A Self-supervised Task-agnostic Embedding for EEG Signals

Abstract—**Brain-Computer Interfaces (BCIs) have great potential for improving the lives of people with disabilities. The success of a BCI system is largely driven by the accuracy of the BCI decoder. This accuracy, in turn, may be limited by the amount of labelled training data available for supervised machine learning algorithms. The success of deep learning algorithms in other computer science areas has not reached the field of BCI decoding due to this lack of abundant labelled data. We use a novel deep learning architecture trained in a self-supervised manner to learn a common vector representation (embedding) of EEG signals that can be used in different BCI tasks. The vector representation is trained using EEG recordings without using any task labels. We validate our embedder using two separate BCI tasks: seizure detection and motor imagery, and assess its usefulness through distance similarity metrics in a clustering approach. The derived embeddings were successful in distinguishing binary classes in both tasks.**

Keywords—Self-supervised learning, Embedding, Transfer Learning, Representation Learning

I. Introduction (*Heading 1*)

Brain-Computer Interface (BCI) systems offer the means to interact with computers and other electronic devices directly through brain signals, aiming to assist individuals with communication and motor function limitations. BCI systems have gained traction in recent years, benefiting from recent advancements in hardware as well as developments in the field of Machine Learning (ML) [1].

Interpreting BCI signals, also called decoding, can be achieved through statistical or supervised ML methods by learning from brain signals or features extracted from them and identifying the common patterns representing different BCI tasks like the intention of moving muscles (motor imagery) or predicting seizures [2]. The usefulness of a BCI system is largely affected by the accuracy of this decoder, which, in turn, can be influenced by the quality and amount of available labelled training data.

In recent years, deep learning methods have dominated fields like natural language processing (NLP) and Computer Vision (CV) [3-7], where large amounts of labelled training data are available. where large amounts of labelled training data are available. Deep learning models for BCI decoding are gaining popularity [8-10] due to their improved performance and reduction in feature engineering efforts through learning alternative representations of data. However, a major challenge in their utility and success is lack of abundant annotated training data.

To address the issue of data sparsity in BCI decoding, we have explored two related ML concepts of Transfer Learning (TL) and Self-Supervised Learning (SSL) in a newly proposed deep neural network architecture. TL aims to share knowledge between ML models to reduce the required amount of training data, and SSL aims to remove the need for supervised (labelled) training data by utilising large volumes of unlabeled data to learn commonly shared features of the data. In our model, we trained a deep neural network that learns a common low-dimensional vector representation of EEG signals (an embedding), which then can be used in downstream tasks of BCI decoding as a feature engineering layer. The main benefit of this embedding is that it is task-agnostic; i.e., the same trained embedder can be used in different BCI tasks by appending a classifier and fine-tuning it using a smaller amount of task-specific labelled data.

Pre-trained embeddings have been successfully used in other computer science areas (CV [11], NLP [12]) for building a feature extraction network that turns high-dimensional input (like images, text, or waves) into low-dimensional vector representations (a small sized array of numbers). Audio processing shares similarities with brain signal processing; PASE+ [13] is a task-agnostic audio embedder that uses known audio features as labels in an SSL paradigm and it is the inspiration behind our proposed architecture.

We trained the network using an SSL paradigm, without providing data labels, reducing the total amount of training data required for the downstream model, and ensuring the task-agnostic nature of the learnt embedder. We used our collection of EEG recordings from an in-clinic epilepsy monitoring program, which is labelled for seizures, but did not use the seizure labels for training.

To assess the utility of the learnt embedder, we used two publicly available datasets in two different BCI domains: seizure detection using the CHB-MIT dataset [14] and motor imagery using the BCI Competition IV 2a [15].

We clustered the embeddings derived from the new data through our model without any up-training or fine-tuning. The clusters were able to distinguish two distinct classes of BCI state in each task. The embeddings generated from the CHB-MIT dataset distinguished between two classes of seizure and no seizure. The embeddings generated from the BCI Competition IV 2a distinguished between two classes of moving the right and left arms.

Our proposed method and architecture can be used to develop a transferable and task-agnostic way to train BCI decoders using unlabelled data. Given the cost and effort associated with annotating BCI data, this model can have a significant contribution in improving the BCI decoding process. The significance of this method is in its potential to achieve a common vector representation of brain signals, which is agnostic to the BCI task. Ultimately, it can help improve the decoder accuracy in domains where not enough labelled training data is available.

