MVCA-Net: Multi-View Convolution Attention Network for measuring EEG rhythms representing Anxiety

Hamidreza Ghonchi[1], Tom Foulsham[2], Saideh Ferdowsi[1]

***Abstract*— Anxiety can significantly impact individuals' daily lives, and can manifest at varying levels from mild to severe. Traditionally, psychologists and psychiatrists assess anxiety primarily through self-report questionnaires. However, advances in computer-aided technologies and neuroimaging techniques offer promising tools to enhance diagnostic accuracy. In this study, we propose a novel deep learning model designed to extract frequency-based features from electroencephalogram (EEG) signals which provide insights into the neural patterns associated with anxiety. Our model consists of a convolutional neural network (CNN), a multi-head attention transformer, and an attention module to effectively capture EEG features distinguishing normal and anxious states. We validated our approach using a publicly available EEG dataset called DASPS, collected from 23 participants, where self-reported anxiety levels were categorized into normal and anxious conditions. The anxious condition was further subdivided into four levels of anxiety based on its severity. The proposed model achieved classification 82.94% accuracy for binary classification (normal vs. anxious) and 74.05% average accuracy for multi-class classification (normal, mild, moderate, and severe anxiety). These results highlight the effectiveness of our approach in leveraging EEG-based frequency features for anxiety assessment across different levels of severity.**

***Index Terms*— Electroencephalogram, Anxiety Detection, Deep Learning, Mental Health, DASPS**

I. Introduction

Anxiety, a widespread mental health condition closely related to stress, poses significant challenges to individuals' lives if left untreated. Since anxiety, especially in severe cases, causes serious issues. Therefore early detection is crucial to reducing the long-term effects of anxiety [1], [2]. In 2019, anxiety disorders and other mental illnesses were not only ranked among the top 25 contributors to excessive global healthcare expenditures, but also identified as some of the most debilitating conditions. Anxiety disorders affect approximately 13% of the global population, and 8% experiencing anxiety or depressive disorders, according to the World Health Organization (WHO) [3]–[5]. This prevalence increased during the COVID-19 pandemic, with cases of anxiety and depression increasing by 25.6% and 27.6%, respectively, in 2020 [6]. Anxiety disrupts cognitive functions, such as memory and attention, and is linked to immune disorders, further affecting daily life [4]. Traditional treatments, including exposure therapy, where patients face fears in controlled settings, remain essential [7]. Recent advances, such as virtual reality exposure therapy [8]–[11] and machine learning-based approaches [12]–[15], offer innovative methods for early detection and management of anxiety, underscoring the urgent need to explore effective interventions to alleviate this growing public health problem.

[1]School of Mathematics, Statistics and Actuarial Science, University of Essex, Colchester, UK. h.ghonchi@essex.ac.uk

[2]Department of Psychology, University of Essex, Colchester, UK.

Self-report questionnaires, such as the Spielberger State-Trait Anxiety Inventory and the Hospital Anxiety and Depression Scale, have been widely used in the past decade as reliable tools for diagnosing anxiety [12]. Although these methods remain essential for assessing anxiety levels, recent advances have integrated physiological and neural signals, including electrocardiogram (ECG), photoplethysmography (PPG), and electroencephalography (EEG), to improve diagnostic precision [16]–[19]. In the past decade, EEG signals have been widely used not only because they are a non-invasive method but also because they offer high temporal resolution recording of brain activity. As shown in various studies, EEG signals are effective in identifying anxiety, as they can detect changes in power spectral density (PSD) associated with generalized anxiety disorder (GAD) [20], [21]. Besides this, emerging techniques such as neuroimaging and machine learning have helped researchers develop reliable solutions in this field [1].

Numerous studies have explored the usage of EEG to identify anxiety and its severity. The publicly available DASPS dataset offers a valuable resource for researchers to examine the application of machine learning in this domain. DASPS, short for "A Database for Anxious States based on Psychological Stimulation" [22], categorizes trials using two primary labels: "normal" and "anxious." Additionally, the dataset further subdivides the anxious category into four levels—"severe anxiety," "moderate anxiety," "light anxiety," and "normal anxiety"—based on participants' self-reports.

Recent studies have used the DASPS dataset to detect and classify anxiety levels. Chatterjee et al. [23] trained several traditional classifiers, such as Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Random Forest (RF), Decision Tree (DT), and Gaussian Naive Bayes (GNB), by extracting time and frequency domain features. They classified anxiety into two and four categories using DASPS. The highest accuracy was obtained using the KNN classifier that achieving 83.8% for both classification scenarios. Similarly, a Chebyshev chaotic map-based technique was introduced by Daneshmand et al. [12]. They classified anxiety into two and four categories using the DASPS dataset with Decision Tree (DT) and K-Nearest Neighbor (KNN) methods, achieving accuracies of 93.75% and 100% for the binary and four-class scenarios, respectively. Many studies have utilized time and frequency features to classify anxiety levels using EEG signals [12], [14], [22]–[24]. Another study that focused on frequency domain features was conducted by Muhammad et al. [24]. They employed various traditional machine learning techniques to detect different states of anxiety using the DASPS dataset. The highest accuracy was achieved using the Random Forest (RF) classifier, with 94.90% for binary classification and 92.74% for four-class classification. Although these studies achieved high accuracy, they relied on manual feature extraction, which can be time-consuming and dependent on domain expertise.

