Journal of Neural Engineering

ACCEPTED MANUSCRIPT • OPEN ACCESS

Supervised autoencoder denoiser for non-stationarity in multi-session EEG-based BCI

To cite this article before publication: Ofer Avin *et al* 2025 *J. Neural Eng.* in press https://doi.org/10.1088/1741-2552/adc48e

IOP Publishing **Journal of Neural Engineering**

Supervised Autoencoder Denoiser for Non-Stationarity in Multi-Session EEG-Based BCI

Avin Ofer*[1], Almagor Ophir*[1], Noah Yoav*[1], Rosipal Roman[2] and Shriki Oren[1]

* *Equal Contribution*
[1] Dept. of Cognitive and Brain Sciences, Ben-Gurion University of the Negev, Beersheba, Israel
[2] Institute of Measurement Science, Slovak Academy of Sciences, Bratislava, Slovakia

Corresponding author: Shriki Oren, shrikio@bgu.ac.il

Abstract

Non-stationarity in EEG signals poses significant challenges for the performance and implementation of brain–computer interfaces (BCIs). In this study, we propose a novel method for cross-session BCI tasks that employs a supervised autoencoder to reduce session-specific information while preserving task-related signals. Our approach compresses high-dimensional EEG inputs and reconstructs them, thereby mitigating non-stationary variability in the data. In addition to unsupervised minimization of the reconstruction error, the objective function of the network includes two supervised terms to ensure that the latent representations exclude session identity information and are optimized for subsequent classification. Evaluation across three different motor imagery datasets demonstrates that our approach effectively addresses domain adaptation challenges, outperforming both naïve cross-session and within-session methods. Our method eliminates the need for data from new sessions, making it fully unsupervised concerning new session data and reducing the necessity for recalibration with each session. Furthermore, the reduction of session-specific information in the reconstructed signals indicates that our approach effectively denoises non-stationary signals, thereby enhancing the accuracy of BCI models. Future applications could extend this model to a broader range of BCI tasks and explore the residual signals to investigate sources of non-stationary brain components and other cognitive processes.

Keywords: Motor-Imagery BCI, EEG, Non-Stationarity, Autoencoder, Denoising

1. 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 or multiple sclerosis [1,2]. Furthermore, BCIs have also been established as a helpful tool for increasing the rehabilitation efficiency of motor skills following a stroke [3-5]. In this paper, we analyze electroencephalogram (EEG) recordings collected across multiple daily sessions of a motor imagery (MI) BCI task.

In the context of multi-session BCI studies, addressing signal non-stationarity poses a significant challenge [6,7]. Decoding algorithms that achieve high performance on a given day tend to underperform during subsequent sessions, owing to fluctuations in underlying neural activity and inconsistencies in EEG headset placement. Several studies showed that neural representation changes over time, without changes in the stimuli. These changes can vary in time spans of between minutes to weeks, and occur on several brain networks [8-12]. In this regard, evaluating the performance of a decoding model using different measures can aid in assessing its effectiveness. The within-session score refers to the performance of a model on a test dataset obtained from the same recording session, without altering the headset, and with minimal time lapse between training and test data acquisition. On the other hand, the cross-session score denotes the model's performance when the training and test sets are derived from distinct sessions, potentially with significant temporal gaps (e.g., different days).

To address EEG signal non-stationarity, recalibration is commonly used, involving training a new classifier for each session, which necessitates additional data collection each time the BCI is used. This time-consuming recalibration is a major obstacle in the wide adoption of this technology. One proposed method to overcome the multi-session non-stationary challenge without training a model from start each day is co-adaptive training [13,14]. In the co-adaptive method, the classifier in constantly updated based on the brain activity during the current session. Another promising approach to reducing recalibration duration is domain adaptation, which involves the adaptation of an algorithm trained in one or more "source domains" to effectively work in a different, yet related, "target domain". An example for this approach is shown with adversarial deep learning models, which have demonstrated good performance in invasive recordings from monkeys [15]. The aim of this method, along with other domain adaptation techniques, is to map the feature domain of subsequent ("target") sessions to better align with the feature domain of the training session on which the model was originally trained ("source"), thus improving the cross-session accuracy of the model. Dynamic domain adaptation is another related method, which was applied to non-invasive human data [16]. Deep learning models such as EEGNet and ConvNet networks have also shown promising results in the cross-session performance for multi-session BCI tasks [17]. Transfer learning between sessions utilizing deep networks has been applied to address this type of classification difficulty [18], but still requires some EEG data from the same session to align with data from the first measurement session ("session 0"). Here, we propose a novel deep neural network architecture that obviates the need for any recalibration data from the new session. It is based on an autoencoder neural network, together with a supervised component.

Autoencoders (AEs) are neural networks that aim to compress high-dimensional inputs and then reconstruct them. The network endeavors to learn weights that minimize the mean squared error (MSE) between the output $\hat{x}$ and the given input x. AEs have been applied to denoise various medical signals, including electrocardiogram (ECG) signals [19], EEG signals [20], and medical images [21]. The fundamental concept is that random noise present in different samples cannot be reconstructed, and therefore, the compressed representation learned by the network will capture the regularities in the data and remove the noise.

We previously utilized a basic AE network to overcome the non-stationarity problem in MI-BCI [22]. In the method proposed in the mentioned paper, the preprocessed EEG signals were replaced with the reconstructed signals as the input to a classifier. This approach improved the accuracy of the BCI task in cross-session tests, indicating that the reconstructed signals contained task-related information and less non-task-related information. Furthermore, the study revealed that the residuals, defined as the difference between the original and reconstructed signals, retained a similar amount of session-specific information as the original signals, whereas the reconstructed signals showed a significant reduction in session-specific information.

