IEEE Open Journal of the Computer Society

Received 9 August 2024; revised 24 October 2024; accepted 10 November 2024. Date of publication 18 November 2024; date of current version 25 November 2024. The review of this article was arranged by Associate Editor Jianzong Wang.

Digital Object Identifier 10.1109/OJCS.2024.3500032

Multimodal EEG-fNIRS Seizure Pattern Decoding Using Vision Transformer

RAFAT DAMSEH [1] (Member, IEEE), ABDELHADI HIRECHE [2], PARIKSHAT SIRPAL [3] (Member, IEEE), AND ABDELKADER NASREDDINE BELKACEM [2] (Senior Member, IEEE)

[1]Department of Computer Science and Software Engineering, United Arab Emirates University, Al Ain Abu Dhabi, United Arab Emirates
[2]Department of Computer and Network Engineering, United Arab Emirates University, Al Ain Abu Dhabi, United Arab Emirates
[3]School of Electrical and Computer Engineering, Gallogly College of Engineering, University of Oklahoma, Norman, OK 73019 USA

CORRESPONDING AUTHOR: ABDELKADER NASREDDINE BELKACEM (e-mail: belkacem@uaeu.ac.ae).

This work was supported by United Arab Emirates University under Grant UPAR-12T055.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Research Ethics Board (REB) of Sainte-Justine Hospital and Centre Hospitalier de l'Université de Montréal, under Application No. 387183, and performed in line with the Declaration of Helsinki.

This article has supplementary downloadable material available at https://doi.org/10.1109/OJCS.2024.3500032, provided by the authors.

ABSTRACT Epilepsy has been analyzed through uni-modality non-invasive brain measurements such as electroencephalogram (EEG) signal, but identifying seizure patterns is more challenging due to the non-stationary nature of the brain activity and various non-brain artifacts. In this article, we leverage a vision transformer model (ViT) to classify three types of seizure patterns based on multimodal EEG and functional near-infrared spectroscopy (fNIRS) recordings. We used spectral encoding techniques to capture temporal and spatial relationships for brain signals as feature map inputs to the transformer architecture. We evaluated model performance using the receiver operating characteristic (ROC) curves and the area under the curve (AUC), demonstrating that multimodal EEG-fNIRS signals improved the classification accuracy of seizure patterns. Our work showed that power spectral density (PSD) features often led to better results than features derived from dynamic mode decomposition (DMD), particularly for seizures with high-frequency oscillations (HFO) and generalized spike-and-wave discharge (GSWD) patterns, with an accuracy of 93.14% and 91.69%, respectively. Low-voltage fast activity (LVFA) seizures achieved consistently high performance in EEG, fNIRS, and multimodal EEG-fNIRS setups. Overall, our findings suggest the effectiveness of using the ViT architecture with multimodal brain data accompanied by appropriate spectral features to classify the neural activity of epileptic seizure patterns.

INDEX TERMS Electroencephalogram (EEG), functional near-infrared spectroscopy (fNIRS), low-voltage fast activity (LVFA), generalized spike-and-wave discharge (GSWD), high-frequency oscillation (HFO), epilepsy, seizure, vision transformer.

I. INTRODUCTION

Epilepsy is a neurological condition causing recurrent epileptic seizures characterized by abnormal electrical activity in the brain [1]. Beyond its physical symptoms, epilepsy has clinical consequences that impact mental and emotional health as well as the general quality of life [2]. Electroencephalogram (EEG) is an non-invasive brain measurement for analyzing and diagnosing the neurodegenerative and psychiatric diseases [3], [4], [5], including epilepsy, where it can help in studying and monitoring epileptic activity. Specifically, EEG analysis is used to identify seizure onset zones, categorize seizure types, and assess the efficacy of the treatment [6]. Examples of EEG data analysis techniques are spike detection, frequency analysis, and time-frequency analysis. These techniques contribute to our understanding of epileptic patterns. Functional near-infrared spectroscopy (fNIRS) is another acquisition method for improving epilepsy analysis and detection [7], [8], [8], [9]. fNIRS measures hemodynamic

brain responses by monitoring changes in oxyhemoglobin (HbO) and deoxyhemoglobin (HbR). fNIRS-based measurements offer a non-invasive tracking of brain activity. Studies have demonstrated that fNIRS-based seizure prediction can outperform EEG, achieving high positive predictive values and accuracy [9], [10]. Moreover, fNIRS allows for continuous and long-term monitoring that is helpful for patients with refractory epilepsy [11]. The emerging combination of both EEG with fNIRS data can further enhance our ability to understand the neurological foundations of epilepsy [12].

There are several patterns of epileptic seizures [13], [14], [15], [16]. EEG patterns characteristic of generalized epilepsy disorders are known as generalized spike-and-wave discharges (GSWDs). They usually have a frequency between 3 and 4 Hz and are composed of a strong spike followed by a gradual wave. An example of generalized seizures that can be detected and diagnosed by GSWDs are absence seizures, which exhibit widespread and synchronized neuronal activity in both hemispheres of the brain. Also, GSWDs help distinguish between generalized and focal epilepsy. This helps guide appropriate treatment strategies and understand the neurophysiological mechanisms of these disorders [13], [17], [18], [19]. High-frequency oscillations (HFOs) in the frequency range of 80–500 Hz are a relatively recent discovery in EEG analysis and are believed to represent localized epileptic activity [15]. HFOs are considered biomarkers of epileptogenic tissue, and their presence can help identify the brain regions responsible for seizure onset. Low-voltage fast activity (LVFA) is typically observed at seizure onset, consisting of rapid, low-amplitude oscillations, and is often considered a hallmark of the transition from the interictal (between seizures) to the ictal (during seizure) phase [16], [20]. Recognizing the LVFA can be useful in pinpointing the seizure onset zone, especially during presurgical evaluations.

Accurate identification and interpretation of seizure patterns are essential for improved diagnosis and treatment planning [21]. Compared to classical machine learning algorithms, deep learning (DL) techniques, e.g., that using convolutional neural networks (CNNs) and transformers [22], [23], [24], excel at automatically extracting useful features from time series data without the need for manual feature engineering [25]. Hence, by processing high-dimensional and combined EEG-fNIRS datasets, such DL algorithms have the potential to recognize intricate seizure patterns even when dealing with nonstationary or noisy data. For example, CNNs have proven to be effective for identifying and classifying EEG signals in [26]. Another deep learning architecture based on recurrent neural networks (RNNs) has been proposed for epileptic seizure detection in [10]. Studies have also explored RNNs to reveal hidden relationships between EEG readings and fNIRS responses [27]. Recent AI models, i.e., language- and vision-based transformers [28], [29], have been explored to improve automated EEG brain data analysis. Limited work has explored their efficiency in processing combined EEG-fNIRS data. Here, inspired by the success of deep transformers in language and vision tasks, we aim to develop a vision-based transformer for detecting various seizure patterns, namely, GSWD, HFO, and LFVA, from synchronized EEG-fNIRS recordings of epilepsy patients.

The main contributions of our work are summarized as follows: 1) We utilized multimodal EEG-fNIRS signals with an established ViT model to detect HFO, LVFA and GSWD epilepsy patterns, 2) we have demonstrated the performance of the ViT model with the integration of spectral analysis techniques (PSD and dynamic mode decomposition, DMD), to decode temporal and spatial relationships in brain signals, and 3) we have shown that combining EEG and fNIRS data with appropriate spectral features significantly improves classification accuracy in seizure pattern recognition. This demonstrates how multimodal data fusion combined with spectral analysis might improve the performance of deep learning models for the investigation of such neurological disorders.

II. RELATED WORK

Previously, other DL architectures have been used to automate EEG data analysis. In [30], the authors presented a graph convolutional neural network (GCN) for decoding time-resolved EEG motor imagery signals. An attention-based bidirectional long short-term memory in conjunction with GCN was proposed for the classification of motor imagery based on EEG [31]. Another study showed a source-level technique that employs CNN to decode four-class motor imagery tasks by integrating spatial, temporal, and frequency data, improving classification accuracy in BCI applications [32].

