Intelligent Rehabilitation and Engineering

SCIEPublish

Systematic Review

Research Progress on Motor Imagery Decoding in Brain-Computer Interface Based on EEG-fNIRS Multimodal Fusion

Zhongpeng Wang [1,2,3], Jing Liu [1], Yijie Zhou [1,2,3,4], Long Chen [1,2,3,*] and Dong Ming [1,2,3]

[1] Academy of Medical Engineering and Translational Medicine, Tianjin University, Tianjin 300072, China; tunerl_wzp1@tju.edu.cn (Z.W.); jing_9@tju.edu.cn (J.L.); yjzhou1994@tju.edu.cn (Y.Z.); richardming@tju.edu.cn (D.M.)
[2] Haihe Laboratory of Brain-Computer Interaction and Human-Machine Integration, Tianjin 300392, China
[3] State Key Laboratory of Advanced Medical Materials and Devices, Tianjin University, Tianjin 300072, China
[4] School of Disaster and Emergency Medicine, Tianjin University, Tianjin 300072, China

* Corresponding author. E-mail: cagor@tju.edu.cn (L.C.)

Received: 12 June 2026; Revised: 14 July 2026; Accepted: 28 July 2026; Available online: 6 August 2026

ABSTRACT: Motor imagery-based brain-computer interface (MI-BCI) decodes subjective motor intentions to achieve proactive output control of external devices, representing a core research direction in BCI with important applications in post-stroke motor rehabilitation. Neural signals for motor imagery can be recorded using electroencephalography (EEG) and functional near-infrared spectroscopy (fNIRS). Combining the high temporal resolution of EEG with the high spatial resolution of fNIRS through multimodal fusion enhances the decoding accuracy of motor imagery commands and provides comprehensive insights into brain dynamics. This paper systematically reviews EEG-fNIRS multimodal fusion strategies and representative algorithms from traditional machine learning (ML) and deep learning (DL) for decoding fused data. We also discuss current challenges and potential solutions to facilitate MI-BCI research and support future BCI-related industries.

Keywords: Brain-computer interface; Electroencephalography; Functional near-infrared spectroscopy; Motor imagery; Multimodal fusion

1. Introduction

Brain-computer interface (BCI) captures and analyzes brain activity to enable direct communication between the brain and external devices [1]. This process bypasses traditional neural and muscular pathways and has proven valuable in robot control [2], mental workload assessment [3], and auxiliary diagnosis of neurological disorders [4]. The growing aging population and rising number of patients with neurological disorders have created an urgent demand for BCI technology in medical and rehabilitation settings [5]. Healthy individuals are also increasingly seeking to enhance their quality of life through BCI, presenting broad market opportunities for widespread BCI adoption. Recent policy incentives have further accelerated the development of the BCI industry chain. According to McKinsey estimates, the global market for BCI medical applications will reach 40 billion to 145 billion between 2030 and 2040, with China playing a significant role [6].

Motor imagery (MI) describes the mental simulation of a movement without its actual execution [7]. For patients with motor impairments, MI-BCI technology decodes the spontaneous brain activity induced by motor imagery, creating an output pathway independent of muscle control that can recognize motor intentions and assist in rehabilitation training. The theory of neural plasticity explains the adaptability of brain structure and function, providing a key physiological basis for the application of MI-BCI in rehabilitation therapy [8]. Decoding MI-EEG signals can accurately identify the subject's motor intentions and enable control over external signals. MI-EEG, due to its non-invasiveness, ease of acquisition, and effective representation of motor intentions, has become a research hotspot both domestically and internationally [9]. Currently, researchers have successfully applied MI-EEG in healthcare [10], robotic arm control [11], and smart homes [12].

Electroencephalogram (EEG) records scalp potential changes from synchronized neuronal activity, capturing millisecond-level neural oscillations with excellent temporal resolution. However, its spatial localization is limited, and the signal is susceptible to physiological and environmental artifacts [13]. Functional near-infrared spectroscopy (fNIRS) monitors cortical blood oxygen concentration changes coupled with neural activity, reflecting brain region activation with relatively high spatial resolution. Its hemodynamic response, however, results in relatively slow temporal dynamics [14].

To better understand neural mechanisms of BCIs, researchers have developed multimodal fusion methods that integrate EEG and fNIRS. This approach effectively combines the high temporal resolution of EEG with the high spatial resolution of fNIRS [15], while also being non-invasive, relatively low-cost, portable, suitable for natural environments, and capable of supporting long-term continuous monitoring [16]. Simultaneously acquiring neural electrical activity signals and local brain hemodynamic responses enables a more comprehensive characterization of brain functional activity, helping to explain neural plasticity processes influenced by BCIs [17]. EEG-fNIRS multimodal fusion analysis has thus become a major topic in neuroscience [18].

This paper systematically reviews EEG-fNIRS research in MI-BCIs from two core perspectives: multimodal fusion strategies and fusion data decoding algorithms. For fusion strategies, we focus on the theoretical principles, technical advantages, and applicable scenarios of data-level, feature-level, and decision-level fusion. For decoding algorithms, we trace the technical evolution and decoding performance of typical methods along two main technical lines: traditional machine learning and deep learning. On this basis, the paper further analyzes the key technical challenges faced by current research and prospects the future development trends in this field, aiming to provide a systematic methodological reference and technical guidance for relevant researchers in the field. A graphical abstract illustrating the overall framework and main content of this review is provided at the beginning of the manuscript.

2. Strategies for EEG-fNIRS Multimodal Fusion

Multimodal data provide complementary information about the same neurophysiological phenomena and have significant advantages over single modalities. Fusion methods play a key role in enhancing the interpretability and practicality of multimodal data, effectively utilizing the strengths of each modality while mitigating their specific confounding signal effects. EEG-fNIRS fusion methods are data analysis techniques for decoding brain activity, offering better spatiotemporal resolution and signal-to-noise ratio (SNR) than when applied individually. This section systematically reviews the different levels of fusion methods in EEG-fNIRS multimodal data processing, including data-level fusion [19], feature-level fusion [20], and decision-level fusion [21].

2.1. Data-Level Fusion

Information fusion organizes source data hierarchically into three abstraction levels: data-level, feature-level, and decision-level fusion. Data level fusion integrates multi sensor information directly at the raw acquisition stage, before feature extraction or pattern recognition. As the lowest level strategy, it preserves the completeness and granularity of the original information to the greatest extent, thereby supporting refined representations of underlying physiological or physical phenomena [22]. This strategy retains modality-specific details that later processing stages would otherwise discard, proving particularly valuable in multimodal BCI and complex sensing systems. However, this high fidelity preservation imposes substantial computational overhead, degrades real time performance, and offers limited interference robustness [23]. Additional difficulties include cross modal data heterogeneity, strict temporal synchronization and spatial alignment requirements, and exponentially increased input dimensionality, which complicate downstream modeling.

Existing data-level fusion approaches fall into two broad categories: deep learning based structured fusion and traditional signal processing based transform-domain fusion. These paradigms address cross modal heterogeneity through fundamentally different mechanisms.

The deep learning paradigm offers structured data-level fusion. Chen et al. proposed the Multi-Channel Fusion Hybrid Network (MCFHNet), which constructs a unified multimodal representation by encoding EEG, oxyhemoglobin (HbO), and deoxyhemoglobin (HbR) signals into three-channel feature maps [24]. The core innovation lies in preserving complete modality-specific information at the raw signal level. MCFHNet achieves 99.641% average classification accuracy on public datasets, substantially outperforming single-modality and feature-level baselines. This performance gain confirms that maintaining fine-grained cross-modal information during early fusion critically enhances decoding accuracy. The approach, however, demands strict sensor synchronization and incurs considerable computational overhead due to exponentially increased input dimensionality.

Traditional signal processing based data-level fusion strategies have also progressed substantially. The core challenge is overcoming feature heterogeneity across different modal signals, which manifests as mismatched sampling rates and amplitude scales, as well as time delays and offsets. Li et al. proposed a wavelet transform based data-level fusion method that separately decomposed EEG and fNIRS signals into four detail levels via the discrete wavelet transform [25]. This method decomposes EEG and fNIRS signals into four wavelet layers and reconstructs fused signals by constructing a weighted fusion rule based on the Fisher values of wavelet coefficients at each layer; common spatial pattern (CSP) features are then extracted and classified using linear discriminant analysis (LDA), ultimately achieving an accuracy rate of 88.1%. This performance significantly surpassed single-modality baselines and traditional feature-level fusion methods while maintaining low computational complexity. Limitations include dependence on prior knowledge for wavelet basis selection and decomposition level determination, as well as insufficient cross-subject generalization validation.

These two paradigms exhibit complementary strengths that reveal a fundamental trade-off in data-level fusion. Deep learning based structured fusion maximizes information preservation and decoding accuracy at the expense of computational resources and synchronization rigor. Signal processing-based transform fusion prioritizes computational efficiency and interpretability while sacrificing flexibility and generalization. This trade-off motivates hybrid strategies that leverage deep learning for feature extraction and signal processing for modality alignment and noise suppression, thereby balancing accuracy and deployability.