However, this approach did not take into account any information regarding the task-related labels and session-related information in the data. Thus, a major question is how to explicitly incorporate this information and construct a compressed representation that better maintains task-related information while removing session-related information. Here, to address these issues, we added two additional components into the loss function: one based on the session label (identifying the session from which the signal was recorded) and another based on the MI task label, to place constraints on the Decoder and Encoder, respectively. By that, our model is working as a supervised representation learning model, which has several examples in other fields, mainly computer vision [23,24]. A similar method for semi-supervised AEs was applied to ECG and EEG data with the classification task only on the latent space and showed an improvement in the reconstruction for task classification [25]. Since AEs are traditionally self-supervised models, we refer to our model as a supervised autoencoder (SAE) due to the inclusion of session and task labels in the training process. In contrast, the basic unsupervised autoencoder will be referred to as uAE.

It is important to emphasize that our main goal is to address the challenges of brain-related non-stationarity in EEG signals, rather than artifact removal. Artifacts resulting from eye movements and muscle activity, are generally consistent across sessions and unlikely to contain session-specific information. In addition, these artifacts are transient, whereas here we focus on removal of continuous components by projecting the data into a lower dimensional space.

The paper provides an overview of the various datasets employed, as well as the EEG pre-processing, the BCI classification model, and the metrics employed to evaluate each approach. The network architecture is also discussed in detail. The findings for each dataset are presented, and recommendations for model use and future applications are suggested.

2. Methods

Subsection 2.1 describes the datasets used in this study, while subsection 2.2 outlines the data preprocessing procedures. The classifier and model architecture are discussed in subsections 2.3 and 2.4, respectively, followed by a detailed explanation of the training procedure in subsection 2.5. Finally, subsections 2.6 through 2.9 present the evaluation of our proposed method using various metrics.

2.1 Data

Our approach was implemented on three distinct datasets. The first dataset, obtained from the Slovak Academy of Sciences (Bratislava), consisted of data collected from a single male participant (subject 201; aged 61) who experienced right-hand hemiplegia resulting from an

ischemic stroke that had occurred four years prior to the study.
In the experimental procedure, a MI task with two distinct categories, right-hand imagery versus idle state, was conducted. The subject participated in a total of 134 sessions over a span of 9 months. Each session comprised 10 trials or more following a specific pattern. Initially, the subject heard the command "Relax" and was instructed to remain in a relaxed state with eyes closed for 21 seconds. Following this, he heard the command "Move" and was prompted to initiate the MI process of the right arm while keeping his eyes closed. The subsequent trial commenced 7.5 seconds after the conclusion of the MI phase. EEG recordings were captured using a g.tec USBamp headset with gel-based electrodes placed over the following specific locations: FC3, C1, C3, C5, O1, FC4, C2, C4, C6, and CP4, in accordance with the international 10–20 system. The collected data were sampled at a rate of 512 Hz and later resampled offline to 128 Hz with an anti-aliasing filter applied and re-referenced to the linked ear. We extracted 6-second segments of the EEG recording before the "Move" command and labeled them as "Idle", and 6-second segments after the "Move" command, designating them as "Right".

The second dataset, which is publicly available on IEEE DataPort [26], comprised 20 participants (11 males), all of whom were right handed and whose average age was 23.2±1.47 years. The experimental procedure for this dataset involved a MI task with four distinct categories, namely, right hand, left hand, both feet, and idle. The participants completed 7 sessions over two weeks, with each session lasting approximately 40 minutes and consisting of 6 blocks. Each block included 40 trials, with 10 trials for each category presented randomly. The structure of each trial involved a pre-cue arrow indicating the next trial for 1.5 seconds, followed by a fixation cross for 1 second. An arrow was then displayed for 5 seconds, prompting the participants to engage in MI, followed by 3 seconds of rest. The EEG recordings were acquired using a 65-channel Synamp2 system (Neuroscan, Inc.) at a sampling rate of 500 Hz. The 26 EEG electrodes were placed based on the international 10–20 system.

The third dataset, which is accessible in an article in *Scientific Data* [27], involved 25 participants (12 females) between 20 and 24 years of age, none of whom had any prior experience with BCI. The task assigned to the participants was a MI task, where each trial commenced with a fixation cross, followed by a left-hand or right-hand movement prompt displayed on the monitor, indicating the next movement to imagine. The total duration of each trial was 7.5 seconds, and each session included 100 trials. Each participant completed the experiment 5 times on 5 different days, with a 2–3-day interval between each session. The EEG headset utilized in this study was a solid electrode cap with Ag/AgCl and 32 electrodes from Wuhan Greentech Technology Co., LTD, with a sampling rate of 250 Hz.

2.2 Pre-processing

All EEG trials were segmented into periods of 6 seconds following the cue to start MI. Data were bandpass filtered between 1–40 Hz. Trials with an amplitude over 250 µV were removed from the dataset. Sessions with less than 10 trials were removed completely.

2.3 Classifier

A commonly used approach for motor imagery-based brain-computer interface (MI-BCI) classification from EEG signals involves feature extraction using common spatial patterns (CSP) [28] and classification through linear discriminant analysis (LDA) [29, 30]. We utilized CSP [28] to identify linear combinations of EEG channels that maximized variance (total power) for one condition while minimizing it for the other. The total power from each CSP component was then extracted and used as input for a multi-class LDA classifier [31].

During the training phase of our model, the network architecture included a component specifically trained to classify the EEG signals based on task labels. However, in the inference phase, we opted to use the CSP-LDA method for classification instead of directly utilizing the neural network. We found that the CSP-LDA method demonstrated superior performance on the data, likely due to the limited amount of data per session and the risk of overfitting with the neural network. Furthermore, this approach facilitates easier comparison with other studies on motor-imagery BCIs, where CSP-LDA is commonly employed.

2.4 Trainable Auto-Encoder Architecture

The proposed trainable architecture considers two training phases: *autoencoder training* and *classifier training*.

Autoencoder Training: The architecture for the initial training phase, as shown in Fig. 1a, utilizes an AE consisting of three layers of 1D convolution for the encoder (with [8, 16, 32] filters per layer) and three layers of 1D transposed convolution for the decoder (with [32, 16, 8] filters per layer). The model was trained with a learning rate of 0.001 using the Adam optimizer [32] over 250 epochs, with Leaky ReLU [33] as the activation function.

This phase consisted of three primary tasks: MI task classification, EEG data reconstruction, and origin session classification.