Few studies have focused on the use of transformer models in fNIRS data. One study addressed the limitations in fNIRS classification by proposing a new framework using a transformer model [33]. Wang et al., proposed an fNIRS-based brain computer interface using a transformer network that uses spatial and temporal information in the associated signals [34]. They achieved high accuracy in movement classification tasks. The ViT model proposed in [29] to process visual data samples was also applied for EEG and fNIRS processing. Such integration interprets EEG and fNIRS signals as visual embedding. One study explored a spectral vision transformer (ViT) for REM sleep behaviour disorder (RBD) and suggested that EEG channels are promising for RBD diagnosis [35]. In [36], a new sleep-stage classification method was proposed using ViT architecture to analyze multichannel EEG sleep data. In [37], the work demonstrates that a ViT model trained on rapid-response EEG data achieved high accuracy in predicting delirium in critically ill patients.

To the best of our knowledge, ViTs have not been investigated in combination with EEG-fNIRS analysis [38]. When dealing with the multimodal nature of EEG and fNIRS data, which could include DMD and power spectral density (PSD) features, the ViT's architecture would be suitable for effectively managing and integrating these diverse data representations. Specifically, ViTs have shown superior performance in handling long-range dependencies and complex

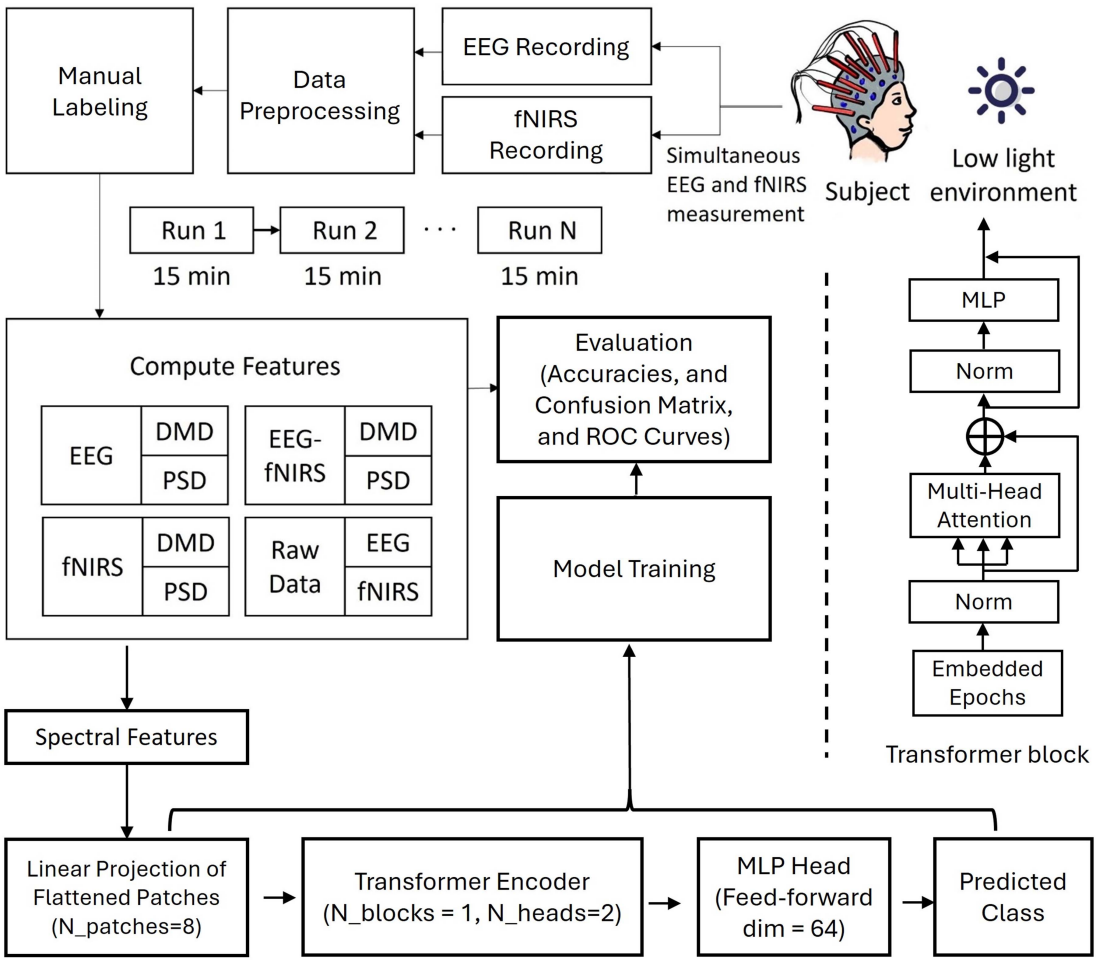

FIGURE 1. **Simultaneous EEG and fNIRS measurements are used to train a vision transformer (ViT) model for seizure pattern classification. Data was manually labelled after being collected in a low-light environment and preprocessed to remove noise. For model training, features are extracted from EEG, fNIRS, and combined EEG-fNIRS signals using raw inputs and spectral encoding techniques. Our classification model is evaluated using accuracy, confusion matrices, and ROC curves.**

hierarchical representations [38], making them well-suited for the intricacies of seizure pattern classification. This study investigated the effectiveness of a customized configuration of ViTs for novel seizure pattern classification, employing concurrent EEG-fNIRS recordings captured longitudinally from epileptic patients.

III. METHODS

A. DATASET COLLECTION AND PREPARATION

The data used in our study have been used in previous studies [10], [27], and were collected from 40 participants (27 males, 13 females) aged between 11 and 62 years (mean age: 32.42 years, standard deviation (SD): 13.97 years) with refractory focal epilepsy who were enrolled for extended EEG-fNIRS recordings. Ethical approval for data collection was obtained from the institutional review boards of Sainte-Justine Hospital and the Center Hospitalier de l'Universite de Montreal. Continuous EEG-fNIRS recordings were also performed. EEG data were recorded at 500 Hz adhering to the standard 10–20 system, utilizing 19 electrodes (Fp1, Fp2, F7, F3, Fz, F4, F8, T7, C3, Cz, C4, T8, P7, P3, Pz, P4, P8, O1, O2). The EEG data were filtered using a bandpass filter of 0.5–100 Hz and a notch filter at 50 Hz. Fig. 1 shows a summary of the framework followed, while Supplemental Fig. 1 depicts a detailed overview of the experimental setup. The baseline (mean) was removed from each EEG channel, and high pass filtering was applied to remove drifting. The EEG data were then scaled between −1 and 1. Although a full HFO envelope is between 80–500 Hz, given our hardware specifications and a sampling frequency (fs) of 500 Hz, we can capture HFO oscillations between 70–250 Hz, adhering to the Nyquist theorem [39].

The fNIRS data were collected at 19.5 Hz using an Imagent Tissue Oximeter, which is sensitive to both HbR and HbO. Baseline was removed from the fNIRS data, and high-pass filtering was applied to address drifting. The fNIRS data were then scaled between −1 and 1. fNIRS data was collected using the Imagent Tissue Oximeter system (ISS Inc.), a multi-channel frequency domain system recording at 19.5 Hz with wavelengths of 690 nm and 830 nm for sensitivity to HbR and

HbO respectively. EEG data was recorded according to the standard 10–20 system using 21 electrodes (in positions Fp1, Fp2, F7, F3, Fz, F4, F8, T7, C3, Cz, C4, T8, P7, P3, Pz, P4, P8, O1, O2) at 500 Hz (Neuroscan Synamps 2TM system). Supplemental Fig. 2 and 3 shows an illustrative EEG and fNIRS recording for representative participants across various epilepsy types (HFO, GSWD, LVFA). Custom-made helmets, taking into consideration different head sizes and shapes were made to fit comfortably. Plastic and polyvinyl chloride manufacturing materials made them rigid and light. The helmets were equipped with a total of 64 light sources, 16 light detectors and 19 EEG electrodes that allowed for stable optical coupling between cortica.lb regions and the scalp. This further prevented inter-optode shifting and movement artifacts to a large extent. The sensitivity of near-infrared light to cortical tissue was maintained by positioning the optodes approximately 3–4 cm apart. Electrodes were placed following the 10–20 EEG instrumentation standard, allowing for full head coverage. The signal-to-noise ratio (SNR) threshold applied in channel analysis was defined as those channels less than 30% of the mean SNR of all channels. fNIRS channels deemed to have SNR were eliminated and not included for analysis which led to an average of 138 channels per patient.