Data-level fusion also finds applications beyond neural signal integration. Wojewoda et al. applied it to a position tracking system for a passive upper-limb rehabilitation robot, fusing data from two laser optical sensors mounted on the robot base and a webcam fixed above the workspace [26]. The optical sensors tracked relative motion at 97 Hz, while the webcam tracked absolute position at 5 Hz, yielding a sampling frequency ratio of approximately 19:1. The system used high-frequency optical measurements as the primary basis for trajectory generation and employed low-frequency absolute measurements from the webcam to periodically correct cumulative drift errors. Experimental results showed position tracking accuracy of 2.8 ± 2.6 mm and orientation tracking accuracy of 0.5 ± 0.4 degree, both significantly superior to either individual sensor subsystem. This system specifically targets home rehabilitation of stroke patients and can track motor recovery trends without professional therapist supervision. Monitoring indicators include movement quality parameters such as path length, path duration, and normalized saccadicity, providing reliable technical support for remote rehabilitation assessment. Among all the included literature, these studies represent cases that employ data-level fusion, reflecting its scarcity in practical rehabilitation device applications and confirming its inherent limitations, including high computational cost and strict synchronization requirements. A comparison of the three fusion levels reveals that multiple constraints shape the choice of fusion strategy, including sensor type, data availability, computational resources, and application scenarios. Biological signal fusion emphasizes complete preservation of physiological information, whereas physical sensor fusion focuses more on noise suppression and drift compensation, with fundamental differences in objectives and methodologies. Notably, the aforementioned studies primarily rely on offline analysis and have not yet addressed online adaptive fusion mechanisms under dynamically changing signal quality, nor provided systematic cross subject generalization evaluations. This issue constitutes the key hurdle that data-level fusion must overcome for translation from laboratory to clinical applications.

As a fundamental fusion strategy, data-level fusion exploits the spatiotemporal complementarity of EEG and fNIRS through structured representation or transform-domain alignment. The choice between deep learning and signal processing paradigms should depend on specific application constraints, including computational resources, real-time requirements, and the availability of synchronized multi-modal data. Closed-loop neurofeedback systems demand online adaptive mechanisms that dynamically adjust fusion weights in response to varying signal quality and task demands. Advancing this direction will drive multimodal BCIs toward more efficient and personalized neuroregulation.

2.2. Feature-Level Fusion

Feature-level fusion occurs at the feature extraction stage, constructing a unified discriminative representation by concatenating, weighting, or jointly encoding features independently extracted from each modality based on criteria such as mutual information. This method leverages prior unimodal feature extraction to map data into a common low-dimensional feature space, alleviating high dimensionality and inter-modality inconsistency while preserving each modality's discriminative information. However, this process may lose some useful information. By systematically optimizing feature complementarity, redundancy, and correlation, this method can significantly improve classification performance and is considered an important fusion strategy in multimodal BCIs.

Feature-level fusion has demonstrated its effectiveness across both offline classification and real time neurofeedback paradigms. In the offline domain, the deep synchronized fusion network (DeepSyncNet) introduces a deep feature level fusion framework that converts raw EEG and fNIRS signals into three dimensional tensors, extracts modality specific features through a four-branch multi scale convolution module, performs element-wise weighted fusion via a gated attention mechanism, and refines the fused representations through temporal pooling alignment and spatiotemporal attention screening [27]. On a public dataset with five fold cross validation, the model achieves 98.40% average accuracy for mental arithmetic and 98.92% for motor imagery tasks, while ablation studies confirm that removing the attention fusion module causes accuracy to drop sharply to 90.48% and 89.93%, respectively, underscoring the critical role of feature-level gated fusion in multimodal integration. In the online setting, Muller et al. developed an EEG-fNIRS bimodal neurofeedback platform that generates a unified neurofeedback score by computing a weighted average of the alpha band event related desynchronization ratio at the C4 electrode and the oxyhemoglobin concentration changes above the right primary motor cortex [28]. A randomized controlled trial demonstrated that the fusion condition produced the strongest correlation between the neurofeedback score and motor imagery vividness, confirming that feature-level fusion effectively integrates electrophysiological and hemodynamic complementary information. Short-channel regression correction and Laplacian spatial filtering further enhance signal quality while maintaining real-time processing capability. These two studies collectively illustrate that feature-level fusion, whether implemented through deep attention mechanisms or weighted averaging strategies, consistently outperforms unimodal approaches by preserving modality-specific discriminative information and enabling cross-modal interactions, though the trade-off between computational efficiency and fusion sophistication remains an open challenge for practical deployment.

Feature-level fusion extracts features from each sensor modality independently during the feature extraction stage, then fuses the resulting feature vectors according to specific association rules before feeding them into a classifier for decision making [29]. Common feature extraction and fusion algorithms include principal component analysis (PCA) and independent component analysis (ICA) for linear dimensionality reduction and source separation, autoencoders for nonlinear feature learning, and dedicated feature fusion neural networks. Among these, autoencoder-based methods have gained particular attention in EEG-fNIRS fusion due to their capacity to jointly learn compact and discriminative representations from heterogeneous modalities, while PCA and ICA remain attractive for their computational efficiency and interpretability in scenarios with limited training data [30]. In EEG-fNIRS bimodal brain-computer interfaces, feature-level fusion has recently undergone a paradigm shift from traditional feature engineering toward deep learning and personalized modeling, forming a multilevel technical framework. At the feature construction and representation learning levels, deep integration methods such as stacked denoising autoencoders perform multilayer nonlinear transformations to jointly learn [31] and fuse higher-order bimodal features, establishing an early paradigm foundation for this field. At the discriminative feature selection level, information-theoretic and statistical criteria driven methods, including mutual information-based feature selection, correlation-filtered channel selection [32], and a two-stage redundancy elimination strategy combining ReliefF with minimum redundancy-maximum relevance [33], effectively reduce feature dimensionality and computational complexity while steadily improving classification performance. These methods share the common goal of feature dimensionality reduction, yet differ in their underlying selection criteria. Mutual information based approaches capture nonlinear feature-target dependencies and suit high dimensional feature spaces with complex interactions. Correlation-filtering prioritizes computational efficiency and suits real-time applications. The Relief-based minimum redundancy maximum relevance (mRMR) strategy balances redundancy elimination with relevance preservation, offering a more comprehensive solution when both feature quality and inter-feature correlations are critical. This diversity of criteria provides flexible options for different data characteristics and application scenarios.

For structured neurophysiological feature representation, task-related tensor decomposition separates task-shared from modality-specific components, endowing fused features with clear neurovascular coupling interpretability [34]; graph neural networks (GNNs) go beyond traditional Euclidean space limitations by revealing the intrinsic joint representation of bimodal features within graph topological structures [35]. Tensor decomposition excels at capturing the multi-dimensional structure of EEG-fNIRS data and preserving physiological interpretability, making it particularly suitable for studies investigating neurovascular coupling mechanisms. In contrast, GNNs are better suited for modeling spatial dependencies across distributed cortical regions, offering greater flexibility in capturing complex inter-channel relationships at the cost of reduced physiological interpretability.

At the fusion strategy decision-making level, combining canonical correlation analysis (CCA) with deep learning demonstrates performance equivalence between feature-level concatenation and decision-level voting [36]; personalized mathematical rules [37] and accuracy adaptive weighting mechanisms [38] offer interpretable and adaptive fusion weight allocation schemes targeting cross-individual generalization and clinical translation, respectively. This evolutionary progression from linear dimensionality reduction to nonlinear representation learning, and from global feature fusion to structured neurophysiological modeling, demonstrates that feature-level fusion continues to evolve toward deeper integration and personalization, with its core objective expanding beyond mere classification accuracy to encompass model interpretability, cross-individual generalization, and clinical applicability. These methods span a spectrum from performance validation to interpretable and adaptive fusion, collectively reflecting the evolution of feature-level fusion from fixed rule-based integration toward context-aware and user-adaptive decision making.

Overall, feature-level fusion has evolved from simple modal concatenation and statistical weighting to a deep integration paradigm that combines structured neurophysiological modeling, graph topology learning, and adaptive decision-making, increasingly focusing on user difference modeling and clinical translatability.

2.3. Decision-Level Fusion

Decision-level fusion independently performs feature extraction and classification for each modality, generating preliminary discrimination results or confidence distributions, then integrates these results through preset fusion rules to produce a final unified output [39]. This approach avoids temporal and spatial alignment issues of heterogeneous signals through a post-hoc integration mechanism and supports independent optimization of each modality, while uncertainty modeling enhances decision robustness [40]. However, decision-level fusion struggles to capture deep complementary information between modalities. Its performance depends on single-modality classifiers, and the system architecture is complex. This approach suits scenarios with significant differences in signal characteristics, though its effectiveness depends on both the fusion strategy and single-modality performance. Commonly used decision-level information fusion algorithms include Dempster-Shafer (D-S) evidence theory, Bayesian decision theory, weighted fusion, voting fusion, *etc*. [41].

Current research on multimodal fusion has progressively evolved from simple linear aggregation to adaptive uncertainty-aware paradigms, driven by the need to handle heterogeneous neural signal characteristics. In studies on multimodal brain signal decision fusion, the collaborative use of EEG and fNIRS has already demonstrated significant advantages. Rabbani et al. used maximum likelihood estimation to fuse classification results from seven deep learning networks, achieving 96% and 82.76% accuracy on two cognitive task datasets [42]. Their results demonstrated that decision fusion effectively improves classification performance and that hybrid convolutional neural network-recurrent neural network (CNN-RNN) structures offer advantages. This approach represents a classical probability-based decision fusion strategy, which relies on the statistical reliability of unimodal classifiers and offers computational simplicity. However, the linear probability fusion mechanism fails to model decision uncertainty and does not address physiological asynchrony between multimodal signals, limiting robustness in dynamic scenarios.

