Simultaneous EEG-fNIRS Data Classification Through Selective Channel Representation and Spectrogram Imaging

CHAYUT BUNTERNGCHIT[1,2], (Graduate Student Member, IEEE), JIAXING WANG[1], AND ZENG-GUANG HOU[1], (Fellow, IEEE)

[1]State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[2]School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China

CORRESPONDING AUTHORS: J. WANG (jiaxing.wang@ia.ac.cn) AND Z.-G. HOU (zengguang.hou@ia.ac.cn)

ABSTRACT The integration of electroencephalography (EEG) and functional near-infrared spectroscopy (fNIRS) can facilitate the advancement of brain-computer interfaces (BCIs). However, existing research in this domain has grappled with the challenge of the efficient selection of features, resulting in the underutilization of the temporal richness of EEG and the spatial specificity of fNIRS data. To effectively address this challenge, this study proposed a deep learning architecture called the multimodal DenseNet fusion (MDNF) model that was trained on two-dimensional (2D) EEG data images, leveraging advanced feature extraction techniques. The model transformed EEG data into 2D images using a short-time Fourier transform, applied transfer learning to extract discriminative features, and consequently integrated them with fNIRS-derived spectral entropy features. This approach aimed to bridge existing gaps in EEG-fNIRS-based BCI research by enhancing classification accuracy and versatility across various cognitive and motor imagery tasks. Experimental results on two public datasets demonstrated the superiority of our model over existing state-of-the-art methods. Thus, the high accuracy and precise feature utilization of the MDNF model demonstrates the potential in clinical applications for neurodiagnostics and rehabilitation, thereby paving the method for patient-specific therapeutic strategies.

INDEX TERMS Brain–computer interfaces, multimodal neuroimaging, short-time Fourier transform, spectrogram imaging.

Clinical and Translational Impact Statement—This study is preclinical research. The high accuracy of MDNF in EEG-fNIRS data analysis paves the way for personalized neurodiagnostics and rehabilitation strategies in clinical settings.

I. INTRODUCTION

ADVANCEMENTS in neuroimaging have spurred remarkable progress in brain-computer interfaces (BCIs), thereby enabling direct communication between the human brain and external devices. This technology holds transformative potential for individuals with motor disabilities and those seeking innovative modes of interaction [1]. Electroencephalography (EEG), with its high temporal resolution, excels at capturing rapid brain dynamics. This facilitates users in controlling external devices through imagined movements or specific cognitive tasks [2]. Functional near-infrared spectroscopy (fNIRS) provides valuable spatial information regarding brain activation, thereby localizing hemodynamic responses associated with different tasks or stimuli [3].

Despite their strengths, EEG and fNIRS have limitations. EEG lacks spatial specificity, rendering it difficult to pinpoint precise brain regions involved in specific cognitive processes. Conversely, fNIRS, while offering spatial localization, exhibits limited temporal resolution, rendering it

challenging to track rapid neural activity changes. The integration of these modalities in hybrid EEG-fNIRS BCIs offers a promising solution, capitalizing on their complementary advantages. Simultaneous EEG-fNIRS recordings provide a more comprehensive view of brain activity, combining EEG's temporal insights with fNIRS's spatial localization [4], [5]. This integrated approach holds the potential for more accurate and robust BCIs across diverse cognitive states and tasks, thus enhancing their adaptability for real-world applications.

Although initial integration efforts encountered challenges in data fusion, co-registration, and synchronization [6], current advancements facilitate seamless EEG-fNIRS recording, opening new research frontiers. Machine learning (ML) and deep learning (DL) are crucial for extracting insights from EEG-fNIRS data in BCI applications. ML algorithms such as support vector machines (SVM) [7] and linear discriminant analysis (LDA) [8] classify cognitive states or intentions, whereas DL models such as convolutional neural networks (CNNs) [9] and recurrent neural networks (RNNs), including long short-term memory (LSTM) [10], further enhance BCI capabilities. CNNs capture spatial patterns in EEG data, while RNNs model temporal dependencies. DL offers advantages in terms of accuracy, real-time control, and communication for motor-impaired individuals by automating feature extraction.

However, the existing ML and DL approaches in hybrid BCIs face significant challenges. A key bottleneck lies in feature selection and representation, which directly impact the classification accuracy and the broader applicability of these BCIs. Traditional approaches often rely on hand-crafted features, a time-consuming process that can introduce biases and may not fully capture the complex, multi-dimensional information within EEG-fNIRS data. Even DL models, despite being capable of automated feature extraction, may struggle with the application of conventional one-dimensional (1D) processing techniques to EEG-fNIRS data. This 1D approach limits the ability to effectively represent the dynamic interplay between the temporal patterns captured by EEG and the spatial information provided by fNIRS.

Furthermore, several existing hybrid BCI systems do not adequately address the synergy between EEG's temporal richness and fNIRS's spatial specificity. There is untapped potential in leveraging these complementary strengths in a unified framework, which could result in improved classification accuracy across a wider range of tasks. In addition, many approaches focus on a single modality (either EEG or fNIRS), missing the opportunity for the creation of a more robust feature space that maximizes the benefits of both.

To address these challenges, this study introduced a novel approach that effectively integrated EEG and fNIRS data, targeting specific brain regions and employing advanced feature extraction techniques. The key contributions of this study are as follows.

1) Development of the multimodal DenseNet fusion (MDNF) model: We developed a customized MDNF model that integrates transformed EEG data along

temporal and spectral dimensions with fNIRS spectral entropy representation, providing a rich, multi-dimensional feature space and enhancing classification accuracy.

2) Strategic selection of EEG channels: EEG channels were selected based on neuroanatomical locations and their associations with cognitive and motor functions. This strategic selection was validated through higher classification outcomes, thus ensuring that key features were effectively included in the model.

3) Preparation of two-dimensional (2D) representation using short-time Fourier transform (STFT) of EEG data: EEG data was transformed into a 2D representation in the both temporal and spectral domains. This facilitated seamless integration with the MDNF model through transfer learning, which ensured the extraction of crucial data features and improved overall model performance.

This study aimed to advance hybrid EEG-fNIRS BCI systems, thereby enhancing the accuracy and applicability across a range of motor imagery (MI) and cognitive tasks, including n-back, discrimination/selection response (DSR), and word generation (WG). The remainder of the paper is organized as follows. Section II reviews the relevant literature. Section III explains the research methodology used in this study. Section IV provides an exposition of the results obtained, a comprehensive discussion, potential applications, and future research directions. Finally, Section V concludes the paper.

II. RELATED WORK

This section provides an overview of prior research on BCI, organized into EEG-based BCI, fNIRS-based BCI, and hybrid EEG-fNIRS.

A. EEG-BASED BCI

EEG-based BCIs facilitate direct communication between the human brain and external devices by recording electrical activity from scalp electrodes. Users perform specific tasks, such as MI or cognitive activities, generating distinct EEG patterns. These patterns are then analyzed to determine the user's intent, facilitating device control. Researchers utilize both open-source datasets and custom experimental setups [16], [17], [18] to advance BCI technology.

Studies have explored various data classification methods, including spectral estimation (e.g., autocorrelation, power spectrum [19], [20]), time-based features (e.g., mean, standard deviation (SD), skewness [21]), and waveform transformations (e.g., STFT, continuous wavelet transformation, and discrete wavelet transformation [22], [23]) for 2D time-frequency feature extraction. Both conventional ML algorithms (e.g., SVM, k-nearest neighbour (k-NN) [24]) and DL methods (e.g., CNNs, deep neural networks (DNNs), RNNs [19], [25], [26]) have been widely used for classification. DL methods, leveraging pre-trained models such as

TABLE 1. Comparison of hybrid EEG-fNIRS BCI methods.

Ref	Method	Features	Classes	Accuracy
[8]	LDA	Pearson correlation coefficient-based	4 classes based on symptom severity	79.31%
[11]	k-NN	FWHTC	MI (left hand vs. right hand)	78.21%
[12]	DNN	Signal frequency band (EEG) HbR and HbO (fNIRS)	MI (left hand vs. right hand)	83.28%
[13]	DNN	Deep features from EEG and fNIRS	n-back (3 classes), DSR, WG, and MI	87, 91, 92, and 91%
[14]	CNN	Deep features from EEG and fNIRS	n-back (3 classes)	89%
[9]	CNN	Bands and band ratios (EEG), Spectral entropy (fNIRS)	n-back (3 classes), DSR, WG, and MI	93.8, 97.2, and 94.5%
[7]	SVM	Bivariate FBC features	0- vs. 2-back, 0- vs. 3-back	77, and 83%
[15]	CNN-LSTM	Nonlinear features from EEG and fNIRS	n-back (0-, 2-, 3-back)	88.41%

DenseNet and ResNet [27], [28], have generally exhibited superior performance in tasks such as depression classification and MI detection.