We focused on synchronizing and aligning multi-modal datasets using EEG and fNIRS data. Multiple consecutive recordings, each lasting approximately 15min., were made, resulting in a compilation of 200 recordings with 50 hrs of length. To address the class imbalance in our dataset, we employed a conditional generative adversarial network (cGAN) as a data-augmentation strategy [40]. This approach is particularly important, given the variable class distributions within our multimodal data. The cGAN algorithm consists of a generator and discriminator designed to augment EEG and fNIRS datasets. The generator creates synthetic samples by combining noise input with conditional labels to replicate the characteristics of the real EEG and fNIRS data. The discriminator, trained to distinguish between real and generated data, iteratively improved the output of the generator. We calculated the number of synthetic samples required for each class based on the class distribution of the EEG and fNIRS data. The cGAN was then trained on our dataset to generate synthetic samples for underrepresented classes to achieve a balanced class distribution.

To ensure that the generated samples have comparable representation quality compared to the original version, we conducted an analysis using central measures of tendency, i.e., entropy, and skewness. Entropy measures the randomness in the data, kurtosis evaluates the tails of the data distribution, and skewness assesses the asymmetry. Our calculations showed that the mean entropy value for real samples was 2.94 with a standard deviation of 0.017. On the other hand, the generated samples had a mean value of 2.91 with a standard deviation of 0.01. This shows that the generated samples have a comparable level of randomness when compared to the real samples. For the kurtosis measurements, the real samples showed values around 2.0, which indicates a distribution with light tails, suggesting fewer extreme outliers. In contrast, the generated samples displayed negative kurtosis values, ranging from approximately -0.15 to -0.66. This negative kurtosis indicates that the generated data have a flatter distribution with fewer outliers than the real data. When skewness values were compared, real samples had values close to zero, indicating a fairly symmetric distribution. The generated samples had skewness values ranging from -0.14 to 0.16. Despite the variability in skewness between real and generated samples, the generated data are not heavily biased in one direction, maintaining the overall shape of the distribution. Our statistical analysis suggests that the generated samples are different from the real samples, ensuring that they do not replicate the same seizure patterns but rather contribute unique variations that can aid in balancing the dataset.

Our analysis utilized a dataset comprising 1,631 epilepsy trial samples classified into three seizure patterns: GSWD with 981 samples, HFO with 534 samples, and LVFA with 116 samples. The distribution across the EEG channels highlighted F3 as relevant electrode for capturing the GSWD and HFO patterns, whereas C4 was predominantly associated with LVFA.

B. MODEL ARCHITECTURE

Here, we will detail the development of our model for seizure pattern classification which is based on the ViT architecture. We will explain the preprocessing and data preparation steps and illustrate our model building blocks. Our model is illustrated in Fig. 1.

1) DATA INPUT AND PREPROCESSING

Our input data is composed of multimodal EEG and fNIRS recordings. We used spectral encoding techniques, namely, DMD and PSD to extract relevant features for our classification model. Each data sample is captured as a two-dimensional array $\mathbf{X} \in \mathbb{R}^{C \times T}$, where C represents the number of channels (e.g., EEG electrodes or fNIRS channels), and T denotes the number of time steps or frequency bins after spectral encoding.

To feed data samples to our ViT model, we divided the input data array $\mathbf{X}$ into a sequence of smaller patches. Specifically, we split $\mathbf{X}$ into N non-overlapping patches of size $P \times P$, where P is the patch size. Each patch is then flattened into a vector. Let $\mathbf{x}_p \in \mathbb{R}^{P^2}$ represent a flattened patch. The sequence of all flattened patches forms the input to the model:

$$\mathbf{X}_p = \left[\mathbf{x}_p^1, \mathbf{x}_p^2, \ldots, \mathbf{x}_p^N\right] \tag{1}$$

Each flattened patch $\mathbf{x}_p^i$ is projected into a D-dimensional embedding space using a trainable linear projection (embedding layer):

$$\mathbf{z}_0^i = \mathbf{E}\mathbf{x}_p^i + \mathbf{e}_{\text{pos}}^i \tag{2}$$

where $\mathbf{E} \in \mathbb{R}^{D \times P^2}$ is the learnable projection matrix, and $\mathbf{e}_{\text{pos}}^i \in \mathbb{R}^D$ is the positional embedding added to retain positional information of the patches.

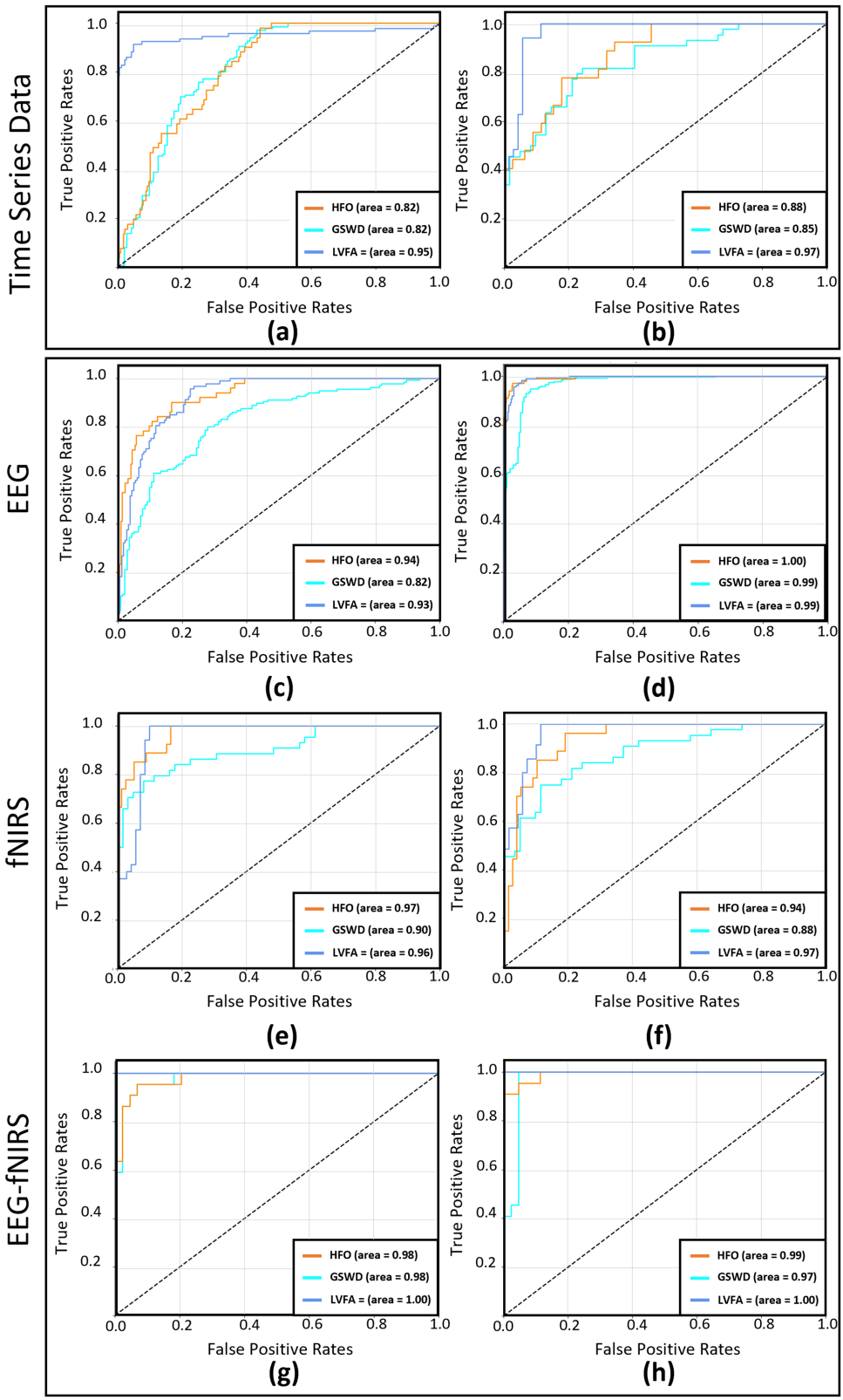

FIGURE 2. ROC curves for different decoding strategies, features, and modalities: (a) ROC curve using time series EEG data as features, (b) ROC curve using time series fNIRS data as features, (c) ROC curve for EEG-only with DMD features, (d) ROC curve for EEG-only with PSD features, (e) ROC curve for fNIRS-only with DMD features, (f) ROC curve for fNIRS-only with PSD features, (g) ROC curve for combined EEG-fNIRS with DMD features, (h) ROC curve for combined EEG-fNIRS with PSD features.