In recent years, with the development of deep learning and evidence reasoning theory, some researchers have begun to explore multimodal fusion methods that can quantify decision uncertainty. To overcome these limitations, recent studies have explored uncertainty-aware decision fusion frameworks. Kang et al. proposed an EEG-fNIRS fusion model based on a two-layer D-S evidence reasoning framework. The model extracts EEG time-frequency features and fNIRS spatiotemporal features through dual-scale convolution and a hybrid attention mechanism, respectively, and introduces two basic belief assignment functions and a Dirichlet distribution for uncertainty modeling, thereby achieving decision integration through two-layer evidence fusion [43]. This method achieved 83.26% average classification accuracy on public datasets, significantly outperforming existing methods and providing a highly reliable technical path for multimodal brain signal fusion. In contrast to maximum likelihood estimation, which relies on linear probability fusion, this evidence-based approach quantifies the uncertainty of each modality's decision via the Dirichlet distribution. It robustly integrates multimodal classification results, even when individual modality decisions prove unreliable, and thus captures the physiological asynchrony between EEG and fNIRS signals that linear fusion methods overlook. This explicit uncertainty management marks a fundamental departure from classical probability-based fusion, moving decision-level fusion beyond deterministic aggregation toward principled uncertainty reasoning.

Though both approaches fall within the decision-level fusion paradigm, their methodological underpinnings diverge considerably. The former hinges on the statistical reliability of independent unimodal classifier outputs, offering computational simplicity and efficiency; the latter explicitly models classification uncertainty through the Dirichlet distribution and realizes progressive belief mass integration via two-tier evidential reasoning, delivering superior accuracy at the expense of heightened computational complexity. Both strategies perform a posteriori ensemble based on final classification outputs without exploiting crossmodal interactions within classifier internals, rendering unimodal classifier performance the primary determinant of the fusion system's performance ceiling. EEG-fNIRS decision-level fusion has demonstrated practical utility across cognitive task decoding, neurorehabilitation, brain-computer interface control, and auxiliary neurological diagnosis, encompassing cognitive load assessment, mental fatigue monitoring, stress state identification, and neurorehabilitation support for individuals with motor impairments.

These practical applications exemplify a broader paradigm shift in multimodal fusion research, transitioning from static rule driven frameworks toward dynamic uncertainty-aware architectures. Investigations by Rabbani et al. and Kang et al. illustrate that sophisticated model ensembles and evidential reasoning approaches not only boost classification performance but also explicitly quantify decision uncertainty. Nevertheless, the field continues to encounter several fundamental challenges, including the design of deep nonlinear fusion architectures that transcend linear aggregation, the development of asynchronous compensation mechanisms to accommodate physiological latency, and the construction of lightweight, interpretable systems suitable for real world deployment. The technical characteristics and relative merits of these three fusion levels are summarized in Table 1.

Table 1. EEG-fNIRS multimodal fusion strategies.

Integration Strategy	Methods	Main Conclusions	Advantages and Disadvantages
Data-level fusion	Multi-channel fusion hybrid network	Unified encoding of three-channel feature maps, attention mechanism enhances feature representation [24].	Offering lossless retention and finest fusion granularity, this method necessitates high synchronization and computational resources to overcome inter-modal discrepancies in rate, scale, and spatiotemporal alignment.
	Wavelet transform fusion	Solves the heterogeneity problem of sampling rate and amplitude scale, with high computational efficiency [25].	
	Raw data joint processing of heterogeneous sensors	Fusing high-frequency laser with low-frequency webcam data achieved high-precision positioning, outperforming unimodal systems [26].	
Feature-level fusion	SDAE ensemble learning	Stacked denoising autoencoder ensemble performs multi-layer nonlinear feature learning and fusion of bimodal signals, exploring the combination of deep features and ensemble strategies to improve robustness and discriminative ability [31].	Map heterogeneous data to a unified feature space to alleviate inter-modal heterogeneity, with relatively strong interpretability and moderate computational efficiency. The feature extraction stage may lose fine-grained information from
	Correlation filter feature selection	Correlation filter optimized channels and feature selection, improving computational efficiency [32].	

	Relief-mRMR feature screening	Feature-level fusion and selection of EEG CSP and fNIRS multi-class common spatial pattern (MCSP) features significantly improved motor imagery classification accuracy [33].	the original signal, and the fusion performance highly depends on initial single-modal feature extraction quality.
	Task-related tensor decomposition	Used for individualized modeling of EEG-fNIRS neurovascular coupling, improving feature extraction accuracy and physiological interpretability [34].	
	GNNs fusion	First fuses graph-structured EEG-fNIRS channels with GNNs, revealing intrinsic joint bimodal features under grasp force task via graph representation learning [35].	
	CCA Deep Learning Fusion	EEG-fNIRS system significantly outperforms unimodal ones, with feature-level concatenation and decision-level soft voting as equally optimal [36].	
	Personalized mathematical rule fusion	Builds user-specific models via interpretable fusion rules to address large inter-user physiological variability and poor generalization in multimodal BCI [37].	
	Parallel deep feature extraction and accuracy-weighted linear fusion	Proposed HEFMI-ICH, first EEG-fNIRS brain hemorrhage dataset with dual-branch depthwise separable convolution (DSConv), accuracy-adaptive weighted fusion for multimodal decoding [38].	
Decision-level fusion	Maximum likelihood estimation integration	Maximum likelihood EEG-fNIRS fusion across seven DL classifiers achieved 96% (CNN-LSTM-GRU) and 82.76% (CNN-LSTM), significantly outperforming unimodal classification [42].	This approach offers modality optimization, classifier, and uncertainty robustness without cross-modal alignment, yet lacks deep complementarity, is unimodal-bounded.
	Two-level decision layer fusion based on D-S evidence reasoning	EEG-fNIRS uncertainty quantification achieved 83.26% accuracy on a public dataset, significantly outperforming traditional decision fusion [43].	

To systematically review the core technological evolution in this field, particularly the specific implementations and performance breakthroughs of feature extraction and fusion methods in motor imagery tasks, as shown in Table 1, three fusion strategies exhibit significant differences in technical principles and applicable scenarios. These studies not only demonstrate the paradigm shift from traditional spatial methods to deep learning but also provide important references for subsequent research.

3. Decoding Algorithms for EEG-fNIRS Fusion Data

In BCI and neuroscience research, classification is the core process to decode neural signals and map them to specific cognitive states. By classifying multimodal data such as EEG and fNIRS, different psychological states of individuals can be effectively identified, providing key support for applications such as cognitive monitoring [44], disease diagnosis [45], assistive interaction [46], and rehabilitation training [47]. By extracting discriminative features from high-dimensional, low signal-to-noise ratio neural signals, the conversion from neural activity to interpretable categories can be achieved, thereby completing the objective quantification of cognitive functions and the closed-loop control of BCIs.

3.1. Traditional Machine Learning Methods

Traditional machine learning methods in BCIs typically follow a serial processing paradigm of feature extraction, feature selection, and classifier design. Unlike deep learning, traditional machine learning relies on manually designed feature descriptors, making performance highly dependent on the quality of feature engineering [48]. This reliance on hand-crafted features not only requires domain expertise but also limits the model's ability to capture complex, nonlinear patterns embedded in EEG and fNIRS signals.

Decoding motor intentions to achieve high degrees of freedom in movement command control is key to improving the precision of BCI device operations [49]. In recent years, significant progress has been made in this field, including three-dimensional drone control based on MI-BCI [50] and robotic arm assistance in daily life [51]. As an important branch of artificial intelligence, machine learning aims to enable computers to automatically discover patterns through data-driven approaches and make predictions and decisions based on what they have learned [52]. In the field of multimodal fusion, traditional machine learning has historically relied on feature engineering and statistical models to achieve modality alignment and fusion, using techniques such as feature concatenation, CCA, and multiple kernel learning [53]. In research on decoding motor intentions based on EEG-fNIRS multimodal fusion, traditional machine learning methods have achieved a series of advances in the two core aspects of feature extraction and classifier design, thanks to their strong interpretability and computational efficiency.

Early research prioritized computational efficiency. Hasan et al. validated the feasibility of hybrid BCI systems under resource-constrained conditions [54]. Subsequent studies progressively introduced more refined feature engineering. Meng et al. demonstrated that Pearson-Fisher coefficient based channel selection effectively integrates signal correlation with feature separability information [55]; Qiu et al. further combined multi-domain features with progressive learning strategies, elevating motor imagery classification accuracy beyond 96% and revealing a strong positive correlation between feature fusion depth and performance gains [56]. Collectively, these studies trace a clear methodological progression from channel level optimization to statistical feature selection and, finally, to multi-domain feature integration, demonstrating that increasing the depth and breadth of feature engineering yields consistent improvements in decoding performance. Tang et al. [57] and Liu et al. [58] respectively explored the CSP-PCA-SVM and PCA-SVM frameworks from feature extraction and dimensionality reduction perspectives, confirming that traditional machine learning approaches maintain competitive performance on medium-scale datasets. In the domain of data-level fusion, Li et al. introduced a wavelet transform-based strategy that provides a novel pathway for cross-modal signal alignment [25]. For fine motor decoding tasks, Yang et al. incorporated Higuchi fractal dimension features into the filter bank CSP framework, achieving 80.47% accuracy in three class grasp posture decoding using k-nearest neighbor (KNN) classification [59]. These four works address distinct aspects of traditional machine learning-based fusion, including feature extraction and dimensionality reduction, data-level cross-modal signal alignment, and fine-grained task-specific decoding. Together, they illustrate the expanding scope of traditional ML methods, from general-purpose feature engineering to task-adaptive and fine motor decoding. The collective findings delineate a clear technical trajectory from single-modal feature engineering toward multi-modal deep integration. Nevertheless, critical bottlenecks persist in cross-subject generalization and real-time deployment.