While the above methods utilized traditional machine learning for anxiety detection and relied on manual feature extraction, deep learning has demonstrated promising performance in EEG analysis [14], [25], [26]. Deep learning not only learns complex patterns in EEG signals but also automatically extracts relevant features, which makes it a more efficient approach. Maheshwari et al. [25] used various EEG datasets to classify different levels of emotions. They trained a rhythm-specific multi-channel convolutional neural network (CNN) using five different EEG frequency bands (i.e., delta, theta, alpha, beta, and gamma). The accuracy they achieved on the DASPS dataset was 53.45% for two-class classification. Similarly, Agarwal et al. [26] developed a new deep learning model using 1D convolutional and long short-term memory (LSTM) layers. This network was trained on several EEG datasets to classify emotional states. In their work, they provided insights into the relationship between different emotions and various scalp regions. Their model achieved an accuracy of around 71% for both binary and multiclass scenarios on the DASPS dataset. In another study, Shikha [14] employed a stacked sparse autoencoder to classify anxiety using EEG signals collected from the DASPS dataset. They extracted features from different domains, including time, frequency, and time-frequency. Their deep learning model achieved an accuracy of 83.93% for classifying normal and anxious trials. In 2024, Ghonchi et al. [27] applied a preprocessing method to EEG signals to transform their representation. Their approach involved generating scalp maps at each time point based on the spatial arrangement of EEG channels. Thereby they preserved spatial information within the data. Using a sliding window technique, they created data segments of varying lengths, ranging from 100 milliseconds to 3 seconds. To classify anxiety levels, they employed a convolutional-recurrent neural network that they achieved accuracies of 94.24% and 92.58% for binary and multi-class classification, respectively. Although the approach by Ghonchi et al. [27] demonstrated high classification accuracy using spatial scalp maps and a convolutional-recurrent network, their method relied on explicit spatial transformations of EEG data. In contrast, our study focuses on the performance of a novel deep learning-based feature extractor for anxiety classification. Instead of generating scalp maps, our model is designed to automatically capture frequency dependencies from raw EEG signals without requiring manually crafted representations. The primary objective of this work is to assess the performance of this novel feature extraction method in distinguishing different anxiety levels.

This paper presents a novel deep learning model for detecting and classifying anxiety states using frequency features and an attention module. Utilizing the DASPS dataset and labels provided by [22], EEG trials are categorized into two and four classes. The proposed model autonomously extracts optimal frequency features directly from raw data and as a result, eliminates the need for manual feature extraction. Although traditional pre-processing steps are vital for noise reduction, they often introduce biases and constraints. By bypassing these steps, the proposed approach preserves the integrity of raw EEG data. The following sections detail the methodology, with Section III outlining the results and Section IV providing the conclusions.

II. Material and Methods

A. Datasets

A DASPS dataset, which stands for "A Database for Anxious States Based on a Psychological Stimulation", available on IEEE DataPort[1] [22], employs an innovative approach by recording Electroencephalogram (EEG) signals from 23 participants as they underwent anxiety induction through face-to-face psychological tasks. The DASPS dataset follows an experimental design consisting of six 30-seconds trials for each participant. Each trial comprises two 15-seconds phases. In the first phase, the participant listens to an emotional scenario narrated by a psychotherapist. In the second phase, the participant attempts to recall the scenario described in the first phase. After completing each trial, the participant rates their feelings using the Self-Assessment Manikin (SAM), which evaluates both valence and arousal. Based on these ratings, the trials were classified as either normal or anxious. This binary classification resulted in 67 normal trials and 71 anxious trials. In a second labeling approach, the anxious trials were further divided into four categories: normal, light, moderate, and severe anxiety. In this classification, 65 trials were categorized as normal, 43 as light anxiety, 15 as moderate anxiety, and 15 as severe anxiety. Further details about this dataset can be found in [22].

B. Data Analysis

The proposed data analysis framework is applied to the DASPS dataset. The beta band ($12 - 30$ Hz) is generally the most relevant for anxiety classification using EEG. However, changes in theta ($4 - 8$ Hz) and alpha ($8 - 12$ Hz) power can also signal anxiety, varying by study and individual differences. Some research suggests a potential link between increased theta activity and anxiety, though it is less prominent [24], [28]. Therefore, For this dataset, a bandpass filter with a frequency range of 4–45 Hz is employed, along with downsampling to 128 Hz, to capture key EEG frequency bands relevant to cognitive and emotional processes. In general, This standardized preprocessing approach enhances the generalizability of models trained on the data. By systematically preparing the dataset, the pipeline facilitates effective feature extraction and deep learning analysis.

