2022 IEEE 16th International Scientific Conference on Informatics

INFORMATICS 2022

November 23–25, 2022, Poprad, Slovakia

PROCEEDINGS

Editors
William Steingartner
Štefan Korečko
Anikó Szakál

Organized by

Slovak Society for Applied Cybernetics and Informatics (SSAKI), affiliated branch at Department of Computers and Informatics FEEI TUKE

Faculty of Electrical Engineering and Informatics, Technical University of Košice

IEEE SMCS Technical Committee on Computational Cybernetics

IEEE Computational Intelligence (CI) Society Chapter of Czechoslovakia Section

ISBN 979-8-3503-1033-7
IEEE catalog number CFP22E80-USB

Using Autoencoders to Denoise Cross-Session Non-Stationarity in EEG-Based Motor-Imagery Brain-Computer Interfaces

Ophir Almagor*
Dept. of Cognitive and Brain Sciences
Ben-Gurion University of the Negev
Beersheba, Israel
ophiralm@post.bgu.ac.il

Ofer Avin*
Dept. of Cognitive and Brain Sciences
Ben-Gurion University of the Negev
Beersheba, Israel
avino@post.bgu.ac.il

Roman Rosipal
Institute of Measurement Science
Slovak Academy of Sciences
Bratislava, Slovakia
roman.rosipal@savba.sk

Oren Shriki
Dept. of Cognitive and Brain Sciences
and Dept. of Computer Science
Ben-Gurion University of the Negev
Beersheba, Israel
shrikio@bgu.ac.il

* Equal contribution

***Abstract*- A major problem in brain-computer interfaces (BCIs) relates to the non-stationarity of brain signals. Consequently, the performance of a classification algorithm trained for an individual subject on a certain day deteriorates during the following days. The traditional approach is to recalibrate the algorithm every session, limiting the wide use of BCIs. Here, we use an autoencoder convolutional neural network to identify a low dimensional representation of the EEG signals from the first day (or days) and show that this allows for stable decoding performance on the following days without resorting to recalibration. Furthermore, we demonstrate that the residual signals, namely the difference between the original and reconstructed EEG, can be used to accurately discriminate among different recording sessions. In line with that, the reconstructed EEG cannot be used to discriminate among recording sessions. This implies that the reconstructed EEG reflects an invariant representation of the subject's intent, whereas the residual signals reflect a non-stationary component, which differs from one session to another. The findings are demonstrated through two different datasets.**

Keywords— electroencephalogram, motor-imagery, brain-computer interface, autoencoders, deep learning, non-stationarity.

I. INTRODUCTION

Brain-computer interfaces (BCIs) can serve as a communication tool for patients who suffer from a severe loss of motor abilities, such as amyotrophic lateral sclerosis (ALS) or multiple sclerosis [1,2]. Furthermore, BCIs have also been established as a helpful tool for increasing the rehabilitation efficiency of motor skills following stroke [3,4]. In this paper, we analyze electroencephalogram (EEG) recordings collected across multiple daily sessions of a motor imagery (MI) BCI task.

A significant challenge in multi-session BCIs is coping with signal non-stationarity [5,6]. A decoding algorithm that obtains high performance on one day, will typically preform lower on other days. This might be due to changes in the underlying neural activity as well as to changes related to the EEG headset mounting on different days. The most common approach to overcome this non-stationarity is recalibration, namely training a new classifier every session. An interesting approach to reduce the duration of recalibration, which demonstrated good performance in invasive recordings from monkeys, is domain adaptation using adversarial models [7]. A related method which was applied to non-invasive human data is dynamic domain adaptation [8]. Other models based on deep learning have also shown promising results using the EEGNet and ConvNet networks [9]. In another study, transfer learning between sessions using deep networks was applied for addressing cross-session classification difficulties [10]. However, these approaches still require some EEG data from the same session in order to align with the data from the first measurement session ('session 0'). Here, we propose a new method, based on autoencoder neural networks, which does not require any recalibration data from the new session.

Autoencoders (AEs) are stochastic neural networks that attempt to compress a high-dimensional input and then reconstruct it. Specifically, given an input x, the network tries to learn weights such that the output $\hat{x}$ will minimize the mean squared error (MSE) loss:

$$\mathcal{L} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2$$

AEs have been used for denoising medical images [11], ECG signals [12], and EEG signals [13]. The underlying premise is that random noise across different samples cannot be reconstructed, and thus the learned compressed representation would denoise the signal. A significant advantage of this approach is that it involves purely unsupervised learning and does not require any data labeling.