- For MI task classification, we applied a fully connected single-layer readout network from the latent layer, representing a compressed feature space of the EEG signals. A Softmax activation function and cross-entropy loss were used with the MI task labels for each signal.

- For EEG data reconstruction, the mean squared error (MSE) loss was computed between the autoencoder's input and output.
- For origin session classification, a separate encoder with the same architecture as the AE encoder was applied to the AE residuals (x - x̂). A fully connected layer was then applied to the latent space of these residuals to classify the origin session.

Both fully connected layers included a dropout layer with a 50% dropout rate to prevent overfitting. By incorporating these losses into the standard AE framework, the compressed representation of the EEG signals retained more task-relevant information while reducing session-specific information, improving the network's ability to classify the origin session of the signals.

The AE loss, $\mathcal{L}$, consisted of three terms:

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

$$(2)\ \ \mathcal{L}_{task} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=0}^{M} y_{ij}^{task}\, log\,(p_{ij}^{task})$$

$$(3)\ \ \mathcal{L}_{session} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=0}^{M} y_{ij}^{session} \log\,(p_{ij}^{session})$$

$$\mathcal{L} = \mathcal{L}_{MSE} + \mathcal{L}_{task} + \mathcal{L}_{session}$$

Where N is the number of samples in the batch, M is the number of motor imagery classes (2 or 4), x is the AE input (EEG signals) and $\hat{x}$ is the model output, y is the relevant label (MI task in the task loss, and session number in the session loss), and p is the probability for the class returned by the fully connected network.

Classifier Training: The subsequent training phase architecture, illustrated in Fig.1b, involves transforming the reconstructed EEG signals into a set of features using CSP. These features are then classified using an LDA classifier. During this phase, the AE weights are fixed, and only CSP and LDA components are trained, as discussed in subsection 2.3. This approach ensures that the classifier can effectively distinguish between the different MI tasks based on the features extracted by the AE and CSP.

2.5 Multi-Session Training

To measure the impact of incorporating multi-session data collected from the same subject, we progressively incorporated more sessions into the training set. The test set consisted of the remaining sessions not used in the training set. For example, if a subject had 7 sessions in total, we began by using session 0 as the training set and sessions 1–6 as the test set. In the subsequent iteration, we employed sessions 0–1 as the training set and sessions 2–6 as the test set, and so forth. The sequence of sessions was randomized for each realization.

2.6 Within-Session Classification

To evaluate the performance of the models on each training set, we divided the pre-processed EEG trials into five folds. Each fold was then used to fit a CSP-LDA classifier on 80% of the training data and tested on the remaining 20% of test data. The mean accuracy score over the folds was then calculated to quantify the performance.

2.7 Cross-Session Classification

The CSP-LDA model was trained on all trials in the training set. Next, the accuracy of this classifier across all trials in each subsequent session in the test set was used to assess performance. The average accuracy over the sessions in the test set was used to measure the performance of the classifier between sessions. This result was considered the naïve between-session performance.

2.8 sAE and Cross-Session Classification

The CSP-LDA model was fitted using the reconstructed EEG of all trials in the training set. For each subsequent session in the test set, data were reconstructed by the AE that was fitted to the training set, and the accuracy of the CSP-LDA model which was trained on the training set reconstructed signals was computed. Thus, this model did not require any data from the test sessions for domain adaptation.

2.9 Classification of the Data Origin Session

To gain a better understanding of the non-stationarity in the EEG signals, a classifier was trained to identify the session from which the data originated. Each trial was labeled with the session number, independent of the MI task label, and the CSP features were extracted. An LDA classifier with 5-fold cross validation was then applied to classify the origin session of the trial data. This procedure was applied to the preprocessed EEG signals, the reconstructed EEG signals after the AE transformation, and the residuals of the AE reconstruction ($x - \hat{x}$), allowing for assessment of whether the signal components removed by the AE indeed contained session-specific information, and whether they reflected the non-stationarity of the EEG.

The division between the training and test sets for each method are summarized in Table 1.

All analysis codes were implemented in Python 3.7. The AE model was implemented using Pytorch. The codes are available at: https://github.com/bci4cpl/NAWD.

3. Results

Our study tackled the issue of signal non-stationarity, a major obstacle in multi-session BCI studies. We examined three separate MI-EEG datasets, each involving unique participants and distinct experimental methodologies. Our approach incorporated a supervised AE model to enhance the decoding algorithm's performance across multiple sessions. In this section, we delve into a comprehensive performance analysis of our model, both within individual sessions and across multiple ones. The analysis showcases the model's adeptness at managing session-related variations in neural activity.

As described in the Methods section, the performance of the model is evaluated using cross-session classification (Section 2.8). The model is then compared to within-session classification and cross-session classification of the raw EEG (Sections 2.6 and 2.7). It is also compared to the cross-session performance following the application of a simple self-supervised AE.

3.1 Stroke Patient Dataset

Figure 2 illustrates the accuracy performance of the different methods as a function of the number of sessions included in the training set. As expected, all methods demonstrated improvement as more sessions were added to the training set; however, the naive cross-session method exhibited slower improvement and required more data to reach the saturation point. The results show that the naive cross-session method consistently achieved an accuracy of approximately 66% with a training set of 30 sessions or more, whereas our proposed method incorporating the sAE achieved an accuracy of around 70% with a training set of 20 sessions or more. This significant improvement of approximately 4% was observed consistently across increasing training set sizes, and highlights the superiority of the sAE over the naive cross-session model in reaching a higher saturation point with less data. Furthermore, the AE model achieved better performance compared to the within session method, which serves as a benchmark for the absence of non-stationary signal noise between the test and training sets. Interestingly, utilizing only the MSE loss in the AE model yielded results similar to those obtained with the sAE, with no significant improvement observed. The classification accuracy for identifying the origin session showed a significant improvement of approximately 9% when using residuals from the AE and sAE models, achieving an accuracy of 70%, compared to the original preprocessed EEG signals, which achieved an accuracy score of 61%. However, when using the reconstructed signals, the accuracy score was considerably lower at approximately 25%. It is worth noting that the chance level for the task was 0.7%. In terms of session classification, both the sAE and AE models performed similarly for both the reconstructed and residual outputs.