C. Multi-View Convolutional Attention

In this section, our Multi-View Convolutional Attention (MVCA-Net) architecture is described. This model is designed to automatically extract and analyze frequency-based features that characterize anxiety. Our proposed model integrates convolutional and transformer layers to extract frequency features. Attention modules were added to the model to dynamically focus on patterns of salient brain signals, as shown in Figure 1.

Frequency feature extraction is essential in time-series data analysis, particularly in EEG signal processing, where different frequency bands correspond to distinct neural activities.

[1] `https://ieee-dataport.org/open-access/dasps-database`

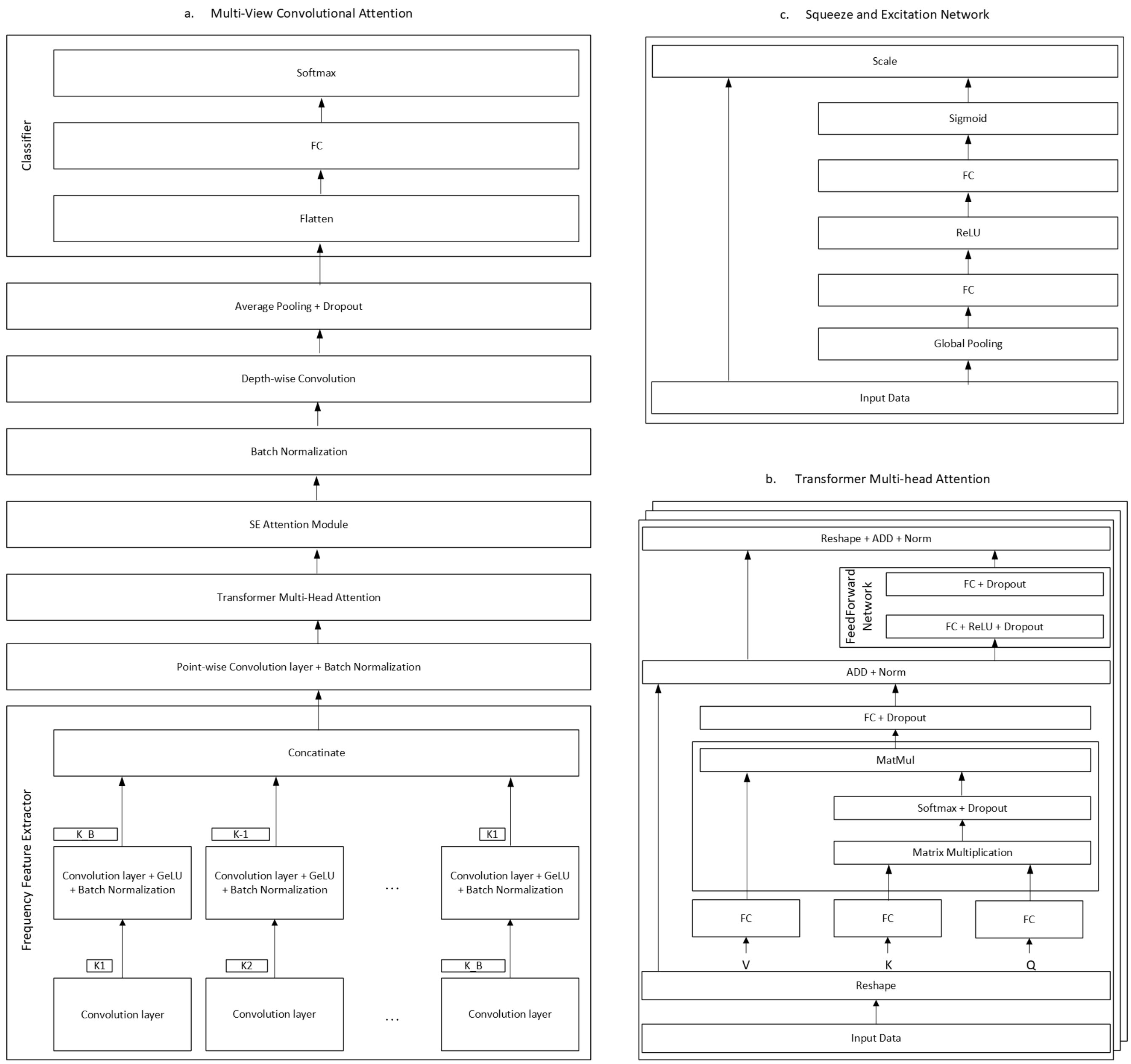

Fig. 1. Overall procedure of the Multi-View Convolutional Attention Network. This network consists three parts: Frequency feature extractor (a), Transformer multi-head attention (b), and SE attention (c).