Despite their successes, EEG-based BCIs are plagued by limited spatial resolution, susceptibility to signal interference, and individual variability, which hinder their broader applicability. To address these challenges, researchers have explored fNIRS as a complementary modality.

B. fNIRS-BASED BCI

Various studies utilize fNIRS data and ML techniques for decoding brain activity patterns. For motor tasks, DL methods (CNNs, LSTM, and Bi-LSTM) have exhibited promise, achieving accuracies of up to 88.5% [29]. Novel feature extraction approaches, such as vector-based phase analysis, have further increased accuracy (98.7% for two-class BCIs [30]). Studies using sparse representation classification [31] and CNN-based time series classification [32] have further demonstrated the potential of fNIRS BCIs, with certain achieving accuracies near or exceeding 98%. Studies have also explored preprocessing techniques, with GLM-based approaches offering potential accuracy improvements [33]. Further, studies utilizing ML algorithms such as random forest and ANNs have also demonstrated high classification accuracy in tasks involving rest, MI, and motor execution [34]. DL approaches, particularly CNNs, have outperformed conventional ML methods for MI task classification [35].

However, fNIRS-based BCIs are plagued by limited measurement depth, and temporal resolution, and generally focusing on cortical regions. Thus, these limitations may necessitate integration with modalities such as EEG for more comprehensive brain activity analysis.

C. HYBRID EEG-fNIRS

Researchers have explored the simultaneous integration of EEG and fNIRS recordings to obtain a more comprehensive understanding of neurological processes. Studies utilizing hybrid frameworks often leverage the complementary strengths of both modalities. For instance, employing SVM with functional brain connectivity (FBC) features derived from combined EEG-fNIRS data has demonstrated promise in classifying cognitive tasks [7]. Similarly, using k-NN for classification and the application of singular value decomposition of Fast Walsh-Hadamard transform coefficients (FWHTC) for feature extraction has also yielded

good results [11]. DL approaches have demonstrated advantages over conventional ML in hybrid systems. DNNs have achieved notable accuracy in classifying cognitive tasks [12], [13]. CNNs excel at capturing spatial patterns within the complex data obtained from EEG and fNIRS recordings [9], [14]. Hybrid models that incorporate both EEG and fNIRS data have demonstrated potential in the classification of cognitive decline and Alzheimer's disease, even identifying relevant brain regions associated with the progression of the disease [8]. Finally, integrating the fNIRS and EEG data using recurrence plots and CNN-LSTM models offers encouraging results for classification tasks [15]. The selection of these models for our study stems from their proven efficacy in handling the multifaceted nature of EEG-fNIRS data. A summary of recent EEG-fNIRS work has been presented in Table 1.

The results presented in Table 1 underscore the potential of hybrid EEG-fNIRS BCIs for diverse applications, including the classification of both cognitive and MI tasks. However, the existing studies often rely on limited datasets or focus on specific tasks. In addition, while demonstrating the feasibility of integrating EEG and fNIRS, many prior approaches may not fully exploit the rich temporal and spatial information available within this combined data.

A key challenge lies in the effective representation of the complex interplay between the temporal dynamics captured by EEG and the spatial information provided by fNIRS. Traditional feature extraction methods, often relying on basic statistical measures, may not adequately capture this interplay.

To address this limitation, this study proposed a novel approach using STFT for EEG signal transformation. STFT decomposes a signal into its constituent time-frequency components, thus providing a more comprehensive representation compared to traditional 1D processing [22]. This approach facilitates the capture of the temporal variations in the EEG signal and the dominant frequencies at different points in time. Studies have shown that STFT-based features can improve classification accuracy in EEG-based BCIs [36], [37], [38], [39]. By incorporating time-frequency information, we aimed to create a richer feature space that better reflects the underlying neural activity patterns associated with various cognitive and MI tasks.

Furthermore, existing DL models for EEG-fNIRS integration often necessitate substantial training data for optimal performance. This can be a significant obstacle,

as collecting large, high-quality EEG-fNIRS datasets can be time-consuming and expensive.

To address this challenge, our study employed transfer learning techniques. Transfer learning involves adapting a pre-trained model on a related task to a new task with limited data. This approach can significantly reduce the training time and improve model performance when the source and target tasks share underlying similarities [40]. In the context of EEG-fNIRS BCIs, pre-trained models on EEG or fNIRS data from related tasks can be fine-tuned for the specific classification problem at hand. This strategy can alleviate the need for massive datasets and potentially enhance the generalizability of our model.

III. METHODOLOGY

This study aimed to create a reliable model for the classification of cognitive and MI tasks using EEG data, transformed into STFT 2D images, integrated with fNIRS data, and processed with the proposed MDNF model. The methodology involves dataset description, STFT-based EEG transformation, channel selection, data filtering, feature extraction, model architecture, and performance evaluation.

A. DATASETS DESCRIPTION

This study employed two publicly available datasets. Dataset 1 [41] included EEG and fNIRS recordings from 26 healthy individuals (9 males, 17 females, average age 26.1±3.5 years). Participants completed three cognitive tasks: n-back (0-back, 2-back, and 3-back), DSR (target vs. non-target symbols), and WG (word generation vs. baseline). The DSR and n-back task sessions comprised a 2 s instruction (I) period, a 40 s task (T) period, and a subsequent 20 s rest (R) period. In the WG task, the task period was shortened to 10 s with 1 s for stop and a rest period of 13-15 s. EEG data, recorded at a 1,000 Hz sampling frequency using a 30-electrode setup, were downsampled to 200 Hz. Simultaneously, fNIRS data, collected with 36 channels, were downsampled to 10 Hz. Further, raw fNIRS data were transformed into changes in oxygenated (ΔHbO) and deoxygenated (ΔHbR) concentrations based on the Beer-Lambert law. The recording configuration has been depicted in Fig. 1(a).

Dataset 2 [42] focused on MI tasks and comprised data from 29 participants (14 males, 15 females, average age 28.5±3.7 years). The dataset included EEG and fNIRS recordings. EEG data was collected from 30 channels using a BrainAmp EEG amplifier, whereas fNIRS data was collected from 36 channels using NIRScout equipment. Participants engaged in kinesthetic MI by visualizing hand-opening or closing while holding a ball (left hand vs. right hand). Each trial comprised distinct phases, including an instruction period, a task phase, and a rest interval, resulting in three sessions in total. Preprocessing involved re-referencing, bandpass filtering, independent component analysis for artefact removal, and downsampling/filtering to enhance data quality. fNIRS data were processed to calculate ΔHbO and
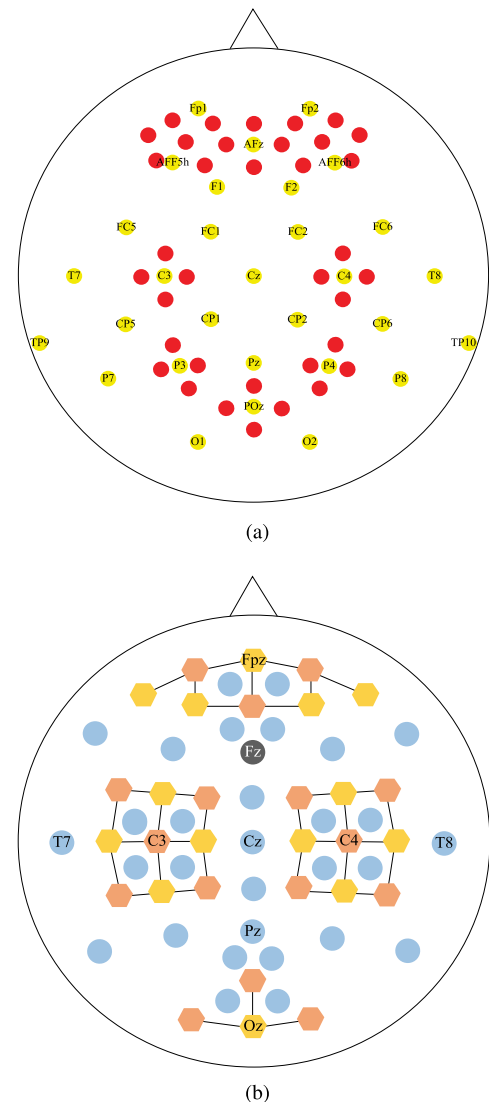

FIGURE 1. (a) Positions of EEG electrodes (yellow) and fNIRS optodes (red) of Dataset 1. (b) Placement of EEG electrodes (blue and black (ground) circles) and fNIRS sources (yellow hexagons) and detectors (orange hexagons) of Dataset 2. Black solid lines denote fNIRS channels.

ΔHbR concentrations. Fig. 1(b) depicts the electrode placement.