II. METHODS

A. Data

We applied our approach to two different datasets. The first dataset was collected at the Slovak Academy of Sciences, Bratislava, and included a single male subject (subject 201; 61 years old), who suffers from right-hand

This research was supported by a CHIST-ERA grant of the EU to OS and RR

hemiplegia, caused by an ischemic stroke. The stroke occurred 4 years before the study.

The experimental protocol involved a MI task with two classes: right-hand imagery vs. idle state. The subject performed the experiment for a total of 134 sessions over a period of 9 months. Each session consisted of 10 trials with the following structure. The subject hears the command 'Relax' and is expected to relax for 21 seconds with eyes closed. After 21 seconds, the subject hears the command 'Move'. He is instructed to start his MI process of the right arm. The subsequent trial starts 7.5 seconds after the end of the MI part. We extracted 6 seconds segments of the EEG recording before the 'Move' command and labeled them as 'Idle', and 6 seconds segments after the 'Move' command and labeled them as 'Right'.

The EEG recording was performed using the g.tec USBamp headset with gel-based electrodes over the following 11 locations FC3, C1, C3, C5, O1, FC4, C2, C4, C6, and CP4 according to the international 10-20 system. Data were collected with a sampling rate of 512 Hz and were offline resampled to 128 Hz, with an anti-aliasing filter applied. Data were also filtered using a band-pass filter in the range 0.1–200 Hz, and a notch filter at 50 Hz.

The second dataset was an open access dataset at IEEE Data port [14] and included 20 subjects (11 males, mean age 23.2 ± 1.47 years, all right-handed). The experimental protocol involved an MI task with four classes (right hand, left hand, both feet, and idle). The subjects participated in 7 sessions within two weeks, where each session lasted around 40 minutes and consisted of 6 blocks. Each block consisted of 40 trials (10 per class) presented in random order. Each trial had the following structure. Pre cue arrow indicated the next trial for 1.5 seconds followed by a fixation cross for 1 second. Then, an arrow indicated to start MI for 5 seconds, followed by 3 seconds of rest. The following analysis was performed with only two classes (left hand and idle) in order to have similar data to subject 201.

The EEG recording was performed using a 65-channel Synamp2 system (Neuroscan, Inc.) with a sampling rate of 500 Hz. Twenty-six EEG electrodes were positioned according to the international 10-20 system.

B. Preprocessing

All EEG trials were segmented into periods of 6 seconds following the cue to start MI. Data were bandpass filtered between 4-40 Hz. Trials with amplitude over 250 μV were removed from the dataset. Sessions with less than 10 trials were removed completely.

C. Classifier

To decode EEG signals, we used filter bank common spatial patterns (FBCSP) [15] and a support vector machine classifier (SVM). The FBCSP method first filters the EEG signals within several frequency bands to correctly identify the subject's optimal frequency band. We used band-pass filters in the 4-40 Hz range in steps of 4 Hz (4-8 Hz, 8-12 Hz, etc...). After filtering, we applied common spatial patterns (CSP) to each filtered signal. The CSP method was used to identify linear combinations of channels that exhibit high variability under one condition and low variability under the other condition [16]. From each CSP signal, we extract a feature that corresponds to the total band power. Subsequently, we apply the mutual-information-based best individual feature (MIBIF) algorithm, a common feature selection algorithm [17,18]. In this algorithm, the mutual information of each feature with the label is computed. The features are then ranked in descending order, and the k first features are selected (we took k to be the number of EEG channels). Lastly, we use an SVM classifier [19] to predict the trial label according to the selected features.

D. Within-Session Classification

For each session, the preprocessed EEG trials were divided into 5 folds (80% train, 20% test). Each fold was fitted with an FBCSP for 80% of the train data and tested on the 20% test data. Performance was quantified by the mean accuracy score over folds.

E. Cross-Session Classification

The FBCSP model was fitted using all trials in the first session (session zero). Then, for each subsequent session, performance was quantified using the accuracy of this classifier across all trials in that session.

F. AE + Cross-Session Classification

An AE was fitted to reconstruct the EEG signals from session zero. The AE architecture was 3 layers of 1D convolution (number of filters per layer – [8, 16, 32]) for the encoder, and 3 layers of 1D transposed convolution (number of filters per layer – [32,16,8]) for the decoder. The learning rate was 0.001 while using Adam optimizer [20] for 250 epochs. The AE loss was MSE between input and output.