3.2 IEEE Dataset

We first evaluated the classification accuracy with respect to the number of training sessions. For the IEEE dataset (4 MI classes), the sAE method substantially outperformed the naive cross-session method (Fig.3a), achieving an accuracy improvement of approximately 10%. Interestingly, our method even outperformed the within-session accuracy by 7%. The improvement was even greater when we removed subjects who performed poorly in the MI task to begin with (Fig. 3b.). Specifically, we excluded from the analysis 11 subjects whose mean accuracy for the within-session was below 30% (leaving N=9 subjects for analysis). The sAE method achieved better results than the MSE AE, with an improvement of approximately 2.5% (Fig. 3b). However, this improvement was only observed when removing the low-performing subjects. To further asses the effectiveness of our proposed method, we compared performance across subjects with the highest and the median WS classification accuracy, Fig.3c and Fig.3d, respectively. For the highest performing subject, our proposed method improved the accuracy by approximately 15% compared to the WS method and by 7% relative to the BS-ASR method (Fig.3c). Similarly, for the median performing subject, our proposed method outperformed the WS and BS-ASR methods by approximately 15% and 10% respectively (Fig.3d).

As discussed in the Introduction, our focus here was on eliminating continuous components related to signal non-stationarity and evaluating the impact on performance. To achieve this, we employed a relatively simple preprocessing pipeline. However, it is valuable to explore the effect of integrating a more advanced artifact removal technique. For the IEEE dataset, we also assessed the impact of applying artifact subspace reconstruction (ASR) [31] on between-session classification. As shown in Fig. 3a and 3b, incorporating ASR during preprocessing improved the performance of naïve between-session classification. However, the performance remained significantly lower compared to that achieved using AEs.

Regarding the classification of the origin session, the residual signals demonstrated comparable accuracy to the original signals, with 82% and 79% accuracy, respectively. In contrast, the reconstructed signals showed significantly lower accuracy with 60% accuracy (Fig. 3e); the chance level was ~14.3%.

There was no notable difference in performance between the uAE and the sAE, as both approaches produced similar outcomes for session classification using the reconstructed and residual signals.

3.3 Shu Dataset

Regarding the Shu dataset, the accuracy of the within-session method exceeded that of the proposed AE cross-session method, with a discrepancy of approximately 5%. Our proposed AE cross-session method performed better than the basic cross-session approach, but the increase in accuracy was marginal, with a difference of around 1% (Fig. 4a). After eliminating participants with a mean within-session accuracy below 60% (N=9), our AE method showed an improvement of 4% over the basic cross-session method; however, this improvement was only evident in the sAE model, whereas in the uAE model, the difference between the two methods was 2% (Fig. 4b).

The classification of origin session showed that the residual signals had similar accuracy to the original preprocessed signals, achieving 95% and 94% accuracy scores, respectively. On the other hand, the reconstructed signals exhibited significantly lower accuracy, with an accuracy of 66% (Fig. 4c); the chance level was 20%. The performance of the uAE and sAE approaches was indistinguishable since both methods yielded comparable results in session classification using both the reconstructed and residual signals.

3.4 Power Spectrum Comparison

To gain further insight into the effect of the AE on EEG signals, we compared the power spectral densities (PSDs) of the original EEG signal, the reconstructed AE EEG signal, and the residual signal (Fig. 5). Our analysis focused on electrodes C3 and C4, which are known to be informative for MI BCI and examined both left and right motor imagery (using the IEEE dataset).

For both electrodes, the PSD of the reconstructed AE signal exhibited typical EEG behavior within the relevant frequency range (<30Hz), while the residual signal exhibited noise-like characteristics.

4. Discussion

This paper introduces a novel approach to address domain adaptation challenges in EEG signal analysis across different sessions. The necessity for domain adaptation arises due to the dynamic nature of brain activity, which makes it non-stationary. Various existing methods for domain adaptation rely on having some data from the target domain (new sessions), employing different techniques to learn a transformation for the new session signals that align with the training data used for model development. In contrast, our proposed method does not depend on training with signals collected from the target domain, making it fully unsupervised with respect to the target domain, and eliminating the need for recalibration with each new session. Instead, it focuses on aligning multiple sessions in a general manner, reducing session-specific information while preserving task-related signals.

Our approach utilizes AEs to extract a low-dimensional representation of the EEG signals that captures the information relevant to the MI task while effectively filtering out session-specific components. Beyond the standard reconstruction error term used in the AE, we introduced two additional objectives during the training process: a task loss and an origin-session loss. The task loss is designed to improve the classification of MI classes, while the origin-session loss seeks to distinguish between sessions using the residual signals. These dual objectives enable the AE to learn more robust feature representations in the latent space, which are less susceptible to session-specific variations and better equipped to handle the inherent non-stationarity of EEG data.

Interestingly, for the stroke patient dataset, our proposed method performed better than the within-session method, which is considered the benchmark for non-stationary data. Additionally, for the IEEE dataset, our AE model yielded higher accuracy scores than the within-session method, suggesting that task-related information is better represented in the reconstructed signals. Notably, the improvement is even more remarkable when subjects with poor performance in the MI task (as determined by the within-session method) were removed. In other words, if the signals are not very informative to begin with, there is not much room for improvement in the model. As for the Shu dataset, our hypothesis that the model can effectively denoise non-stationary signals for MI classification was only partially supported. While the observed improvement was modest, it was nevertheless evident. It is worth noting that the variability between MI-EEG datasets can be substantial due to differences in recording hardware, instructions, experimental protocol, and other factors.