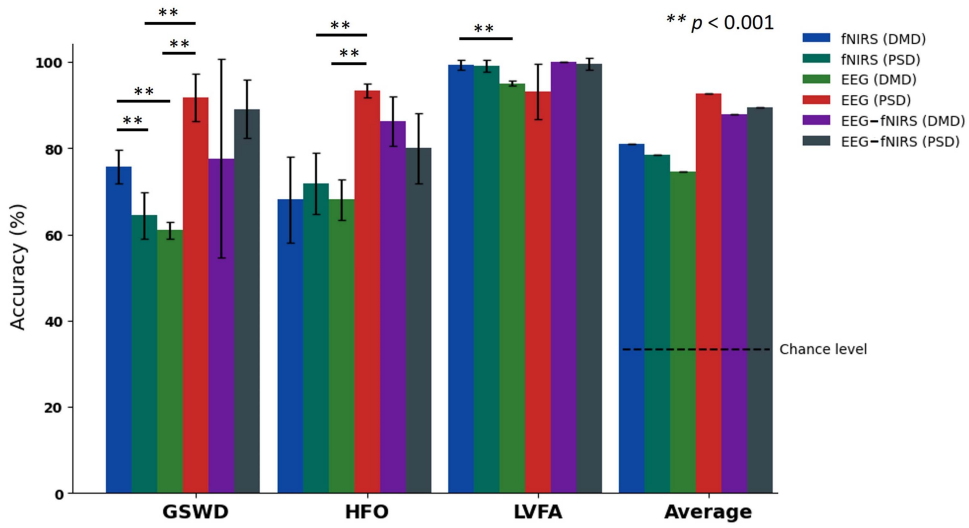

FIGURE 3. Classification rate (the means ± SD) for epilepsy seizure pattern classes, GSWD, HFO, and LVFA, and modality, fNIRS, EEG, and multimodal EEG-fNIRS, using DMD and PSD schemes. **p < 0.001 using t-test. The chance level illustrated in a dashed line for decoding 3 classes is 33.33%.

The output is a sequence of embedded patches:

$$\mathbf{Z}_0 = \left[\mathbf{z}_0^1, \mathbf{z}_0^2, \ldots, \mathbf{z}_0^N\right] \tag{3}$$

2) TRANSFORMER ENCODER

The sequence of embedded patches $\mathbf{Z}_0$ is then passed through a stack of L Transformer encoder layers. Each layer consists of multi-head self-attention (MSA) and a feed-forward network (FFN), with layer normalization (LN) and residual connections. For $l = 1, \ldots, L$:

$$\mathbf{Z}'_l = \text{MSA}\left(\text{LN}\left(\mathbf{Z}_{l-1}\right)\right) + \mathbf{Z}_{l-1} \tag{4}$$

$$\mathbf{Z}_l = \text{FFN}\left(\text{LN}\left(\mathbf{Z}'_l\right)\right) + \mathbf{Z}'_l \tag{5}$$

The MSA block allows the model to jointly attend to information from different representation subspaces. For each head $h = 1, \ldots, H$, the attention output is computed as:

$$\text{Attention}(\mathbf{Q}_h, \mathbf{K}_h, \mathbf{V}_h) = \text{softmax}\left(\frac{\mathbf{Q}_h \mathbf{K}_h^\top}{\sqrt{d_k}}\right)\mathbf{V}_h \tag{6}$$

where $\mathbf{Q}_h$, $\mathbf{K}_h$, and $\mathbf{V}_h$ are the projections of the input $\mathbf{Z}_{l-1}$. Each FFN consists of two linear layers with a GELU activation in between:

$$\text{FFN}(\mathbf{x}) = \text{GELU}(\mathbf{x}\mathbf{W}_1 + \mathbf{b}_1)\mathbf{W}_2 + \mathbf{b}_2 \tag{7}$$

3) CLASSIFICATION HEAD

After the transformer encoder layers, a global average pooling is applied to aggregate the information. We used global average pooling over the sequence dimension as follows:

$$\mathbf{z}_{\text{avg}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{Z}_L^i \tag{8}$$

Finally, a fully connected layer with softmax activation maps the pooled representation to class probabilities:

$$\hat{\mathbf{y}} = \text{softmax}\left(\mathbf{W}_{\text{out}}\mathbf{z}_{\text{avg}} + \mathbf{b}_{\text{out}}\right) \tag{9}$$

Our design allows the ViT model to capture the multimodal sequence data in EEG and fNIRS recordings. Next, we will explain how we integrated the spectral encoding schemes in the classification framework.

C. INTEGRATION OF SPECTRAL ENCODING

Dynamic Mode Decomposition: DMD is a powerful technique for analyzing the dynamics of complex systems, including the brain's electrical and hemodynamic activity [41]. In the context of epilepsy and EEG-fNIRS data, DMD can help in understanding the neural mechanisms and temporal dynamics associated with epileptic seizures. By decomposing EEG-fNIRS data into dynamic modes, one can identify spatiotemporal patterns and interactions between different brain regions. This decomposition is important for pinpointing the onset and propagation of epileptic activity. Moreover, DMD analysis can be used to filter out noise to improve the prediction of seizure events and real-time monitoring and intervention in epilepsy management [42]. Hence, here we investigate the use DMD analysis as a preprocessing phase of EEG-fNIRS data samples fed to the transformer model. The DMD process applied on the input data samples are described below.

We first performed a truncated Singular Value Decomposition (SVD) on the EEG-fNIRS signal matrix $\mathbf{X}$ of shape (channels, data). This decomposition yields three components: the matrix of left singular vectors ($\mathbf{U}$), the diagonal matrix of singular values ($\mathbf{\Sigma}$), and the conjugate transpose of the matrix of right singular vectors ($\mathbf{V}^*$):

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$$

Following SVD, we selected the first r columns of $\mathbf{U}$, forming the matrix $\mathbf{U}_r$, and the corresponding rows of $\mathbf{V}^*$, forming $\mathbf{V}_r^*$. This step results in a low-rank approximation of the EEG-fNIRS signal matrix $\mathbf{X}$ expressed as:

$$\mathbf{X}_r \approx \mathbf{U}_r \mathbf{\Sigma}_r \left(\mathbf{V}_r^*\right)^*$$

A matrix $\mathbf{A}$ is computed by multiplying the subsequent EEG-fNIRS signal matrix $\mathbf{X}_{r+1}$ with the pseudo-inverse of $\mathbf{X}_r$:

$$\mathbf{A} = \mathbf{X}_{r+1} \left(\mathbf{X}_r^+\right)$$

We computed the eigenvalues (λ_i) and corresponding eigenvectors ($\mathbf{v}_i$) of matrix $\mathbf{A}$ to capture the temporal dynamics of the EEG-fNIRS signal. Then, the spatial patterns or modes ($\boldsymbol{\phi}_i$) are determined as follows:

$$\boldsymbol{\phi}_i = \frac{1}{\lambda_i} \mathbf{X}_r \mathbf{v}_i$$

Power Spectral Density: The PSD features used in our classification framework are computed using the Welch's method. This method divides the signal into overlapping segments, applies a window to each, and computes the periodogram of each segment. The PSDs from these segments are then averaged to produce the final PSD estimate for each channel. For a given EEG-fNIRS signal $x[n]$, the PSD $P_{xx}(f)$ is calculated as:

$$P_{xx}(f) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n] w[n] e^{-j2\pi fn} \right|^2$$

where $w[n]$ is the window function and N is the number of points in each segment. These PSD features provide a frequency-based representation of the EEG-fNIRS signals. The PSD features matrix is then concatenated with that obtained by DMD, i.e., the matrix $\mathbf{X}$ or its approximated form $\mathbf{X}_r$.