The FBCSP model was then fitted using the reconstructed EEG of all trials in session zero. For each subsequent session, data were reconstructed by the AE that was trained on session zero, and the accuracy of the FBCSP model which was trained on session zero was computed. Thus, this model did not require any data from the test sessions.

The classification workflow is illustrated in Fig. 1.

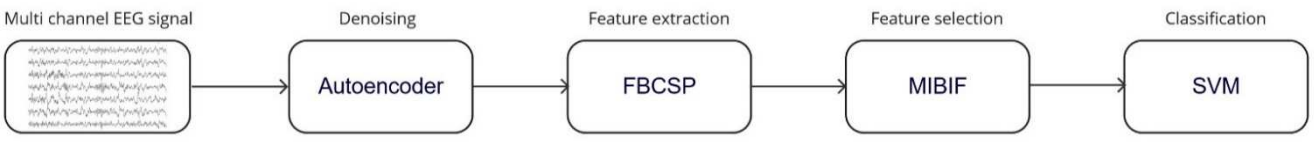

Figure 1: Workflow chart of our proposed algorithm.

G. Multi-Session Training

We applied the same analysis as described for session zero, but gradually combined more sessions together and used them as the training set for the cross-session and the AE cross-session conditions. We then tested the model's accuracy score over the rest of the sessions. The benchmark for the within-session score was 5-fold cross-validation over all the sessions that were not included in the training set (namely, the test set).

H. Classification of the data origin session

To gain insight into the nature of the non-stationarity in the EEG signals, we trained a classifier to identify the session from which the data originated. Specifically, we labeled each trial with the session number, regardless of the MI task label. We then used CSP features and a linear discriminant analysis (LDA) with 5 folds CV to classify from which session the trial data originated. We applied the above procedure to the original EEG signals, the reconstructed EEG signals after the AE transformation, and the residuals of the AE reconstruction ($x - \hat{x}$). This allowed us to assess whether the signal components that the AE removed indeed contain session-specific information, namely, whether they reflect the non-stationarity of the EEG.

All analysis codes were implemented in Python 3.7. The AE model was implemented using Pytorch and the FBCSP was implemented using the FBCSP toolbox (https://fbcsptoolbox.github.io/). The codes are available at: https://github.com/bci4cpl/Non-Staionarity-Autoencoder-denoising.

III. Results

Figure 2 presents the results from the analyses of subject 201, the stroke patient. The mean accuracy score of the cross-session classifier, across all 134 sessions, was 61.6%, whereas the mean accuracy of the AE + cross-session model was 68.5%. The accuracy of the within-session model was at a chance level, due to the low amount of data per session. We address this problem by testing the within-session over the entire test set.

As evident in Fig. 2, although there is an improvement in the cross-session score as more session data are added to the training set, the AE cross-session model constantly outperforms the simple cross-session model. Additionally, the AE cross-session method achieved comparable performance to the within-session cross-validation model, despite using a substantially smaller amount of training data.

We next examined the ability to identify the session from which the data originated. When using the original signals, the accuracy of the trained classifier across the 134 sessions was 67.6%. The accuracy when using the reconstructed de-noised signals was 15.7%, whereas the accuracy when using the residual signals was 72.7%.

Subsequently, we applied the analyses to the IEEE dataset (Fig. 3). The presented performance reflects the mean across subjects, where only subjects with the entire 7-day recording after trial rejection were analyzed (N=17). The mean over 7 days within-session accuracy score was 71.3%, with 0.4% standard error. For the cross-session, the accuracy score was 55.3% with 0.29% standard error, whereas the AE cross-session score was 59.6% with 0.35% standard error. Some subjects had low accuracy scores when using the within-session method, and neither of the cross-session methods had good performance in these cases. We checked the mean across subjects only in cases where the mean within-session score was above 75%. The improvement of the AE cross-session model compared to the simple cross-session model was superior – 67.4% accuracy for the AE cross-session model, against 57.1% using the simple cross-session model.