Recent research has extended multimodal fusion from EEG-fNIRS dual modal configurations to more complex architectures integrating three or more modalities. Mateen N et al. pioneered the application of a genetic algorithm-optimized SVM framework to a hybrid EEG-EMG (electromyography)-fNIRS BCI system, attaining 98.17% classification accuracy in EEG-EMG tasks with a highly parsimonious feature set while improving average classification performance by 4 to 5 percentage points, thereby furnishing an efficient solution for real-time low-latency systems [60]. This multimodal expansion aligns with the broader trajectory toward individualized modeling. Laura Dipietro et al. constructed a multimodal predictive system that synthesizes clinical, sensor, and neuroimaging data for stroke rehabilitation outcome forecasting [61]. Through feature extraction, PCA-based dimensionality reduction, and Elastic Net regression, they elevated the coefficient of determination from 0.83 to 0.89, underscoring the substantial value of cross-modal data fusion in regression-based predictive tasks. Whereas the classification-oriented studies above concentrate on discrete motor intention decoding, this regression-based approach addresses a distinct clinical question, rehabilitation outcome prediction, illustrating that multimodal fusion confers advantages across both classification and predictive paradigms. Typical classifiers and their performance of traditional machine learning methods are shown in Table 2.

In summary, traditional machine learning methods follow a clear developmental trajectory in EEG-fNIRS multimodal fusion research. Channel selection and statistical feature engineering laid the foundational framework for subsequent methodological advances. The shift toward multi-domain feature integration expanded the information space available for classification, whereas data-level reconstruction strategies resolved cross-modal signal alignment challenges that feature-level approaches could not fully address. More recently, subject-specific feature selection has emerged as a means to mitigate the inter-individual variability that undermines cross-subject generalization. Deep learning methods excel at automatic feature extraction and have delivered substantial performance improvements, frequently surpassing traditional approaches by notable margins in large sample offline studies. Nevertheless, traditional machine learning methods retain distinct advantages in small-sample scenarios, real-time systems, and model interpretability. These complementary strengths indicate that hybrid approaches, such as integrating deep learning-derived features with interpretable traditional classifiers or adaptively selecting fusion strategies based on data availability, may offer a promising direction for future BCI systems.

Table 2. Decoding algorithms of traditional machine learning.

Methods	Main Contributions	Advantages and Disadvantages	References
Linear discriminant analysis	Multimodal fusion classification combining CSP and wavelet-fused signals achieved 88.1% accuracy; LDA further improves stability under small-sample conditions.	Simple computation, stable performance, good interpretability; poor adaptability to nonlinear distributed data, strong independence assumption between features.	[25]
Decision tree	Used for motor intention classification, used in parallel with KNN to build a hybrid classification system.	Strong interpretability, no need for data normalization, high computational efficiency; prone to overfitting, sensitive to noise, relatively weak generalization ability.	[54]
Support vector machine	Filter bank CSP-PCA fusion achieved 92.25% (self-collected) and 96.90% (public) accuracy; genetic algorithm-based feature selection further improved accuracy by 4–5%.	Strong generalization for high-dimensional features and nonlinear problems via kernel functions; however, long training time for large-scale data and hyperparameter tuning relies on experience.	[57,60]
KNN	Used for motor intention classification, achieving low-complexity classification tasks; combined with filter bank common spatial pattern and higuchi fractal dimension to decode grasping gestures.	Simple and low-overhead, suitable for real-time and small-sample scenarios, but limited high-dimensional expressivity; accuracy highly feature-quality dependent.	[59]
Elastic net regression	Used for stroke rehabilitation prediction tasks, integrating clinical, sensor, and neuroimaging data, increasing the predicted coefficient of determination from 0.83 to 0.89.	L1/L2 regularization suits high-dimensional sparse features with feature selection; mainly regression-oriented, limited classification use.	[61]

3.2. Deep Learning Methods

Deep learning is a machine learning method based on artificial neural networks, and its core is the automatic extraction of multi-level feature representations from data through multiple layers of non-linear transformations [62]. It learns complex patterns from large-scale data and is suitable for various types of data, such as images [63], speech [64], and time-series signals [65], showing powerful feature learning and classification capabilities in fields like BCIs. In research on motor intention decoding based on EEG-fNIRS multimodal fusion, deep learning methods have achieved significant breakthroughs in decoding performance due to their end-to-end feature learning capabilities.

Early research primarily concentrated on optimizing single-modality network architectures, with CNN networks, long short-term memory (LSTM) networks, bidirectional long short-term memory (Bi-LSTM) networks, temporal convolutional network (TCN), and gated recurrent units (GRU) laying the foundational backbones for subsequent multimodal extensions. A key challenge then emerged: how to effectively integrate heterogeneous features from different modalities. To address this, Ciaran C et al. pioneered a dual-CNN subnet structure that separately extracts EEG and fNIRS features and fuses them at an early stage before feeding into a GRU classifier, establishing a paradigm for independent feature extraction and cross-modal integration in bimodal systems [66]. However, this architecture still relied on manually designed branches. Mughal et al. advanced the field by transforming raw signals into recurrence plot features and employing a CNN-LSTM network for joint spatiotemporal modeling, achieving 88.41% average accuracy and 92.4% peak performance across modalities, thereby validating the potential of non-primitive representations to enhance cross-modal discriminability [67]. To overcome the difficulty of simultaneously capturing time-frequency characteristics, Liu Jinrui et al. combined wavelet packet energy entropy for frequency-domain extraction with a Bi-LSTM network for time-domain modeling, and further integrated these features via a one-dimensional CNN for deep fusion, attaining 95.31% accuracy on the public HYGRIP dataset [68]. Concurrently, optimization strategies emerged to improve parameter tuning efficiency. Majid Nour et al. integrated multi-band analysis with the grey wolf optimization algorithm to fine-tune the fully connected layers of a CNN, achieving 99.85% classification accuracy in bilateral hand MI tasks [69]. These advancements collectively indicate that deep learning in EEG-fNIRS fusion has evolved beyond isolated network design toward a diversified paradigm where architectural innovation, representation enhancement, and optimization-driven refinement jointly exploit the complementary characteristics of bimodal signals.

As research continues to advance, deep learning methods have achieved a series of significant breakthroughs in two key areas: decision fusion and cross-subject evaluation. In terms of decision fusion, Md. Hasin Raihan Rabbani et al. pioneered the introduction of maximum likelihood estimation into the EEG-fNIRS multi-task classification framework. After systematically comparing the performance of CNN, LSTM, and GRU networks, as well as their combined architectures, in single-modality classification, they derived fused decisions by multiplying the decision likelihood ratios, achieving a peak accuracy of 96% in cross-subject cognitive task identification. Building on this foundation, subsequent research gradually shifted focus from fusion strategies themselves toward model generalization capabilities in cross-subject scenarios. Arif et al. constructed a dual-branch CNN architecture comprising an EEG temporal branch and an fNIRS spatial branch, with a fusion classification layer performing decision-making [70]. They pioneered the introduction of a cross-subject evaluation framework into EEG-fNIRS multimodal analysis, achieving an F1 score of 65.05% under cross-subject conditions, substantially outperforming traditional machine learning and mainstream deep learning methods. Bunterngchit et al. approached the problem from the perspective of loss function optimization [71]. By combining a temporal convolutional network with an LSTM and designing a custom loss function, they achieved accuracy rates exceeding 99% across various cognitive tasks, reaching approximately 97.48% in motor imagery tasks, significantly surpassing existing cross-subject classification methods. These studies reveal a clear trajectory from the exploration of decision fusion strategies toward optimization of cross-subject generalization capabilities. Nevertheless, current cross-subject evaluations rely predominantly on macro-level metrics such as accuracy and F1 scores, lacking in-depth analysis of feature distribution shifts when models transfer across subjects. Moreover, the evaluation protocols across studies remain unstandardized. Future research should prioritize establishing standardized cross-subject evaluation benchmarks and exploring the potential of domain adaptation and domain generalization methods to mitigate the impact of individual differences on fusion model performance.