D. LOSS FUNCTION AND OPTIMIZATION

To train our ViT model, we used a categorical cross-entropy loss function for multiclass classification, which is commonly picked for such tasks. The categorical cross-entropy loss is defined as

$$L(y, \hat{y}) = -\sum_{i=1}^{M} y_i \log(\hat{y}_i)$$

where y is the true label in a one-hot encoded form, $\hat{y}$ is the predicted probability distribution from the softmax layer of the model, and M is the number of classes. Adam optimizer was used to train the model. This optimizer has an adaptive learning rate computed for model parameters from the first and second moments of the gradients. The general update rule for the Adam optimizer is given as follows:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \hat{m}_t$$

where θ represents the parameters of the model, η is the step size, ϵ is a small scalar used to prevent division by zero, $\hat{m}_t$ is the bias-corrected estimate of the first moment (the mean) of the gradients, and $\hat{v}_t$ is the bias-corrected estimate of the second moment (the uncentered variance) of the gradients.

IV. EXPERIMENTS AND RESULTS

A. NETWORK TRAINING AND TESTING

In our training, we used an early stopping of 25 epochs to prevent overfitting. Our model assessment was done using a 10-fold cross-validation to ensure that the model's performance was robust and independent of a particular split in the data. In each fold, the model is trained on a subset of the data and tested on the remaining subset. The performance across all folds was averaged to provide a reliable estimate of the model's performance.

B. EVALUATION METRICS ANALYSIS

The receiver operating characteristic (ROC) curve and the area under the curve (AUC) of the ROC were used to evaluate the model's performance for each class. The ROC curve is a graphical representation of the true positive rate (TPR) versus the false positive rate (FPR) at various threshold settings. The AUC provides a single measure of the model performance across all classification thresholds, with a higher AUC indicating better performance. For the averaged folds in the cross-validation, a confusion matrix was computed to provide detailed insights into the true versus predicted labels. In our analysis, we comparatively evaluated models trained on data samples from individual EEG recordings, individual fNIRS recordings, and combined EEG-fNIRS recordings acquired concurrently. In each of these trials, we evaluated the model architectures by integrating either the PSD- or DMD-extracted feature maps that were fed to our ViT Model. We then conducted an ablation study with model training that did not have any spectral encoding as features were extracted to the ViT layers.

C. ANALYSIS WITHOUT SPECTRAL ENCODING

We trained our ViT models without integrating spectral encoding layers into the architecture. As depicted in Fig. 2(a), the ROC curve for EEG time series as a feature showed a low AUC of 0.82 for the HFO and GSWD classes. For the LVFA class, the AUC was 0.95. On the other hand, Fig. 2(b) presents the ROC curve for fNIRS time series as features. The LVFA achieved an AUC of 0.97, indicating a consistent performance with the fNIRS modality. The AUC for GSWD was 0.85, and HFO reported an AUC of 0.88. The confusion matrices computed when training on EEG and fNIRS time series features are depicted in Supplemental Fig. 4(a) and (b), respectively. The LVFA class showed a true positive rate of 86.58%, whereas HFO and GSWD demonstrated a lower accuracy for the EEG time series data. For the fNIRS case, the LVFA's performance was high with a true positive rate of 94.25%, whereas HFO and GSWD exhibit lower true positive rates than the other modalities. It should be noted that the HFO class showed the lowest true positive rates of 28.33% and

36.06% for the EEG and fNIRS cases, respectively, compared to other seizure patterns.

D. EEG-BASED ANALYSIS

In this section, we present the analysis employed to assess the predictive accuracy of three seizure patterns, i.e., HFO, GSWD, and LVFA using EEG-only modality. Fig. 2(c) illustrates the ROC curve for EEG decoding using DMD features. The classification of the HFO class showed high performance, with an AUC of 0.94, indicating its discriminative DMD features in this modality. Comparatively, classifying LVFA also showed a strong discriminative ability with an AUC of 0.93, whereas in GSWD cases with an AUC of 0.82, the model demonstrated good but slightly lower performance. On the other hand, Fig. 2(d) presents the ROC curve for EEG-only decoding with PSD features, where an AUC of 1.00 for HFO signifies perfect classification capability. An LVFA AUC of 0.99 was similarly impressive, closely mirrored a perfect score, was similarly impressive. The GSWD also performed well, with an AUC of 0.99. The classification performance was further examined using confusion matrices for different feature sets and modalities. Supplemental Fig. 4(c) shows the confusion matrix for EEG-only decoding with DMD features. The HFO displayed a balanced classification with notable true-positive and true-negative rates of 67.06% and 32.74%, respectively. The GSWD class had a higher misclassification rate, whereas LVFA showed a remarkable true-negative rate of 94.95% true negative rate, indicating its strong specificity in this context. Supplemental Fig. 4(d), shows that EEG-only decoding with PSD features exhibited high classification performance on the HFO and LVFA classes, with high true-positive rates of 93.14% and 93.87%, respectively. The model also performed well on the GSWD pattern, with a true-positive rate of 91.69%.

E. FNIRS-BASED ANALYSIS

Here, we describe the classification performance when our transformer model is trained only on fNIRS recordings. Fig. 2(e) shows the ROC curve for fNIRS-only decoding using DMD features. Similar to the EEG-only analysis, HFO and LVFA exhibited strong performance, with AUCs of 0.97 and 0.96, respectively. The performance on the GSWD samples maintained a high level of discriminative power, as evident from its AUC of 0.90. Fig. 2(f) shows the ROC curve for fNIRS-only decoding with PSD features. Both HFO and GSWD achieved near-perfect AUC scores of 0.98, whereas LVFA achieved a perfect AUC of 1.00. From the confusion matrices of fNIRS-only decoding with DMD features (Supplemental Fig. 4(e)), LVFA stood out with a 94.28% true negative rate, whereas HFO showed lower sensitivity and specificity values, as evident from its true positive and negative rates. On the other hand, Supplemental Fig. 4(f) illustrates fNIRS-only decoding with PSD features, where the LVFA achieved a high true positive rate of 99.10 %, highlighting its substantial performance. The HFO and GSWD groups showed lower true-positive rates of 71.84% and 64.55%, respectively.

F. COMBINED EEG-FNIRS ANALYSIS

In this section, we present the classification performance of different seizure patterns by integrating both EEG and fNIRS signals with the DMD and PSD spectral encoding schemes. In Fig. 2(g), the ROC curve for the combined EEG-fNIRS decoding with DMD features indicates that the LVFA classification was perfect, with an AUC of 1.0. The model exhibited high performance on the HFO and GSWD classes, with an AUC of 0.98 for both. Fig. 2(h), shows the combined ROC curve from EEG-fNIRS decoding with PSD features. HFO, GSWD, and LVFA showed noticeable discriminative abilities, with AUCs of 0.99, 0.97, and 1.00, respectively, suggesting that the combined modality and PSD features can effectively enhance classification accuracy. The confusion matrices generated after combined EEG-fNIRS training are depicted in Supplemental Fig. 4(g) and (h). Both encoding schemes, DMD and PSD features, exhibited high true positive rates, especially for the LVFA class, which demonstrated near-perfect classification abilities with true negative rates of 100% and 99.55%.