Traditional methods rely on predefined filters, but modern deep learning architectures provide a data-driven approach to automatically extract these features [29]–[31].

The MVCA-Net proposed in this paper consists of three distinct blocks. The initial block incorporates frequency-specific features using convolutional kernels of varying sizes. It enables to extract multi-frequency characteristics from the input data as shown in Figure 1-a. The kernel sizes correspond to various temporal resolutions, which allows the model to focus on different frequency bands. Smaller kernel sizes extract high-frequency components (rapid signal changes), while larger kernel sizes capture low-frequency components (slower trends in the data). The kernels are shown as K_b in Figure 1-a. we define a kernel list that dynamically generates various kernel sizes based on the frequency range of the input signals. The kernel list is computed as 1.

$$K_b = \left[\frac{f}{2b} \right], b = 1, 2, 3, ..., B \tag{1}$$

Which f is the frequency rate and B is the target number of frequency domain features. This module uses two convolutional layers for each kernel size K_b, which works in parallel. The first convolutional layer filters primarily across the temporal dimension. The second convolutional layer takes the output of the first layer and tries to refine the frequency features. The intention behind this sequential filtering within each path is to allow the network to first extract features at one frequency scale $k1, K2, ..., K_B$ and then subsequently process or contextualize these features using a different scale $K_B, K_{B-1}, ..., K1$, enabling the learning of hierarchical frequency characteristics directly within each parallel branch before aggregation. Then batch normalization normalizes the output to improve stability and speed during training. After that, all the features extracted concatenate to prepare for further layers.

After extracting frequency features, another layer performs a point-wise convolution with a kernel and a stride of 1. This layer refines the extracted features without altering the input's spatial dimensions. It combines information across channels, emphasizes critical patterns while suppressing irrelevant ones, and adjusts the feature representation to match the required dimensions for subsequent processing. This transformation prepares the data for the next block which is Transformer, by optimizing feature representations, ensuring dimensional compatibility, and enhancing the input's ability to capture attention-based relationships effectively.

After the point-wise convolutional layer, a transformer multi-head attention is introduced to refine the representation of the frequency features, which serve as input to this module as shown in Figure 1-b.

The module begins by restructuring the input EEG data (originally organized as [Batch, Filter, Channel, TimeSample]) to merge the Filter and Channel dimensions into a unified embedding dimension. This reshaping step transforms the data into a format compatible with attention-based processing (structured as [Batch, Embedding, TimeSample]). Next, the module initializes three distinct linear layers to independently project the input into Query (Q), Key (K), and Value (V) matrices for multi-head self-attention. The number of attention heads (num_heads) determines how the embedding dimension (embed_dim) is partitioned, requiring embed_dim to be divisible by num_heads to ensure equal feature distribution across all heads. Finally, the reshaped EEG data are passed through the linear layers to generate the Q, K, and V matrices. These are fed into the self-attention mechanism, enabling the model to learn dependencies across time samples and spatial/spectral features in the EEG signals.

The self-attention mechanism is implemented using a scaled dot-product attention function. This part computes the similarity between Query and Key using matrix multiplication, scales the result, and applies a softmax function to generate attention weights. These weights determine the importance of different time steps in the EEG sequence. The final attended representation is obtained by multiplying the attention weights by the Value (V) matrix. This step allows the model to focus on the most relevant frequency features while suppressing the less important ones. After attention is computed, the output is passed through a feedforward network, followed by a dropout layer. The processed features are then added to the reshaped format of the input data.

The module processes the self-attention output through a feedforward network (FFN) to enhance feature extraction. The FFN consists of two linear layers with a ReLU activation function between them. These allow the model to learn complex representations of EEG signals. A dropout layer is applied to prevent overfitting before mapping the features back to their original embedding dimension. Following this, a residual connection is added, where the FFN output is summed with the original input and normalized to stabilize the training. Finally, the tensor is reshaped back to its original EEG format in order to ensure that the output maintains the relationships necessary for subsequent processing.

Following the multi-head attention mechanism, a squeeze-and-excitation (SE) attention block is employed to further enhance feature representation by focusing on the most informative channels, as shown in Figure 1-c. SE attention operates by adaptively recalibrating channel-wise feature responses, allowing the model to focus on critical frequency components while suppressing less relevant ones.

The SE attention block begins with a global pooling operation, where the input features are aggregated across the spatial and temporal dimensions to generate a compact, channel-wise descriptor. This global representation captures the overall importance of each channel in the frequency feature map. Next, the descriptor is passed through a pair of fully connected layers with a non-linear activation function, modeling the complex interdependencies among the channels. The output of fully connected layers is a set of channel-wise attention weights, which are rescaled using a sigmoid activation function to ensure values are between 0 and 1. These weights are then applied to the original input features via element-wise multiplication, effectively highlighting critical channels while diminishing the impact of less significant ones. By integrating the SE attention block, the model gains the ability to refine its focus on relevant frequency features, further enhancing its capacity to capture subtle patterns in the data.