Classification of the signal origin session was performed per subject, and we report the mean across subjects. The mean accuracy for using the original signals as input was 75.6%

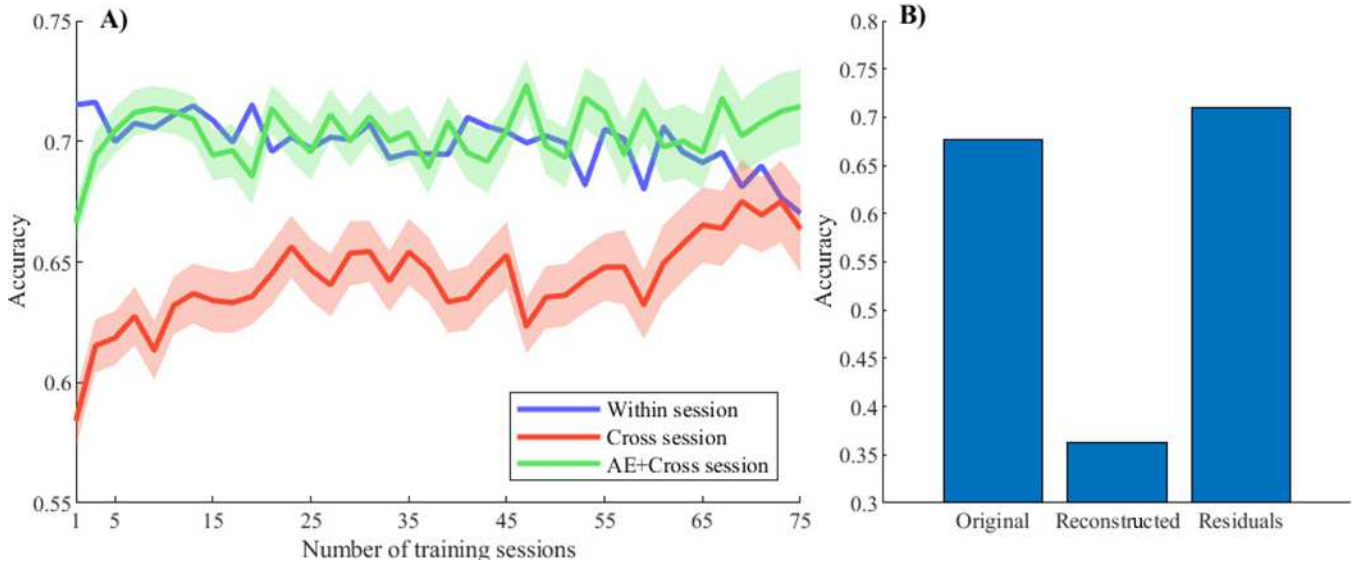

Figure 2: Analysis of longitudinal data from a stroke patient. **A)** Mean accuracy of each model. The cross-session (red) and AE + cross-sessions (green) models were trained on an increasing size of the training set. The within-session benchmark (blue) is the score of the cross validation on the test set (no inference on new sessions). The shaded area represents the standard error of the accuracy across sessions. **B)** Mean accuracy of origin session classification using the original signals, reconstructed de-noised, and residuals signals.

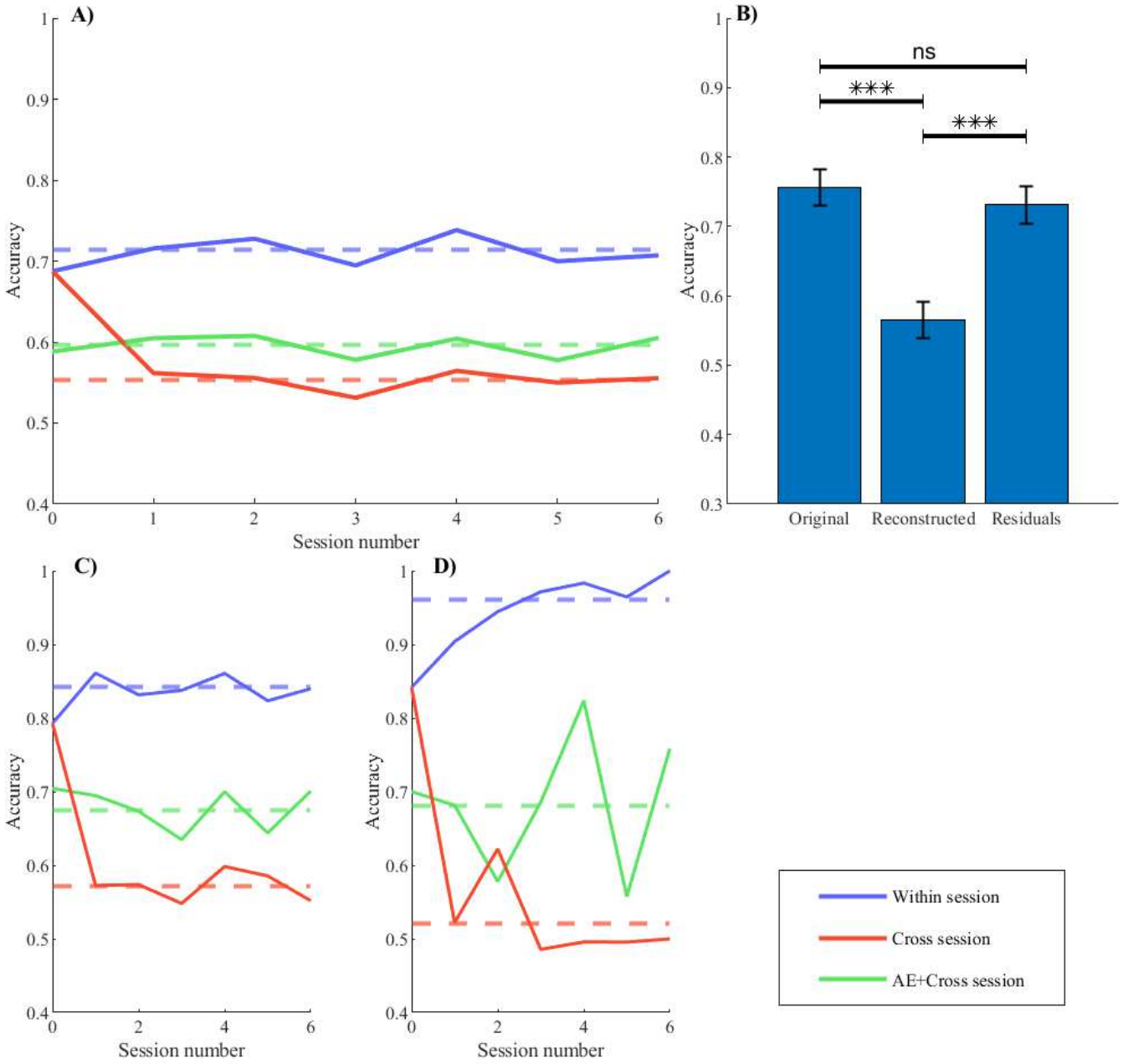

Figure 3: Analysis of the IEEE dataset. **A)** Mean accuracy over subjects (N=17) of within-session (blue), cross-session (red), and AE + cross-session models (green)**. B)** Mean accuracy and standard error of origin session classification using the original signals, reconstructed de-noised, and residuals signals. ($*** P < 0.001$). **C)** Mean accuracy over subjects with mean within-session accuracy > 75% (N=8) of within-session, cross-session, and AE + cross-session models. **D)** Accuracy of the subject with highest within-session mean accuracy of within-session, cross-session, and AE + cross-session models.

with 2.5% standard error. When using the de-noised reconstructed signals, accuracy dropped to 56.5%, with 2.6% standard error, whereas when using the residuals of the AE, accuracy was 73%, with 2.7% standard error. Student t-test was performed to assess whether the difference between the conditions is significant. No significant difference was found between the original and residual signal conditions ($p = 0.503$). However, a significant difference was found between the original and reconstructed signals ($p < 0.0001$), and between the residuals and the reconstructed signals ($p < 0.001$).

IV. Discussion

We presented a novel approach for overcoming signal non-stationarity in EEG-based MI BCIs. Our method is based on a convolutional AE network, which finds a low-dimensional compressed representation of the EEG signals and eliminates components related to cross-session variability. The main motivation behind the proposed approach is that it relies on a purely unsupervised learning approach to denoise the non-stationary component. Furthermore, it does not require any additional data collection from the target session. Our results suggest that the AE denoising method can help create stable BCIs, which may have a wide range of applications.

For the stroke patient data, which contained over 130 sessions, the results remained constantly better than the naïve use of the same classification model trained with the data of past sessions. Furthermore, the findings are consistent across a different dataset with 20 subjects. This is an improvement over most current methods for stable BCIs, which require some data from the test session. An advantage of the AE approach is that it is purely unsupervised and only relies on the assumption that the non-stationary signal cannot be learned by the AE due to the randomness of the noise components among sessions. We also found that when the within-session classification obtains better performance, our proposed model has "more room" for improvement compared to simple cross-session classification.

Lastly, the fact that the reconstructed signal performs poorly in classifying the session from which the data originated, compared to the original and the residuals signals, suggests that the residuals removed from the signals contain session-specific information. Analyzing these signals can provide further insights into the sources of non-stationarity. Future work can look for AE architectures that perform better than the simple AE we used. Another future direction would be to examine the application of the approach to other BCI paradigms, such as a P300 BCI or emotion recognition.

REFERENCES

[1] J. J. Daly and J. R. Wolpaw, "Brain–computer interfaces in neurological rehabilitation," The Lancet Neurology, vol. 7, no. 11, pp. 1032–1043, Nov. 2008.

[2] J. Wolpaw, N. Birbaumer, D. McFarland, G. Pfurtscheller and T. Vaughan, "Brain–computer interfaces for communication and control", 2022.

[3] A. Zimmermann-Schlatter, C. Schuster, M. A. Puhan, E. Siekierka, and J. Steurer, "Efficacy of motor imagery in post-stroke rehabilitation: a systematic review," Journal of NeuroEngineering and Rehabilitation, vol. 5, no. 1, Mar. 2008.

[4] S. de Vries and T. Mulder, "Motor imagery and stroke rehabilitation: a critical discussion," Journal of Rehabilitation Medicine, vol. 39, no. 1, pp. 5–13, 2007.

[5] P. Shenoy, M. Krauledat, B. Blankertz, R. P. N. Rao, and K.-R. Müller, "Towards adaptive classification for BCI," Journal of Neural Engineering, vol. 3, no. 1, pp. R13–R23, Mar. 2006.

[6] A. Arieli, A. Sterkin, A. Grinvald, and A. Aertsen, "Dynamics of Ongoing Activity: Explanation of the Large Variability in Evoked Cortical Responses," Science, vol. 273, no. 5283, pp. 1868–1871, Sep. 1996.

[7] A. Farshchian, J. A. Gallego, J. P. Cohen, Y. Bengio, L. E. Miller, and S. A. Solla, "Adversarial Domain Adaptation for Stable Brain-Machine Interfaces," arXiv:1810.00045 [cs, q-bio, stat], Jan. 2019, Accessed: Sep. 22, 2022.

[8] J. Ma, B. Yang, W. Qiu, Y. Li, S. Gao, and X. Xia, "A large EEG dataset for studying cross-session variability in motor imagery brain-computer interface," Scientific Data, vol. 9, no. 1, p. 531, Sep. 2022.

[9] Z. Li *et al*., "Dynamic Domain Adaptation for Class-Aware Cross-Subject and Cross-Session EEG Emotion Recognition," in *IEEE Journal of Biomedical and Health Informatics*, 2022.

[10] D. Wu, Y. Xu, and B.-L. Lu, "Transfer Learning for EEG-Based Brain–Computer Interfaces: A Review of Progress Made Since 2016," IEEE Transactions on Cognitive and Developmental Systems, vol. 14, no. 1, pp. 4–19, Mar. 2022.

[11] L. Gondara, "Medical Image Denoising Using Convolutional Denoising Autoencoders," IEEE Xplore, 2016.

[12] H.-T. Chiang, Y.-Y. Hsieh, S.-W. Fu, K.-H. Hung, Y. Tsao, and S.-Y. Chien, "Noise Reduction in ECG Signals Using Fully Convolutional Denoising Autoencoders," IEEE Access, vol. 7, pp. 60806–60813, 2019.

[13] N. M. N. Leite, E. T. Pereira, E. C. Gurjão, and L. R. Veloso, "Deep Convolutional Autoencoder for EEG Noise Filtering," IEEE Xplore, Dec. 01, 2018.

[14] Q. Zhou, "EEG dataset of 7-day Motor Imagery BCI," ieee-dataport.org, Nov. 2020, Accessed: Sep. 22, 2022.

[15] Kai Keng Ang, Zheng Yang Chin, Haihong Zhang and Cuntai Guan, "Filter Bank Common Spatial Pattern (FBCSP) in Brain-Computer Interface," 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), 2008.

[16] C. Guger, H. Ramoser, and G. Pfurtscheller, "Real-time EEG analysis with subject-specific spatial patterns for a brain-computer interface (BCI)," IEEE Transactions on Rehabilitation Engineering, vol. 8, no. 4, pp. 447–456, 2000.

[17] K. K. Ang and C. Quek, "Rough set-based neuro-fuzzy system," IEEE Xplore, Jul. 01, 2006.

[18] A. K. Jain, P. W. Duin, and Jianchang Mao, "Statistical pattern recognition: a review," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 1, pp. 4–37, 2000.

[19] W. S. Noble, "What is a support vector machine?," Nature Biotechnology, vol. 24, no. 12, pp. 1565–1567, Dec. 2006.

[20] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," arXiv.org, 2014.