Recent research has progressively shifted the frontier of EEG-fNIRS fusion from basic architectural design toward addressing the intertwined challenges of data efficiency, cross-modal alignment, and clinical deployability. The prohibitive cost of acquiring labeled multimodal data constitutes a primary bottleneck. Euijin Jung et al. addressed this limitation through the EEG-fNIRS representation learning model, which combines a masked autoencoder for modality-specific feature extraction with contrastive learning for shared representation discovery, followed by few-shot linear probing [72]. This framework achieves fully supervised level performance with only 4 to 8 labeled samples per class, reducing labeled data requirements by approximately 90%. However, data efficiency alone cannot resolve the inherent spatiotemporal misalignment between EEG and fNIRS signals, which poses a second critical obstacle to effective fusion. Mutian Liu et al. tackled this issue via the spatial temporal alignment network, which incorporates an fNIRS guided spatial alignment layer and an EEG guided temporal alignment layer coupled with a cross attention mechanism, attaining 69.65% accuracy for motor imagery and 85.14% for mental arithmetic in subject-specific evaluations. Concurrently with these alignment efforts, architectural innovations have further extended the performance frontier. Zhizheng Yuan et al. introduced a capsule dynamic graph convolutional network that achieves 91.72% average accuracy on the HYGRIP dataset, surpassing single-modality EEG by approximately 8% and fNIRS by approximately 20% [73]. These methodological advances collectively establish the foundation for the most consequential challenge, translating laboratory grade fusion systems into real-world clinical practice. Chayut Bunterngchit et al. took a decisive step in this direction by developing the graph attention convolutional LSTM network, which integrates multiscale spatial feature extraction, graph attention based channel relationship modeling, and bidirectional temporal dynamics capture [74]. Validated on an EEG dataset from 50 stroke patients, the model achieved 99.52% classification accuracy with 97.43% leave-one-subject-out generalization and an inference latency of 33 to 56 milliseconds. These complementary lines of inquiry target distinct bottlenecks in EEG-fNIRS fusion, including labeled data scarcity, spatiotemporal misalignment, feature extraction from dynamic graph structures, and clinical translation using lightweight, real-time architectures. Their collective contributions illustrate that the field is moving toward holistic solutions that simultaneously address data efficiency, representation learning, and deployment constraints. This demonstration of clinical grade performance underscores a broader trajectory: the field has matured beyond isolated algorithmic benchmarks toward integrated frameworks that simultaneously address data scarcity, modality heterogeneity, and deployment constraints.

Transfer learning is a machine learning method that improves the learning efficiency of a target domain by reusing knowledge from a source domain. Its core lies in breaking the traditional assumption in modeling that training and testing data must follow the same distribution [75]. In EEG-fNIRS multimodal motor imagery decoding studies, due to significant individual differences in EEG and near-infrared signals, and the high cost of synchronous acquisition and annotation of multimodal data, cross-subject generalization faces the dual challenges of small sample sizes and distribution shift.

Two complementary technical approaches to transfer learning have emerged in EEG-fNIRS cross-subject decoding. One approach concentrates on source domain selection strategies. Chen et al. employed the Wasserstein distance to separately select the optimal source domains for EEG and fNIRS modalities and achieved source-target statistical distribution alignment through deep correlation alignment and CCA, attaining 74.87% accuracy on intracerebral hemorrhage patient data [38]. This method, however, depends critically on the quality and accessibility of source domain data. The alternative approach emphasizes direct alignment of cross-domain distributions. Yu and Zhang proposed a two-stage alignment strategy incorporating Euclidean alignment at the data-level to reduce inter-subject covariance distribution discrepancies and multi-constraint feature alignment to simultaneously match marginal and conditional distributions between source and target domains in a shared subspace, while introducing a selective pseudo-label update strategy to prevent error accumulation [76]. Unlike Chen's method, this strategy circumvents external source domain selection and achieves cross-subject generalization through explicit feature-space distribution alignment, though pseudo-label quality remains a critical factor influencing alignment performance.

Departing from these strategy-level solutions, Liu approaches the problem from an architectural perspective, embedding domain adaptation mechanisms directly into the network hierarchy [77]. Liu finds that shallow layers favor global marginal distribution alignment, while deeper layers require finer-grained conditional adaptation. This observation holds particular value for multimodal fusion scenarios, where domain discrepancies arising from distinct sensor modalities demand differentiated adaptation strategies across network layers. The Liu framework adopts maximum mean discrepancy for distribution alignment. This metric offers high computational efficiency but captures complex distribution shifts in high-dimensional feature spaces less effectively than adversarial alternatives. The identical label space assumption further constrains its applicability in cross-task transfer settings.

These studies reveal that transfer learning in EEG-fNIRS fusion has evolved from single-point domain adaptation strategies toward multi-level, multi-constraint joint alignment frameworks. Significant challenges persist, however, including the precision of inter-domain distribution discrepancy measurement and the reliability of alignment under completely unlabeled target domain conditions. More broadly, deep learning continues to drive methodological innovation across EEG-fNIRS multimodal BCIs, with research trends shifting from accuracy-driven approaches toward enhanced generalization, cross-subject adaptation, and clinical applicability [78]. Table 3 summarizes the performance of deep learning methods in EEG-fNIRS decoding.

Table 3. Decoding algorithms for deep learning.

Methods	Main Contributions	Advantages and Disadvantages	References
Spatial-temporal alignment network (STA-Net)	fNIRS-guided spatial and EEG-guided temporal alignment with cross-attention achieves task accuracies of 69.65%, 85.14%, and 79.03%.	Bidirectional guided alignment resolves modal spatiotemporal heterogeneity, yet task-specific performance varies notably across three tasks.	[21]
CNN + LSTM + GRU	Maximum likelihood decision fusion yields the highest cross-subject cognitive task accuracy of 96%.	Decision-level fusion circumvents modality alignment, yet linear probability fusion lacks decision uncertainty modeling.	[42]
CNN + GRU	Dual CNN nets each extract EEG and fNIRS features; fusion to GRU classifier achieves first independent dual-modality feature extraction and early fusion.	The dual-branch structure effectively extracts modality-specific features; mainly aimed at speech tasks, limited application scenarios.	[66]
CNN-LSTM	EEG-fNIRS recursive graph features with CNN-LSTM spatiotemporal fusion achieve 88.41% integrated accuracy and up to 92.4% maximum.	Recursive graphs effectively represent the spatiotemporal structure of signals; constructing recursive graphs increases computational complexity.	[67]
Bi-LSTM + 1DCNN	Wavelet packet energy entropy extracts frequency features, Bi-LSTM extracts temporal features, and one-dimensional convolutional neural network (1DCNN) performs deep fusion, achieving 95.31% accur	Time-frequency domain feature extraction is sufficient, and the dual-branch structure achieves feature complementarity; the model structure is complex, and real-time performance needs to be evaluated.	[68]

CNN	Using multi-bandwidth analysis and grey wolf optimization to tune the fully connected layer, this method first introduces a cross-subject evaluation framework, achieving 99.85% accuracy on public datasets.	Extremely high accuracy, cross-subject evaluation is close to practical application; the optimization algorithm has high computational overhead, and cross-subject performance is relatively low.	[69]
TCN + LSTM	TCN-LSTM with a custom loss function achieves above 99% classification accuracy for cognitive tasks and 97.48% for motor imagery decoding tasks.	Temporal convolution effectively captures local dependencies; the model is sensitive to hyperparameters and has weak interpretability.	[71]
Masked autoencoder (MAE) + Contrastive learning	Only 4–8 labeled samples per class are needed to achieve fully supervised performance, reducing labeling requirements by about 90%.	Breaks the dependency on paired data, significantly reducing labeling costs; pre-training computational cost is relatively high.	[72]
Capsule dynamic graph convolutional network	Achieves 91.72% accuracy on the HYGRIP dataset, approximately 8% higher than unimodal EEG and 20% higher than fNIRS.	Retains spatial hierarchical relationships and adaptively learns the topological structure, model complexity is high, training is difficult.	[73]
Graph attention convolutional LSTM network (GACL-Net)	Achieves 99.52% accuracy on a stroke patient dataset, inference speed 33-56ms.	Fast inference enables real-time clinical rehabilitation, primarily validated on stroke patient cohorts.	[74]

4. Summary and Outlook

Over the past decade, EEG-fNIRS based MI-BCI decoding has advanced considerably. Researchers have moved from single data-level fusion toward feature-level fusion, decision-level fusion, and multi-level integration, resulting in increasingly sophisticated fusion frameworks. The paradigm shift from traditional machine learning to deep learning has significantly improved classification accuracy and model generalization, allowing fuller utilization of multimodal complementary advantages. Nevertheless, several key bottlenecks remain before clinical application and daily use become feasible.

The study of EEG-fNIRS based motor intention decoding has developed over several decades. With the continuous updates in fusion methods and decoding algorithms, it is expected to overcome traditional bottlenecks such as limited command sets and less natural brain-machine interaction. However, it should be noted that some problems in the current research still need to be addressed:

1. The decoding accuracy of fine motor intentions is insufficient. MI-BCIs mostly rely on simplified paradigms, and the coding mechanisms for multi-dimensional movement vectors and dynamic neural responses are still unclear. Currently, decoding motor intentions based on scalp EEG signals still faces bottlenecks. The weak signals of fine movements and error-related potentials (ErrP) result in overall accuracy that has not yet reached practical levels, making most systems still rely on simplified action paradigms to achieve basic control functions. At the same time, research on the coding mechanisms of multi-dimensional movement parameters, such as direction, speed, and force, remains limited and has not systematically revealed the dynamic EEG response patterns throughout the process from motor planning to execution.
2. Multimodal hardware synchronization and signal fusion are difficult, and differences in spatiotemporal characteristics restrict the effectiveness of fusion. Existing linear fusion methods lack adaptability, limiting the efficiency of information collaboration. EEG and fNIRS face issues such as mismatched sampling frequencies and spatial registration errors during hardware integration, and significant differences in signal time-domain and frequency-domain characteristics increase the complexity of multimodal fusion. The commonly used linear weighted fusion methods in current research are difficult to adapt to different task states and individual characteristics, and a framework capable of dynamically adjusting fusion weights has not been established, restricting the collaborative expression ability of multimodal information in motor intention decoding.