Following the SE attention mechanism, a depthwise convolution is applied to the data. The depthwise convolution ensures that the spatial dependencies between channels are preserved while maintaining a low computational cost. Furthermore, using the depthwise convolution method helps the model focus on local attention, specifically emphasizing the relationship between adjacent frequencies. This local attention reduces the computational cost compared to a full-attention mechanism. The depthwise convolution operates on each channel independently, which is suitable for spatial information processing in EEG data. In the following, batch normalization and average pooling are applied before the final dropout layer to enhance the performance of the depthwise convolutional operations.

The classifier block includes a Dense Layer with 2 or 4 neurons and a softmax activation function. 2 refers to normal and anxious classes, and 4 refers to normal, light, moderate, and severe anxiety classes.

III. Experimental Results

In this section, the performance of MVCA-Net is evaluated to detect and classify different levels of anxiety using the DASPS dataset. We conducted three experiments on our model: one using only frequency feature extraction, another incorporating frequency features with an SE attention module, and the third utilizing the complete MVCA-Net model. We also compare our results with existing methods used for this dataset. A summary of all the parameters and setups used in this study is presented in Table I. For the model setup, we

TABLE I
PARAMETERS AND SETUPS USED FOR TRAINING MODEL.

	Parameter Name	Value
Data preparation	Band-pass filter	4-45 Hz
	Normalization	Z-score algorithm
	down sample	128 Hz
Proposed Model	$Kernel(k)$	$1, 9, 17, 25, 33, 41, 49, 57$
	F_1	2
	F_2	1
	B	8
	$Activations$	$GeLU$
	$Number of heads$	4
	$Number of layers$	3
	$Transformer dropout$	0.5
	$Transformer embed\_dim$	112
	D	14
	$Pooling$	4
	$Classifier$	$Softmax$
	$Loss$	$CrossEntropy$
	$Optimizer$	$Adam$
	$Epochs$	150
	$Batchsize$	32
	$K-fold$	5

TABLE II
THE COMPARISON OF THE ACCURACIES ON DIFFERENT CLASSES ON DASPS DATASET TO CHECK THE EFFECT OF TRANSFORMER AND ATTENTION MODULE.

Model	# Classes	Accuracy
Feature extractor	2 classes	$75.31\% \pm 1.97\%$
	4 classes	$63.91\% \pm 1.39\%$
Feature extractor + SE Attention	2 classes	$76.47\% \pm 1.74\%$
	4 classes	$64.73\% \pm 1.52\%$
MVCA-Net	**2 classes**	$\mathbf{82.94\% \pm 1.22\%}$
	4 classes	$\mathbf{74.05\% \pm 3.96\%}$

employed cross-entropy as the loss function and the Adam optimizer. We trained the model over 150 epochs and utilized k-fold cross-validation with $k = 5$. This approach ensures robust evaluation and enhances the reliability of our findings in classifying anxiety levels.

To evaluate the performance of the MVCA-Net, we also conducted several experiments using the DASPS dataset. These experiments involve classifying anxiety states using a frequency feature extractor, a frequency feature extractor combined with an SE attention module, and the complete MVCA-Net model as reported in Table II.

In deep learning models, attention mechanisms and transformers play a critical role in addressing the limitations of traditional feature extractors. The sequential processing and dynamic weighting of features inherent in attention mechanisms ensure that critical frequency components are emphasized. Meanwhile, transformers excel at capturing global dependencies. They are able to characterise complex patterns of the model that span across time and frequency domains. Together, these components significantly improve the performance of the model, particularly in scenarios involving multiple classes, where the relationships between features become more intricate. This improvement demonstrates the transformative impact of attention and transformers in deep learning applications for frequency-based data.

Table II presents the impact of incorporating SE attention mechanisms and transformer-based modules on classification accuracy in two- and four-class scenarios. The baseline model, consisting solely of a frequency feature extractor, achieves $75.31\% \pm 1.97\%$ accuracy for the two-class classification and $63.91\% \pm 1.39\%$ for the four-class classification. Although these results are satisfactory for simpler tasks, they are limited in their ability to capture intricate patterns in the data. This is because, in EEG signals, the repetitive and cyclical nature of the data can create significant similarities between different segments. It makes it difficult for the model to accurately identify distinguishing features between classes. These similarities often introduce ambiguity, especially in multi-class classification scenarios, where the boundaries between classes become less distinct. As evident in Table II, the model performs better in extracting meaningful frequency features in the two-class experiment, as the reduced number of classes simplifies the decision boundaries and allows the model to focus on more prominent and distinguishable patterns within the data.

When a SE attention mechanism is added to the frequency feature extractor, the accuracy improves to $76.47\% \pm 1.74\%$ for two classes and $64.73\% \pm 1.52\%$ for four classes. This improvement can be attributed to the ability of attention to dynamically prioritize relevant frequency features while suppressing less important ones. The most significant boost in accuracy is observed when a transformer module is introduced along with the frequency feature extractor and SE attention mechanism. In this configuration, the model achieves $82.94\% \pm 1.22\%$ accuracy for two classes and $74.05\% \pm 3.96\%$ for four classes. This substantial improvement underscores the ability of transformers to model long-range dependencies and complex relationships within the frequency features. The multihead attention mechanism of the transformer enables the model to attend to multiple aspects of the input features simultaneously and extracts diverse patterns across the temporal and spectral domains.

The results in Table II highlight the differences between different states of anxiety, through EEG. The DASPS dataset focuses on the more nuanced domain of affective states—particularly anxiety. The drop in accuracy for four-class classification means that the increased number of classes introduces a greater challenge. This drop in accuracy when distinguishing between four classes suggests that subtle differences between anxiety levels may require richer spatio-temporal or connectivity features for accurate classification.

We conducted a comparative analysis of our proposed model against methods previously employed in the DASPS dataset for anxiety classification. Table III presents a comprehensive summary of these methods. A review of the table reveals that most of the previous studies on this dataset mainly relied on feature extraction techniques combined with

TABLE III

THE COMPARISON OF DIFFERENT METHODS.

State-of-the-arts	# Classes	Features	Classifier	Accuracy
Traditional Machine Learning				
Chatterjee et al. [23]	2 classes 4 classes	Hjorth parameters and 4 EEG band powers	KNN	83.8% 83.8%
Jin et al. [7]	2 classes 4 classes	Time-Domain and Frequency-Domain analysis	Random Forest	78.34% 70.45%
Daneshmand et al. [12]	2 classes 4 classes	Innovative Chebyshev chaotic map-based features	KNN	93.75% 100%
Muhammad et al. [24]	2 classes 4 classes	Asymmetry index, rational index, and mean power	Random Forest	94.90% 92.74%
Deep Learning				
Baghdadi et al. [22]	2 classes 4 classes	Time, Frequency, and Time-Frequency features	Stacked Sparse Autoencoder	83.50% 74.60%
Maheshwari et al. [25]	2 Classes	No features extraction	CNN	53.45%
Agarwal et al. [26]	2 classes: Valence 2 classes: Arousal	No features extraction	1D CNN-LSTM	71.93% 71.63%
Shikha et al. [14]	2 classes	Time, Frequency, and Time-Frequency features	Stacked Sparse Autoencoder	83.98%
Ghonchi et al. [27]	2 classes 4 classes	No feature extraction	Convolution-Recurrent Neural Network	94.24% 92.58%
Our work	**2 classes** **4 classes**	**frequency features extracted by proposed model**	**MVCA-Net**	**82.94%** **74.05%**

traditional machine learning algorithms. Among these, the approaches of Muhammad et al. [24] and Daneshmand et al. [12] demonstrated the highest accuracy within the scope of traditional machine learning frameworks. In particular, two studies ([25], [26]) adopted a different approach by directly feeding raw EEG data into deep learning models for anxiety classification. They incorporated various preprocessing steps beforehand. When utilizing deep learning to extract features directly from raw EEG data, our findings align with those of previous studies, particularly those that employed deep learning models and raw data, such as [25] and [26]. The consistency of our results with these works underscores the effectiveness of deep learning in analyzing raw EEG data, even when compared to methods relying on pre-extracted features or traditional machine learning algorithms. This compatibility further validates the robustness of our proposed model and the significance of leveraging raw data for enhanced performance in EEG-based tasks. The relatively lower accuracy of MVCA-Net compared to [27] can be attributed to its exclusive reliance on frequency-domain features. The proposed model focuses solely on extracting frequency features from raw EEG signals. While frequency analysis is essential for understanding EEG signal characteristics, it does not capture spatial relationships between electrodes or the temporal dynamics of neural activity. Unlike other approaches, our method neither involves handcrafted feature extraction nor alters the data representation; instead, it places the burden of extracting frequency features entirely on the model itself. As demonstrated in [27], integrating additional modalities, such as spatial information or time-dependent patterns, can significantly enhance the performance of anxiety classification models. On the other hand, this approach significantly reduces runtime and makes MVCA-Net a more efficient alternative, as shown in Table IV.

TABLE IV

COMPARING THE AVERAGE RUNTIME OF MVCA-NET AND [27]

Model	Preprocessing Runtime	Epoch Runtime
Ghonchi et al. [27]	65 seconds	79 seconds
MVCA-Net	**0.9 second**	**8 second**

Table IV compares the runtime performance of MVCA-Net with the method proposed by Ghonchi et al. [27] on the DASPS dataset. While the accuracy of MVCA-Net is lower in comparison, this model is still in its early stages and currently relies only on frequency-domain features. Despite this limitation, the results demonstrate a significant reduction in computational cost. MVCA-Net achieves a preprocessing runtime of just 0.9 seconds compared to 65 seconds in Ghonchi et al.'s method, and an epoch runtime of 8 seconds versus 79 seconds. These improvements indicate the potential of MVCA-Net for real-time applications. Future enhancements, such as incorporating additional spatial and temporal features, could further improve its classification performance.

IV. Conclusion

This study introduced a novel deep learning approach to effectively extract frequency features from EEG signals for anxiety detection and severity classification. The proposed framework follows a structured data processing pipeline, including filtering, model training, and classification. Initially, a bandpass filter was applied to remove unwanted noise from the raw EEG signals, followed by downsampling to 128 Hz. This preprocessing step not only standardizes the data but also reduces the computational cost associated with model training. The MVCA-Net proposed in this paper consists of three main components: a frequency feature extractor, a Transformer module, and an SE attention mechanism. Each component plays a crucial role in capturing different aspects of the EEG signals for anxiety classification. The model's performance was evaluated using a 5-fold cross-validation strategy which demonstrates a significant improvement over previously published deep learning methods. Our approach successfully extracted meaningful features from EEG data, achieving reliable classification accuracy. Although this model achieves lower accuracy compared to the previous work [27], it is still in progress, and the current stage focuses solely on frequency features. Notably, while our method performs competitively with traditional machine learning techniques, it offers the advantage of reducing reliance on extensive preprocessing and handcrafted feature extraction, making it a more automated and efficient alternative.

References

[1] K. Sardeshpande and V. R. Thool, "Psychological stress detection using deep convolutional neural networks," in *Computer Vision and Image Processing: 4th International Conference, CVIP 2019, Jaipur, India, September 27–29, 2019, Revised Selected Papers, Part II 4*. Springer, 2020, pp. 180–189.

[2] M. Jyothirmayi, M. Sidharth, B. Vijay, R. Khandelwal, K. Raju, and P. Dinesh, "Early detection and analysis of anxiety in adolescents," *Mathematical Models and Computer Simulations*, vol. 15, no. 5, pp. 956–967, 2023.

[3] H. Mo, S. C. Hui, X. Liao, Y. Li, W. Zhang, and S. Ding, "A multimodal data-driven framework for anxiety screening," *IEEE Transactions on Instrumentation and Measurement*, 2024.

[4] W. H. Organization, *World mental health report: Transforming mental health for all*. World Health Organization, 2022.

[5] P.-E. Bruand, T. Gelle, J. A. D. Tine, A. F. Ndao, L. Kleinebreil, D. Moussaoui, P.-M. Preux, F. Boumediene *et al.*, "Evaluating the effects of mental health e-learning on the knowledge, attitudes and practices of allied and healthcare professionals in senegal," *Clinical Epidemiology and Global Health*, 2024.

[6] W. H. Organization *et al.*, "Mental health and covid-19: early evidence of the pandemic's impact: scientific brief, 2 march 2022," World Health Organization, Tech. Rep., 2022.

[7] N. Jin, Y. Wu, J. Park, Z. Qin, and H.-N. Liang, "Brain-metaverse interaction for anxiety regulation," in *2023 9th International Conference on Virtual Reality (ICVR)*. IEEE, 2023, pp. 385–392.

[8] D. Kamińska, K. Smółka, and G. Zwoliński, "Detection of mental stress through eeg signal in virtual reality environment," *Electronics*, vol. 10, no. 22, p. 2840, 2021.

[9] M. A. Rahman, D. J. Brown, M. Mahmud, M. Harris, N. Shopland, N. Heym, A. Sumich, Z. B. Turabee, B. Standen, D. Downes *et al.*, "Enhancing biofeedback-driven self-guided virtual reality exposure therapy through arousal detection from multimodal data using machine learning," *Brain Informatics*, vol. 10, no. 1, p. 14, 2023.

[10] M. Yadav, M. N. Sakib, E. H. Nirjhar, K. Feng, A. H. Behzadan, and T. Chaspari, "Exploring individual differences of public speaking anxiety in real-life and virtual presentations," *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1168–1182, 2020.

[11] J. Šalkevicius, R. Damaševičius, R. Maskeliunas, and I. Laukienė, "Anxiety level recognition for virtual reality therapy system using physiological signals," *Electronics*, vol. 8, no. 9, p. 1039, 2019.

[12] F. Daneshmand-Bahman and A. Goshvarpour, "Anxiety recognition using a new eeg signal analysis approach based on sample density in a chebyshev chaotic map," in *Affective Computing in Healthcare: Applications based on biosignals and artificial intelligence*. IOP Publishing Bristol, UK, 2023, pp. 1–1.

[13] A. Arsalan and M. Majid, "A study on multi-class anxiety detection using wearable eeg headband," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 12, pp. 5739–5749, 2022.

[14] Shikha, M. Agrawal, M. A. Anwar, and D. Sethia, "Stacked sparse autoencoder and machine learning based anxiety classification using eeg signals," in *Proceedings of the First International Conference on AI-ML Systems*, 2021, pp. 1–7.

[15] L. S. Mokatren, R. Ansari, A. E. Cetin, A. D. Leow, O. Ajilore, H. Klumpp, and F. T. Y. Vural, "Eeg classification based on image configuration in social anxiety disorder," in *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 2019, pp. 577–580.

[16] E. H. Nirjhar, A. Behzadan, and T. Chaspari, "Exploring bio-behavioral signal trajectories of state anxiety during public speaking," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 1294–1298.

[17] R. Sharma, T. Guha, and G. Sharma, "Multichannel attention network for analyzing visual behavior in public speaking," in *2018 ieee winter conference on applications of computer vision (wacv)*. IEEE, 2018, pp. 476–484.

[18] M. Elgendi, V. Galli, C. Ahmadizadeh, and C. Menon, "Dataset of psychological scales and physiological signals collected for anxiety assessment using a portable device," *Data*, vol. 7, no. 9, p. 132, 2022.

[19] A. Arza, J. M. Garzón-Rey, J. Lázaro, E. Gil, R. Lopez-Anton, C. de la Camara, P. Laguna, R. Bailon, and J. Aguiló, "Measuring acute stress response through physiological signals: towards a quantitative assessment of stress," *Medical & biological engineering & computing*, vol. 57, pp. 271–287, 2019.

[20] L. Malviya, S. Khandelwal, and S. Mal, "Mental stress detection using eeg extracted frequency bands," in *Soft Computing: Theories and Applications: Proceedings of SoCTA 2021*. Springer, 2022, pp. 283–293.

[21] Z. Shen, G. Li, J. Fang, H. Zhong, J. Wang, Y. Sun, and X. Shen, "Aberrated multidimensional eeg characteristics in patients with generalized anxiety disorder: a machine-learning based analysis framework," *Sensors*, vol. 22, no. 14, p. 5420, 2022.

[22] A. Baghdadi, Y. Aribi, R. Fourati, N. Halouani, P. Siarry, and A. M. Alimi, "Dasps: A database for anxious states based on a psychological stimulation," *arXiv preprint arXiv:1901.02942*, 2019.

[23] D. Chatterjee, R. Gavas, and S. K. Saha, "Detection of mental stress using novel spatio-temporal distribution of brain activations," *Biomedical Signal Processing and Control*, vol. 82, p. 104526, 2023.

[24] F. Muhammad and S. Al-Ahmadi, "Human state anxiety classification framework using eeg signals in response to exposure therapy," *Plos one*, vol. 17, no. 3, p. e0265679, 2022.

[25] D. Maheshwari, S. K. Ghosh, R. Tripathy, M. Sharma, and U. R. Acharya, "Automated accurate emotion recognition system using rhythm-specific deep convolutional neural network technique with multi-channel eeg signals," *Computers in Biology and Medicine*, vol. 134, p. 104428, 2021.

[26] R. Agarwal, M. Andujar, and S. Canavan, "Classification of emotions using eeg activity associated with different areas of the brain," *Pattern Recognition Letters*, vol. 162, pp. 71–80, 2022.

[27] H. Ghonchi, T. Foulsham, and S. Ferdowsi, "Assessing neural patterns of anxiety using deep learning: An eeg study," in *2024 32nd European Signal Processing Conference (EUSIPCO)*. IEEE, 2024, pp. 1596–1600.

[28] W. Liu, G. Li, Z. Huang, W. Jiang, X. Luo, and X. Xu, "Enhancing generalized anxiety disorder diagnosis precision: Mstcnn model utilizing high-frequency eeg signals," *Frontiers in Psychiatry*, vol. 14, p. 1310323, 2023.

[29] W. Chen, C. Wang, K. Xu, Y. Yuan, Y. Bai, and D. Zhang, "D-fast: Cognitive signal decoding with disentangled frequency-spatial-temporal attention," *IEEE Transactions on Cognitive and Developmental Systems*, 2024.

[30] E. Eldele, Z. Chen, C. Liu, M. Wu, C.-K. Kwoh, X. Li, and C. Guan, "An attention-based deep learning approach for sleep stage classification with single-channel eeg," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 809–818, 2021.

[31] Y. Ding, N. Robinson, C. Tong, Q. Zeng, and C. Guan, "Lggnet: Learning from local-global-graph representations for brain–computer interface," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.