Predicting local fMRI activations from EEG: a Feasibility Study Using Both Classical and Modern Machine Learning Pipelines

Tomer Amit
The School of Computer Science
Tel Aviv University
Tel Aviv, Israel
tomeramit1@mail.tau.ac.il

Taly Markovits
Sagol school of neuroscience
Tel Aviv University
Tel Aviv, Israel
talym@mail.tau.ac.il

Guy Gurevitch
Sagol school of neuroscience
Tel Aviv University
Tel Aviv, Israel
guygu4@gmail.com

Talma Hendler
Sagol school of neuroscience
Tel Aviv University
Tel Aviv, Israel
talma@tlvmc.gov.il

Lior Wolf
The School of Computer Science
Tel-Aviv University
Tel Aviv, Israel
wolf@cs.tau.ac.il

***Abstract*—fMRI's clinical use is limited by cost, while EEG is more accessible but lacks spatial detail and deep brain coverage. Research aims to predict deep brain activations from combined fMRI and EEG data. We compare classical machine learning and a CNN-transformer pipeline for this mapping across multiple brain regions. As we show, in the first dataset, which is heavily tilted toward visual perception, the activations in the Hippocampus cannot be recovered reliably from EEG, using either pipeline. However, in other regions, predictability is much higher, and in those cases, the deep learning pipeline obtains better predictions. In a second dataset that is based on musical feedback while the visual is blocked, both pipelines yield improved results in the Hippocampus.**

***Index Terms*—fMRI fingerprinting, EEG time series analysis, Self-supervised learning.**

I. Introduction

Functional Magnetic Resonance Imaging (fMRI) measures and map brain activity through blood flow changes. It offers high spatial resolution and detailed information about brain regions, including deep ones [1], [2]. Electroencephalography (EEG) provides high temporal resolution but limited spatial resolution, especially for deep brain regions [3]. EEG equipment is significantly less expensive to purchase, maintain, and operate compared to MRI machines. Predicting fMRI signals from EEG data aims to merge the strengths of both modalities with lower cost, providing a more comprehensive understanding of brain activity, cognitive processes, and their applications in research and clinical contexts. It also represents an interdisciplinary approach to unraveling the intricacies of brain function. Previous studies on the feasibility of an EEG to fMRI mapping have been conducted. [4], [5] have shown its feasibility using Ridge regression, and [6] have used Partial Least Squares regression (PLS).

In this work, we repeat these studies, which rely on a classical machine learning pipeline, and also employ a modern deep learning pipeline for predicting fMRI signals from EEG data. The deep learning approach follows recent work in ECG mapping to other types of measurements [7] and employs contrastive learning as an unsupervised pre-training phase, followed by finetuning for the specific goal of fMRI signal prediction using a regression objective.

We evaluated both the classical machine learning pipeline and the modern deep learning one on two paired fMRI and EEG datasets. As we show, in the first dataset, effective prediction is feasible, only on some of the brain regions. Where it is possible, there is an improved performance of the deep learning pipeline. In the second dataset, better results in the Hippocampus are obtained by both pipelines.

Our main claims are: (1) The novel deep learning pipeline outperforms the classical one when the signal is strong enough; (2) There is a significant impact of the data collection setup on the models' correlations to the ground truth; and (3) Self-supervised pre-training leads to superior prediction results.

II. Related work

Self-supervised learning, has gained significant research attention in recent years by allowing create a general representation from unlabeled data in different domains [8]–[14].

Contrastive learning become popular as an effective self-supervised learning paradigm. This approach leverages the inherent structure within unlabeled data by contrasting positive pairs against negative pairs. In computer vision, the pre-training of SimCLR [15] is achieved by augmenting the same image as positive examples while taking the rest of the images as negative examples in a large batch size for training. In the field of signal processing, [12] introduces a method that encodes speech audio into latent representations and then masks spans of these representations. Similarly [16] pre-train their models on large, unlabeled EEG datasets, following [12] at a high level. Then, they fine-tune these models for a specific classification tasks, such as sleep stage classification in their work. Most similar to our work, [7] apply temporal and contextual contrastive losses to the latent features produced by their encoder. They pre-train their model on unlabeled ECG data, then fine-tune it using a classification head for various tasks, such as glucose levels classifcation and emotion classifcation.