G. COMPARATIVE ANALYSIS

We conducted a comparative analysis in Supplemental Table 3 to study the performance of our classification model when compared to other previous methods used for the detection and classification of epileptic seizure patterns. In [43], the authors used energy features to detect different types of HFOs (ripples, fast ripples, fast ripples during ripples) achieving a sensitivity of 87%. Wu et al. also focused on the detection of HFO (ripples and fast ripples), scoring a sensitivity of 92.4% [44]. Statsenko et al. examined generalized seizure detection using deep learning methods, i.e., CNNs and RNNs, where they achieved a sensitivity of 87.7% [45]. Ullah et al. used a pyramidal 1D-CNN for detecting epileptic seizures, reporting an accuracy of up to 99.1% with a sensitivity reaching 99% [46]. Sirpal et al. combined EEG and fNIRS data in seizure detection, achieving a sensitivity of 89. 7% and a specificity of 95. 5% [10]. Rosas-Moreno et al. focused on pre-ictal state prediction with a CNN and fNIRS, reporting a sensitivity of 95.24% and an accuracy of 96.9% to 100% [47] while Tsipouras et al. assessed the impact of different frequency sub-bands using the spectral information of EEG signal and achieved an accuracy of up to 91% on a five-classes classification task [48]. In comparison, our model, particularly the combined EEG-fNIRS with PSD scheme, achieved significantly higher performance, with a sensitivity of 99.55% for HFOs and LVFA and 89.95% for GSWD with an accuracy of 87.3% for GSWD and 83.3 and 100% for HFO and LVFA respectively as shown in Supplemental Table 1.

H. STATISTICAL ANALYSIS

The similarities and differences in classification performances when using different signal features of epileptic seizures were

examined using statistical analysis techniques, i.e., t-tests, p-values, two-way analysis of variance (ANOVA), and multiple comparison tests of the means (Tukey's test). Fig. 3 presents the mean accuracies obtained when using different modalities, fNIRS, EEG, and combined EEG-fNIRS, to classify the three classes of epilepsy, GSWD, HFO, and LVFA, with DMD and PSD schemes. For the GSWD class, the highest mean accuracy was achieved by the EEG-PSD training at approximately 91.7% (SD: 5.5), followed by the EEG-fNIRS-PSD training at approximately 89% (SD: 6.8). Significant differences ($p < 0.001$) were found for the GSWD pattern using fNIRS-DMD vs. fNIRS-PSD and EEG-DMD vs. EEG-PSD. This suggests that the combination of EEG with PSD and fNIRS with DMD is more efficient for detecting GSWD than EEG-DMD or fNIRS-PSD. For the HFO class, the EEG-PSD modality had a mean accuracy of 93.3 (SD: 1.6). EEG-fNIRS-PSD and EEG-fNIRS-DMD achieved 79.9% (SD: 8.1) and 86.4% (SD: 7.5), respectively. Other training setups achieved accuracies ranging between 68.1% (SD: 9.9) and 71% (SD: 7). We also observed a significant difference in HFO within EEG-DMD vs. EEG-PSD and EEG-PSD vs. fNIRS-PSD, which adds to the significance of combining EEG with PSD and also reflected that EEG may have more features for the HFO class. The Accuracies for the LVFA class ware considerably higher than those for the other classes, i.e., between 93.1% (SD: 6.45) for EEG-PSD and 100% (SD: 0) for EEG-fNIRS-DMD. No significant differences were observed in the EEG-fNIRS modality comparisons, but fNIRS-DMD showed significant differences compared to EEG-DMD, highlighting the effectiveness of combining fNIRS with DMD features.

Supplemental Table 2 shows a comparison of p-values obtained from t-tests across different classes and brain measurement modalities. Analysis of variance for fNIRS data using DMD features revealed a significant main effect of epileptic seizure type ($F(2, 29) = 61.9354$, $p < 0.0001$), indicating substantial differences in classification accuracy between the seizure types examined. Tukey's HSD test further demonstrated significant differences between seizure classes, most notably between the GSWD and LVFA classes (mean difference $= -23.7450$, $p < 0.0001$), and between the HFO class and the LVFA class (mean difference $= -31.2810$, $p < 0.0001$), suggesting a superior accuracy when distinguishing LVFA from the other seizure types. Similarly, PSD feature analysis of the fNIRS data showed a significant effect of epileptic seizure type ($F(2, 29) = 112.51$, $p < 0.0001$). This highlights the distinct differences in classification accuracy among the seizure types. Significant disparities were observed in all pairwise comparisons using Tukey's HSD test, with the largest discrepancy noted between the GSWD and LVFA classes (mean difference $= -34.69$, $p < 0.0001$). This shows the effectiveness of classification on some patterns of seizures, suggesting the need for optimizing predictive models through the selection of pattern-specific features. For EEG data analyzed with DMD features, ANOVA indicated a significant main effect of epileptic seizure type ($F(2, 29) = 326.38$, $p < 0.0001$). Tukey's HSD test revealed significant differences between the seizure classes, especially between the GSWD and LVFA classes (mean difference $= -34.1250$, $p < 0.0001$), and the HFO class and the LVFA class (mean difference $= -27$, $p < 0.0001$), indicating the ability to differentiate LVFA from other types of seizure based on EEG data.

The analysis of EEG data using PSD features yielded a non-significant main effect of the epileptic seizure type ($F(2, 29) = 0.28$, $p = 0.75$). No significant differences were observed in the post hoc comparisons, indicating that the classification accuracy did not vary significantly across the seizure types within this training setup. The integrated analysis of fNIRS and EEG data using DMD features showed a limited effect on the epileptic seizure type ($F(2, 29) = 6.06$, $p = 0.00$), with little to no difference identified between the GSWD and LVFA classes (mean difference $= -22.36$, $p = 0.0067$) in Tukey's test. Moreover, when analyzing the integrated fNIRS and EEG data with PSD features, a pronounced effect of epileptic seizure type was noted ($F(2, 29) = 22.44$, $p < 0.0001$). Tukey's post-hoc analysis highlighted noticeable differences between all seizure classes, particularly between the GSWD class and the LVFA class (mean difference $= -10.45$, $p = 0.0037$). This suggests that the integration of the fNIRS and EEG modalities when employing PSD features significantly contributed to the predictive accuracy of the models, reinforcing the potential of multimodal approaches in enhancing the performance of predictive models for epileptic seizure classification.

V. DISCUSSION

A. FINDINGS

Our findings further confirm the capability of EEG and fNIRS signals to effectively classify specific HFO, LVFA and GSWD patterns associated with different aspects of epileptic activity. Models trained on EEG channels outperformed multichannel fNIRS as show in Supplemental Fig. 4, benefiting from their precise temporal resolution and ability to capture rapid neural dynamics. Conversely, fNIRS provides valuable spatial information, facilitating the identification of specific brain regions involved in each seizure type. Notably, our study did not prioritize the selection of the most relevant electrodes for decoding, potentially affecting the classification accuracy of fNIRS compared to EEG signals. Analysis of the confusion matrices of the models indicates that the LVFA consistently delivers high specificity across different modalities and feature sets. The HFO pattern showed robustness, particularly with PSD features where it achieved accuracy between 71% and 93% for the different modalities, whereas the GSWD, although slightly less consistent, still demonstrated moderate predictive power where it achieved accuracy between 64% and 91% for the different modalities.

Our study's results align with previous work on HFO detection using a CNN model that achieved an accuracy of 92.43% for detecting ripples and 87.85% for fast ripples, showing the high precision possible with automated systems in classifying epileptogenic zones [49]. Additionally, transformer-based

models for HFO detection using MEG data achieved an accuracy of 96.15%, further highlighting the potential of using advanced architectures for enhanced classification [50].

In summary, our findings highlight the importance of feature and modality selection in enhancing classification performance. It has been shown that the integration of spectral features in our ViTs substantially improves performance. Surprisingly, EEG-only training with PDS feature decoding showed a consistent performance, with the highest accuracies in the case of the GSWD and HFO patterns but more prone to sensitivity of different cut-off points as per ROC curves. from previous work with fNIRS prediction using multimodal autoencoder from EEG [27], and while Combined EEG-fNIRS setups helped improve the LFVA classification while maintaining a high level of accuracy for other patterns with a high improvement in sensitivity towards all cut-off points as shown in Fig. 2.

