then building a hierarchical graph convolution network model to train the features. This improves the recognition performance by exploring the deeper information between EEG channels. When only two modal signals, EEG+HbO/EEG+HbR, are used for fusion, ablation studies with various modules demonstrated that the suggested MA-HGCN is the most effective and that the model still works well when compared to other multimodal fusion techniques. Future research aims to achieve higher recognition accuracies through fusion experiments with various graph convolutional layer types and new EEG and fNIRS features. Additionally, migration learning is employed in conjunction with the model in cross-subject experiments to enhance the model's cross-subject recognition.

Acknowledgements

This work was supported by Doctoral Supervised Research Programme, Guangzhou Xinhua University.

Declarations

All authors declare that they have no conflicts of interest.

References

Ahn, S., Nguyen, T., Jang, H., Kim, J.G. and Jun, S.C. (2016) 'Exploring neuro-physiological correlates of drivers' mental fatigue caused by sleep deprivation using simultaneous EEG, ECG, and fNIRS data', *Frontiers in Human Neuroscience*, Vol. 10, No. 3, p.219.

Ali, M.U., Kim, K.S., Kallu, K.D., Zafar, A. and Lee, S.W. (2023) 'OptEF-BCI: an optimization-based hybrid EEG and fNIRS – brain computer interface', *Bioengineering*, Vol. 10, No. 5, p.608.

Bahdanau, D., Cho, K. and Bengio, Y. (2014) *Neural Machine Translation by Jointly Learning to Align and Translate*, arXiv preprint, *arXiv:* 1409.0473v7.

Bao, G., Yang, K., Tong, L., Shu, J., Zhang, R., Wang, L., ... and Zeng, Y. (2022) 'Linking multi-layer dynamical GCN with style-based recalibration CNN for EEG-based emotion recognition', *Frontiers in Neurorobotics*, Vol. 16, No. 2, p.834952.

Barch, D.M., Braver, T.S., Nystrom, L.E., Forman, S.D., Noll, D.C. and Cohen, J.D. (1997) 'Dissociating working memory from task difficulty in human prefrontal cortex', *Neuropsychologia*, Vol. 35, No. 10, pp.1373–1380.

Chen, G., Liu, Y. and Zhang, X. (2024) 'EEG-fNIRS-based emotion recognition using graph convolution and capsule attention network', *Brain Sci*., Vol. 14, No. 8, p.820.

Chiarelli, A.M., Croce, P., Merla, A. and Zappasodi, F. (2018) 'Deep learning for hybrid EEG-fNIRS brain-computer interface: application to motor imagery classification', *Journal of Neural Engineering*, Vol. 15, No. 3, p.036028.

Cicalese, P.A., Li, R., Ahmadi, M.B., Wang, C., Francis, J.T., Selvaraj, S., Schulz, P.E. and Zhang, Y. (2020) 'An EEG-fNIRS hybridization technique in the four-class classification of alzheimer's disease', *Journal of Neuroscience Methods*, Vol. 336, No. 6, p.108618.

Cui, X., Bray, S. and Reiss, A.L. (2010) 'Functional near infrared spectroscopy (NIRS) signal improvement based on negative correlation between oxygenated and deoxygenated hemoglobin dynamics', *Neuroimage*, Vol. 49, No. 6, pp.3039–3046.

Deligani, R.J., Borgheai, S.B., McLinden, J. and Shahriari, Y. (2021) 'Multimodal fusion of EEG-fNIRS: a mutual information-based hybrid classification framework', *Biomedical Optics Express*, Vol. 12, No. 3, pp.1635–1650.

Duan, R.N., Zhu, J.Y. and Lu, B.L. (2013) *6th International IEEE/EMBS Conference on Neural Engineering (NER).*

Hu, J., Shen, L. and Sun, G. (2018) 'Squeeze-and-excitation networks', *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.7132–7141.

Kim, Y., Soh, J.W., Park, J., Ahn, B., Lee, H.S., Moon, Y.S. and Cho, N.I. (2019) 'A pseudo-blind convolutional neural network for the reduction of compression artifacts', *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 30, No. 4, pp.1121–1135.

Kipf, T.N. and Welling, M. (2016) 'Semi-supervised classification with graph convolutional networks', *arXiv preprint arXiv*:1609.02907.

Kloetzer, M. and, Mahulea, C. (2020) 'Path planning for robotic teams based on LTL specifications and Petri Net models', *Discrete Event Dynamic Systems-Theory and Applications*, Vol. 30, No. 1, pp.55–79.

Kumar, C., Rahimi, N., Gonjari, R., McLinden, J., Hosni, S.I., Shahriari, Y. and Shao, M. (2023) 'Context-aware multimodal auditory bci classification through graph neural networks', in *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, pp.1–4.

Lazeyras, F.O., Blanke, O., Perrig, S., Zimine, I., Golay, X., Delavelle, J., Michel, C.M., De Tribolet, N., Villemure, J.G. and Seeck, M. (2000) 'EEG-triggered functional MRI in patients with pharmacoresistant epilepsy', *Journal of Magnetic Resonance Imaging*, Vol. 12, No. 1, pp.177–185.