Our work is differs from previous studies in several key aspects. First, we adopt the extensive set of contrastive losses introduced by [7], with specific adaptations for EEG being multi-channel data. Second, to handle our pre-processed EEG data, we use a smaller encoder in terms of both the number of layers and features compared to [16]. Third, we use the same dataset in the pre-training and fine-tuning phases. Finally, our downstream task is regression, compared to the classification tasks in [7], [16].

III. Datasets

fMRI data was acquired on a 3T GE scanner using T2*-weighted EPI (TR/TE/flip: 3000/35/90; FOV: 20x20cm; 39 3mm axial slices). Preprocessing was performed using fMRIPrep included coregistration, normalization, unwrapping, noise component extraction, segmentation, and skull-stripping. ROI time series were extracted using NiftiLabelsMasker with zscore standardization and a frequency band of 0.01-0.1 Hz, without smoothing. The first 10 frames were discarded to account for T2* equilibration effects.

EEG data was recorded continuously during scanning via a BrainAmp ExG MRI-compatible system, with sampling rate of 5000 Hz and 30 or 31 EEG channels. The preprocessing included gradient artifacts using the FASTR algorithm [17], downsampling to 250Hz and ICA components decomposition. Additional steps involved band-pass filtering, removal of noisy segments, and correction of outliers. We next represented the preprocessed EEG time series in the time-frequency domain in each channel, extracting the log-power of eight frequency bands from the time series of each channel. The band power estimation was performed in sliding windows of 1 [sec] and an overlap of 0.5 [sec], resulting in a time-series with the sampling rate of 2 Hz. The division into bands is [0-2; 2-4; 4-8; 8-12; 12-16; 16-20; 20-25; 25-40] total of 8. To account for the hemodynamic response of fMRI data, -12 to -2.5 seconds was taken, 19 features per channel.

A. Datasets:

The study utilized two datasets of synchronized EEG and fMRI. The first dataset (D_1) obtained from [18] included data from 32 healthy subjects, while they were watching 40 nine-second film clips. The second dataset (D_2) sourced from [19] included data from 26 healthy subjects, who passively listened to pleasurable musical excerpts with their eyes closed. A total of eight excerpts, lasting three minutes each, in two runs.

Formally, the datasets are defined as follows: $D_1 = \{(X_i, Y_i)\}_{i=1}^{28512}$, where $X_i \in \mathbb{R}^{19\times30\times8}$ and $Y_i \in \mathbb{R}^1$. $D_2 = \{(X_i, Y_i)\}_{i=1}^{62590}$, where $X_i \in \mathbb{R}^{19\times31\times8}$ and $Y_i \in \mathbb{R}^1$. In both datasets, Y_i represents the mean of fMRI values in the specific target region, while X_i is a three-dimensional tensor composed of (#EEG features) $\times$ (#EEG channels) $\times$ (#frequency bands).

B. Brain regions:

Hippocampus is a curved structure located in the medial temporal lobe, and is associated with memory formation and spatial navigation. **Pallidum** is part of the basal ganglia, located in the subcortical regions of the brain, involved in motor control and various cognitive functions. **Precuneus** is a region located in the parietal lobe, and its role has been studied extensively in neuroscience. **Inferior parietal** lobule is located in the parietal lobe, and involved in various cognitive functions. **Visual** is a region located in the occipital lobe, which is primarily associated with visual processing.

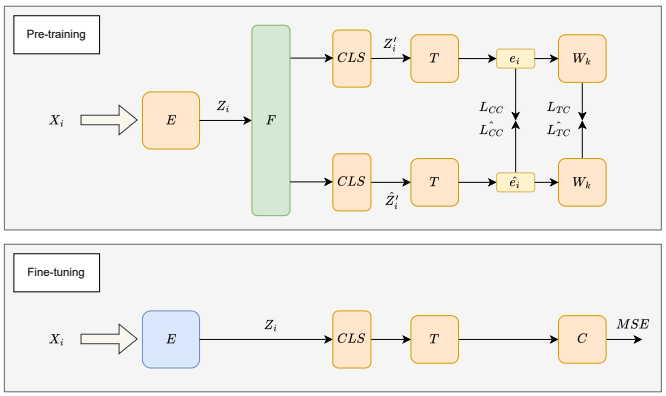

Fig. 1. The figure below depicts the fMRI signal prediction deep learning pipeline with pre-training and fine-tuning phases. Orange boxes denote trainable modules, while blue indicating frozen ones. Yellow represents vector outputs, and green denotes simple operations functions. During pre-training, $X_{i_{dl}}$ is passed though E, followed by F, where masking and cropping are applied along with addition of the CLS token at the beginning of the resulting vectors. Subsequently, T is applied, producing e_i, $\hat{e_i}$, both projected using W_k. Then optimization is performed for all objectives. In the fine-tuning phase, $X_{i_{dl}}$ passes through the frozen module E, and the CLS token is added. T is then applied with a linear layer C on top. Finally, optimization is executed using the mean square error (MSE).

IV. Method

In the classical machine learning pipeline, regularized regression algorithms are applied on the flatten vectors $X_{i_{ml}} \in \mathbb{R}^{4560}$, $X_{i_{ml}} \in \mathbb{R}^{4712}$, for the first and second dataset, respectively. In the deep learning pipeline we keep the time-step dimension and work on input vectors $X_{i_{dl}} \in \mathbb{R}^{19\times240}$, $X_{i_{dl}} \in \mathbb{R}^{19\times248}$, for the first and second dataset, respectively.

The deep learning pipeline is composed of two main modules as in [16]: E, a convolution-based encoder, and T, a transformer. We train both networks in a two-step process similar to [7], first pre-training and then fine-tuning. See Fig. 1 for illustration. All models in both pipelines predicts the fMRI activations from EEG signals.

Pre-training This phase consists of two tasks: predicting future time-steps and matching masked and unmasked latent features. The CNN encoder E, processes $X_{i_{dl}}$, producing the latent vector $Z_i = E(X_{i_{dl}})$, where $Z_i = [z_i^1, ..., z_i^T]$. Then, within function F, the following steps are executed: (1) Applying Masking of length M to a copy of Z_i, initiating from each time-step with a random probability of P, resulting in the vector $\hat{Z}_i$. (2) Sampling $t \sim \{1, ..., T-K\}$ where K denotes the number of future time-steps to predict in the contrastive loss. (3) Cropping Z_i, $\hat{Z}_i$ at index t. Lastly, a trainable CLS token is added at the beginning of the vectors like in [9], [20]. In total, $CLS(F(Z_i)) = [Z'_i, \hat{Z}'_i] =$

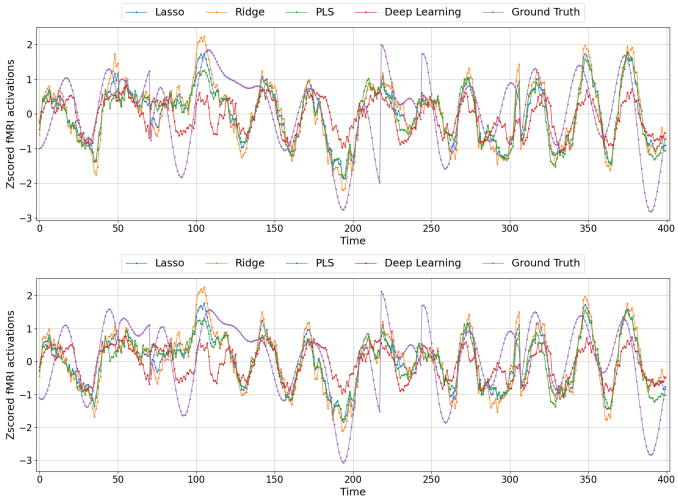

Fig. 2. Comparison of the different methods on subject 1 in the left and right visual region, respectively.