B. FUTURE DIRECTIONS

One of the challenges in AI lies in using small datasets to generalize model performances. It is to be noted that large datasets are often unavailable in some domains, e.g., healthcare. This study aimed to extract suitable spatiotemporal features from EEG and fNIRS data for training a deep-learning model to classify three types of epileptic seizures (HFO, GSWD, and LVFA). We explored the use of a transformer-based model, i.e., ViT, known for effectively handling sequence data. Our work demonstrated high accuracy in automating seizure classification, which can ultimately enhance epilepsy care and associated clinical outcomes. Due to the limitation of the small data size involved here, we implemented an augmentation procedure to enable better training of the ViT model. Nevertheless, benchmarking other augmentation techniques developed specifically for EEG and fNIRS data could help to achieve better classification performance for some types of seizure patterns. The ability to decode epilepsy and seizure patterns using variants of ViT models and multimodal EEG-fNIRS recordings represents an opportunity for future research in this direction. Developing approaches to merge these data modalities can help in relevant studies on seizure detection and classification. On the other hand, working on specific designs of ViTs with layers that can collect distinctive features from hemodynamic and electrophysiological signals is another research direction to investigate. Possible improvements in the ViT architecture could be related to developing a cross-modal attention mechanism that can emphasize the salient regions of EEG and fNIRS signals. In addition, integrating other sequence modeling layers in the classification framework, e.g., with recurrent neural networks, could help capture more accurate and complete cognitive events with their dynamics during seizures in a spatiotemporal perspective. One important aspect that could be explored to build on the work presented in this article is adding explainability blocks in the prediction model, such as SHAP values or Grad-CAM. This can help provide clinicians with more meaningful insights into how a patient's brain activity is influencing their epilepsy. It is known that personalized treatment plans can improve clinical outcomes. A future direction to explore could be related to studying learning techniques applied to each patient's profile to enhance individualized assessments in real clinical environments. For translational work that builds on this work, early detection of seizure patterns in real-time is important. Developing a learning framework with reduced inference time while maintaining comparable prediction performance could be explored. Lastly, our findings could confirmed by validating on larger cohorts through collaboration with large multicenter, clinic-based data sets. This will ensure that models are trained on multiple case types and conditions.

VI. CONCLUSION

This study explored the use of a ViT-based model to classify epileptic seizure patterns using combined EEG and fNIRS data, incorporating spectral encoding schemes such as DMD and PSD. Our trained model demonstrated high accuracy in handling long-range dependencies when trained on PSD-encoded EEG data. When trained on raw time series of EEG or fNIRS inputs alone, the performance was of a lower accuracy. We have shown that combining EEG with fNIRS data with DMD and PSD spectral encoding improved the classification accuracy of some seizure patterns. The findings suggest that further exploration of the synergy between EEG and fNIRS, along with optimized preprocessing and augmentation techniques, could enhance the detection and classification performances, potentially aiding in diagnosing various neurological disorders.

REFERENCES

[1] R. D. Thijs, R. Surges, T. J. O'Brien, and J. W. Sander, "Epilepsy in adults," *Lancet*, vol. 393, no. 10172, pp. 689–701, 2019.

[2] T. P. Suurmeijer, M. F. Reuvekamp, and B. P. Aldenkamp, "Social functioning, psychological functioning, and quality of life in epilepsy," *Epilepsia*, vol. 42, no. 9, pp. 1160–1168, 2001.

[3] A. N. Belkacem, N. Jamil, J. A. Palmer, S. Ouhbi, and C. Chen, "Brain computer interfaces for improving the quality of life of older adults and elderly patients," *Front. Neurosci.*, vol. 14, 2020, Art. no. 692.

[4] N. Jamil, A. N. Belkacem, S. Ouhbi, and A. Lakas, "Noninvasive electroencephalography equipment for assistive, adaptive, and rehabilitative brain–computer interfaces: A systematic literature review," *Sensors*, vol. 21, no. 14, 2021, Art. no. 4754.

[5] A. N. Belkacem, N. Jamil, S. Khalid, and F. Alnajjar, "On closed-loop brain stimulation systems for improving the quality of life of patients with neurological disorders," *Front. Hum. Neurosci.*, vol. 17, 2023, Art. no. 1085173.

[6] C. Binnie and P. Prior, "Electroencephalography," *J. Neurol., Neurosurgery, Psychiatry*, vol. 57, no. 11, 1994, Art. no. 1308.

[7] M. Ferrari and V. Quaresima, "A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application," *Neuroimage*, vol. 63, no. 2, pp. 921–935, 2012.

[8] D. A. Boas, C. E. Elwell, M. Ferrari, and G. Taga, "Twenty years of functional near-infrared spectroscopy: Introduction for the special issue," *Neuroimage*, vol. 85, pp. 1–5, 2014.

[9] E. Guevara et al., "Prediction of epileptic seizures using fNIRS and machine learning," *J. Intell. Fuzzy Syst.*, vol. 38, no. 2, pp. 2055–2068, 2020.

[10] P. Sirpal, A. Kassab, P. Pouliot, D. K. Nguyen, and F. Lesage, "fNIRS improves seizure detection in multimodal EEG-fNIRS recordings," *J. Biomed. Opt.*, vol. 24, no. 5, pp. 051408–051408, 2019.