B. DATA TRANSFORMATION USING STFT

The first step involved transforming raw EEG data into images using STFT. This offers several advantages in the analysis of time-series data, such as EEG and fNIRS signals. STFT facilitated the visualization of the temporal evolution of frequency components, effectively encapsulating the spectral content of the data over time. This transformation simplifies the data representation, rendering it more amenable to DL models designed for image processing. It facilitates the extraction of intricate patterns and relationships in the time-frequency domain, which can be crucial for tasks such as event classification or feature extraction. In addition,

transforming data into images can enhance interpretability and facilitate the application of CNNs, which are proficient at learning hierarchical features from image-like data. Thus, STFT-driven image representations bridge the gap between time-series data and modern deep-learning techniques, offering a powerful tool for the comprehensive classification of complex signals. STFT decomposes the data into overlapping segments of time and frequency representation, creating spectrograms. The translation-invariant nature of STFT images facilitates the identification of patterns despite temporal shifts in EEG data. To calculate the STFT, we divided the EEG data into overlapping segments of 1 s with a 0.5 s overlap. The STFT of an EEG is computed using (1).

$$STFT[n, k] = \sum_{i=1}^{N} x[m]w[n-m]e^{-j2\pi k \frac{m}{N}} \quad (1)$$

where n represents the time index, k represents the frequency index, $x[m]$ is the EEG signal at index m, and w is the window function. The number of samples N in each segment is determined by the sampling frequency and window size. The small chunks of 1 s with an overlap of 0.5 s have been used for the generation of STFT. The number of samples is determined by the sampling frequency and the selected window sizes.

C. CHANNEL SELECTION

Although the use of all available EEG channels for training classifiers may appear ideal, it can be computationally expensive and susceptible to overfitting. To address these limitations, we strategically selected a subset of informative channels based on their anatomical locations and their association with cognitive and motor functions. We strategically selected a subset of 10 EEG channels (Fp1, Fp2, AFz, F1, F2, C3, C4, CP1, CP2, and Pz) for Dataset 1, and a similar subset for Dataset 2 (AFp1, AFp2, AFF5h, F3, F4, FCC5h, Cz, CCP3h, CCP5h, and CCP6h). This selection is based on a thorough understanding of the neuroscience underpinning the tasks.

For Dataset 1, the channels Fp1, Fp2, and AFz, located in the prefrontal cortex, are essential for higher-order cognitive functions such as decision-making, planning, and working memory. Similarly, for Dataset 2, the channels AFp1, AFp2, and AFF5h target the same cognitive functions. The fronto-central cortex (channels F1, F2, C3, and C4 for Dataset 1; F3, F4, and FCC5h for Dataset 2) is implicated in attention, motor preparation, and sensory integration, which are crucial for all tasks and particularly for the MI task's visuomotor components. Finally, the parietal cortex (channels CP1, CP2, and Pz for Dataset 1; CCP3h, CCP5h, and CCP6h for Dataset 2) supports spatial awareness, attention, and sensory processing, which are vital for the cognitive tasks at hand [43].

For fNIRS, a simplified spectral entropy-based representation was used. Consequently, no channel selection was performed. This simplified representation facilitated the

model in maintaining lower complexity even when utilizing all available channels.

This strategic selection ensures that brain activity from essential cognitive and motor areas is represented while also facilitating a more efficient and focused analysis. The chosen combination reflects a balance between covering a broad spectrum of brain functions and maintaining computational efficiency. Although alternative channel configurations may be viable, the presented selection was justified by its effectiveness in capturing a comprehensive array of brain processes, as evidenced by enhanced precision in task categorization outcomes. through the adoption of this method, which is deeply rooted in an understanding of brain functionalities, we substantially improved the capacity of the classifier to identify relevant patterns and render accurate categorizations based on EEG data, thereby optimizing the classification process for cognitive task identification.

To further validate the rationale behind this channel selection, an experimental approach was employed. The proposed model, incorporating the selected channels and their corresponding features (band values and band-ratios), was used to evaluate the channel selection strategy presented in [9]. To assess both cognitive and MI states, channels were grouped based on their anatomical locations and associated brain regions. Leveraging data from five diverse individuals, the evaluation results confirmed that the selected channel groups (Tables 2 and 3) for both cognitive and MI states yielded accuracy levels comparable to utilizing all available channels. This finding underscores the effectiveness of the chosen channels in representing a comprehensive range of brain activity while maintaining computational efficiency.

D. DATA FILTERING

Following the selection of specific EEG channels, the subsequent step was data filtering. A 4th-order Butterworth bandpass filter was devised to extract EEG data within the 1-45 Hz frequency range. This filtration process preserved the relevant activity information while diminishing unwanted artefacts and extraneous data. The filtering procedure was applied to the distinct channels individually. In addition, a distinct low-pass filter was implemented for processing fNIRS data, setting the upper limit at 0.2 Hz. This low-pass filter attenuated frequencies above this threshold, ensuring that only the slower hemodynamic changes and relevant signals associated with fNIRS data were retained [44]. Consequently, the filtering process optimized the fNIRS data for subsequent analysis, aligning with the specific characteristics of this modality.

E. FEATURE EXTRACTION

Two distinct approaches were employed to extract the spectral features from the EEG and fNIRS data for subsequent classification. For EEG data, STFT was utilized to convert the time-series data into spectrograms (STFT images). These spectrograms effectively captured the temporal dynamics of

TABLE 2. Model-based selection of channels group for cognitive states in Dataset 1.

Subgroup	Channels	Accuracy	Precision	Recall	F1 score
Frontal channels	Fp1, Fp2, AFF5h, AFF6h, AFz, F1, F2	0.588	0.602	0.589	0.575
Central channels	C3, C4, Cz	0.63	0.63	0.63	0.63
Parietal channels	Pz, P3, P4, P7, P8	0.678	0.678	0.678	0.678
Temporal channels	T7, T8, TP9, TP10	0.792	0.794	0.793	0.792
Occipital channels	O1, O2, POz	0.725	0.761	0.726	0.716
Motor and cognitive integration	Fp1, Fp2, F1, F2, C3, C4, Pz, CP1, CP2	0.666	0.687	0.665	0.656
Sensorimotor integration	C3, C4, CP1, CP2, Pz, P3, P4, T7, T8	0.745	0.746	0.745	0.745
Cognitive and emotional processing	Fp1, Fp2, AFz, F1, F2, Cz, Pz, O1, O2	0.649	0.662	0.65	0.643
Visual and auditory processing	T7, T8, TP9, TP10, O1, O2, POz, P3, P4	0.711	0.715	0.711	0.71
Executive and sensory processing	Fp1, Fp2, AFz, F1, F2, Cz, CP1, CP2, Pz, POz	0.713	0.715	0.713	0.712
Selected subgroup	**Fp1, Fp2, AFz, F1, F2, C3, C4, CP1, CP2, Pz**	**0.816**	**0.816**	**0.816**	**0.816**
All channels	30 channels	0.867	0.869	0.867	0.867

TABLE 3. Model-based selection of channels group for MI states in Dataset 2.

Subgroup	Channels	Accuracy	Precision	Recall	F1 score
Frontal channels	AFp1, AFp2, AFF1h, AFF2h, AFF5h, AFF6h, F3, F4	0.589	0.659	0.567	0.506
Central channels	Cz, FCC3h, FCC4h, FCC5h, FCC6h	0.592	0.626	0.573	0.534
Parietal channels	Pz, P3, P4, P7, P8	0.619	0.617	0.615	0.615
Occipital channels	POO1, POO2, PPO1h, PPO2h	0.539	0.562	0.512	0.401
Motor and cognitive integration	AFp1, AFp2, AFF5h, F3, F4, Cz, CCP3h, CCP5h, CCP6h	0.634	0.634	0.627	0.626
Sensorimotor integration	Cz, CCP3h, CCP4h, CCP5h, CCP6h, Pz, P3, P4, T7, T8	0.619	0.775	0.595	0.53
Cognitive and emotional processing	AFp1, AFp2, AFF5h, F3, F4, Cz, Pz, POO1, POO2	0.64	0.64	0.641	0.64
Visual and auditory processing	T7, T8, POO1, POO2, P3, P4	0.574	0.777	0.547	0.443
Executive and sensory processing	AFp1, AFp2, AFF5h, F3, F4, Cz, CCP3h, CCP4h, CCP5h, CCP6h, Pz, POO1	0.634	0.633	0.632	0.632
Selected subgroup	**AFp1, AFp2, AFF5h, F3, F4, FCC5h, Cz, CCP3h, CCP5h, CCP6h**	**0.726**	**0.754**	**0.715**	**0.711**
All channels	30 channels	0.732	0.731	0.731	0.731

brain activity within a specific frequency range. Fig. 2(a) shows sample STFT images for channels Fp1 (channel 1) and CP1 (channel 8) of participant 1 during the WG task. For fNIRS data, owing to the inherently lower spectral resolution of fNIRS (limited to 0-0.2 Hz), STFT images were deemed unsuitable for capturing the nuanced spectral information. To address this, the fNIRS data were segmented into smaller time windows (2-5 s) and subjected to spectral entropy calculations. The idea of selecting the 2-5 s varying windows involves ensuring that the output shape of the data remains consistent, with the same number of frames for each signal (despite varying time intervals). Spectral entropy measures the complexity and unpredictability contained in the data across the frequency range of fNIRS. It serves as an alternative for the spectral representation of the data via the analysis of the quantified ratio of spatial information encoded along the frequency bands. The spectral entropy is calculated as follows.

$$H = -\sum_f \hat{P}(f) \log(\hat{P}(f)) \qquad (2)$$

where H is the spectral entropy, $\hat{P}(f)$ represents the normalized power spectral density of fNIRS data, and f is the frequency.