II. Method

A. Overview

Figure 1 provides an overview of our approach. The experiments were run in two stages:

1. In Stage 1, we trained a novel neural network architecture using the "SVHM" dataset in an SSL manner (both the model and dataset described

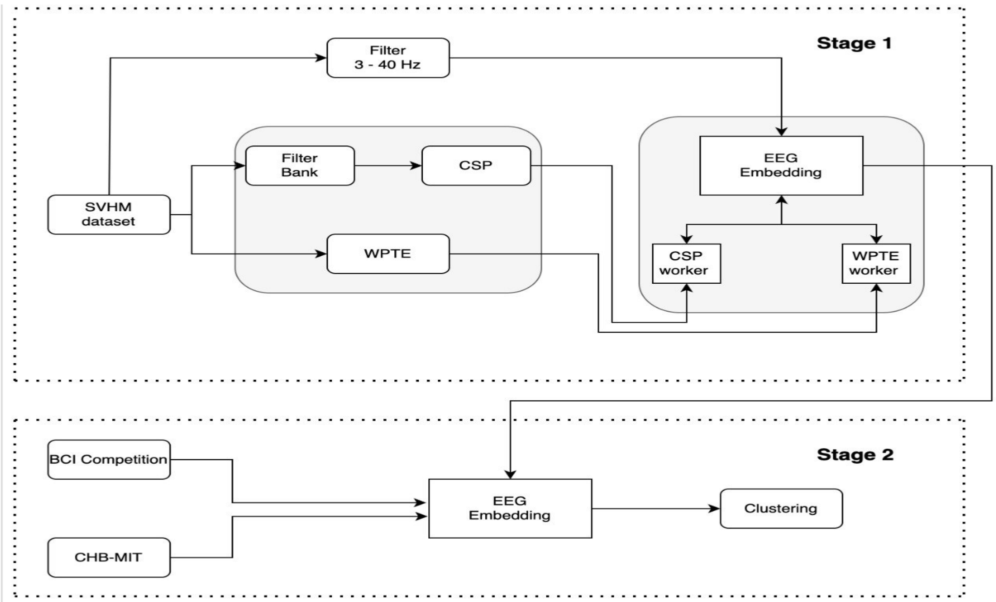

Fig. 1. Overall System Design

below). The goal of this step was to create a task-agnostic vector representation learner (embedder) for the EEG signals.

2. In Stage 2, we evaluated the embedder in two distinct downstream tasks of seizure detection (using the CHB-MIT dataset) and motor imagery (using the BCI competition dataset) using clustering.

B. Dataset

Stage 1 – Clinical Epilepsy monitoring dataset (SVHM dataset)

To train the embedder in Stage 1, we used a dataset of large volume without the need for a specific BCI task or label. To this end, we collected the 24/7 epilepsy monitoring EEG recordings of four patients with epilepsy who were monitored at St Vincent's Hospital, Melbourne (SVHM), Australia. Although this data is annotated by clinicians on seizure activities, we did not use the labels in Stage 1 training and used both the ictal and pre-ictal signals. The montage used in this experiment was the International 10-20 System for 21 channels with the 256 Hz sample rate. We used over 286 hours of recording. Each trial was divided into 4 seconds and was high-pass filtered for DC removal.
The Office of Research Ethics and Integrity at University of Melbourne approved this research (ID14613).

Stage 2 – BCI Competition IV Dataset 2a and CHB-MIT

For Stage 2, we utilised two distinct labelled datasets to demonstrate the task-agnostic nature of the embedder. The BCI Competition IV-2a dataset [15] consists of recordings with 22 EEG electrodes on nine distinct participants. Signals between 0.5 Hz and 100 Hz were bandpass filtered before being sampled at 250 Hz to get the data. We used left and right hand movement in a binary setting. The dataset consists of two sessions per individual, each of which has 288 trials. Each trial lasted 7.5 s, but we used the 3 s motor imagery component of each trial per the stated protocol, up-sampled the data to 256 Hz, and then DC from the EEG data using a high-pass filter with a cutoff frequency of 2 Hz [15].

Another dataset used in Stage 2 was CHB-MIT [14], which is recorded from 23 pediatric patients with intractable epilepsy at Boston Children's Hospital, USA. The montage is the International 10-20 System for electrode positioning in a bipolar configuration. The EEG is 16 bits at 256 Hz sampling rate. Trained specialists annotated seizure events on this data. Due to the different channel numbers among these patients, we used the data from 10 patients that had the same montage as the SVHM dataset. We studied 347 hours of inter-ictal activity and 1 hour of ictal activity from 51 annotated seizures.

C. Model Architectures

We set out to learn a task-agnostic embedder that would learn to represent high-dimensional EEG signals with a lower-dimensional vector through an SSL training regime. This model on the high-level was inspired by the PASE+ model [10], designed for audio processing. We redesigned the model for BCI decoding and to adapt to the multi-channel nature of the EEG signals, namely redesigned the SincNet layer, the vector representation layer, and all the workers. We also incorporated the topological position of EEG electrodes in our montage.

Figure 2 shows the overall architecture of our neural network with its two main components: the Embedder and the Workers. The Embedder component learns the vector representation of the EEG signals that maximises the learning objective. The Worker component performs known transformations on the signals (called features) and uses them as self-learned labels to guide the learning objective.

Embedder

The embedder layer uses a Convolutional Neural Network (CNN) architecture to learn a vector representation of the EEG signal. CNNs have a long history of being used in BCI decoders as they provide a feature extraction capability inherently built in the model [16].

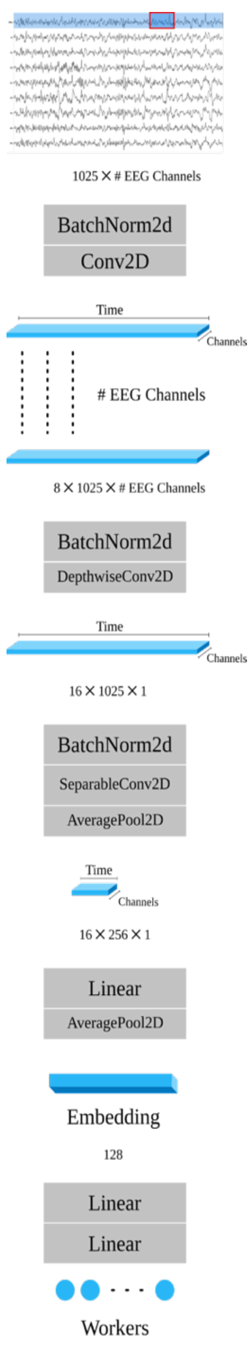

Fig. 2. Neural Network Architecture

We used EEGNet, a compact CNN architecture for EEG-based BCIs [3], as our embedder architecture (see Fig. 2). EEGNet uses separable convolutions made up of a depthwise convolution followed by a pointwise convolution [17] to learn temporal features and the best way of combining them. The first layer of the network extracts features along the time domain using multiple 1-dimensional kernels. In order to use EEGNet as an embedder, we replaced the last fully-connected classification layer with two fully-connected layers, creating a fixed-size embedding vector of 128 elements. This 128-element vector is a lower dimensional representation of the input EEG signal.

Worker Layer

The Worker layer comprised several regressor components that were fed by the embedding vectors generated by the embedder layer. They solved a self-supervised regression task, where the ground-truth labels were created by known transformations (features) rather than explicitly provided labels.

The Worker components were designed as small and shallow networks with few learnable parameters in order to delegate the learning task to the transferable and task-agnostic component (Embedder). Each component of the Worker layer corresponds to a feature.

The network was trained with batch sizes of 32, over 200 epochs, and with a learning rate of 0.001 using stochastic gradient descent.

Features

Features act as known signal transformations that the SSL system uses to generate implicit labels for an unlabeled dataset. Each Worker component in the Worker layer corresponds with a feature.

Similar to their roles in traditional ML, features extract information relevant to the ML task from raw input data. In the context of EEG brain signals, this could take various forms to encode information about time, frequency, time-frequency, and spatial information of the signals.

Features are selected based on domain knowledge of EEG signals, and can be easily expanded without changing the model architecture. In this work, we used Wavelet Packet Tsallis Entropy (WPTE) [18], which is a time-frequency wavelet transform model projecting signals into a space of mutually orthogonal wavelet basis functions, and Filter Bank Common Spatial Pattern (FBCSP) [19], which is a spatial filtering where signals are transformed into a variance matrix representing the discrimination between classes.

As the Embedder and the Worker were jointly trained in an end-to-end manner, the learnt embeddings incorporated elements of all these features.

Stage 2 - Clustering

Stage 2 was used to assess the utility of the Stage 1 trained embedder. To this end, we passed the data from the two public datasets through our learnt embedder and used the vector representations as an input in a clustering system. We ran a k-means++ clustering algorithm with Euclidean distance to assess the quality of these generated embeddings. No up-training or fine-tuning was performed in this stage.

III. Results And Discussion

Clustering EEG signals using our learnt embedder was able to distinguish two separate clusters of seizure and no seizure for the CHB-MIT dataset (Fig 3, top). Similarly, in the motor imagery task, it distinguished left and right hand movement in the BCI Competition dataset (Fig 3, bottom).

Fig 3 shows the clustering evaluation using three methods: The Error elbow, Calinski score, and silhouette score. The elbow method (Fig 3a, d) shows that 2 is the optimal number of clusters formed by our embedder. Both the Calinski and the silhouette score evaluate clusters based on their between-cluster and within-cluster characteristics. The Clainski score (Fig 3b, e) shows that the 2 clusters lead to the highest between-cluster variance and lowest in-cluster variance, and the silhouette score (Fig 3c, f) shows a high degree of in-cluster cohesion in each of the two clusters.

These results validate the embedder's capability in distinguishing classes in different tasks, using unlabeled data as training.

IV. Conclusion

We propose an embedding architecture that is trainable using an SSL paradigm to learn low-dimensional, task-agnostic representations of EEG signals. The utility of this embedder was verified in two distinct downstream BCI tasks of motor imagery and seizure detection.

The Worker layer of our architecture plays a significant role in the quality of embeddings and can be further expanded by adding more features. This layer also helps in keeping the embedding agnostic by incorporating features that are useful in various BCI tasks. Thus, expanding the Workers can expand the downstream tasks achievable by the embedding.

Overall, our proposed method has the potential to transform BCI decoding through reducing the need for large labelled datasets for different BCI tasks.

References

[1] McFarland, D.J.; Wolpaw, J.R. Brain-computer interfaces for communication and control. Commun. ACM 2011, 54, 60–66. [CrossRef]

[2] Rasheed, S. (2021). A review of the role of machine learning techniques towards brain–computer interface applications. *Machine*

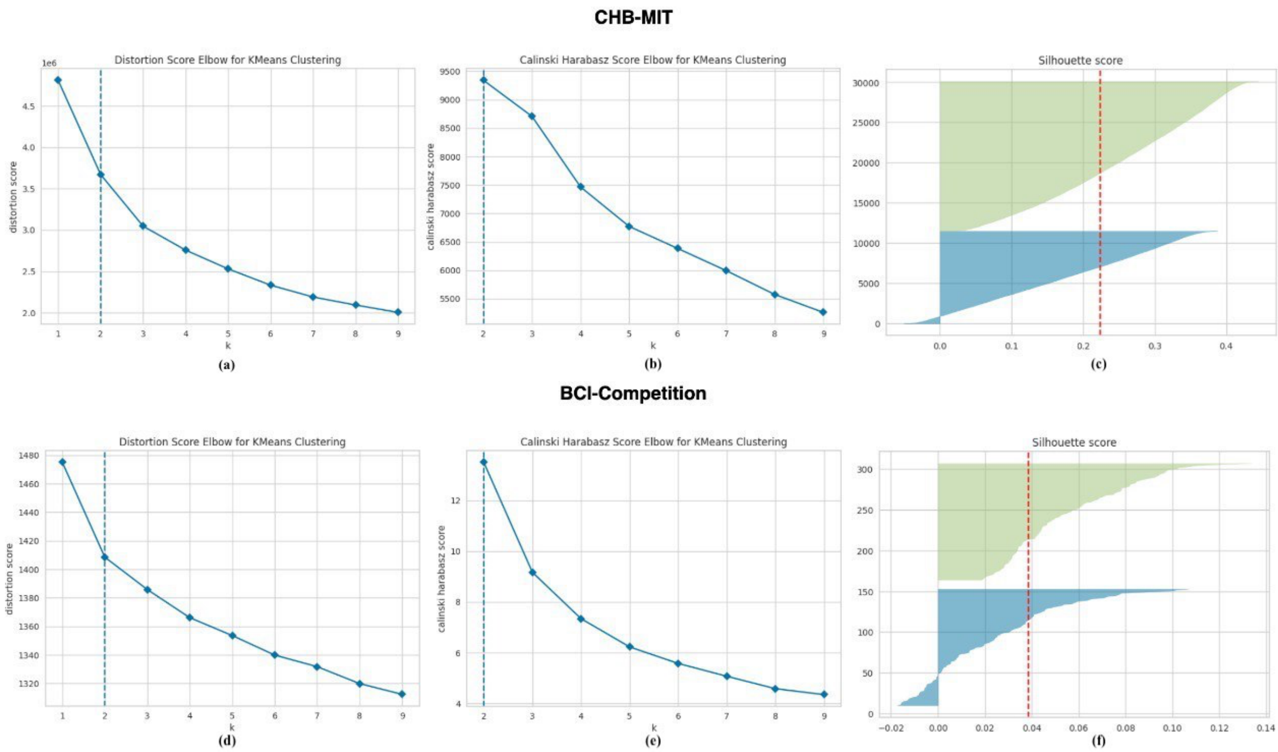

Fig. 2. Clustering Performance Over Two BCI Tasks

Learning and Knowledge Extraction, *3*(4), 835–862. https://doi.org/10.3390/make3040042

[3] M. Hersche, T. Rellstab, P. D. Schiavone, L. Cavigelli, L. Benini, and A. Rahimi, “Fast and Accurate Multiclass Inference for MI-BCIs Using Large Multiscale Temporal and Spectral Features,” in Proc. IEEE EUSIPCO, 2018, pp. 1690–1694.

[4] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, “EEGNet: a compact convolutional neural network for EEG-based brain computer interfaces,” Journal of Neural Engineering, vol. 15, no. 5, p. 056013, 2018.

[5] D. Li, J. Wang, J. Xu, and X. Fang, “Densely Feature Fusion Based on Convolutional Neural Networks for Motor Imagery EEG Classification,” IEEE Access, vol. 7, pp. 132 720–132 730, 2019.

[6] Y. Zhao, S. Yao, S. Hu, S. Chang, R. Ganti, M. Srivatsa, S. Li, and T. Abdelzaher, “On the improvement of classifying EEG recordings using neural networks,” in Proc. IEEE Big Data, 2017, pp. 1709–1711.

[7] M.-A. Li, J.-F. Han, and L.-J. Duan, “A Novel MI-EEG Imaging With the Location Information of Electrodes,” IEEE Access, vol. 8, pp. 3197– 3211, 2020.

[8] Jiang, X., Zhao, J., Du, B., & Yuan, Z. (2021). Self-supervised contrastive learning for EEG-based sleep staging. 2021 International Joint Conference on Neural Networks (IJCNN). [CrossRef]

[9] Kostas, D., Aroca-Ouellette, S., & Rudzicz, F. (2021). BENDR: Using Transformers and a contrastive self-supervised learning task to learn from massive amounts of EEG Data. Frontiers in Human Neuroscience, 15. [CrossRef]

[10] Pascual, S., Ravanelli, M., Serrà, J., Bonafonte, A., & Bengio, Y. (2019). Learning problem-agnostic speech representations from multiple self-supervised tasks. Interspeech 2019. [CrossRef]

[11] Kim, H. E., Cosa-Linan, A., Santhanam, N., Jannesari, M., Maros, M. E., & Ganslandt, T. (2022). Transfer learning for medical image classification: A literature review. *BMC Medical Imaging*, *22*(1). https://doi.org/10.1186/s12880-022-00793-7

[12] Wang, H., Li, J., Wu, H., Hovy, E., & Sun, Y. (2022). Pre-trained language models and their applications. *Engineering*. https://doi.org/10.1016/j.eng.2022.04.024

[13] Pascual, S., Ravanelli, M., Serrà, J., Bonafonte, A., & Bengio, Y. (2019). Learning problem-agnostic speech representations from multiple self-supervised tasks. Interspeech 2019. [CrossRef]

[14] Goldberger, A., et al. "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation [Online]. 101 (23), pp. e215–e220." (2000).

[15] C. Brunner, R. Leeb, G. R. Muller-Putz, A. Schl ¨ ogl, and G. Pfurtscheller, ¨ “BCI competition 2008 - Graz data set A.”

[16] Saeidi, M., Karwowski, W., Farahani, F. V., Fiok, K., Taiar, R., Hancock, P. A., & Al-Juaid, A. (2021). Neural decoding of EEG signals with Machine Learning: A Systematic Review. *Brain Sciences*, *11*(11), 1525. https://doi.org/10.3390/brainsci11111525

[17] F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” CoRR, vol. abs/1610.02357, 2016. [Online]. Available: [CrossRef]

[18] K. Qian et al., “A bag of wavelet features for snore sound classification,” Ann. Biomed. Eng., vol. 47, no. 4, pp. 1000–1011, 2019.

[19] Wei Wu, Xiaorong Gao, & Shangkai Gao. (2005). One-versus-the-rest(ovr) algorithm: An extension of common spatial patterns(CSP) algorithm to multi-class case. 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference. [CrossRef]