The observed trend in the within-session score for the Shu and IEEE datasets reveals that as additional data are incorporated into the training set, the score progressively decreases. We posit that this decline stems from the inclusion of data originating from multiple sessions, leading to increased variability unrelated to the task at hand (non-stationary variability). Consequently, this variability impedes the model's ability to learn effectively. This can explain the fact that our method outperformed even the within-session benchmark in the IEEE and stroke patient datasets. Conversely, our proposed method exhibited an opposing pattern, demonstrating improvement over time. Even within the Shu dataset, which comprises only 5 sessions, it is plausible that when provided with a greater number of sessions, the AE method will outperform the within-session method. This observation supports another advantage of AE domain adaptation over alternative methods for different domain adaptation scenarios. Unlike other approaches that solely apply the transformation to the test set, our method applies it to both the training and test sets. This enables the model to enhance its learning process by eliminating non-stationary signals from the training set prior to learning.

The significant decrease in accuracy of origin session classification for reconstructed signals, as compared to the original preprocessed signals, lends support to the notion that the loss function applied to the compressed residual signals helped to remove session-specific information from the signals. In conjunction with the improvement in task accuracy, we can reasonably infer that our method effectively denoised non-stationary signals that impede the accuracy of the BCI model.

To gain further insight, we examined the effect of incorporating the ASR method for artifact removal. While it enhanced the performance of the between-session classifier, it was still substantially below the performance obtained with the uAE and sAE. Together with the observation that the AE removes session-specific information, this indicates that the AE does much more than artifact removal.

Our proposed AE model can be applied to a wider range of BCI tasks, as it is simple to use and does not require additional data collection. Our method is compatible with more complex classification models, which can be integrated due to the fact that our method only denoises EEG signals. While state-of-the-art models that address the cross-session domain adaptation in BCI can perform better than our proposed model [35], they all suffer from the need to collect new data each session. In contrast, our model's greatest advantage lies in its ability to achieve high performance using only the provided training set, without the need for data collection or recalibration. Domain adaptation for cross-session recordings in the neural activity of monkeys was also demonstrated during motor tasks with single cell recordings, but not in the BCI context [36]. Our work shows that the idea of unsupervised domain adaptation can be implemented with EEG and in the BCI context. Additionally, the residual signals may contain information that is unrelated to the MI task, but could be studied to explore non-stationary brain dynamics, the subject's mental state, and other brain processes. Such research topics can be addressed using our AE model, which can be viewed not just as a denoising tool for non-stationary signals, but also as a means of disentangling information-carrying signals and non-stationary components.

It is important to emphasize that our primary objective was to address the issue of brain dynamics non-stationarity between sessions, aiming to eliminate the need for system recalibration with each new session. To this end, our goal was to demonstrate the relative contribution of the AE in overcoming non-stationarity, rather than obtaining the highest absolute performance in BCI classification. To achieve this, we used a basic preprocessing approach and employed CSP and multiclass LDA for classification. Clearly, incorporating better artifact removal in the preprocessing stage and applying more sophisticated state-of-the-art classification methods, would improve the absolute performance. For example, in a previous publication [22] we used Filter-Bank CSP (FBCSP) for feature extraction, which led to improved performance on the same data. Here, for considerations related to computational cost, we used a simple CSP approach. Because the focus was on relative improvements, we also did not attempt to compare the absolute performance of our method with state-of-the-art bench marks on the public datasets.

While our previous work, as illustrated in Fig. 3, achieved higher accuracy, it did not effectively address the non-stationarity challenge. In contrast, the current study demonstrates that our supervised AE approach is more robust to these session-specific fluctuations. Furthermore, this study analyzes three different MI datasets and focuses on offline systems. Extending this work to online systems, using more sophisticated classifiers, and other tasks, such as mental imagery BCI, remains a goal for future research. This extension aims to further generalize our findings and validate the effectiveness of our proposed methodology across broader applications.