This approach quantifies the complexity of the signal's frequency spectrum, providing a more comprehensive representation of the underlying brain activity. Notably, in contrast

to EEG processing, all 72 channels encompassing both HbO and HbR data were retained for the fNIRS analysis.

The preprocessing stage culminated in a data transformation process tailored to each modality. For EEG data, this entailed converting the time-series data into spectrograms via the STFT, with each image standardized to a resolution of 128×128 pixels and three color channels. Conversely, fNIRS data was reshaped into samples defined by a fixed number of frames and channels, ensuring consistency across varying time windows. Table 4 provides a comprehensive overview of these transformations.

F. DATA SPLITTING
To prevent data leakage across trials, a meticulous data splitting procedure was implemented. For EEG data, each trial was transformed into a set of images via the STFT, with one image generated per selected channel. This group of images, representing a single trial, was treated as a unified sample during the splitting process, guaranteeing that all images from a given trial remained together. In the case of fNIRS data, a framing operation with overlapping windows was applied. This transformed the data from its original trial-based structure (T) into a format where each trial was represented by a sequence of windows (T_{W_i}), with i denoting the index of the corresponding window. Crucially, each group of windows derived from a single trial was assigned the same activity class label. Subsequently, the data split was performed across trials, ensuring that all windows belonging to a specific trial
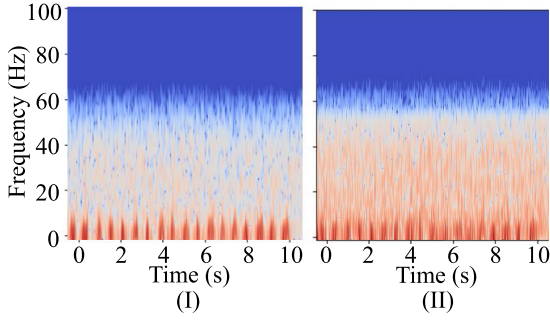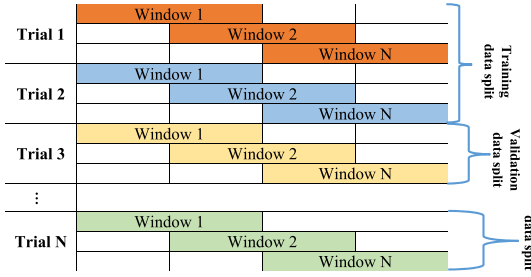
FIGURE 2. (a) Spectrograms of EEG data for the WG task from participant 1. **(I)** Channel Fp1 showing frontal lobe activity, and **(II)** channel CP1 indicating parietal lobe activity, both demonstrating the time-frequency characteristics captured by the STFT transformation. **(b)** Data splitting procedure. Each trial is segmented into overlapping windows. Trials are then assigned to training, validation, and test sets, ensuring all windows within a trial remain together.

remained within the same set (training, validation, or testing). This approach is visually illustrated in Fig. 2(b).

G. MODEL ARCHITECTURE

The proposed MDNF model for the EEG data classification involved DenseNet201, which is a DL architecture characterized by its convolutional blocks with skip connections as a basenet. This choice is guided by several factors, including its intricate structure, efficient parameter usage, and high accuracy owing to its foundation on the ImageNet dataset. The proficiency of the model in recognizing complex patterns within spectrograms aligns seamlessly with the representation of EEG data. Notably, the capacity of the model to capture intrinsic characteristics without requiring manual feature extraction is a significant benefit, facilitating it in the detection of activity patterns associated with different brain areas [45].

Furthermore, because it has been extensively trained on a varied range of pictures from the ImageNet database, DenseNet201 is particularly well-suited for instances when the EEG dataset is restricted. In terms of architecture, the strategy entails the development of ten separate DenseNet201 models, known as basenets, each catering to one of the ten specified EEG channels. The outputs of these basenets pass through pooling layers and are then flattened. These 10-flattened outputs are integrated through concatenation and subsequently fed into a dense layer equipped with

DN: DenseNet201 M base layer, **A:** Average pooling M layers (3, 3), **FM:** Flatten M layers, **CO:** Concatenate layer, **FU:** Fully connected layer 2048 filters, **C:** Conv 2D layer, **R:** Activation ReLu, **M:** Max pooling layer, **F:** Flatten layer, **D:** Dense layer 512 Filters, **DO:** Dropout 0.5, **O:** Output layer (N classes)
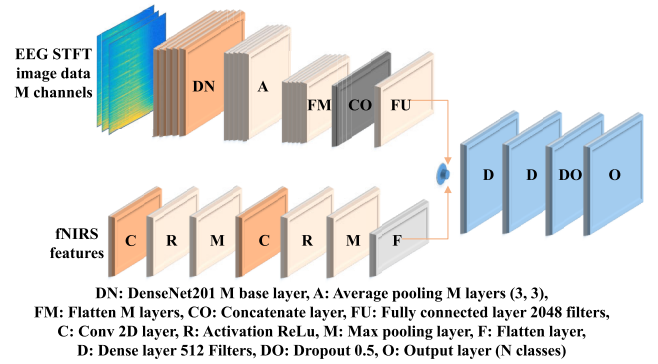
FIGURE 3. Architecture of the proposed MDNF model. This diagram illustrates the flow of data from EEG and fNIRS inputs through the DenseNet201 basemodel, highlighting the key layers and connections used for feature extraction and integration.

Algorithm 1 Pseudocode of the Proposed MDNF

Require: EEG input channels: $X_{\text{EEG}}^{(i)}$ for $i = 1$ to 10, each with shape (128, 128, 3), labels y, nearest neighbors k, fNIRS Input Channels: $X_{\text{fNIRS}}^{(j)}$ for $j = 1$ to 72

1: Process begins
2: **while** $i \leq$ EEG$_{channels}$ **do**
3:     $A_{\text{EEG}}^{(i)} \leftarrow$ Pass $X_{\text{EEG}}^{(i)}$ through a DenseNet201 basemodel
4: **end while**
5: **while** $j \leq$ fNIRS$_{channels}$ **do**
6:     $A_{\text{fNIRS}}^{(j)} \leftarrow$ Pass $X_{\text{fNIRS}}^{(j)}$ through Eqs. (3) - (9)
7: **end while**
8: $EEG_{\text{concatenated}} \leftarrow [A_{\text{EEG}}^{(1)}, A_{\text{EEG}}^{(2)}, \ldots, A_{\text{EEG}}^{(10)}]$
9: $fNIRS_{\text{concatenated}} \leftarrow [A_{\text{fNIRS}}^{(1)}, A_{\text{fNIRS}}^{(2)}, \ldots, A_{\text{fNIRS}}^{(72)}]$
10: $Data_{\text{concatenated}} \leftarrow fNIRS_{\text{concatenated}}, EEG_{\text{concatenated}}$
11: Pass through Eqs. (10) - (15)
12: Output classes

2048 filters. This step ensures compatibility with the outputs generated by the fNIRS models. The fNIRS model comprises flattened, convolution, pooling, and flattened layers, culminating in concatenation with the EEG model's outputs. The final stage encompasses two fully connected layers and a dropout layer, facilitating the transformation of outputs into the required classification categories. This intricate architecture harnesses the strengths of both the EEG and fNIRS models, thereby culminating in a comprehensive and powerful system for accurate EEG data classification. The architecture is shown in Fig. 3 and follows a series of mathematical annotations to reach the output classes. The pseudocode of the method is presented in Algorithm 1.