3. Neurofeedback exhibits individual response differences and non-responder phenomena, and hemodynamic signal delays affect feedback synchrony and training outcomes. Approximately 30% of participants show non-responder characteristics during neurofeedback training, and differences in individual anatomy, cognitive strategies, and task engagement significantly impact the consistency of training effects. In addition, the hemodynamic responses measured by fNIRS have a delay of several seconds, leading to temporal mismatches in multimodal feedback systems, which may reduce the efficiency of neural plasticity induction during training [79].
4. The equipment has poor portability, and its application scenarios are limited. The system has a single function and is difficult to support complex tasks, making clinical translation a significant challenge. Currently, EEG acquisition devices are usually large in size and complicated to operate, which restricts the promotion and application of MI-BCIs in daily environments. At the same time, due to limitations in decoding algorithm performance and paradigm flexibility, the existing systems are relatively single in function. Especially in populations such as stroke patients, the system's effectiveness and tolerability have not been fully validated, and changes in neurovascular coupling mechanisms under pathological conditions may further affect signal quality and intervention outcomes [80].

Beyond the specific challenges outlined above, future EEG-fNIRS integration will deepen along three dimensions: multimodal fusion strategies, model architecture innovation, and system application loops. Further exploration of multi-level fusion mechanisms should include collaborative optimization of feature-level, decision-level, and spatiotemporal-level fusion, combined with cross-modal attention, spatiotemporal joint modeling, meta-learning, and transfer learning to enhance model generalization across subjects, tasks, and scenarios. As discussed in Section 3.2, transfer learning offers a promising pathway to address cross-subject generalization, though challenges such as negative transfer and computational efficiency remain unresolved. Promoting lightweight, embedded system design will enable wearable, low-latency, long-term neural signal monitoring, providing feasible pathways for daily cognitive enhancement, rehabilitation training, and real-time brain-computer interaction.

In terms of data resources, evaluation systems, and clinical translation, large-scale, multi-center, and multi-pathological open datasets should be constructed, and standardized evaluation protocols and performance benchmarks should be established to promote fair model comparison and clinically reproducible validation. In addition, further integration of multiple physiological signals such as eye movements and electromyography, is needed to achieve more comprehensive state recognition and coordinated regulation. By deepening interdisciplinary collaboration among neuroscience, computational science, clinical medicine, and rehabilitation engineering, EEG-fNIRS technology can be advanced from the laboratory to clinical diagnosis, personalized neural modulation, intelligent rehabilitation systems, and immersive interactive applications, forming a complete closed loop of 'signal acquisition—analysis—intervention—evaluation' and laying the foundation for precise neural health management and next-generation human-machine intelligent integration systems.

Statement of the Use of Generative AI and AI-Assisted Technologies in the Writing Process

During the preparation of this manuscript, the authors used DeepSeek in order to improve the clarity and check the grammar. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Acknowledgments

The authors thank their colleagues for constructive discussions that improved the manuscipt.

Author Contributions

Conceptualization, Z.W. and L.C.; Methodology, J.L.; Investigation, J.L.; Data Curation, Y.Z.; Writing—Original Draft Preparation, Z.W. and J.L.; Writing—Review & Editing, L.C., J.L. and D.M.; Visualization, Y.Z.; Supervision, L.C. and D.M.; Project Administration, L.C. and D.M.; Funding Acquisition, L.C. and D.M. All authors have read and agreed to the published version of the manuscript.

Ethics Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

No new data were created or analyzed during this study. Data sharing is not applicable to this article.

Funding

This work was supported by the National Key Research and Development Program of China (2023YFF1205200), the National Natural Science Foundation of China (62376190, 62476193), the Autonomous Project of Haihe Laboratory of Brain-Computer Interaction and Human-Machine Integration (25HHNJSS00007), and the Tianjin Science and Technology Program-State Key Laboratory Major Special Project (25ZXZSSS00020).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Rodríguez-Obregón DE, Moreno HA, Carrera IG, Lopez-Lopez L. A Decade-Long Systematic Review of EEG-Driven Gait-Intention Interfaces for Lower-Limb Rehabilitation and Prosthetics. *Appl. Mech. Mater.* **2025**, *929*, 107–118. DOI:10.4028/p-rpE8vv
2. Palanichamy C, Thirumoorthi SP, Lakshminarayanan K, Madathil D, Rahman MH. Multimodal brain-computer interface for robotic control: Integration of real-time gaze tracking and EEG-based motor imagery. *Med. Biol. Eng. Comput.* **2026**, *64*, 781–792. DOI:10.1007/s11517-025-03489-6
3. Zhang M, Yang Y, Zhao Y, Sui C, Sui Y, Jiang Y, et al. The application of integrating electroencephalograph-based emotion recognition technology into brain–computer interface systems for the treatment of depression: A narrative review. *Adv. Technol. Neurosci.* **2024**, *1*, 188–200. DOI:10.4103/ATN.ATN-D-24-00018
4. Zhang Y, Liu D, Li T, Zhang P, Li Z, Gao F. CGAN-rIRN: A data-augmented deep learning approach to accurate classification of mental tasks for a fNIRS-based brain-computer interface. *Biomed. Opt. Express* **2023**, *14*, 2934–2954. DOI:10.1364/BOE.489179
5. Guo JC. Frequent policies boost the accelerated breakthrough of the brain-computer interface industry chain. *Securities Daily*, 6 February 2025. DOI:10.28096/n.cnki.ncjrb.2025.000509
6. Qin X. Beijing and Shanghai introduce action plans, brain-computer interface industry may explode. *China Business News*, 10 February 2025. DOI:10.38300/n.cnki.nzgjy.2025.000199
7. Chai YY. Research on Feature Extraction and Classification Methods for Motor Imagery EEG Signals. Master's Thesis, Beijing University of Chemical Technology, Beijing, China, 2024.
8. Xu GY. Research on Classification and Recognition of Motor Imagery EEG Signals in Brain-Computer Interface. Master's Thesis, Xi'an Technological University, Xi'an, China, 2024.