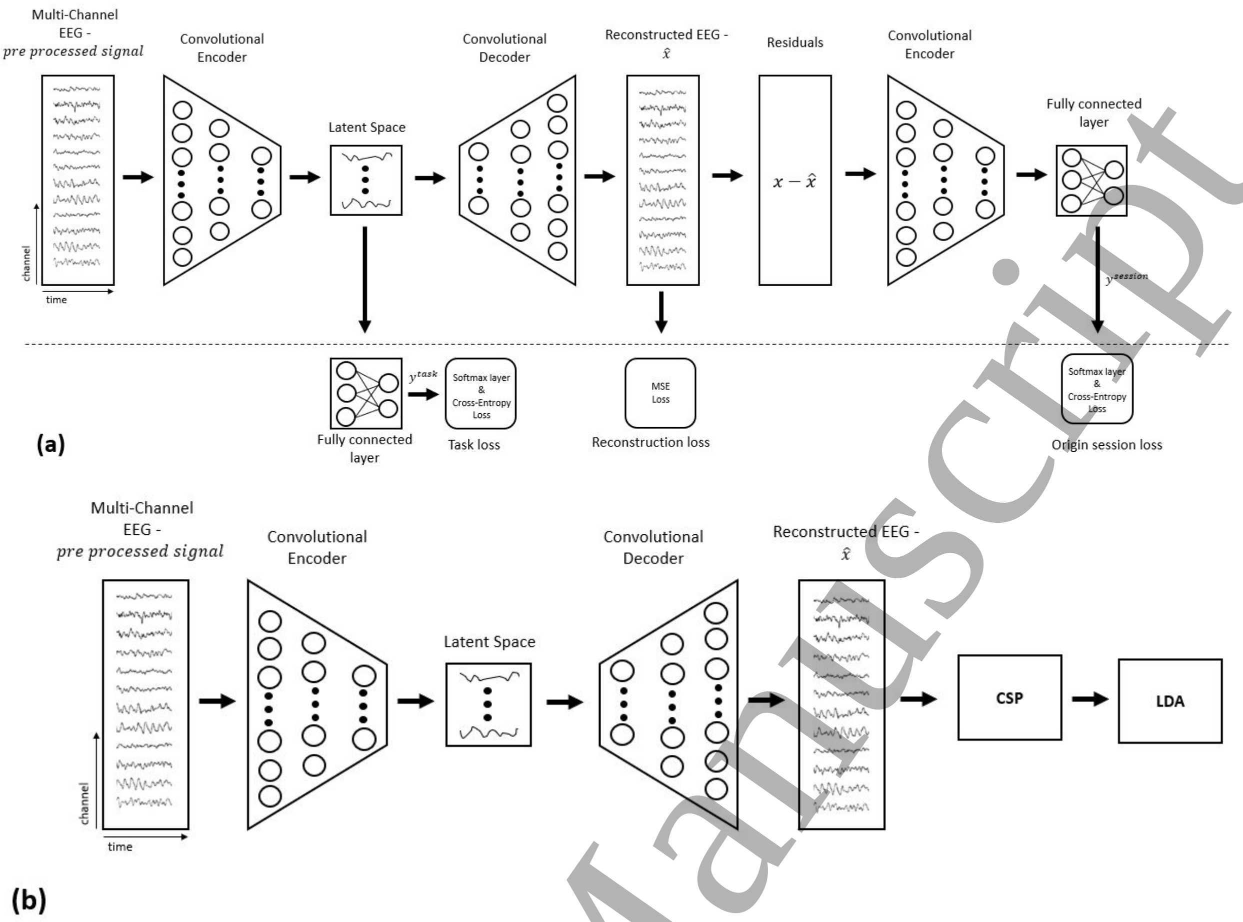

Figure 1. Model Architecture. *a) Architecture for supervised autoencoder (SAE) training. The task loss corresponds to the cross-entropy for predicting task labels (e.g., left/right motor imagery) using the latent space. The origin session loss represents the cross-entropy for predicting the session of data origin, based on the auto-encoder (AE) residuals. The reconstruction loss is the mean squared error between the input and output of the AE. The model is trained to minimize the sum of these three loss terms. b) Architecture for common spatial pattern (CSP) and linear discriminant analysis (LDA) training and inference. The AE component of the network from part a is used to reconstruct the signals as a preprocessing step before applying the CSP-LDA classifier.*

Method	Training set	Test set
Cross-Session	Sessions 1 to K	Sessions K+1 to L
Within-Session	80% of trials from the same session	The remaining 20% of trials from the same session
Data Origin Session	80% of trials from different sessions, labeled with session number	The remaining 20% of trials from different sessions, labeled with session number

Table 1. *Overview of training and test set divisions for each method. The 'Cross-Session' method uses data from different sessions, the 'Within-Session' method focuses on data from the same session, and the 'Data Origin Session' method emphasizes the session from which the data originated.*

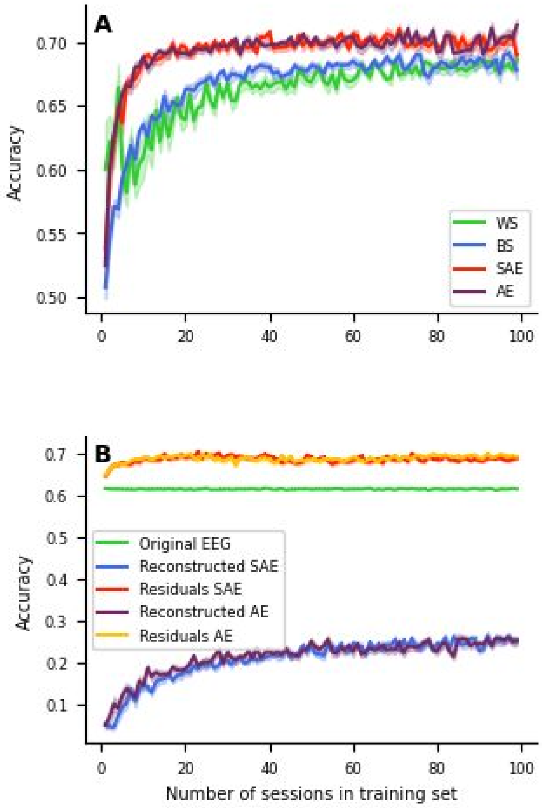

Figure 2. Stroke patient results. *Mean over 10 realizations; shaded are standard error over realizations. A) Classification accuracy for MI task using the within-session (WS) method (green), between session (BS) method (blue), unsupervised auto-encoder (uAE) method (purple), and supervised auto-encoder (sAE) method (red). B) Classification accuracy for the session from which the data originated. Original preprocessed signals (green), reconstructed signals from the uAE (purple) and sAE (blue), and the residuals from the AE (yellow), and sAE (red).*

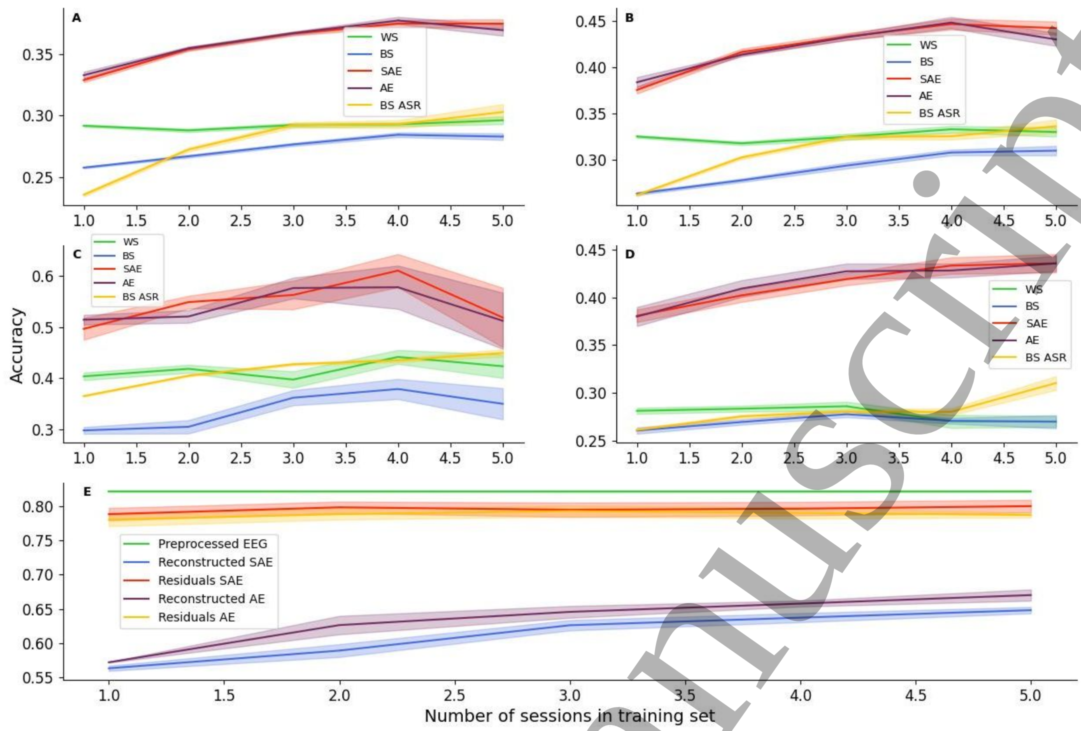

Figure 3. IEEE dataset results. *Mean classification accuracy over 10 realizations, with shaded areas representing the standard error. A) Classification accuracy for the MI task, using the within-session (WS) method (green), between-session (BS) method (blue), between-session with artifact-subspace reconstruction (BS-ASR) method (yellow), and the unsupervised (uAE) and supervised (sAE) auto-encoder methods in purple and red, respectively. B) Classification accuracy for the MI task considering only subjects who achieved WS accuracy above 30%. C) Classification accuracy for the subject with the highest WS accuracy. D) Classification accuracy for the subject with median WS accuracy. E) Classification accuracy for the session from which the data originated. Preprocessed EEG signals (green), the reconstructed signals from the uAE (purple) and sAE (blue), and the residuals from the uAE (yellow) and sAE (red).*

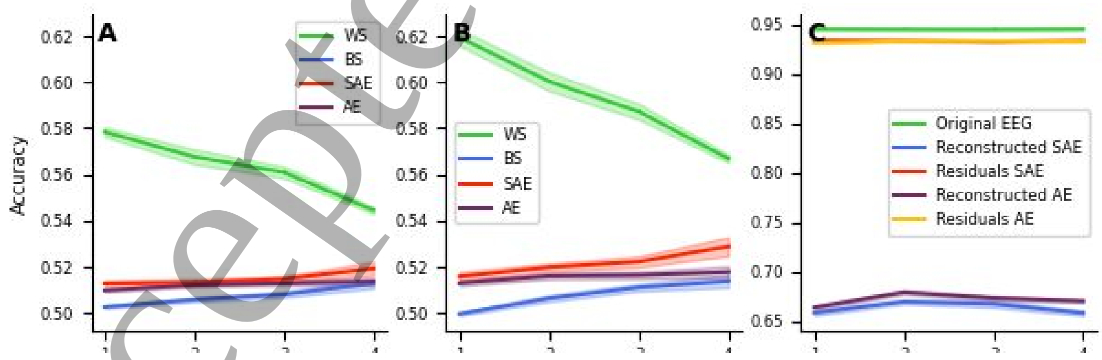

Figure 4. Shu dataset results. *Mean over 10 realizations; shaded are standard error over realizations. A) Classification accuracy for the MI task, using the within-session (WS) method (green), between-session (BS) method (blue), and the uAE (purple) and sAE (red) methods. B Classification accuracy for the MI task considering only subjects who achieved WS method accuracy above 30%. C) Classification accuracy for the session from which the data originated. Original preprocessed signals (green), the reconstructed signals from the uAE (purple) and sAE (blue), and the residuals from the uAE (yellow) and sAE (red)*

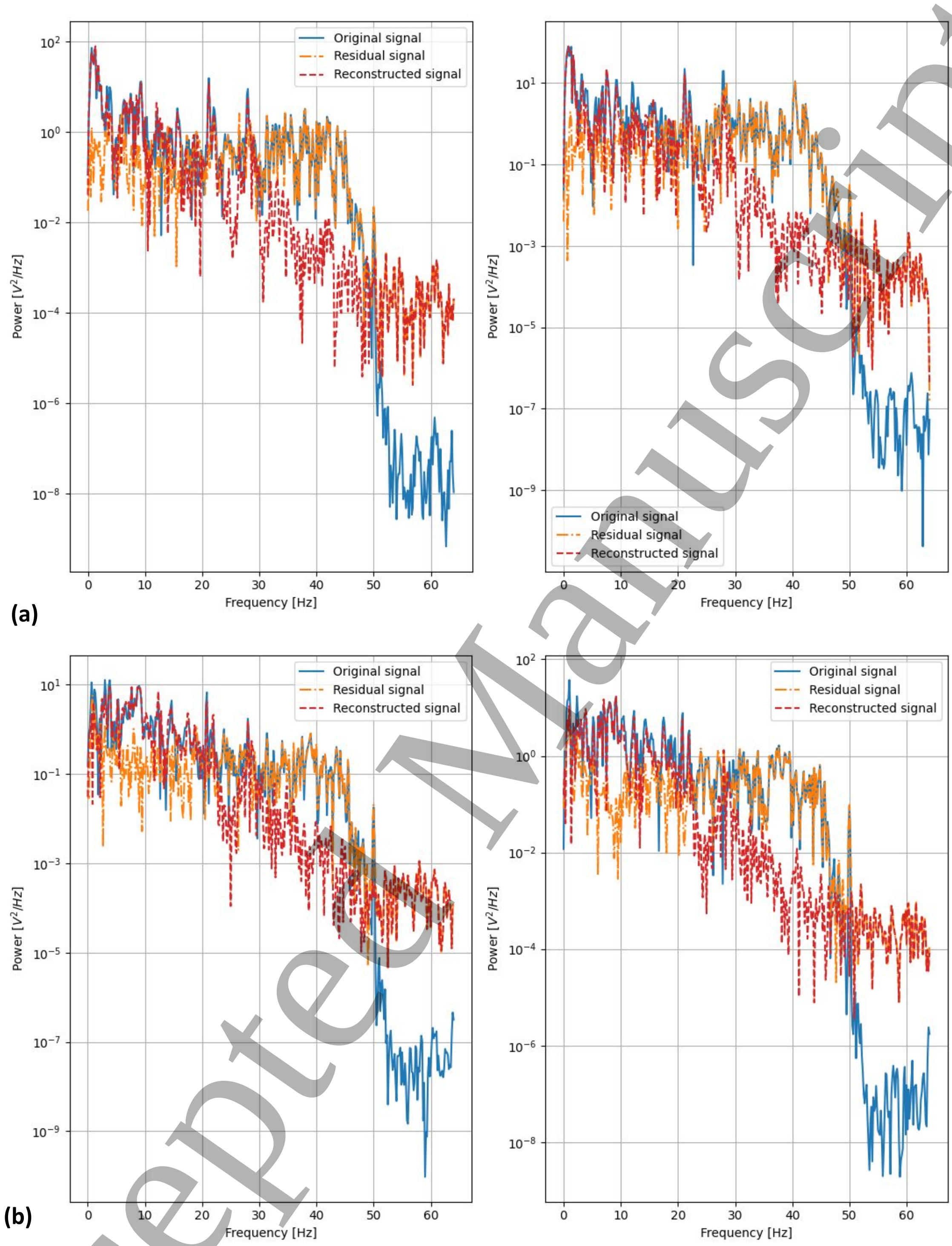

Figure 5. Power spectrum results of electrodes C3 and C4. *A) Power spectrum of electrodes C3 (left) and C4 (right) for right hand MI. B) Power spectrum of electrodes C3 (left) and C4 (right) for left hand MI. The analysis was based on the IEEE dataset.*

References

[1] 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.

[2] 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.

[3] 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.

[4] J. Wolpaw, N. Birbaumer, D. McFarland, G. Pfurtscheller and T. Vaughan, "Brain–computer interfaces for communication and control", *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767-791, Jun. 2022.

[5] R. Mane, T. Chouhan, and C. Guan, “BCI for stroke rehabilitation: motor and beyond,” *Journal of Neural Engineering*, vol. 17, no. 4, 2020.

[6] 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.

[7] A. Arieli, A. Sterkin, A. G therinvald, 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.

[8] D. Deitch, A. Rubin, and Y. Ziv, “Representational drift in the mouse visual cortex,” *Current Biology*, vol. 31, no. 19, pp. 4327-4339.e6, Oct. 2021.

[9] M. E. Rule, A. R. Loback, D. V. Raman, L. N. Driscoll, C. D. Harvey, and T. O’Leary, “Stable task information from an unstable neural population,” *eLife*, vol. 9, p. e51121, Jul. 2020.

[10] L. N. Driscoll, N. L. Pettit, M. Minderer, S. N. Chettih, and C. D. Harvey, “Dynamic reorganization of neuronal activity patterns in parietal cortex,” *Cell*, vol. 170, no. 5, pp. 986-999.e16, Aug. 2017.

[11] C. Clopath, T. Bonhoeffer, M. Hübener, and T. Rose, “Variance and invariance of neuronal long-term representations,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 372, no. 1715, p. 20160161, Mar. 2017.

[12] 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.

[13] A. Abu-Rmileh, E. Zakkay, L. Shmuelof, and O. Shriki, “Co-adaptive training improves efficacy of a multi-day EEG-based motor imagery BCI training,” *Frontiers in Human Neuroscience*, vol. 13, p. 362, Oct. 2019.

[14] J. Faller, R. Scherer, U. Costa, E. Opisso, J. Medina, and G. R. Müller-Putz, “A co-adaptive brain-computer interface for end users with severe motor impairment,” *PLoS ONE*, vol. 9, no. 7, p. e101168, Jul. 2014.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] L. Gondara, “Medical image denoising using convolutional denoising autoencoders,” IEEE Xplore, 2016.

[22] O. Almagor, O. Avin, R. Rosipal, and O. Shriki, “Using autoencoders to denoise cross-session non-stationarity in EEG-based motor-imagery brain-computer interfaces,” *IEEE Xplore*, Nov. 01, 2022.

[23] G. Wang, Y. Tang, L. Lin, and P. H. S. Torr, “Semantic-aware auto-encoders for self-supervised representation learning,” *openaccess.thecvf.com*, 2022.

[24] M. Tschannen, O. Bachem, and M. Lucic, "Recent advances in autoencoder-based representation learning," *arXiv:1812.05069 [cs, stat]*, Dec. 2018, Available: https://arxiv.org/abs/1812.05069

[25] A. Gogna, A. Majumdar and R. Ward, "Semi-supervised stacked label consistent autoencoder for reconstruction and analysis of biomedical signals," in *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 9, pp. 2196–2205, Sept. 2017.

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

[27] 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.

[28] 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.

[29] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol. 7, no. 2, pp. 179–188, Sep. 1936.

[30] Shang-Lin Wu, Chun-Wei Wu, N. R. Pal, Chih-Yu Chen, Shi-An Chen and Chin-Teng Lin, "Common spatial pattern and linear discriminant analysis for motor imagery classification," *2013 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB)*, Singapore, pp. 146-151.

[31] Gaur, Pramod, et al. "A multi-class EEG-based BCI classification using multivariate empirical mode decomposition based filtering and Riemannian geometry." Expert Systems with Applications 95 (2018): 201-211.

[32] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," arXiv.org, 2014.
.

[33] Maas, Andrew L., Awni Y. Hannun, and Andrew Y. Ng. "Rectifier nonlinearities improve neural network acoustic models." *Proc. icml*. Vol. 30. No. 1. 2013.

[34] Chang, Chi-Yuan, et al. "Evaluation of artifact subspace reconstruction for automatic EEG artifact removal." 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2018.

[35] Ma, X., Rizzoglio, F., Perreault, E. J., Miller, L. E. & Kennedy, A. Using adversarial networks to extend brain computer interface decoding accuracy over time. http://biorxiv.org/lookup/doi/10.1101/2022.08.26.504777 (2022)

[36] J. Jude, M. G. Perich, L. E. Miller, and M. H. Hennig, "Robust alignment of cross-session recordings of neural population activity by behaviour via unsupervised domain adaptation," *arXiv.org*, Feb. 16, 2022.