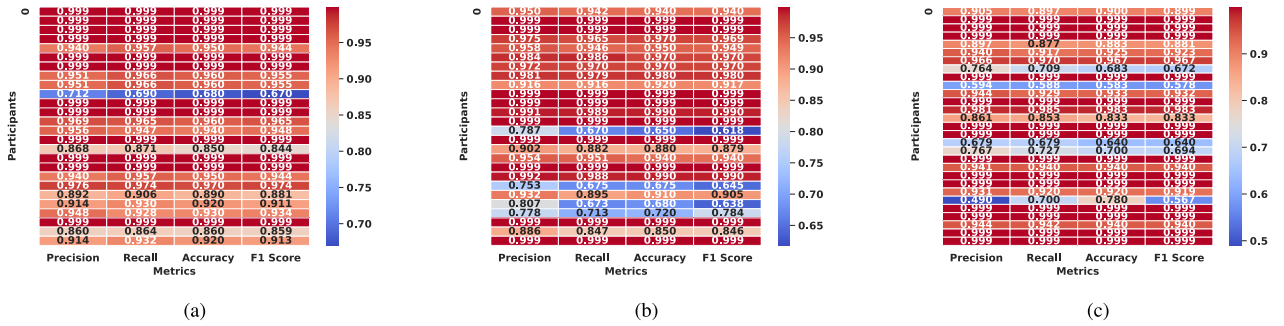$$Z_1 = X_{\text{fNIRS}} * W_1 + B_1 \tag{3}$$
$$A_1 = \text{Activation}(Z_1) \tag{4}$$
$$A_2 = \text{MaxPool}(A_1) \tag{5}$$
$$Z_3 = A_2 * W_2 + B_2 \tag{6}$$
$$A_3 = \text{Activation}(Z_3) \tag{7}$$

TABLE 4. Summary of data transformation and input shapes for EEG and fNIRS.

Task	Experimental period (s)			Sampling frequency (Hz)		Data processing		Model inputs	
	I	T	R	EEG	fNIRS	EEG	fNIRS	EEG	fNIRS
n-back	2	40	20	Raw=1000, Resampled=200	Raw=10.4, Resampled=10	STFT images (x-axis: 0-40 s, y-axis: 0-100 Hz)	Signal framing: window size = 2 s, overlap = 2.5 s, frames = 17	2D images (128 x 128 x 3) for the selected channels	Spectral entropy features calculated on windows, samples, channels, and frames
DSR									
WG	2	10	13-15			STFT images (x-axis: 0-10 s, y-axis: 0-100 Hz)	Signal framing: window size = 2 s, overlap = 1 s, frames = 11		
MI			15-17		Raw=12.5, Resampled=10				

FIGURE 4. Intra-subject performance metrics heatmap on the HOV approach for (a) n-back, (b) WG, and (c) MI tasks. The heatmaps illustrate the precision, recall, accuracy, and F1 score for each subject, providing a visual representation of the model's performance across different cognitive and motor tasks. These results highlight the model's ability to consistently classify EEG-fNIRS data with high accuracy for individual subjects.

$$A_4 = \text{MaxPool}(A_3) \tag{8}$$

$$A_5 = \text{Flatten}(A_4) \tag{9}$$

where X_{fNIRS} is the input data from the fNIRS sensor, W_1 represents the weights matrix for the first layer, B_1 is the bias vector for the first layer, Z_1 is the linear combination output of the first layer, A_1 is the activation output of the first layer, A_2 is the output after max pooling operation on A_1, W_2 represents the weights matrix for the second layer, B_2 is the bias vector for the second layer, Z_3 represents the linear combination output of the second layer, A_3 is the activation output of the second layer, A_4 is the output after max pooling operation on A_3, and A_5 is the flattened output.

We concatenated the outputs of the fNIRS sub-model (A_5) and the concatenated EEG output to obtain $A_{\text{Concatenated}}$. The rationale for concatenating EEG and fNIRS features at the middle stage of the model is based on several factors for ensuring that both modalities are integrated in an impactful manner. Performing concatenation following the 2D CNN layers facilitates the individual treatment of each modality by the model and consequently captures the modality-specific characteristics along temporal and spatial specificity before combining them. This yields a comprehensive feature set, thereby ensuring that key information from both modalities has been extracted.

In addition, both modalities share different frequency bands, sampling rates, and absolute amplitude levels. Therefore, their concatenation at an early stage, such as before the input of the model, could result in the introduction of bias towards the modality with higher amplitude values and a diverse spectrum. Conversely, if the modalities were combined after the classification stage, it could result in a loss of complementary information, which is the primary advantage of combining both modalities.

These data were then passed through fully connected layers as follows:

$$Z_{\text{Dense1}} = \text{FC}(A_{\text{Concatenated}}^{(1)}) \tag{10}$$

$$A_{\text{Dense1}} = \text{Activation}(Z_{\text{Dense1}}) \tag{11}$$

where $A_{\text{Concatenated}}$ is the concatenation of A_5 and the concatenated EEG output, Z_{Dense1} is the linear combination output of the first fully connected layer, and A_{Dense1} is the activation output of the first fully connected layer.

$$Z_{\text{Dense2}} = \text{FC}(A_{\text{Dense1}}) \tag{12}$$

$$A_{\text{Dense2}} = \text{Activation}(Z_{\text{Dense2}}) \tag{13}$$

where Z_{Dense2} is the linear combination output of the second fully connected layer and A_{Dense2} is the activation output of the second fully connected layer.

We applied dropout with a probability of 0.5 to A_{Dense2} to get A_{Dropout}.

$$Z_{\text{Output}} = \text{FC}(A_{\text{Dropout}}) \tag{14}$$

$$A_{\text{Output}} = \text{Activation}(Z_{\text{Output}}) \tag{15}$$

TABLE 5. Simulation parameters of the model.

Simulation Parameters	Values
Basemodel	DenseNet201
Epochs	50
Train test split	80:20
Validation data	30%
Batch size	08
Learning rate	1e-4
Optimizer	Adam
Loss function	Categorical cross entropy
Metrics	Accuracy

TABLE 6. Average and SD of intra-subject performance metrics for n-back, WG, and MI tasks using HOV and CV approaches.

Task		n-back		WG		MI	
Approach		HOV	CV	HOV	CV	HOV	CV
Accuracy	Average	0.951	0.943	0.922	0.931	0.917	0.907
	SD	0.0697	0.0656	0.1102	0.0755	0.1334	0.1317
Precision	Average	0.953	0.946	0.943	0.911	0.917	0.908
	SD	0.0642	0.0674	0.076	0.1092	0.1563	0.1177
Recall	Average	0.955	0.941	0.922	0.911	0.918	0.907
	SD	0.066	0.0711	0.1098	0.1096	0.135	0.1203
F1 score	Average	0.949	0.94	0.92	0.909	0.909	0.899
	SD	0.0722	0.0736	0.1162	0.1153	0.1566	0.136

where A_{Dropout} is the output after applying dropout to A_{Dense2}, Z_{Output} is the linear combination output of the output layer, and A_{Output} is the activation output of the output layer, representing the output classes (n-back, DSR, WG, or MI).

The simulation parameters that have been used as part of model training and testing outcomes are presented in Table 5.

The input to the model was presented in the form of 2D EEG images and processed fNIRS data. The RGB images, which represent time and frequency information, were standardized to a shape of $128 \times 128 \times 3$. In each of these images, the information along the time axis was limited to the duration of the individual tasks. The frequency axis was bounded by the Nyquist criterion, that is, 100 Hz (with a 200 Hz sampling frequency). For the fNIRS data, the 2D shape was determined by the segmentation into windows and the inclusion of channel information for each sample.

H. PERFORMANCE EVALUATION

We assessed the model performance using several metrics: accuracy, precision, recall, and the F1 score. These metrics, calculated from a confusion matrix and particularly relevant for BCI evaluation, offered a comprehensive picture of the model's efficacy in classifying cognitive and MI tasks.

IV. RESULTS AND DISCUSSION

This section presents the experimental results, a comparison of our proposed model to prior studies, and a discussion of its clinical translational implications and future directions.

A. EXPERIMENTAL RESULTS

The outcomes were generated using the hold-out validation (HOV) and cross-validation (CV) approaches. The prescribed methodology was implemented in Google Colaboratory, using the Python programming language. Performance assessment metrics for the proposed approach, which considered the integration of EEG-fNIRS data, were evaluated for each participant across all three cognitive and MI tasks. The intra-subject performance metrics results for the n-back, WG, and MI tasks on the HOV approach were detailed in heatmap format, as illustrated in Figs. 4(a), 4(b), and 4(c), respectively. The precision, recall, accuracy, and F1 score, computed across all 26 participants in Dataset 1 and 29 participants in Dataset 2, are presented as comprehensive performance metrics.

In addition, the effectiveness of our channel selection approach was validated through experimental results. Consequently, using the selected 10 channels, the model achieved high classification accuracy. This high performance indicated that the chosen channels effectively captured the necessary features induced by the tasks, justifying the exclusion of the remaining channels. Thus, these results demonstrate that the strategic selection of EEG channels based on neuroanatomical relevance was appropriate and sufficient.

To facilitate better comparison with benchmarking studies, CV has been employed using the k-fold approach. The data were passed through 5 folds, and their average performance metrics were analyzed. It was found that the performance metrics of the model using CV were on par with the HOV case, which exhibited only minor variations in performance metrics. These outcomes are depicted in Table 6.