9. Murtaza H, Ahmed M, Murtaza G, Zafar SN, Nazir T. A novelty towards neural signatures—Unveiling the inter-subject distance metric for EEG-based motor imagery. *Biomed. Signal Process. Control* **2025**, *105*, 107552. DOI:10.1016/j.bspc.2025.107552
10. AL-Quraishi MS, Ahmed G, AbouOmar MS, Ouerdane F, Eltayeb A, Alyazidi NM, et al. A Brain–Machine Interface Framework for Motor Imagery Classification Using Functional Connectivity and Ensemble Learning: Toward Real-Time Applications in Healthcare and Industry. *Arab. J. Sci. Eng.* **2026**, *51*, 15973–15987. DOI:10.1007/s13369-025-10616-w
11. Du PF. Research on Motion Planning of Upper Limb Rehabilitation Robot Based on Motor Imagery EEG Signals. Master's Thesis, Anhui University of Science and Technology, Huainan, China, 2025.
12. Cariello S, Sanalitro D, Micali A, Buscarino A, Bucolo M. Brain–Computer-Interface-Based Smart-Home Interface by Leveraging Motor Imagery Signals. *Inventions* **2023**, *8*, 91. DOI:10.3390/inventions8040091
13. Al-qaysi ZT, Zaidan BB, Zaidan AA, Suzani MS. A review of disability EEG based wheelchair control system: Coherent taxonomy, open challenges and recommendations. *Comput. Methods Programs Biomed.* **2018**, *164*, 221–237. DOI:10.1016/j.cmpb.2018.06.012
14. Melissant C, Ypma A, Frietman EEE, Stam CJ. A method for detection of Alzheimer's disease using ICA-enhanced EEG measurements. *Artif. Intell. Med.* **2005**, *33*, 209–222. DOI:10.1016/j.artmed.2004.07.003
15. Yan H, Zhou Z, He X, Li J, Yuan Z, Wang H. Dynamic graph-temporal convolutional neural network for EEG-fNIRS multimodal motor imagery/execution decoding. *Control Theory Appl.* **2024**, *36*, 1–8. Available online: https://link.cnki.net/urlid/44.1240.TP.20241130.0849.002 (accessed on 27 July 2026).
16. Blanco R, Koba C, Crimi A. Investigating the interaction between EEG and fNIRS: A multimodal network analysis of brain connectivity. *J. Comput. Sci.* **2024**, *82*, 102416. DOI:10.1016/j.jocs.2024.102416
17. Zhang J, Zhang ZY, Wang YL, Zhou B, Xia CY, Su M, et al. Unveiling the upper-limb functional recovery mechanisms in stroke patients using brain-machine interfaces: A near-infrared functional imaging-based study. *Sci. Rep.* **2025**, *15*, 39704. DOI:10.1038/s41598-025-23267-6
18. Balendra, Negi PC, Sharma N, Sharma S. Scalogram sets based motor imagery EEG classification using modified vision transformer: A comparative study on scalogram sets. *Biomed. Signal Process. Control* **2025**, *104*, 107640. DOI:10.1016/j.bspc.2025.107640
19. Codina T, Blankertz B, Lühmann AV. Multimodal fNIRS–EEG sensor fusion: Review of data-driven methods and perspective for naturalistic brain imaging. *Imaging Neurosci.* **2025**, *3*, IMAG-a. DOI:10.1162/IMAG.a.974
20. Deligani RJ, Borgheai SB, McLinden J, Shahriari Y. Multimodal fusion of EEG-fNIRS: A mutual information-based hybrid classification framework. *Biomed. Opt. Express* **2021**, *12*, 1635. DOI:10.1364/BOE.413666
21. Liu M, Yang B, Meng L, Zhang Y, Gao S, Zan P, et al. STA-Net: Spatial–temporal alignment network for hybrid EEG-fNIRS decoding. *Inf. Fusion* **2025**, *119*, 103023. DOI:10.1016/j.inffus.2025.103023
22. Zhou H, Zhao Y, Liu Y, Lu S, An X, Liu Q. Multi-Sensor Data Fusion and CNN-LSTM Model for Human Activity Recognition System. *Sensors* **2023**, *23*, 4750. DOI:10.3390/s23104750
23. Faisal M, Sahoo S, Hazarika J. STeCANet: Spatio-temporal cross attention network for brain computer interface systems using EEG-fNIRS signals. *J. Neural Eng.* **2025**, *22*, 066035. DOI:10.1088/1741-2552/ae2954
24. Chen J, Wang D, Hu B, Yi W, Xu M, Chen D, et al. MCFHNet: Multi-Channel Fusion Hybrid Network for Efficient EEG-fNIRS Multi-modal Motor Imagery Decoding. In Proceedings of the 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Glasgow, UK, 11–15 July 2022; pp. 4821–4825. DOI:10.1109/EMBC48229.2022.9871385
25. Li LZ, Meng M, Gao YY, Ma YL. A data fusion method for hybird EEG-fNIRS BCI base on wavelet transform. *Chin. J. Sens. Actuators.* **2023**, *36*, 1064–1072. Available online: https://chinatransducers.seu.edu.cn/ch/reader/view_abstract.aspx?file_no=cg220191&flag=1 (accessed on 27 July 2026).
26. Wojewoda KK, Culmer PR, Jackson AE, Levesley MC. Position tracking of a passive rehabilitation robot. In Proceedings of the 2016 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER), Chengdu, China, 19–22 June 2016; pp. 1–6. DOI:10.1109/CYBER.2016.7574785
27. Li H. DeepSyncNet: Deep synchronized fusion network for EEG-fNIRS multimodal brain-computer interfaces. *Biomed. Signal Process. Control* **2026**, *113*, 109100. DOI:10.1016/j.bspc.2025.109100
28. Muller CO, Prampart T, Bannier E, Corouge I, Maurel P. Evaluating the effects of multimodal EEG-fNIRS neurofeedback for motor imagery: An experimental platform and study protocol. *PLoS ONE* **2025**, *20*, e0331177. DOI:10.1371/journal.pone.0331177
29. Amin SU, Alsulaiman M, Muhammad G, Mekhtiche MA, Shamim Hossain M. Deep Learning for EEG motor imagery classification based on multi-layer CNNs feature fusion. *Future Gener. Comput. Syst.* **2019**, *101*, 542–554. DOI:10.1016/j.future.2019.06.027

30. Bunterngchit C, Wang J, Su J, Wang Y, Liu S, Hou ZG. AMFN: Autoencoder-led Multimodal Fusion Network for EEG–fNIRS Classification. *Procedia Comput. Sci.* **2024**, *250*, 8–14. DOI:10.1016/j.procs.2024.11.002
31. Yang S, Yin Z, Wang Y, Zhang W, Wang Y, Zhang J. Assessing cognitive mental workload via EEG signals and an ensemble deep learning classifier based on denoising autoencoders. *Comput. Biol. Med.* **2019**, *109*, 159–170. DOI:10.1016/j.compbiomed.2019.04.034
32. Ali MU, Zafar A, Kallu KD, Masood H, Mannan MM, Ibrahim MM, et al. Correlation-filter-based channel and feature selection framework for hybrid EEG-fNIRS BCI applications. *IEEE J. Biomed. Health Inform.* **2023**, *28*, 3361–3370. DOI:10.1109/JBHI.2023.3294586
33. Guo M, Feng L, Chen X, Li M, Xu G. A novel strategy for differentiating motor imagination brain-computer interface tasks by fusing EEG and functional near-infrared spectroscopy signals. *Biomed. Signal Process. Control* **2024**, *95*, 106448. DOI:10.1016/j.bspc.2024.106448
34. Lin J, Lu J, Shu Z, Han J, Yu N. Subject-Specific Modeling of EEG-fNIRS Neurovascular Coupling by Task-Related Tensor Decomposition. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2024**, *32*, 452–461. DOI:10.1109/TNSRE.2024.3355121
35. Zhu Z, Han J, Zhang Z, Wannawas N, Faisal AA. Identifying the Nature of Grip Force Signals in EEG & fNIRS with Multi-Modal Graph Fusion Network. In Proceedings of the 2025 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Copenhagen, Denmark, 14–18 July 2025; pp. 1–7. DOI:10.1109/EMBC58623.2025.11254624
36. Gao T, Chen C, Liang G, Ran Y, Huang Q, Liao Z, et al. Feature fusion analysis approach based on synchronous EEG-fNIRS signals: Application in etomidate use disorder individuals. *Biomed. Opt. Express* **2025**, *16*, 382. DOI:10.1364/BOE.542078
37. Nia AF, Tang V, Maso Talou GD, Billinghurst M. Decoding emotions through personalized multi-modal fNIRS-EEG Systems: Exploring deterministic fusion techniques. *Biomed. Signal Process. Control* **2025**, *105*, 107632. DOI:10.1016/j.bspc.2025.107632
38. Shi J, Chen D, Zhao X, Zhao Z, Li S, Xu Y, et al. HEFMI-ICH: A hybrid EEG-fNIRS motor imagery dataset for brain-computer interface in intracerebral hemorrhage. *Sci Data* **2025**, *12*, 1816. DOI:10.1038/s41597-025-06100-7
39. Fu C, Qian F, Su K, Su Y, Wang Z, Shi J, et al. HiMul-LGG: A hierarchical decision fusion-based local–global graph neural network for multimodal emotion recognition in conversation. *Neural Netw.* **2025**, *181*, 106764. DOI:10.1016/j.neunet.2024.106764
40. Zhang YT, Wang XY, Guo JQ, Liu ZX. Multimodal sentiment analysis of book reviews based on decision-level fusion. *J. Libr. Inf. Sci.* **2025**, *10*, 72–79. Available online: https://d.wanfangdata.com.cn/periodical/kjqbkfyjj202505011 (accessed on 27 July 2026).
41. Cai H, Qu Z, Li Z, Zhang Y, Hu X, Hu B. Feature-level fusion approaches based on multimodal EEG data for depression recognition. *Inf. Fusion* **2020**, *59*, 127–138. DOI:10.1016/j.inffus.2020.01.008
42. Rabbani MHR, Islam SMR. Deep learning networks based decision fusion model of EEG and fNIRS for classification of cognitive tasks. *Cogn. Neurodyn.* **2024**, *18*, 1489–1506. DOI:10.1007/s11571-023-09986-4
43. Kang RL, Li YR, Shi WX, Li JX. Decision-level fusion method for EEG-fNIRS motor imagery classification based on Dempster-Shafer evidential reasoning. *Acta Electron Sin.* **2025**, *53*, 941–950. Available online: https://link.cnki.net/urlid/11.2087.TN.20250508.1349.010 (accessed on 27 July 2026).
44. Li Y, Zhu L, Huang A, Zhang J, Yuan P. Multimodal MBC-ATT: Cross-modality attentional fusion of EEG-fNIRS for cognitive state decoding. *Front. Hum. Neurosci.* **2025**, *19*, 1660532. DOI:10.3389/fnhum.2025.1660532
45. Wang H, Huang X, Xu L, Guo S, Gong C, Mengcheng C, et al. Analysis of the analgesic mechanism of TENS-WAA in colonoscopy using the EEG-fNIRS system: A study protocol for a randomised controlled trial. *BMJ Open* **2025**, *15*, e103681. DOI:10.1136/bmjopen-2025-103681
46. Berger A, Horst F, Müller S, Steinberg F, Doppelmayr M. Current State and Future Prospects of EEG and fNIRS in Robot-Assisted Gait Rehabilitation: A Brief Review. *Front. Hum. Neurosci.* **2019**, *13*, 172. DOI:10.3389/fnhum.2019.00172
47. Wu F, Liu Y, Jiang H, Li L, Feng C, Sun J, et al. Integrated EEG–fNIRS for Characterizing Cortical Responses and Neurovascular Coupling in Automated and Discrete Gait Tasks. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2025**, *33*, 3805–3814. DOI:10.1109/TNSRE.2025.3610690
48. Moreno-Castelblanco SR, Vélez-Guerrero MA, Callejas-Cuervo M. Lower-Limb Motor Imagery Recognition Prototype Based on EEG Acquisition, Filtering, and Machine Learning-Based Pattern Detection. *Sensors* **2025**, *25*, 6387. DOI:10.3390/s25206387
49. Atilla F, Postma M, Alimardani M. Gamification of motor imagery brain-computer interface training protocols: A systematic review. *Comput. Hum. Behav. Rep.* **2024**, *16*, 100508. DOI:10.1016/j.chbr.2024.100508