$[[CLS, z_i^1, ..., z_i^t], [CLS, \hat{z}_i^1, ..., \hat{z}_i^t]]$. Then, T is applied on both vectors: $T(Z_i'), T(\hat{Z}_i) = e_i, \hat{e_i}$. where $e_i, \hat{e_i} \in \mathbb{R}^d$ are the output embeddings of the CLS token, and d is the dimension of T. Our objectives are:

$$L_{TC} = -\frac{1}{K}\sum_{k=1}^{K} log\frac{exp((W_k(e_i))^T z_i^{t+k})}{\sum_{n\in N_{t,k}} exp((W_k(e_i))^T z_i^n)} \quad (1)$$

$$\hat{L}_{TC} = -\frac{1}{K}\sum_{k=1}^{K} log\frac{exp((W_k(\hat{e_i}))^T z_i^{t+k})}{\sum_{n\in N_{t,k}} exp((W_k(\hat{e_i}))^T z_i^n)} \quad (2)$$

$$L_{CC} = -\frac{1}{N}\sum_{i=1}^{N} log\frac{exp(sim(e_i, \hat{e_i}))}{\sum_{j\neq i} exp(sim(e_i, \hat{e_i}))} \quad (3)$$

$$\hat{L}_{CC} = -\frac{1}{N}\sum_{i=1}^{N} log\frac{exp(sim(e_i, \hat{e_i}))}{\sum_{j\neq i} exp(sim(\hat{e_i}, e_i))} \quad (4)$$

where $sim(a, b) = a^T b/||a||||b||$ is the cosine similarity metric, and N is the size of the batch.

To calculate L_{TC}, $\hat{L_{TC}}$, we use additional K linear layers (W_k). The pre-training objective is the sum $L_{pre-train} = \lambda_1(L_{TC} + \hat{L}_{TC}) + \lambda_2(L_{CC} + \hat{L}_{CC})$ where λ_1, λ_2 are hyper-parameters.

Fine-tuning During this phase, the layers of E are frozen, and the output of T is fed to an additional linear layer C. The fine-tuning objective is $L_{\text{finetune}} = MSE(C(T(CLS(E(X_i)))), y_i)$, where the function $CLS(Z_i)$ adds the CLS token at the beginning. We use another linear layer C as a final step, and then optimize using mean square error (MSE) as a loss function.

Architecture The network E has three blocks, each one of them is composed of a convolution layer, a dropout layer, a Group Normalization [21], and a GELU [22] activation function, like in [16]. The convolution layers channel size is 256, kernel sizes are (3, 5, 7) padding (1, 2, 3) and stride of 1 for all convolution layers. T has a dimension of $d = 512$, 2 layers, 2 heads. Since we apply pre-processing on the EEG data, our networks E, and T are much lighter than [16]. E does not down sample the time dimension, and has smaller number of convolution blocks and features. T has lower dimension and smaller number of heads and transformer layers.

TABLE I

THE PEARSON CORRELATION SCORES AND P-VALUES BETWEEN THE MODELS' OUTPUTS AND THE MEAN FMRI SIGNAL VALUES IN VARIOUS BRAIN REGIONS FOR THE FIRST DATASET.

Brain region	Left brain Pearson correlation			
	Ridge	PLS	Lasso	DL pipeline
Hipp.	0.025 ± 0.02	0.024 ± 0.03	0.019 ± 0.02	-0.010 ± 0.02
Pallidum	0.008 ± 0.03	0.038 ± 0.04	0.070 ± 0.05	0.076 ± 0.03
Precuneus	0.035 ± 0.04	0.036 ± 0.04	0.043 ± 0.05	0.006 ± 0.03
I-parietal	0.061 ± 0.04	0.088 ± 0.04	0.073 ± 0.04	0.108 ± 0.06
Visual	0.297 ± 0.05	0.333 ± 0.06	0.325 ± 0.05	0.353 ± 0.06

Brain region	Right brain Pearson correlation			
	Ridge	PLS	Lasso	DL pipeline
Hipp.	0.025 ± 0.03	0.030 ± 0.03	0.019 ± 0.02	0.004 ± 0.03
Pallidum	0.018 ± 0.04	0.029 ± 0.03	0.067 ± 0.05	0.087 ± 0.03
Precuneus	0.064 ± 0.04	0.067 ± 0.03	0.072 ± 0.03	0.079 ± 0.04
I-parietal	0.039 ± 0.02	0.077 ± 0.03	0.091 ± 0.04	0.123 ± 0.03
Visual	0.290 ± 0.05	0.325 ± 0.05	0.319 ± 0.05	0.342 ± 0.06

Brain region	Left brain p-values			
	Ridge	PLS	Lasso	DL pipeline
Hipp.	0.250 ± 0.36	0.107 ± 0.13	0.270 ± 0.31	0.293 ± 0.32
Pallidum	0.245 ± 0.37	0.187 ± 0.34	0.112 ± 0.29	0.015 ± 0.04
Precuneus	0.082 ± 0.23	0.048 ± 0.12	0.065 ± 0.18	0.181 ± 0.19
I-parietal	0.054 ± 0.09	0.000 ± 0.00	0.019 ± 0.04	0.013 ± 0.04
Visual	0.000 ± 0.00	0.000 ± 0.00	0.000 ± 0.00	0.000 ± 0.00

Brain region	Right brain p-values			
	Ridge	PLS	Lasso	DL pipeline
Hipp.	0.159 ± 0.23	0.233 ± 0.32	0.269 ± 0.29	0.262 ± 0.38
Pallidum	0.282 ± 0.29	0.374 ± 0.33	0.125 ± 0.26	0.001 ± 0.00
Precuneus	0.051 ± 0.15	0.040 ± 0.10	0.018 ± 0.05	0.000 ± 0.00
I-parietal	0.098 ± 0.22	0.001 ± 0.00	0.002 ± 0.00	0.000 ± 0.00
Visual	0.000 ± 0.00	0.000 ± 0.00	0.000 ± 0.00	0.000 ± 0.00

TABLE II

THE PEARSON CORRELATION SCORES BETWEEN THE MODELS' OUTPUTS AND THE MEAN FMRI SIGNAL VALUES FOR THE SECOND DATASET.

Brain region	Ridge	PLS	Lasso	DL pipeline
Hipp. left	0.117 ± 0.02	0.153 ± 0.03	0.133 ± 0.03	0.155 ± 0.02
Hipp. right	0.103 ± 0.02	0.134 ± 0.03	0.117 ± 0.02	0.135 ± 0.02

V. EXPERIMENTS

To evaluate the effectiveness of the deep learning pipeline and compare it against other baselines, we utilize a repeated 10-fold cross-validation, allocating 20% of the data for testing in each fold. The data is stratified by patients, ensuring that all data related to a specific patient is either in the train or test set. To optimize hyperparameters within each fold, we conduct an inner 5-fold cross-validation. We measure the Pearson correlation score between the fMRI signals and the different model's prediction in all experiments.

In the machine learning pipeline, we perform a grid search to find best hyperparameters. In Lasso (Regression) and Ridge

TABLE III
PEARSON CORRELATION SCORES FROM THE ABLATION STUDY ON THE FIRST DATASET, ASSESSING DIFFERENT ALTERNATIVE METHODS.

Brain region	Left brain		
	Fully Supervised	Unfreeze CNN	DL pipeline
Hipp.	-0.031 ± 0.040	-0.010 ± 0.018	-0.010 ± 0.019
Pallidum	0.060 ± 0.031	0.078 ± 0.040	0.076 ± 0.030
Precuneus	0.019 ± 0.022	0.008 ± 0.032	0.006 ± 0.030
I-parietal	0.093 ± 0.034	0.103 ± 0.056	0.108 ± 0.056
Visual	0.318 ± 0.043	0.367 ± 0.060	0.353 ± 0.064

Brain region	Right brain		
	Fully Supervised	Unfreeze CNN	DL pipeline
Hipp.	0.000 ± 0.032	0.012 ± 0.024	0.004 ± 0.030
Pallidum	0.038 ± 0.018	0.082 ± 0.038	0.087 ± 0.034
Precuneus	0.052 ± 0.032	0.085 ± 0.033	0.079 ± 0.040
I-parietal	0.115 ± 0.033	0.119 ± 0.037	0.123 ± 0.031
Visual	0.315 ± 0.040	0.353 ± 0.058	0.342 ± 0.063

(Regression), with alpha values in (0.001, 0.01, 0.02), and in PLS with number of components varies in (5, 7, 10). In all three methods we set a maximum of 2000 iterations. In the deep learning pipeline, we choose the number of epochs for both pre-training and fine-tuning phases, both with maximum of twelve epochs, based on the best validation score.

A comparison between the predictions of different methods and the ground truth for the first dataset is shown in Fig. 2.