Upon examining Fig. 4, the average and SD of intra-subject performance metrics for the n-back, WG, and MI tasks were calculated, as presented in Table 6. Notably, for the DSR task, all performance metrics reached 0.999±0.0001 and 0.996±0.0013 for every subject in the HOV and CV approaches, respectively. The model achieved high accuracy while ensuring that no instances of overfitting occurred. This was confirmed by passing the model through two validation schemes. In addition, the performance on the test (unseen) dataset was on par with the training and validation data, further affirming the model's normal convergence. These metrics served as comprehensive indicators of the model's performance in classifying EEG-fNIRS data. The average performance metric across both datasets has been further presented as the bar graph in Fig. 5(a). These results demonstrate the model's potential for classifying motor-related brain activity.

Notably, certain deviations were observed in the results. Specifically, for four subjects in the WG task, one subject in the n-back task, and five subjects in the MI task, as depicted in Fig. 4, lower accuracy values and other performance metrics were recorded. This could be attributed to the inherent characteristics of the data collected from these particular participants, which may lack crucial distinguishing features in the time-frequency domain. Moreover, the classification rates for the WG and MI tasks were lower compared to the n-back and DSR tasks. There could be several reasons for this. First, the WG task involved complex cognitive processes, including language production and semantic retrieval. These
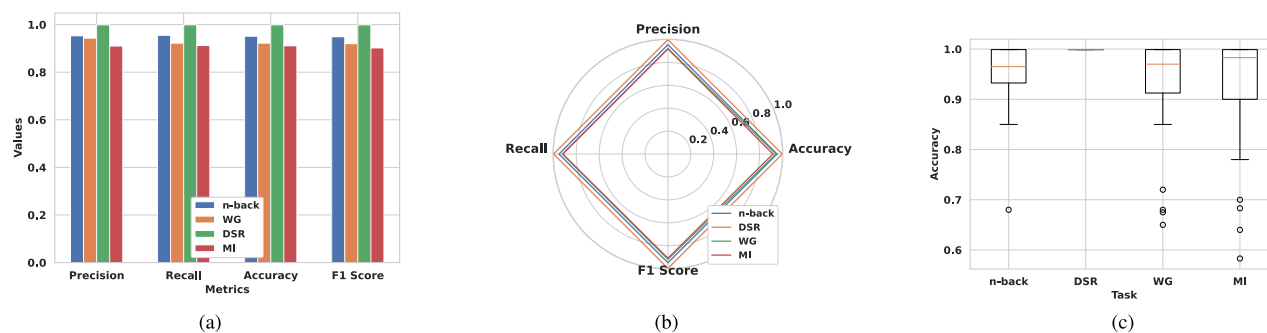
FIGURE 5. Additional results on the HOV approach. (a) Average intra-subject performance metrics (accuracy, precision, recall, and F1 score) for both datasets. (b) Radar chart comparing performance metrics across all tasks. (c) Box plots showing the distribution of accuracy for each task.

processes could result in more variable and less distinctive neural patterns, rendering the modelling challenging. Similarly, the MI tasks involve complex neural activations that can be subtle and harder to distinguish. Such variability can result in relatively lower classification outcomes. In the future, this could be overcome through the development of advanced models with improved handling of attention mechanisms.

Nevertheless, the model consistently demonstrates a high degree of accuracy across a majority of cases, signifying the robustness of the adopted approach. Thus, the proposed model outperformed existing architectures and methods in terms of accuracy and showcased its ability to consistently achieve optimal outcomes for a range of classification tasks involving EEG-fNIRS data processing. The comprehensive evaluation of accuracy, precision, recall, and F1 score provides a holistic view of the model's effectiveness, thus underscoring its potential for advancing our understanding of neural processes in cognitive and MI tasks.

To gain a holistic perspective on the model's performance across different cognitive and MI tasks, we employed radar plots to visualize the average accuracy, precision, recall, and F1 score for both datasets as depicted in Fig. 5(b). Remarkably, the radar plot illustrated that the DSR task exhibited the highest average performance across all metrics, followed by the n-back, WG, and MI tasks. This ordering aligned with the complexity of these tasks, with DSR necessitating a higher level of cognitive engagement. The radar plot provides an informative summary of the model's capabilities, offering a concise comparison of its performance.

To delve deeper into the accuracy distribution across both datasets, we constructed box plots. These plots exclusively focused on accuracy as a metric and revealed the variability in classification performance within each group. Although the DSR task consistently attained exceptional accuracy, the box plots for the n-back, WG, and MI tasks indicated variations in the performance among participants as shown in Fig. 5(c). Interestingly, these variations can be attributed to individual differences in task-related brain activity. Thus, the box plots confirmed the model's robustness and highlighted the potential for tailoring BCI systems to individual users based on their unique neural responses. This underscores

the adaptability and personalization capabilities of our approach.

Further analysis of the task-dependent performance indicated that our model exhibited exceptional accuracy in the DSR task, a notable achievement compared to its performance in the n-back and WG tasks. This distinction could stem from the DSR task's relative simplicity, where participants are required to discriminate between basic visual stimuli. Such simpler tasks may elicit more distinct and classifiable neural patterns in both EEG and fNIRS data. Conversely, the n-back and WG tasks, which engage higher-level cognitive functions, may produce more nuanced and complex brain activity, thereby presenting a greater challenge for accurate classification.

The efficacy of STFT and the MDNF architecture is central to the success of the proposed model in EEG-fNIRS-based BCI tasks. The STFT transformation of EEG data into spectrograms revealed the critical interplay between time and frequency dimensions, which is essential for recognizing patterns linked to various cognitive states. Meanwhile, the MDNF architecture, leveraging a DenseNet201 backbone, adeptly learns the intricate relationships present in this multimodal data. The capacity of the model to merge spatial information from fNIRS with the detailed temporal dynamics captured by STFT-based EEG features is a significant factor behind its high classification accuracy.

B. MODEL'S COMPUTATIONAL COMPLEXITY

The proposed model balanced between accuracy and computational efficiency. While STFT image generation (60-80 min in each subject on a CPU) and training (43-50 min in each subject on a T4 GPU) were somewhat time-consuming, testing was highly efficient (0.5 min in each subject on a T4 GPU). This computational profile is well-suited for offline analysis. However, real-time BCI applications may require further optimization. The exploration of more compact versions of the MDNF model could provide a solution, minimizing computational demands without sacrificing performance. In addition, the leveraging of specialized neural network hardware could significantly accelerate processing
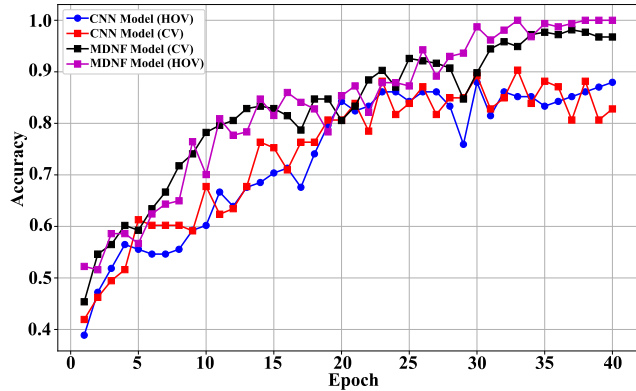
FIGURE 6. Performance comparison between the MDNF model and a simplified CNN architecture for the n-back task on Subject 1. The evaluation involves results from HOV and CV approaches, demonstrating the superior accuracy of the MDNF model.

speeds, rendering our model a viable option for real-time applications wherein swift response times are essential.

C. OPTIMIZATION AND GENERALIZATION

To mitigate the risk of overfitting, several measures were taken. The foremost measure was the reduction of the number of EEG channels based on anatomical locations, resulting in a 10-channel representation instead of 30 channels. The fNIRS data were transformed into a reduced-dimensional representation via the extraction of a single value of spectral entropy from each frame (window) of data. Pooling layers were introduced in the model architecture to discard any redundant information that could result in overfitting. The overall data was divided into training, validation, and test sets. Ensuring that the performance of the validation and test datasets was on par with the training dataset confirmed that the model did not experience overfitting. However, although the performance of the model on individual subjects produced promising results, its generalization across multiple subjects remains a question and will be explored in future work.

D. MODEL BENCHMARKING AND ABLATION EXPERIMENTS

To further analyze the effectiveness of the model, the performance of the model was analyzed by replacing the DenseNet base models with simplified CNN layers to assess the impact of transfer learning in the current MDNF model. The outcomes of both models under CV and HOV are presented in Fig. 6. The performance outcomes are depicted for Subject 1 under the n-back classes. The MDNF model achieved an accuracy of over 97% for both HOV and CV. By excluding the DenseNet layers and using a simplified CNN model, the accuracy was reduced to less than 90% for both HOV and CV cases.

To ensure the prevention of overfitting in the established models, the performance was evaluated using both HOV and CV. In addition, the performance on the test dataset (unseen data) was compared to the training and validation

TABLE 7. Comparison with prior studies using the same datasets.