[11] D. K. Nguyen et al., "Non-invasive continuous EEG-fNIRS recording of temporal lobe seizures," *Epilepsy Res.*, vol. 99, no. 1-2, pp. 112–126, 2012.
[12] 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, 2021, Art. no. 102595.
[13] H. Laufs, U. Lengler, K. Hamandi, A. Kleinschmidt, and K. Krakow, "Linking generalized spike-and-wave discharges and resting state brain activity by using EEG/FMRi in a patient with absence seizures," *Epilepsia*, vol. 47, no. 2, pp. 444–448, 2006.
[14] X. Fu et al., "Integrating optimized multiscale entropy model with machine learning for the localization of epileptogenic hemisphere in temporal lobe epilepsy using resting-state FMRI," *J. Healthcare Eng.*, vol. 2021, no. 1, 2021, Art. no. 1834123.
[15] J. Engel Jr and F. L. da Silva, "High-frequency oscillations–where we are and where we need to go," *Prog. Neurobiol.*, vol. 98, no. 3, pp. 316–318, 2012.
[16] M. De Curtis and V. Gnatkovsky, "Reevaluating the mechanisms of focal ictogenesis: The role of low-voltage fast activity," *Epilepsia*, vol. 50, no. 12, pp. 2514–2525, 2009.
[17] P. Gloor, "Generalized epilepsy with spike-and-wave discharge: A reinterpretation of its electrographic and clinical manifestations1: The 1977 William G. Lennox lecture, American epilepsy society," *Epilepsia*, vol. 20, no. 5, pp. 571–588, 1979.
[18] K. J. Staley, A. White, and F. E. Dudek, "Interictal spikes: Harbingers or causes of epilepsy?," *Neurosci. Lett.*, vol. 497, no. 3, pp. 247–250, 2011.
[19] L. Chauviere et al., "Changes in interictal spike features precede the onset of temporal lobe epilepsy," *Ann. Neurol.*, vol. 71, no. 6, pp. 805–814, 2012.
[20] S. A. Weiss et al., "ICTAL onset patterns of local field potentials, high frequency oscillations, and unit activity in human mesial temporal lobe epilepsy," *Epilepsia*, vol. 57, no. 1, pp. 111–121, 2016.
[21] C. Cheng, Y. Zhang, L. Liu, W. Liu, and L. Feng, "Multi-domain encoding of spatiotemporal dynamics in EEG for emotion recognition," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 3, pp. 1342–1353, Mar. 2023.
[22] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[23] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, "Transformers in vision: A survey," *ACM Comput. Surv.*, vol. 54, no. 10s, pp. 1–41, 2022.
[24] A. Hireche, R. Damseh, P. Sirpal, and A. N. Belkacem, "EEG-based epileptic seizure pattern decoding using vision transformer," in *Proc. IEEE 15th Int. Conf. Innov. Inf. Technol.*, 2023, pp. 55–60.
[25] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: A review," *J. Neural Eng.*, vol. 16, no. 3, 2019, Art. no. 031001.
[26] J. Cao, J. Zhu, W. Hu, and A. Kummert, "Epileptic signal classification with deep EEG features by stacked CNNS," *IEEE Trans. Cogn. Devel. Syst.*, vol. 12, no. 4, pp. 709–722, Dec. 2020.
[27] P. Sirpal, R. Damseh, K. Peng, D. K. Nguyen, and F. Lesage, "Multimodal autoencoder predicts fNIRS resting state from EEG signals," *Neuroinformatics*, vol. 20, no. 3, pp. 537–558, 2022.
[28] A. Vaswani et al., "Attention is all you need," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 6000–6010.
[29] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.
[30] Y. Hou et al., "GCNS-Net: A graph convolutional neural network approach for decoding time-resolved EEG motor imagery signals," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 6, pp. 7312–7323, Jun. 2024.
[31] Y. Hou et al., "Deep feature mining via the attention-based bidirectional long short term memory graph convolutional neural network for human motor imagery recognition," *Front. Bioeng. Biotechnol.*, vol. 9, 2022, Art. no. 706229.
[32] Y. Hou, L. Zhou, S. Jia, and X. Lun, "A novel approach of decoding EEG four-class motor imagery tasks via scout ESI and CNN," *J. Neural Eng.*, vol. 17, no. 1, 2020, Art. no. 016048.
[33] Z. Wang, J. Zhang, Y. Xia, P. Chen, and B. Wang, "A general and scalable vision framework for functional near-infrared spectroscopy classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 1982–1991, 2022.
[34] Z. Wang, J. Zhang, X. Zhang, P. Chen, and B. Wang, "Transformer model for functional near-infrared spectroscopy classification," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 6, pp. 2559–2569, Jun. 2022.
[35] K. M. Gunter, A. Brink-Kjær, E. Mignot, H. B. Sørensen, E. During, and P. Jennum, "SViT: A spectral vision transformer for the detection of REM sleep behavior disorder," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 9, pp. 4285–4292, Sep. 2023.
[36] H. Yao, T. Liu, R. Zou, S. Ding, and Y. Xu, "A spatial-temporal transformer architecture using multi-channel signals for sleep stage classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 3353–3362, 2023.
[37] M. A. Mulkey, H. Huang, T. Albanese, S. Kim, and B. Yang, "Supervised deep learning with vision transformer predicts delirium using limited lead EEG," *Sci. Rep.*, vol. 13, no. 1, 2023, Art. no. 7890.
[38] K. Han et al., "A survey on vision transformer," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 87–110, Jan. 2023.
[39] F. H. Simozo, J.-B. Destro-Filho, and L. O. M. Junior, "Detrended-fluctuation-analysis (DFA) and high-frequency-oscillation (HFO) coefficients and their relationship to epileptic seizures," in *Proc. NEUROTECHNIX*, 2014, pp. 99–105.
[40] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*.
[41] P. P. Mitra and B. Pesaran, "Analysis of dynamic brain imaging data," *Biophysical J.*, vol. 76, no. 2, pp. 691–708, 1999.
[42] 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, 2022, Art. no. 5865.
[43] N. Sciaraffa, M. A. Klados, G. Borghini, G. Di Flumeri, F. Babiloni, and P. Aricò, "Double-step machine learning based procedure for hfos detection and classification," *Brain Sci.*, vol. 10, no. 4, 2020, Art. no. 220.
[44] M. Wu, H. Qin, X. Wan, and Y. Du, "HFO detection in epilepsy: A stacked denoising autoencoder and sample weight adjusting factors-based method," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1965–1976, 2021.
[45] Y. Statsenko et al., "Automatic detection and classification of epileptic seizures from eeg data: Finding optimal acquisition settings and testing interpretable machine learning approach," *Biomedicines*, vol. 11, no. 9, 2023, Art. no. 2370.
[46] I. Ullah et al., "An automated system for epilepsy detection using EEG brain signals based on deep learning approach," *Expert Syst. Appl.*, vol. 107, pp. 61–71, 2018.
[47] R. Rosas-Romero et al., "Prediction of epileptic seizures with convolutional neural networks and functional near-infrared spectroscopy signals," *Comput. Biol. Med.*, vol. 111, 2019, Art. no. 103355.
[48] M. G. Tsipouras, "Spectral information of EEG signals with respect to epilepsy classification," *EURASIP J. Adv. Signal Process.*, vol. 2019, no. 1, pp. 1–17, 2019.
[49] R. Zuo et al., "Automated detection of high-frequency oscillations in epilepsy based on a convolutional neural network," *Front. Comput. Neurosci.*, vol. 13, 2019, Art. no. 6.
[50] J. Guo et al., "Transformer-based high-frequency oscillation signal detection on magnetoencephalography from epileptic patients," *Front. Mol. Biosci.*, vol. 9, 2022, Art. no. 822810.

RAFAT DAMSEH (Member, IEEE) received the B.S. degree in telecommunication engineering from Yarmouk University, Irbid, Jordan, in 2011, the M.Sc. degree in electrical and computer engineering from Concordia University, Montreal, QC, Canada, in 2015, and the Ph.D. degree in biomedical engineering from Polytechnique Montréal, Montreal, in 2020., From 2020 to 2021, He was a Postdoctoral Researcher with Harvard Medical School, Boston, MA, USA. He is currently an Assistant Professor with the Department of Computer Science and Software Engineering, United Arab Emirates University, Al-Ain, United Arab Emirates. His research interests include medical image computing and reconstruction and computational neuroimaging.

ABDELHADI HIRECHE received the master's degree from the University of Chlef, Ouled Fares, Algeria, in 2021. Since 2023, he has been a Research Assistant with the College of Information Technology, UAE University, Al Ain, United Arab Emirates. His research interests include the fields of artificial intelligence, robotics, and brain-computer interfaces.

PARIKSHAT SIRPAL (Member, IEEE) received the Ph.D. degree in electrical engineering from Universitéde Montréal, Montréal, QC, Canada. He is currently with the Gallogly College of Engineering, The University of Oklahoma, Norman, OK, USA. His research interests include brain-computer interface (BCI) systems, multimodal analysis of brain activity, and the development of predictive models for brain disorders.

ABDELKADER NASREDDINE BELKACEM (Senior Member, IEEE) received the Ph.D. degree in information processing from the Tokyo Institute of Technology, Tokyo, Japan, in 2015. He is currently an Associate Professor with the United Arab Emirates University (UAEU), Al Ain, United Arab Emirates. Prior to joining the UAEU, he was a specially appointed Researcher with the Department of Neurosurgery, Osaka University Medical School, Suita, Japan, and an Assistant Professor with the Endowed Research Department of Clinical Neuroengineering, Global Center for Medical Engineering and Informatics, Osaka University, Suita, from 2015 to 2019. His research interests mainly include brain computer/machine interface using human magneto- and electro-encephalography (MEG/EEG), human robot interaction, applied artificial intelligence, robotics, and computational neuroscience. Dr. Belkacem was one of the recipients of the prestigious international award "MIT Technology Review Innovators Under 35 MENA" and brings more than 10 years of experience in brain machine interface (BMI). One of his works was adopted on a cover page of IEEE Trans. Neural. Syst. Rehabil. Eng., Vol 26-6, 2018. He has contributed to many international Technical Program Committees such as IEEE International Conference on Systems, Man, and Cybernetics as Co-Chair of BMI Workshops since SMC2018. He is a Guest Associate Editor of BMI on Journal of *Frontiers in Human Neuroscience*. Since 2021, he has been the Vice Chair for the IEEE P2725.1 Working Group.