Implementation details for the deep learning pipeline In the pre-training phase, we choose a masking probability of $P = 0.1$, a masking length of $M = 2$, and a future time-step prediction length of $K = 5$ due to short time-step length compare to [7]. Additional parameters include a learning rate of $4e^{-5}$, and $\lambda_1 = 1$, $\lambda_2 = 0.7$ as in [7]. In the fine-tuning phase, we use a learning rate of e^{-5}. In both phases we use a batch size of 128, Adam optimizer [23] with a weight decay of $3e^{-4}$, $\beta_1 = 0.9$, and $\beta_2 = 0.99$.

Results As can be seen in Tab. I, the results for the Hippocampus in the first dataset, across all methods are poor, with a very low correlation and a standard deviation of the same order of magnitude. This outcome was expected, due to noise activation produced from the visual regions. In the Pallidum, Precuneus, and Inferiorparietal, the results are slightly better. In these cases, the deep learning pipeline outperforms the classical methods in all cases, except for the left Precuneus. In the Visual region, as expected, a much better correlation is observed in all methods. In this case, the deep learning pipeline outperforms others on both sides.

In the second dataset, the results for the Hippocampus are better for both pipelines compared to the first dataset. Additionally, all P-values are 0.

Analysis (1) Predicting fMRI activations from EEG in the visual region is easier than other regions during visual experiment. In the first dataset, the p-values of all methods on both visual sides are 0, and the correlation values are the highest.

(2) Predicting fMRI activations from EEG in deeper brain regions is harder. In the first dataset, the p-values of most other regions, which are deeper within the brain than the Visual region, vary between low values in the Pallidum, Precuneus, and Inferiorparietal, to high values in the Hippocampus. However, during experiments with a musical feedback-based task in the second dataset, where noise from visual regions is reduced compared to the first dataset, the p-value decreases to 0 in the Hippocampus. This indicates greater reliability in predictions under these conditions. Nonetheless, the correlation values remain lower than those of the visual regions in the first dataset, further supporting our claim.

(3) The predictability in different brain regions is affected by the task paradigm. In the first dataset, the p-values are higher in the Hippocampus, comparing the second dataset, where experimenting with a musical feedback-based task, are 0.

(4) Where predictability is low – meaning, with low Pearson correlation values – the results of all methods are poor, and there is no advantage to the deep learning method. In the first dataset, the p-values of all methods in the Hippocampus, where predictability is low, are higher than 0.1.

(5) Where predictability is high, i.e., have high Pearson correlation values, the deep learning pipeline is more reliable than other methods. In regions where the correlation values are above 0.1 for the deep learning pipeline – such as in the Inferiorparietal and Visual on both sides in the first dataset – the p-values are same (0) or better than other methods in those regions. The same holds for the second dataset, where the Hippocampus p-value is 0.

Ablation study We compare the deep learning pipeline with partial variants, reported in Tab. III. In the "Fully Supervised" variant, we trained our networks without pre-training, and during the finetuning, both networks E and T kept unfrozen. In the "Unfreeze CNN" variant, we perform pre-training as usual. However, during fine-tuning, we keep network E unfrozen.

As can be seen, the Fully Supervised variant underperforms in comparison to the other variants. The results of the "Unfreeze CNN" variant are comparable to the adopted deep learning pipeline, with slightly improved results in some of the regions. This last result is surprising in light of previous work [7], [16] but the differences are not large.

VI. CONCLUSIONS

Employing EEG in place of fMRI can open the door for both diagnostic and therapeutic applications. Our experiments show that for the dataset at hand, which is not very large but is as large as most fMRI-based studies, and is also heavily tilted toward visual stimuli, one can reliably predict the activations in multiple brain regions but not in the Hippocampus.

It is evident from the results that deep learning methods are favorable over classical ones, even in the regime of limited-sized datasets. As the ablation study shows, this is enabled by the use of self-supervised learning. Other recent advancements in deep learning can improve results further. For example, in addition to the efforts reported here, we have been experimenting with using diffusion models for this prediction task [24], [25], still without significant success.

ACKNOWLEDGMENTS

This work has was supported by the Israel Science Foundation (Grant No. 2923/20) within the Israel Precision Medicine Partnership program, and a grant from the Tel Aviv University Center for AI and Data Science (TAD).

REFERENCES

[1] G. H. Glover, "Overview of functional magnetic resonance imaging," *HHS Author Manuscripts*, 2011.

[2] R. Bradley and Buchbinder, "Functional magnetic resonance imaging," *ScienceDirect, Handbook of Clinical Neurology*, 2016.

[3] B. Andrea, B. Franceschiello, and M. M. M. 3, "Electroencephalography," *Curr Biol*, 2019.

[4] Y. Meir-Hasson, S. Kinreich, I. Podlipsky, T. Hendler, and N. Intrator, "An eeg finger-print of fmri deep regional activation," *Neuroimage*, vol. 102, pp. 128–141, 2014.

[5] Y. Meir-Hasson, J. N. Keynan, S. Kinreich, G. Jackont, A. Cohen, I. Podlipsky-Klovatch, T. Hendler, and N. Intrator, "One-class fmri-inspired eeg model for self-regulation training," *PLoS One*, vol. 11, no. 5, p. e0154968, 2016.

[6] N. Singer, G. Poker, N. Dunsky-Moran, S. Nemni, S. R. Balter, M. Doron, T. Baker, A. Dagher, R. J. Zatorre, and T. Hendler, "Development and validation of an fmri-informed eeg model of reward-related ventral striatum activation," *NeuroImage*, vol. 276, p. 120183, 2023.

[7] N. Lalzary and L. Wolf, "Dual contrastive learning for self-supervised ecg mapping to emotions and glucose levels," in *2023 IEEE SENSORS*. IEEE, 2023, pp. 1–4.

[8] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[9] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[10] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever *et al.*, "Improving language understanding by generative pre-training," 2018.

[11] S. Schneider, A. Baevski, R. Collobert, and M. Auli, "wav2vec: Unsupervised pre-training for speech recognition," *arXiv preprint arXiv:1904.05862*, 2019.

[12] A. Baevski, Y. Zhou, A. Mohamed, and M. Auli, "wav2vec 2.0: A framework for self-supervised learning of speech representations," *Advances in neural information processing systems*, vol. 33, pp. 12 449–12 460, 2020.

[13] S. Gidaris, P. Singh, and N. Komodakis, "Unsupervised representation learning by predicting image rotations," *arXiv preprint arXiv:1803.07728*, 2018.

[14] S. Atito, M. Awais, and J. Kittler, "Sit: Self-supervised vision transformer," *arXiv preprint arXiv:2104.03602*, 2021.

[15] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International conference on machine learning*. PMLR, 2020, pp. 1597–1607.

[16] D. Kostas, S. Aroca-Ouellette, and F. Rudzicz, "Bendr: using transformers and a contrastive self-supervised learning task to learn from massive amounts of eeg data," *Frontiers in Human Neuroscience*, vol. 15, p. 653659, 2021.

[17] R. K. Niazy, C. F. Beckmann, G. D. Iannetti, J. M. Brady, and S. M. Smith, "Removal of fmri environment artifacts from eeg data using optimal basis sets," *Neuroimage*, vol. 28, no. 3, pp. 720–737, 2005.

[18] Admon, R and Vaisvaser, S, N. Erlich, T. Lin, I. Shapira-Lichter, E. Fruchter, T. Gazit, and T. Hendler, "The role of the amygdala in enhanced remembrance of negative episodes and acquired negativity of related neutral cues," *Biological Psychology*, vol. 139, pp. 17–24, 2018.

[19] N. Singer, P. Gilad, D.-M. Netta, N. Shlomi, R. Shira, D. Balter, Maayan, B. Travis, A. Dagher, Z. Robert, J, and H. Talma, "Development and validation of an fmri-informed eeg model of reward-related ventral striatum activation," 2022.

[20] D. Alexey, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv: 2010.11929*, 2020.

[21] Y. Wu and K. He, "Group normalization," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.

[22] D. Hendrycks and K. Gimpel, "Gaussian error linear units (gelus)," *arXiv preprint arXiv:1606.08415*, 2016.

[23] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[24] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics," in *Proceedings of the 32nd International Conference on Machine Learning*, 2015, pp. 2256–2265.

[25] Y. Li, X. Lu, Y. Wang, and D. Dou, "Generative time series forecasting with diffusion, denoise, and disentanglement," in *Advances in Neural Information Processing Systems*, A. H. Oh, A. Agarwal, D. Belgrave, and K. Cho, Eds., 2022. [Online]. Available: https://openreview.net/forum?id=rG0jm74xtx