Proposed method	Task	Accuracy
SVM [7]	n-back	83%
CNN-LSTM [15]	n-back	88.41%
DNN [13]	n-back	87%
	WG	92%
	DSR	91%
	MI	91%
k-NN with FWHTC [11]	MI	78.21%
Multimodal multitask neural network [46]	MI	82.11%
Y-shaped neural network [47]	MI	76.21%
*p*th-order polynomial fusion [48]	MI	77.53%
STFT-MDNF	n-back	95.1% (HOV)
	WG	93.1% (CV)
	DSR	99.9% (HOV)
	MI	91.7% (HOV)

performance and found to be consistent. As evident in Fig. 6, the training curves converged smoothly without any indication of divergence. These outcomes confirmed that the performance of the model does not exhibit any signs of overfitting.

E. COMPARISON WITH PRIOR STUDIES

Our model demonstrated significant progress in classifying cognitive and MI tasks using EEG-fNIRS data. In the n-back, WG, and DSR tasks, it achieved accuracies of 95.1% (HOV), 93.1% (CV), and 99.9% (HOV) respectively. It consistently outperformed previous studies across the n-back, WG, and DSR tasks using the same dataset [7], [13], [15]. Similarly, the MI task accuracy was 91.7% (HOV), which exceeded other studies using the same dataset [11], [13], [46], [47], [48]. A comparison with prior studies is presented in Table 7. This enhancement in performance is attributed to the hybrid data advantage, where the integration of EEG and fNIRS data offers a more comprehensive spatio-temporal perspective than the use of either modality alone. The informative dimensions derived from STFT-based spectrograms and spectral entropy effectively discerned the subtle patterns critical for distinguishing between cognitive states. In addition, the excellent ability of the MDNF architecture to navigate the complex interplay within this multimodal data significantly contributed to our model's success.

F. CLINICAL TRANSLATIONAL IMPLICATIONS OF MDFN

The impressive performance of the proposed model in cognitive and MI task classification using EEG-fNIRS data holds significant promise for clinical translation. By accurately discriminating between cognitive states, the model offers valuable insights into the neural mechanisms underlying attention, memory, decision-making, and other complex brain functions [49]. This could advance our fundamental understanding of cognition by facilitating researchers in linking specific task performance metrics (e.g., faster reaction times on the n-back task) to the model's identified neural patterns. Moreover, this discriminatory power suggests potential applications in the diagnosis and management of neurological disorders. Through the detection of subtle alterations in brain patterns associated with cognitive and motor functions, the

model could enable the early identification of conditions such as Parkinson's disease, stroke, or dementia [50]. This early detection could result in timely interventions, potentially slowing disease progression and improving patient outcomes.

The adaptability of the model to limited EEG data, coupled with its accuracy, opens up possibilities for real-time brain monitoring using wearable devices, paving the way for noninvasive BCIs that could transform healthcare with early interventions, continuous monitoring in real-world settings, and assistive technologies for individuals with disabilities or those undergoing neurorehabilitation. In addition, the capacity of the MDNF model to identify specific neural patterns linked to cognitive or motor deficits could inform personalized rehabilitation strategies, thereby optimizing the therapy outcomes by targeting the most affected neural circuits.

Future research should focus on the validation of the model's generalizability across diverse datasets and experimental conditions. The use of interpretability techniques would provide deeper insights into the model's decision-making processes and potentially reveal new biomarkers for cognitive disorders [51]. Furthermore, exploring calibration strategies or adaptive algorithms could tailor the model to individual users' brain patterns, which could enhance its accuracy and broaden applications into personalized medicine or neurorehabilitation. Consequently, rigorous validation in cross-subject classification studies is necessary to ensure the model's generalizability across diverse patient populations, which is a crucial step towards the integration of the MDFN model into routine clinical practice.

V. CONCLUSION

The evaluation of our proposed model demonstrated a significant advancement in EEG-fNIRS data processing and classification. Its superior accuracy, precision, recall, and F1 scores across tasks such as n-back, WG, DSR, and MI exhibited its effectiveness in capturing complex time-frequency patterns within neural data. Surpassing the performance of existing approaches, the model offers exceptional potential for the accurate decoding of cognitive and MI tasks. This study has far-reaching implications for advancing cognitive neuroscience and enabling new clinical diagnostics, personalized treatment strategies, and future neurocognitive research. With the continued evolution of the field, the model presents exciting opportunities for diverse applications. Thus, continued refinement of the architecture, exploration of its generalizability, and increased focus on interpretability will undoubtedly result in even greater discoveries regarding the human brain and cognitive processes. This research signifies a major step forward in our quest to understand the intricacies of the mind.

REFERENCES

[1] M. F. Mridha, S. C. Das, M. M. Kabir, A. A. Lima, M. R. Islam, and Y. Watanobe, "Brain–computer interface: Advancement and challenges," *Sensors*, vol. 21, no. 17, p. 5746, Aug. 2021.

[2] M. Abtahi et al., "Merging fNIRS-EEG brain monitoring and body motion capture to distinguish Parkinsons disease," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 6, pp. 1246–1253, Jun. 2020.

[3] P. Pinti et al., "The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience," *Ann. New York Acad. Sci.*, vol. 1464, no. 1, pp. 5–29, 2020.

[4] R. Li, D. Yang, F. Fang, K.-S. Hong, A. L. Reiss, and Y. Zhang, "Concurrent fNIRS and EEG for brain function investigation: A systematic, methodology-focused review," *Sensors*, vol. 22, no. 15, p. 5865, Aug. 2022.

[5] Z. He et al., "Advances in multimodal emotion recognition based on brain–computer interfaces," *Brain Sci.*, vol. 10, no. 10, p. 687, Sep. 2020.

[6] A. P. Buccino, H. O. Keles, and A. Omurtag, "Hybrid EEG-fNIRS asynchronous brain–computer interface for multiple motor tasks," *PLoS ONE*, vol. 11, no. 1, Jan. 2016, Art. no. e0146610.

[7] J. Cao, E. M. Garro, and Y. Zhao, "EEG/fNIRS based workload classification using functional brain connectivity and machine learning," *Sensors*, vol. 22, no. 19, p. 7623, Oct. 2022.

[8] P. A. Cicalese et al., "An EEG-fNIRS hybridization technique in the four-class classification of Alzheimer's disease," *J. Neurosci. Methods*, vol. 336, Apr. 2020, Art. no. 108618.

[9] C. Bunterngchit, J. Wang, T. Chearanai, and Z.-G. Hou, "Enhanced EEG-fNIRS classification through concatenated convolutional neural network with band analysis," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2023, pp. 1–6.

[10] S. Kumar, A. Sharma, and T. Tsunoda, "Brain wave classification using long short-term memory network based OPTICAL predictor," *Sci. Rep.*, vol. 9, no. 1, p. 9153, Jun. 2019.

[11] E. Ergün and Ö. Aydemir, "A hybrid BCI using singular value decomposition values of the fast Walsh–Hadamard transform coefficients," *IEEE Trans. Cogn. Develop. Syst.*, vol. 15, no. 2, pp. 454–463, Jun. 2023.

[12] A. M. Chiarelli, P. Croce, A. Merla, and F. Zappasodi, "Deep learning for hybrid EEG-fNIRS brain–computer interface: Application to motor imagery classification," *J. Neural Eng.*, vol. 15, no. 3, Jun. 2018, Art. no. 036028.

[13] M. Saadati, J. Nelson, and H. Ayaz, "Multimodal fNIRS-EEG classification using deep learning algorithms for brain–computer interfaces purposes," in *Advances in Neuroergonomics and Cognitive Engineering*. Cham, Switzerland: Springer, 2020, pp. 209–220. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-20473-0_21

[14] M. Saadati, J. Nelson, and H. Ayaz, "Convolutional neural network for hybrid fNIRS-EEG mental workload classification," in *Advances in Neuroergonomics and Cognitive Engineering* (Advances in Intelligent Systems and Computing). Cham, Switzerland: Springer, 2020, pp. 221–232. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-20473-0_22

[15] N. E. Mughal et al., "EEG-fNIRS-based hybrid image construction and classification using CNN-LSTM," *Frontiers Neurorobotics*, vol. 16, Aug. 2022, Art. no. 873239.

[16] Y. Wang, J. Wang, W. Wang, J. Su, and Z.-G. Hou, "Calibration-free transfer learning for EEG-based cross-subject motor imagery classification," in *Proc. IEEE 19th Int. Conf. Autom. Sci. Eng. (CASE)*, Aug. 2023, pp. 1–6.

[17] C. Wang, L. Liu, W. Zhuo, and Y. Xie, "An epileptic EEG detection method based on data augmentation and lightweight neural network," *IEEE J. Transl. Eng. Health Med.*, vol. 12, pp. 22–31, 2024.