Morioka, H., Kanemura, A., Morimoto, S., Yoshioka, T., Oba, S. and Kawanabe, M. (2014) 'Decoding spatial attention by using cortical currents estimated from electroencephalography with near-infrared spectroscopy prior information', *Neuroimage*, Vol. 90, No. 7, pp.128–139.

Nguyen, T., Ahn, S., Jang, H., Jun, S.C. and Kim, J.G. (2017) 'Utilization of a combined EEG/NIRS system to predict driver drowsiness', *Scientific Reports*, Vol. 7, No. 1, p.43933.

Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S. and Burgess, P.W. (2020) 'The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience', *Annals of the New York Academy of Sciences*, Vol. 1464, No. 1, pp.5–29.

Pouliot, P., Tremblay, J., Robert, M., Vannasing, P., Lepore, F., Lassonde, M., Sawan, M., Nguyen, D.K. and Lesage, F. (2012) 'Nonlinear hemodynamic responses in human epilepsy: a multimodal analysis with fNIRS-EEG and fMRI-EEG', *Journal of Neuroscience Methods*, Vol. 204, No. 2, pp.326–340.

Qiu, L., Zhong, Y., He, Z. and Pan, J. (2022) 'Improved classification performance of EEG-fNIRS multimodal brain-computer interface based on multi-domain features and multi-level progressive learning', *Front. Hum. Neurosci.*, Vol. 16, No. 8, p.973959.

Qu, J., Cui, L., Guo, W., Bu, L. and Wang, Z. (2024) 'Development of a novel machine learning-based approach for brain function assessment and integrated software solution', *Advanced Engineering Informatics*, Vol. 60, No. 5, p.102461.

Rabbani, M.H.R. and Islam, S.M.R. (2021) 'Integration of decision fusion and feature fusion on EEG and fNIRS signal', *Comp. Methods Prog. Biomed. Update*, Vol. 10, No. 1, p.26.

Ren, B., Ren, P., Luo, W. and Xin, J. (2024) 'A brain network analysis model for motion sickness in electric vehicles based on EEG and fNIRS signal fusion', *Sensors*, Vol. 24, No. 20, p.6613.

Sammer, G. (2007) 'Relationship between regional hemodynamic activity and simultaneously recorded EEG-theta associated with mental arithmetic-induced workload', *Human Brain Mapping*, Vol. 28, No. 8, pp.793–803.

Sargent, A., Heiman-Patterson, T., Feldman, S., Shewokis, P.A. and Ayaz, H. (2018) 'Mental fatigue assessment in prolonged BCI use through EEG and fNIRS', *Neuroergonomics*, Vol. 36, No. 5, pp.315–316.

Shin, J., von Luhmann, A., Blankertz, B., Kim, D-W., Jeong, J., Hwang, H-J. and Muller, K-R. (2017) 'Open access dataset for EEG+NIRS single-trial classification', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 25, No. 10, pp.1735–1745.

Sun, H., Hou, H., Song, W., Hu, H., Liu, N. and Shankar, A. (2024) 'Brain computer interface with EEG signals to improve feedback system in higher education based on augmentative and alternative communication', *International Journal of Information and Communication Technology*, Vol. 24, No. 2, pp.128–144.

Sun, Y., Ayaz, H. and Akansu, A.N. (2020) 'Multimodal affective state assessment using fNIRS+ EEG and spontaneous facial expression', *Brain Sciences*, Vol. 10, No. 2, p.85.

Trakoolwilaiwan, T., Behboodi, B., Lee, J., Kim, K. and Choi, J.W. (2017) 'Convolutional neural network for high accuracy functional near infrared spectroscopy in a brain-computer interface: three-class classification of rest, right, and left hand motor execution', *Neurophoton*, Vol. 5, No. 1, p.011008.

Uchitel, J., Vidal-Rosas, E.E., Cooper, R.J. and Zhao, H. (2021) 'Wearable, integrated EEG-FNIRS technologies: a review', *Sensors*, Vol. 21, No. 18, p.6106.

Verma, P., Heilinger, A., Reitner, P., Grünwald, J., Guger, C. and Franklin, D. (2019) 'Performance investigation of brain-computer interfaces that combine EEG and fNIRS for motor imagery tasks', in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, IEEE, pp.259–263.

Yang, S., Zhang, Y.E.J., Xia, X. and Xu, X. (2024) 'Preserving node similarity adversarial learning graph representation with graph neural network', *Engineering Reports*, Vol. 6, No. 10, p.e12854.

Zeng, D., Huang, K., Xu, C., Shen, H. and Chen, Z. (2021) 'Hierarchy graph convolution network and tree classification for epileptic detection on electroencephalography signals', *IEEE Transactions on Cognitive and Developmental Systems*, Vol. 99, No. 4, pp.10–22.

Zhang, G., Yu, M., Liu, Y. J., Zhao, G., Zhang, D. and Zheng, W. (2021) 'SparseDGCNN: Recognizing emotion from multichannel EEG signals', *IEEE Transactions on Affective Computing*, Vol. 14, No. 1, pp.537–548.

Zheng, F., Hu, B., Zhang, S., Li, Y. and Zheng, X. (2021) 'EEG emotion recognition based on hierarchy graph convolution network', in *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE.