A Hybrid Common Spatial Pattern with Attention-Based Convolutional Neural Networks for Motor Imagery EEG

[1]M. Abdullah Azzam, [1]Muhammad Afiq Che Man,
[1]Nur Amirah Abd Hamid, [1]Yuan You Tiew,
[1,2]Ibrahim Shapiai
[1]Malaysia Japan International Institute of Technology
(MJIIT)
[2]Centre for Artificial Intelligence and Robotics (CAIRO)
Universiti Teknologi Malaysia, Jalan Sultan Yahya Petra,
54100 Kuala Lumpur, Malaysia

Abstract* - Motor imagery on electroencephalogram (EEG) signals is widely used in brain-computer interface (BCI) systems with many exciting applications. Recently, many deep learning classifiers have been adopted, especially Convolutional Neural Networks (CNNs) in BCI application. However, CNNs suffer from the loss of salient features during training, causing the spatial invariant problem that affects the performance. This study develops a framework using CSP and Short-Time Fourier Transform (STFT) with Attention Convolutional Neural Network (CSP-STFT CNN) for EEG BCI classification. The features from CSP are translated into the spatial domain using STFT as input to attention-based CNN as the classifier. This framework uses attention-based CNNs to classify the collected spatial images across different test subjects. Finally, the performance of the CSP-STFT CNN is validated on BCI benchmark datasets, Competition III dataset Iva. The proposed CSP-STFT offers a promising result; the classifier achieved better performance in terms of classification accuracy, averaging 80% across all five subjects for Competition III dataset IVa. The precision and recall are excellent too, ranging around 0.8-0.9. In general, the proposed CSP-STFT CNN can offer richer joint spatiotemporal features as inputs to classifiers, whereas using an Attention-CNN classifier improves upon the earlier problems suffered by CNNs.

Keywords - Common Spatial Pattern, Short Time Fourier Transform, CNN, Attention Mechanism

I. INTRODUCTION

The Brain Computer Interface (BCI) EEG refers to BCI technology that utilizes EEG to measure and analyze brain activity for the purpose of controlling a computer or other external device [1]. EEG is a non-invasive technique that involves placing electrodes on the scalp to detect the electrical activity generated by the brain. In BCI EEG, the EEG signals are recorded and processed in real-time to extract features that can be used to control a computer or device. These features can include signals associated with different types of brain activity, such as alpha, beta, gamma, and delta waves. The motor imagery (MI) EEG, having been widely used in BCI systems as part of neuroimaging and rehabilitation, are considered under the event related potential (ERP) derivative of EEG [2]. In the motor imagery paradigm, by having the user to imagine the execution of a specific movement with a designated limb, the command is encoded by altering the rhythmic activity in locations concerning the sensorimotor cortex that would typically correspond to this limb [3]. After recording the signal, the BCI system would proceed to decode the intended command correctly. Nonetheless a significant problem in EEG-based BCI systems is the limited quality and resolution of the signal due to volume conduction effects, a low signal-to-noise ratio, and the non-stationary nature of EEG [4]. In order to improve the quality and get a better motor imagery signal, the noise and artefact signals should be eliminated through filtering process. For the EEG signals, there are a multitude of methods for filtering them such as temporal, spectral, and spatial filtering.

MI-BCI using EEG involves detecting and interpreting electrical signals produced by the brain when a person imagines performing a specific motor task. This type of BCI typically involves the use of an EEG headset or cap that is placed on the scalp to measure electrical activity in the brain. During a motor imagery task, the person is typically instructed to imagine performing a specific movement, such as moving their left hand or right foot [2]. The EEG headset then detects and records the electrical activity in the person's brain that is associated with this motor imagery. Similarly, to conventional BCI, the imagery EEG then is analyzed using signal processing algorithms and machine learning techniques to classify the different types of motor imagery tasks that the person is imagining.

CNN have shown promising results in the analysis of EEG signals [5]. The use of CNN allows for the automatic learning of relevant features from raw EEG data, which can be used for a variety of tasks such as classification, segmentation, and feature extraction. One of the main advantages of CNN is their ability to exploit spatial and temporal relationships within the data [6]. This is particularly useful for EEG signals, which have both spatial and temporal dimensions. CNN can be used to learn spatial patterns such as extracted features from CSP in the scalp topography of EEG signals [6]. There are various approaches to improve the overall BCI system utilizing CSP algorithms on the extraction part such as integration with STFT and classifier with attention mechanism applied together with CNNs to boost the classification [7]. It is important to determine the right combination of CSP and STFT and process it as inputs to Attention-CNNs, as well as tuning the Attention-CNNs so that it will be able to take in the features and classify them well. However, an improper feature extraction process could further reduce the quality of the image representation and eliminate important information from the EEG signal. Besides, the wrong arrangement and definition in the model building stage will also hinder the model from learning better.

II. RELATED WORKS

Earlier, machine learning is commonly used to classify the extracted features of the motor imagery EEG signal. For instance, the motor imagery EEG signal was classified using linear discriminant analysis (LDA) [8]. LDA is probably the most popular baseline classifier for BCI-EEG-based applications because it has a shallow computational requirement that makes it suitable for online BCI systems. In addition to that, this classifier is simple to be used and generally provides good results [9]. Meanwhile, several other classifiers such as linear SVM has been employed for kernel CSP (Ma et al., 2018). Besides SVM, extreme learning machines have been employed for classifying motor imagery EEG signals [10]. There are also tree-based algorithms that are employed as classifiers such as XGBoost with gradient boosting as its feature selection process [11].

With the rise of deep learning to prominence in computer vision, speech recognition, and recommendation systems, among others, it is natural that the techniques would spill over into the medical field. In fact, in motor imagery, many of techniques based on deep learning are beginning to be employed as classifiers [6]. It is without surprise given that the motor imagery study has been ongoing for decades with many datasets being produced, allowing these deep learning classifiers to be able to perform even better classification with little to no handcrafted that is involved in feature extraction [12]. Nonetheless, while CNNs are being introduced in motor imagery classification, CNNs suffer from the loss of salient features during training, causing the spatial invariant problem that affects the performance.

In recent years, researchers began to study the utilization of attention mechanisms in the computer vision domain area [13]. Attention networks in computer vision are defined as a neural attention mechanism that equips a neural network with the ability to focus on features of its inputs. Even though attention mechanism is still a new approach in the computer vision domain, many studies have employed attention mechanisms in their works for image classification and object localization. Attention mechanism could alleviate the ability of CNN to be spatial invariance to the input images with the small pixel size. In medical image analysis, incorporating the attention mechanism into the CNNs architecture is widely used to extract and exploit the local features. Local features are hardly exploited in medical images due to small regions of interest (ROI) (e.g., lesion, hippocampal, etc.). Examples of attention mechanism has been utilized in ultrasound detection[14] using self-attention and attention-gated models respectively. Similar feat is achieved with chest X-ray images [15] involving the implementation of saliency map extraction, recurrent attention and incorporating attention to a pre-trained VGG model. To our best, there are no previous studies that implement the idea of utilizing attention mechanisms in motor imagery classification with the aforementioned feature extraction, thus making our respective model become the first model to put forth the concept into implementation.

III. CSP-STFT CNN

The BCI system process was modified via the proposal of a feature extraction process now incorporating CSP and STFT, as well as a feature classification subsystem comprising pre-trained CNNs augmented with attention mechanism, presented in Figure 1. As such, the proposed BCI framework introduces two novelties for the feature extraction and classification processes (labelled in the figure as the shaded boxes in red dotted lines) as a unified platform named CSP-STFT CNN to better handle motor imagery classification.

A. CSP-STFT CNN

In this study, BCI Competition III set IVa by the Berlin Institute of Technology which commonly used by several researchers [16, 17]. The data sets are two classes motor imagery of EEG signal with 118 channels.

Signal pre-processing is carried out using a bandpass filter to remove unnecessary frequency bands, which may carry artefacts, i.e., a power line artefact or a muscular artefact in the signal. The EEG filter is applied in each channel using a 3rd- or 4th-order bandpass Butterworth filter with a frequency cut-off of 8 Hz to 13 Hz. The order number of the Butterworth filter is different between the datasets, depending on the characteristic of the motor imagery EEG signals. The bandpass Butterworth filter for the EEG signal had previously been used in previous studies as well[18]. Generally, the EEG signal has five types of frequency bands, which are delta (0.4-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), and gamma (above 30 Hz) [18]. The motor imagery signal predominantly works in the alpha band (8 Hz to 13 Hz) frequency, so in the signal pre-processing design of this study, the alpha band range is used as the frequency filter cut-off.

DOI 10.5013/IJSSST.a.24.02.03 ISSN: 1473-804x online, 1473-8031 print

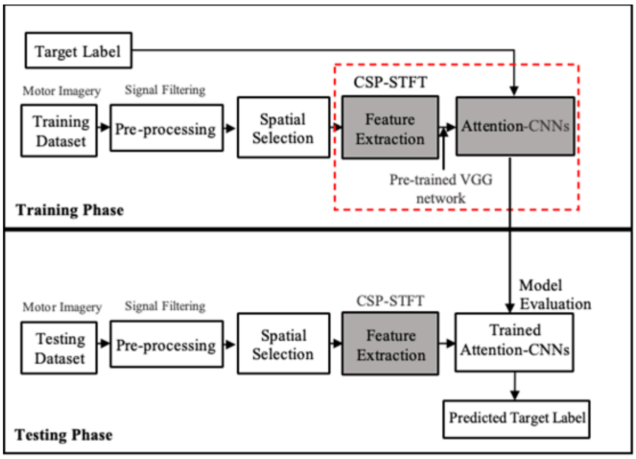

Figure 1. The proposed CSP-STFT CNN

The CSP algorithm is computed by simultaneous diagonalization of two covariance matrices. The normalized covariance matrix can be calculated using Equation (1):

$$C_y = \frac{1}{n_y}\sum \frac{E_j(y)E_j^T(y)}{trace\,(E_j(y)E_j^T(y))} \tag{1}$$

where Ej denotes EEG signal for the j-th trial, ny represent a number of trials for the y class, and y is a class (e.g. left and right). T is the transpose operator and trace is diagonally sum operation of matrix elements. Conventional CSP is employed to generate a Q-dimensional feature vector ŷ is formed from the variance of Ẑ, using Equation (2):

$$\hat{y}_q = \log\left(\frac{var(\hat{Z}_q)}{\sum_{q=1}^{Q} var(\hat{Z}_q)}\right) \tag{2}$$

where ŷq = q-th component of ŷ, Ẑq = q-th row of Ẑ, and var(Ẑq) = variance of the Ẑq vector. The selected pattern from CSP will be transformed into STFT. The STFT is then set to image-based EEG at size of . Once the features is extracted from CSP and transform to image-based EEG using STFT, the generated image is paritioned into training and testing data for classification. For classification, CNN with attention mechanism is employed.

For this work, the standard VGG network (Simonyan & Zisserman, 2014) depth of 16 weight layers (VGG-16) with very small (3x3) convolution filters is utilized as baseline architecture in this work. However, what is different here is that a little modification is added to optimize the network architecture in the classification task of motor imagery. The idea is to utilize all convolutional layers in the VGG-16 in extracting the features of the input, which will be significantly useful for subsequent attention module. We determine the attention module as an essential component in our work to identify the salient spatial context of visual features in the filtered EEG spectrum images. As a result, the feature maps extracted from the convolutional layers are expected to be employed more efficiently, as well as any potential irrelevant noise being suppressed further.

For this work, the standard VGG network (Simonyan & Zisserman, 2014) depth of 16 weight layers (VGG-16) with very small (3x3) convolution filters is utilized as baseline architecture in this work. However, what is different here is that a little modification is added to optimize the network architecture in the classification task of motor imagery. The idea is to utilize all convolutional layers in the VGG-16 in extracting the features of the input, which will be significantly useful for subsequent attention module. We determine the attention module as an essential component in our work to identify the salient spatial context of visual features in the filtered EEG spectrum images. As a result, the feature maps extracted from the convolutional layers are expected to be employed more efficiently, as well as any potential irrelevant noise being suppressed further.

For this work, the standard VGG network (Simonyan & Zisserman, 2014) depth of 16 weight layers (VGG-16) with very small (3x3) convolution filters is utilized as baseline architecture in this work. However, what is different here is that a little modification is added to optimize the network architecture in the classification task of motor imagery. The idea is to utilize all convolutional layers in the VGG-16 in extracting the features of the input, which will be significantly useful for subsequent attention module. We determine the attention module as an essential component in our work to identify the salient spatial context of visual features in the filtered EEG spectrum images. As a result, the feature maps extracted from the convolutional layers are expected to be employed more efficiently, as well as any potential irrelevant noise being suppressed further.

TABLE I. DESCRIPTION ON SELECTED CONVOLUTIONAL LAYER, S

Convolutional layer, *s*	Number of channels, *n*	Pixel size of feature vector, f_i^s
7	256	112 x 112
10	512	56 x 56
13	512	28 x 28

The main idea of attention module is briefly described in Figure 3.7 where to jointly attend each f_i^s at each layer s to global feature vector g r to perceive the relevant features that are compatible to the global scale features extracted by g. Thus, we define the idea as compatibility score $C(\mathcal{F}^s, g) = \{c_i^s\}_{i=1}^{n}$ where c_i^s refers to the gating unit in the grid attention block as in Equation (3).

$$c_i^s = \Psi\sigma_1(W_f f_i^s + W_g g + b_g) + b_\psi \tag{3}$$

where contains a set of learnable parameter of linear transformations Ψ, W_f, W_g and bias term b_g, b_ψ. σ_1 is a rectified linear unit (ReLU), a non-linear activation function. , W_f and W_g epresents the learnable weight of local feature vectors at spatial location i and learnable weight of global feature vector respectively. We then perform softmax operation in Equation (4) on the compatibility score to obtain the normalized attention coefficient, α_i^s:

$$\alpha_i^s = \frac{e^{c_i^s}}{\sum_i e^{c_i^s}} \tag{4}$$

At each chosen layer s, weighted attended feature vectors g^s defined in Equation (5) is computed to be used for subsequent classification.

$$g^s = \sum_{i=1}^{n} \alpha_i^s f_i^s \tag{5}$$

B. Parameter Setup

Initially, all the setting parameters of the proposed method are predefined. In specific, all setting parameters are summarized in Table II.

TABLE II. SETTING PARAMETERS FOR CSP-STFT CNN

Parameters	Values/ Description
Image-based EEG size	224 x 224
Temporal Filter	3rd Order Bandpass Butterworth Filter between 8Hz and 15Hz
Feature extraction method	CSP-STFT
Feature Classification method	Attention-CNNs
Number of CSP Features (output)	6
Ratio of training and testing sets	80 : 20
Number of epochs	150 with early stopping minimum 60

C. Performance Evaluation

In order to validate the proposed technique, the following performance evaluation index are used as given in the following Equations:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$precision = \frac{TP}{TP + FP} \tag{7}$$

$$recall = \frac{TP}{TP + FN} \tag{8}$$

where, TP, TN, FP and FN are represented true positive, true negative, false positive and false negative, respectively.

IV. RESULTS AND DISCUSSION

Based on the testing results of CSP-STFT-Attention-CNN on Dataset 1 shown in Table 3, an average accuracy of 0.83 was obtained—an 15% increase compared to the benchmarked method of Integrated Selection-CSP-ELM [19]. In fact, subjects al and av had the highest train accuracy for the framework. Meanwhile, improvements could also be seen across other subjects aa, aw, and ay in terms of training and testing accuracies. It is interesting to see such a method with a modified version of CSP would slightly perform better than CSP-STFT CNN framework Precision and recall are relatively high across most of the subjects with averages around 0.814 and 0.912 respectively as seen in Table IV.

TABLE III. CLASSIFICATION ACCURACY FOR CSP-STFT CNN ON BCI COMPETITION III SET IVA

Subjects	Integrated Selection-CSP-ELM	CSP-STFT-Attention-CNN
aa	(0.80) 0.61	(0.81) 0.81
al	(0.98) 0.87	(0.96) 0.98
av	(0.94) 0.65	(0.82) 0.76
aw	(0.99) 0.60	(0.94) 0.87
ay	(0.78) 0.65	(0.94) 0.75
Average	**(0.90) 0.68**	**(0.89) 0.83**

TABLE IV. PRECISION AND RECALL SCORES FOR CSP-STFT CNN ON BCI COMPETITION III SET IVA

Subjects	Precision	Recall
aa	0.70	0.89
al	1.0	0.95
av	0.73	0.89
aw	0.91	0.83
ay	0.73	1.0
Average	**0.81**	**0.91**

As shown in Figure 2, the training and validation accuracies keep improving throughout the epochs albeit some oscillations every now and then. In accordance to that, the training and validation losses tend to go down as well as the training progresses and epochs increase. This paradigm is true for all subjects with the exception of subject ay. For this particular subject only as depicted on Figure 2(e), while the training accuracy improved to 0.8 to 1.0 range of accuracies, the validation accuracy does not truly go beyond 0.6.

DOI 10.5013/IJSSST.a.24.02.03 ISSN: 1473-804x online, 1473-8031 print

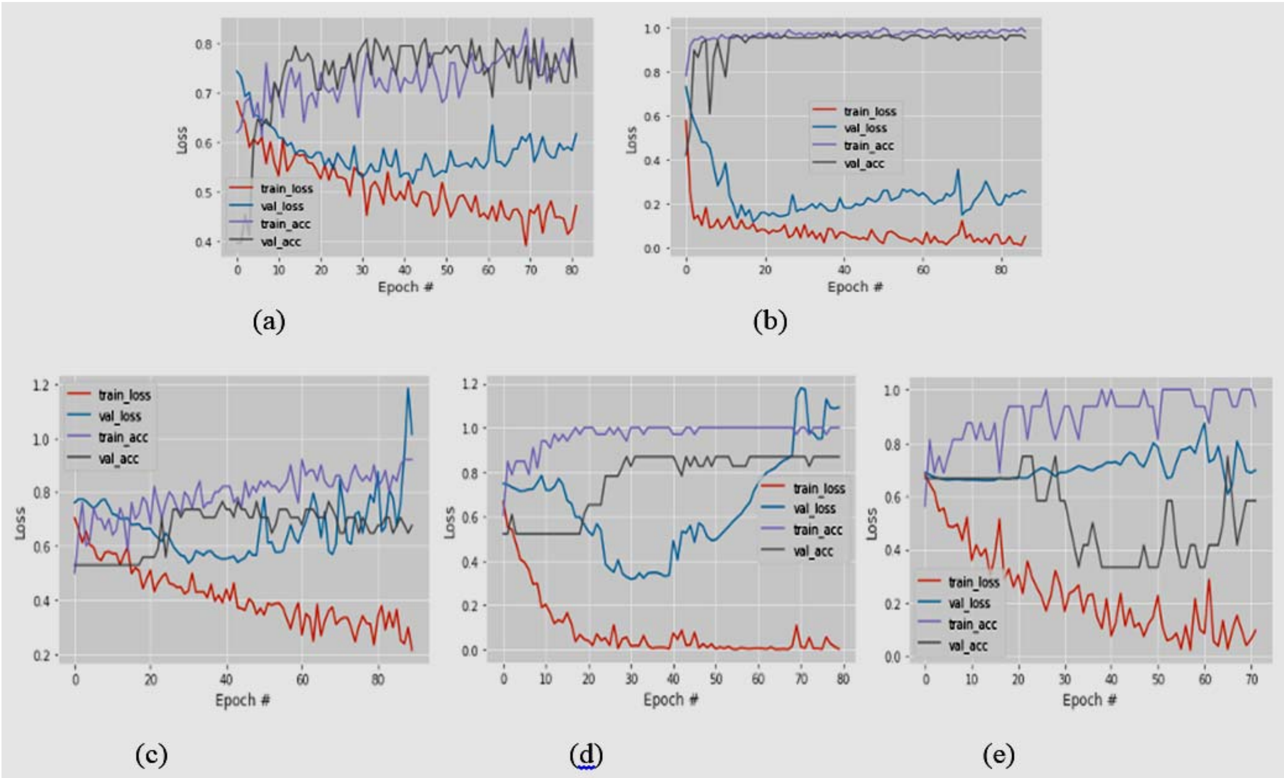

Figure 2. Training and validation loss & accuracy for all subjects 1 (a) subject aa (b) subject al (c) subject av (d) subject aw and (e) subject ay

On the other hand, for Figure 2(c) and Figure 2(d), the validation losses do not go down forever, in fact after several epochs, they actually go up; whereas for Figure 2(e), the validation loss has remained relatively high, around 0.6 since epoch 1. This simply implies that the model has reached a point of overfitting in which it can not improve upon unseen dataset, which is the validation/test data in this case. In theory, with longer epochs, it is certainly doable to achieve higher accuracies with finer tuned parameters. Nonetheless this is not the case for Dataset 1 as obvious results can be seen in the first 50 epochs before the training begins to settle down.

V. CONCLUSION

Conventional CSP algorithm generates a set of spatial filters that can define multi-dimensional data as a set of uncorrelated components as discriminative features. This is further enriched in this study as a combination of CSP-STFT is proposed as feature extraction. STFT is utilized to allow the extraction of joint time-frequency features from the spatial output of CSP. Then an Attention-CNN model is paired with the feature extraction method to better capture the spectrum images. This research aimed to investigate and evaluate the new BCI framework with various benchmark datasets and parameter evaluations. A new classifier technique based on attention mechanism to boost CNNs was proposed to address it. CNNs have been struggling as it lacks the ability to be spatially invariant to the input data as well as its convolutional layers only allowing for specific features to be learnt at a layer. As such, attention mechanism is brought over to boost its performance. By using a pre-trained VGG16 CNN architecture, attention is attached to the top layer of the VGG16 layers with rescaled feature to allow selective learning of the data. The proposed CSP-STFT CNN framework was validated and evaluated using the defined evaluation parameters such as accuracy, precision and recall. From the test results, the BCI system using the proposed method yielded decent results and has better performance than some of the benchmarked BCI systems. This shows that for the framework, undoubtedly useful features are extracted and processed, allowing the classifier to further improve upon its accuracy. In future, not just in CSP-based application but also the ones that do not use CSP as its feature extraction method. Therefore, whether or not classifier could also be generalized as a major classifier method could be ascertained, and thus the applicability of the method can be proven in common cases.

ACKNOWLEDGMENT

The authors would like thanks to TNB Research sdn. Bhd. and Universiti Teknologi Malaysia (UTM) for funding this research through Grant R.K130000.7643.4C425 and Center for Artificial Intelligent and Robotics (CAIRO).

REFERENCES

[1] I. Volosyak, F. Gembler, and P. Stawicki, "Age-related differences in SSVEP-based BCI performance," Neurocomputing, vol. 250, pp. 57-64, 2017.

[2] N. Padfield, J. Zabalza, H. Zhao, V. Masero, and J. Ren, "EEG-based brain-computer interfaces using motor-imagery: Techniques and challenges," Sensors, vol. 19, no. 6, p. 1423, 2019.

[3] M. Lotze and U. Halsband, "Motor imagery," Journal of Physiology-paris, vol. 99, no. 4-6, pp. 386-395, 2006.

[4] W. Samek, C. Vidaurre, K.-R. Müller, and M. Kawanabe, "Stationary common spatial patterns for brain–computer interfacing," Journal of neural engineering, vol. 9, no. 2, p. 026013, 2012.

[5] R. T. Schirrmeister et al., "Deep learning with convolutional neural networks for EEG decoding and visualization," Human brain mapping, vol. 38, no. 11, pp. 5391-5420, 2017.

[6] H. Altaheri et al., "Deep learning techniques for classification of electroencephalogram (EEG) motor imagery (MI) signals: A review," Neural Computing and Applications, pp. 1-42, 2021.

[7] L. Li and N. Sun, "Attention-Based DSC-ConvLSTM for Multiclass Motor Imagery Classification," Computational Intelligence and Neuroscience, vol. 2022, 2022.

[8] X. Ma, S. Qiu, C. Du, J. Xing, and H. He, "Improving EEG-based motor imagery classification via spatial and temporal recurrent neural networks," in 2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC), 2018: IEEE, pp. 1903-1906.

[9] S. Aggarwal and N. Chugh, "Review of machine learning techniques for EEG based brain computer interface," Archives of Computational Methods in Engineering, pp. 1-20, 2022.

[10] Q. She, K. Chen, Y. Ma, T. Nguyen, and Y. Zhang, "Sparse representation-based extreme learning machine for motor imagery EEG classification," Computational intelligence and neuroscience, vol. 2018, 2018.

[11] A. Tiwari and A. Chaturvedi, "A multiclass EEG signal classification model using spatial feature extraction and XGBoost algorithm," in 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2019: IEEE, pp. 4169-4175.

[12] S. U. Amin, H. Altaheri, G. Muhammad, M. Alsulaiman, and W. Abdul, "Attention based Inception model for robust EEG motor imagery classification," in 2021 IEEE international instrumentation and measurement technology conference (I2MTC), 2021: IEEE, pp. 1-6.

[13] J. Sun, J. Jiang, and Y. Liu, "An introductory survey on attention mechanisms in computer vision problems," in 2020 6th International Conference on Big Data and Information Analytics (BigDIA), 2020: IEEE, pp. 295-300.

[14] J. Schlemper et al., "Attention gated networks: Learning to leverage salient regions in medical images," Medical image analysis, vol. 53, pp. 197-207, 2019.

[15] E. Pesce, S. J. Withey, P.-P. Ypsilantis, R. Bakewell, V. Goh, and G. Montana, "Learning to detect chest radiographs containing pulmonary lesions using visual attention networks," Medical image analysis, vol. 53, pp. 26-38, 2019.

[16] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms," IEEE Transactions on biomedical Engineering, vol. 58, no. 2, pp. 355-362, 2010.

[17] W. Zhang, C. Tan, F. Sun, H. Wu, and B. Zhang, "A review of EEG-based brain-computer interface systems design," Brain Science Advances, vol. 4, no. 2, pp. 156-167, 2018.

[18] S. Daud and R. Sudirman, "Butterworth bandpass and stationary wavelet transform filter comparison for electroencephalography signal," in 2015 6th international conference on intelligent systems, modelling and simulation, 2015: IEEE, pp. 123-126.

[19] H. Fauzi, M. I. Shapiai, N. A. Setiawan, J. Jaafar, and M. Mustafa, "Channel selection for common spatial pattern Based on energy calculation of motor imagery EEG signal," in 2017 International conference on control, electronics, renewable energy and communications (ICCREC), 2017: IEEE, pp. 33-39.