[18] W. Li, C. Fang, Z. Zhu, C. Chen, and A. Song, "Fractal spiking neural network scheme for EEG-based emotion recognition," *IEEE J. Transl. Eng. Health Med.*, vol. 12, pp. 106–118, 2024.

[19] M. A. Rahman, A. Anjum, M. M. H. Milu, F. Khanam, M. S. Uddin, and M. N. Mollah, "Emotion recognition from EEG-based relative power spectral topography using convolutional neural network," *Array*, vol. 11, Sep. 2021, Art. no. 100072.

[20] J. Wang, W. Wang, S. Ren, W. Shi, and Z.-G. Hou, "Neural correlates of single-task versus cognitive-motor dual-task training," *IEEE Trans. Cogn. Develop. Syst.*, vol. 14, no. 2, pp. 532–540, Jun. 2022.

[21] K. Fu, J. Qu, J. Chai, and Y. Dong, "Classification of seizure based on the time-frequency image of EEG signals using HHT and SVM," *Biomed. Signal Process. Control*, vol. 13, pp. 15–22, Sep. 2014.

[22] L. P. A. Arts and E. L. van den Broek, "The fast continuous wavelet transformation (fCWT) for real-time, high-quality, noise-resistant time–frequency analysis," *Nature Comput. Sci.*, vol. 2, no. 1, pp. 47–58, Jan. 2022.

[23] S. Khatun, R. Mahajan, and B. I. Morshed, "Comparative study of wavelet-based unsupervised ocular artifact removal techniques for single-channel EEG data," *IEEE J. Transl. Eng. Health Med.*, vol. 4, pp. 1–8, 2016.

[24] M.-P. Hosseini, A. Hosseini, and K. Ahi, "A review on machine learning for EEG signal processing in bioengineering," *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 204–218, 2021.

[25] A. Vahid, M. Mückschel, S. Stober, A.-K. Stock, and C. Beste, "Applying deep learning to single-trial EEG data provides evidence for complementary theories on action control," *Commun. Biol.*, vol. 3, no. 1, p. 112, Mar. 2020.

[26] G. Xu et al., "A deep transfer convolutional neural network framework for EEG signal classification," *IEEE Access*, vol. 7, pp. 112767–112776, 2019.

[27] N. Pusarla, A. Singh, and S. Tripathi, "Learning DenseNet features from EEG based spectrograms for subject independent emotion recognition," *Biomed. Signal Process. Control*, vol. 74, Apr. 2022, Art. no. 103485.

[28] X. Qiu, F. Yan, and H. Liu, "A difference attention ResNet-LSTM network for epileptic seizure detection using EEG signal," *Biomed. Signal Process. Control*, vol. 83, May 2023, Art. no. 104652.

[29] H. Hamid, N. Naseer, H. Nazeer, M. J. Khan, R. A. Khan, and U. S. Khan, "Analyzing classification performance of fNIRS-BCI for gait rehabilitation using deep neural networks," *Sensors*, vol. 22, no. 5, p. 1932, Mar. 2022.

[30] H. Nazeer et al., "Enhancing classification accuracy of fNIRS-BCI using features acquired from vector-based phase analysis," *J. Neural Eng.*, vol. 17, no. 5, Oct. 2020, Art. no. 056025.

[31] A. Gulraiz, N. Naseer, H. Nazeer, M. J. Khan, R. A. Khan, and U. S. Khan, "LASSO homotopy-based sparse representation classification for fNIRS-BCI," *Sensors*, vol. 22, no. 7, p. 2575, Mar. 2022.

[32] T. Ma et al., "CNN-based classification of fNIRS signals in motor imagery BCI system," *J. Neural Eng.*, vol. 18, no. 5, Oct. 2021, Art. no. 056019.

[33] A. von Lühmann, A. Ortega-Martinez, D. A. Boas, and M. A. Yücel, "Using the general linear model to improve performance in fNIRS single trial analysis and classification: A perspective," *Frontiers Human Neurosci.*, vol. 14, p. 30, Feb. 2020.

[34] S. B. Erdoğan, E. Özsarfati, B. Dilek, K. S. Kadak, L. Hanoĝlu, and A. Akın, "Classification of motor imagery and execution signals with population-level feature sets: Implications for probe design in fNIRS based BCI," *J. Neural Eng.*, vol. 16, no. 2, Apr. 2019, Art. no. 026029.

[35] A. Janani, M. Sasikala, H. Chhabra, N. Shajil, and G. Venkatasubramanian, "Investigation of deep convolutional neural network for classification of motor imagery fNIRS signals for BCI applications," *Biomed. Signal Process. Control*, vol. 62, Sep. 2020, Art. no. 102133.

[36] X. Lu, A. Wen, L. Sun, H. Wang, Y. Guo, and Y. Ren, "An epileptic seizure prediction method based on CBAM-3D CNN-LSTM model," *IEEE J. Transl. Eng. Health Med.*, vol. 11, pp. 417–423, 2023.

[37] M. Rashed-Al-Mahfuz, M. A. Moni, S. Uddin, S. A. Alyami, M. A. Summers, and V. Eapen, "A deep convolutional neural network method to detect seizures and characteristic frequencies using epileptic electroencephalogram (EEG) data," *IEEE J. Transl. Eng. Health Med.*, vol. 9, pp. 1–12, 2021.

[38] T. H. Shovon, Z. A. Nazi, S. Dash, and M. F. Hossain, "Classification of motor imagery EEG signals with multi-input convolutional neural network by augmenting STFT," in *Proc. 5th Int. Conf. Adv. Elect. Eng. (ICAEE)*, 2019, pp. 398–403.

[39] O. Ali, M. Saif-ur-Rehman, S. Dyck, T. Glasmachers, I. Iossifidis, and C. Klaes, "Enhancing the decoding accuracy of EEG signals by the introduction of anchored-STFT and adversarial data augmentation method," *Sci. Rep.*, vol. 12, no. 1, p. 4245, Mar. 2022.

[40] Z. Wan, R. Yang, M. Huang, N. Zeng, and X. Liu, "A review on transfer learning in EEG signal analysis," *Neurocomputing*, vol. 421, pp. 1–14, Jan. 2021.

[41] J. Shin, A. von Lühmann, D.-W. Kim, J. Mehnert, H.-J. Hwang, and K.-R. Müller, "Simultaneous acquisition of EEG and NIRS during cognitive tasks for an open access dataset," *Sci. Data*, vol. 5, no. 1, pp. 1–16, Feb. 2018.

[42] J. Shin et al., "Open access dataset for EEG+NIRS single-trial classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 10, pp. 1735–1745, Oct. 2017.

[43] E. Guttmann-Flury, X. Sheng, and X. Zhu, "Channel selection from source localization: A review of four EEG-based brain–computer interfaces paradigms," *Behav. Res. Methods*, vol. 55, no. 4, pp. 1980–2003, Jul. 2022.

[44] P. Pinti, F. Scholkmann, A. Hamilton, P. Burgess, and I. Tachtsidis, "Current status and issues regarding pre-processing of fNIRS neuroimaging data: An investigation of diverse signal filtering methods within a general linear model framework," *Frontiers Hum. Neurosci.*, vol. 12, p. 505, Jan. 2019.

[45] Y. Huang, J. Yang, P. Liao, and J. Pan, "Fusion of facial expressions and EEG for multimodal emotion recognition," *Comput. Intell. Neurosci.*, vol. 2017, pp. 1–8, Sep. 2017.

[46] Q. He, L. Feng, G. Jiang, and P. Xie, "Multimodal multitask neural network for motor imagery classification with EEG and fNIRS signals," *IEEE Sensors J.*, vol. 22, no. 21, pp. 20695–20706, Nov. 2022.

[47] Y. Li, X. Zhang, and D. Ming, "Early-stage fusion of EEG and fNIRS improves classification of motor imagery," *Frontiers Neurosci.*, vol. 16, Jan. 2023, Art. no. 1062889.

[48] Z. Sun, Z. Huang, F. Duan, and Y. Liu, "A novel multimodal approach for hybrid brain–computer interface," *IEEE Access*, vol. 8, pp. 89909–89918, 2020.

[49] M. K. Yeung and V. W. Chu, "Viewing neurovascular coupling through the lens of combined EEG–fNIRS: A systematic review of current methods," *Psychophysiology*, vol. 59, no. 6, p. 14054, Jun. 2022.

[50] Z. Liu, J. Shore, M. Wang, F. Yuan, A. Buss, and X. Zhao, "A systematic review on hybrid EEG/fNIRS in brain–computer interface," *Biomed. Signal Process. Control*, vol. 68, Jul. 2021, Art. no. 102595.

[51] D. Rosenbaum et al., "Neuronal correlates of spider phobia in a combined fNIRS-EEG study," *Sci. Rep.*, vol. 10, no. 1, p. 12597, Jul. 2020.

• • •