50. Gong J, Liu H, Zhao L, Maeda T, Cao J. EEG-Powered UAV Control via Attention Mechanisms. *Appl. Sci.* **2025**, *15*, 10714. DOI:10.3390/app151910714
51. Du M, Chen L, Li Y, Xia L, Liu Y, Guo M, et al. Effect of exercise-based interventions on stroke rehabilitation: An umbrella review of systematic reviews and meta-analyses. *J. NeuroEng. Rehabil.* **2025**, *22*, 240. DOI:10.1186/s12984-025-01781-y
52. Slama K, Yahyaouy A, Riffi J, Mahraz MA, Tairi H. Comprehensive review of machine learning and deep learning techniques for epileptic seizure detection and prediction based on neuroimaging modalities. *Vis. Comput. Ind. Biomed. Art* **2025**, *8*, 27. DOI:10.1186/s42492-025-00208-8
53. Zhang T, Li CL, Meng L, Wang H, Su YC. Review of multimodal data fusion. *Henan Sci. Technol.* **2025**, *52*, 37–42. DOI:10.19968/j.cnki.hnkj.1003-5168.2025.20.008
54. Hasan MAH, Khan MU, Mishra D. A computationally efficient method for hybrid EEG-fNIRS BCI based on the Pearson correlation. *BioMed Res. Int.* **2020**, *2020*, 1838140. DOI:10.1155/2020/1838140
55. Meng M, Dai LY, Ma YL, Gao YY. Multimodal EEG-fNIRS channel selection method based on PF coefficient. *Chin. J. Sens. Actuators.* **2022**, *35*, 84–91. Available online: https://chinatransducers.seu.edu.cn/ch/reader/view_abstract.aspx?flag=1&file_no=cg210208&journal_id=cgjsxb (accessed on 27 July 2026).
56. Qiu L, Zhong Y, He Z, Pan J. Improved classification performance of EEG-fNIRS multimodal brain-computer interface based on multi-domain features and multi-level progressive learning. *Front. Hum. Neurosci.* **2022**, *16*, 973959. DOI:10.3389/fnhum.2022.973959
57. Tang X, Yang C, Sun X, Zou M, Wang H. Motor Imagery EEG Decoding Based on Multi-Scale Hybrid Networks and Feature Enhancement. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2023**, *31*, 1208–1218. DOI:10.1109/TNSRE.2023.3242280
58. Liu HD, Xu BJ, Li MQ. EEG-fNIRS feature fusion based on PCA. *Mod. Electron. Technol.* **2023**, *46*, 29–33. DOI:10.16652/j.issn.1004-373x.2023.19.006
59. Yang JC. Research on Decoding of Upper Limb Motor Parameters and Neurofeedback Based on VR Cues. Master's Thesis, Kunming University of Science and Technology, Kunming, China, 2023.
60. Mateen N, Naeem M, Khan MJ, Yousaf T, Ali A, Altabey WA, et al. Subject based feature selection for hybrid brain computer interface using genetic algorithm and support vector machine. *Results Eng.* **2025**, *27*, 105649. DOI:10.1016/j.rineng.2025.105649
61. Dipietro L, Eden U, Teixeira P, Tirasawasdichai N, Warinpramote J, Pundik S, et al. A multimodal machine learning approach to predict Fugl-Meyer scores and motor recovery potential in stroke rehabilitation: Toward precision-based therapies. *Inf. Sci.* **2025**, *721*, 122564. DOI:10.1016/j.ins.2025.122564
62. Huang X, Meng H, Li Z. Deep learning for epileptic seizure prediction from EEG signals: A review. *Biomed. Signal Process. Control* **2026**, *117*, 109518. DOI:10.1016/j.bspc.2026.109518
63. Zheng G, Wang H, Chai X, Xin X, Li F, Li H, et al. Interpretable deep learning for multicenter gastric cancer T staging from CT images. *NPJ Digit. Med.* **2025**, *9*, 2. DOI:10.1038/s41746-025-02002-5
64. Ancilin J, Lyernisha SR. PV2ES-BiRTM: Variational auto encoder sampling-based deep learning framework for emotion recognition from speech signals. *Knowl. Based Syst.* **2026**, *332*, 114885. DOI:10.1016/j.knosys.2025.114885
65. Baig MAA, Ratyal NI, Amin A, Jamil U, Khalid HM, Zia MF. A multi-modal deep learning framework for power quality disturbance classification: An integration of 1D time-series signals and 2D scalograms. *Comput. Electr. Eng.* **2025**, *128*, 110716. DOI:10.1016/j.compeleceng.2025.110716
66. Cooney C, Folli R, Coyle D. A Bimodal Deep Learning Architecture for EEG-fNIRS Decoding of Overt and Imagined Speech. *IEEE Trans. Biomed. Eng.* **2022**, *69*, 1983–1994. DOI:10.1109/TBME.2021.3132861
67. Mughal NE, Khan MJ, Khalil K, Javed K, Sajid H, Naseer N, et al. EEG-fNIRS-based hybrid image construction and classification using CNN-LSTM. *Front. Neurorobot.* **2022**, *16*, 873239. DOI:10.3389/fnbot.2022.873239
68. Liu JR, Song T, Shu ZL, Han JD, Yu NB. An EEG-fNIRS time-frequency feature fusion and collaborative classification method for motor decoding. *Chin. J. Sci. Instrum.* **2022**, *43*, 165–173. DOI:10.19650/j.cnki.cjsi.J2209389
69. Nour M, Öztürk Ş, Polat K. A novel classification framework using multiple bandwidth method with optimized CNN for brain–computer interfaces with EEG-fNIRS signals. *Neural Comput. Applic.* **2021**, *33*, 15815–15829. DOI:10.1007/s00521-021-06202-4
70. Arif A, Wang Y, Yin R, Zhang X, Helmy A. EF-Net: Mental State Recognition by Analyzing Multimodal EEG-fNIRS via CNN. *Sensors* **2024**, *24*, 1889. DOI:10.3390/s24061889
71. Bunterngchit C, Wang J, Su J, Wang Y, Liu S, Hou ZG. Temporal attention fusion network with custom loss function for EEG–fNIRS classification. *J. Neural Eng.* **2024**, *21*, 066016. DOI:10.1088/1741-2552/ad8e86
72. Jung E, An J. EFRM: A Multimodal EEG–fNIRS Representation-learning Model for few-shot brain-signal classification. *Comput. Biol. Med.* **2025**, *199*, 111292. DOI:10.1016/j.compbiomed.2025.111292

73. Yuan Z, Li Y, Zhang H, Liu X, Li S, Zhu Y, et al. Decoding Hybrid EEG-fNIRS Upper Limb Motor Execution with Capsule Dynamic Graph Convolutional Neural Network. In Proceedings of the 2025 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Copenhagen, Denmark, 14–18 July 2025; pp. 1–7. DOI:10.1109/EMBC58623.2025.11251587
74. Bunterngchit C, Baniata LH, Baniata MH, ALDabbas A, Khair MA, Chearanai T, et al. GACL-Net: Hybrid Deep Learning Framework for Accurate Motor Imagery Classification in Stroke Rehabilitation. *Comput. Mater. Contin.* **2025**, *83*, 517–536. DOI:10.32604/cmc.2025.060368
75. Barón Díaz D, Drommershausen A, Grünberger A, Holtmann D. Transfer Learning Approaches in Bioprocess Engineering: Opportunities and Challenges. *Biotechnol. Bioeng.* **2026**, *123*, 1417–1431. DOI:10.1002/bit.70186
76. Yu B, Zhang L. A Multi-Constrained Transfer Learning for Cross-Subject Decoding of Motor Imagery-Based BCI. *Mathematics* **2026**, *14*, 1314. DOI:10.3390/math14081314
77. Liu FD. Research on Classification of Motor Imagery EEG Signals Based on Deep Transfer Learning. Master's Thesis, Changchun University of Technology, Changchun, China, 2025.
78. Balsabti SM, Al-Gburi RM, gaib R, Mustafa A, Ahmed SK, Issa AM, et al. Advances in deep learning for multimodal brain imaging: A comprehensive survey. *Neurosci. Inform.* **2026**, *6*, 100252. DOI:10.1016/j.neuri.2025.100252
79. Sun X, Dai C, Wu X, Han T, Li Q, Lu Y, et al. Current implications of EEG and fNIRS as functional neuroimaging techniques for motor recovery after stroke. *Med. Rev.* **2024**, *4*, 492–509. DOI:10.1515/mr-2024-0010
80. Pan YD, Kang RL, Li YR. Cross-subject classification method based on EEG-fNIRS fusion. *J. Nanjing Univ. Inf. Sci. Technol.* **2025**, *17*, 1–13. DOI:10.13878/j.cnki.jnuist.20250510001