ECAI 2024
U. Endriss et al. (Eds.)

doi:10.3233/FAIA241030

EEG-Based fMRI Digital Twin: Towards a Cheap and Ecological Approach to Measure Subcortical Brain Activity

Nikolay Dagaev[b], Ilia Semenkov[a,b] and Alexei Ossadtchi[a,b,c,*]

[a]Artificial Intelligence Research Institute (AIRI), Moscow, Russia
[b]HSE University, Moscow, Russia
[c]LLC "Life Improvement by Future Technologies Center" (LIFT), Moscow, Russia
ORCID (Nikolay Dagaev): https://orcid.org/0000-0002-1627-5231, ORCID (Ilia Semenkov): https://orcid.org/0000-0003-1515-7062, ORCID (Alexei Ossadtchi): https://orcid.org/0000-0001-8827-9429

Abstract. Electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) are the two most commonly used non-invasive methods for studying brain function, having different but complementary strengths: high temporal resolution of the former and high spatial resolution of the latter. Crucially, fMRI is vital for studying subcortical areas, as those are practically out of reach for EEG. At the same time, EEG is cost-effective and, thus, preferable to fMRI if comparable information can be extracted. Here we present an EEG-to-fMRI neural network with an interpretable module for feature extraction. Using the EEG-fMRI dataset, we show that our model allows us to predict the detailed resting state Blood Oxygenation Level Dependent (BOLD) activity of seven bilaterally symmetric subcortical structures solely from multichannel EEG data. Preliminary results reported here show a performance level significantly above chance and exceeding the state-of-the-art accuracy typically reported for a single structure such as the amygdala or striatum. These findings pave the road toward the creation of low-cost mobile scanners of subcortical activity with improved usability, EEG-based fMRI digital twin technology, with a broad range of applications – from fundamental neuroscience through diagnostics to neurorehabilitation and affective neurointerfaces. The demo video is presented in https://youtu.be/IOOwb7Wt2sY.

1 Introduction

Electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) are the most popular and successfully used non-invasive methods for measuring brain activity. Both have their strengths and weaknesses. For example, EEG's interpretability is limited as it records mixed activity of spatially extended populations of neurons. At the same time, the EEG achieves millisecond-scale temporal resolution and opens a window for exploring rapid cortical processes. On the other hand, fMRI measuring Blood Oxygenation Level Dependent (BOLD) signal furnishes much higher spatial resolution as compared to EEG but is limited by the pace of the underlying hemodynamics with characteristic response times on the order of a second when imaging the entire brain volume. These modalities are also very different from both the researcher's and subject's perspectives. EEG devices are cheap, compact, ecological, and affordable even for everyday in-home use while fMRI scanners are bulky, expensive, and require a human to stay fixed in a horizontal plane for extended periods to ensure the recordings are of reasonable quality. Yet, despite these very clear limitations, the fMRI technology is nowadays unique in its ability to reliably capture the activity of deep cortical (e.g. hippocampus) and subcortical structures (e.g. basal ganglia) in a non-invasive manner. Limited access to the fMRI explains the fact that human subcortical structures remain the least studied brain territories both structurally and functionally [17]. At the same time, it is hard to overestimate their role in cognitive processes including memory, attention, and reward mechanisms [4], motor functions [2], and affective cognition [12]. Also, the subcortical structures increasingly become the main suspects whose deficiencies in structure [6] and function [15] are implicated in a broad range of neuropsychiatric disorders. Therefore, there is a pressing demand for ecological and affordable tools for functional visualization of the workings of subcortical brain regions such as basal ganglia, cerebellum, and thalamus.

One possible approach is to use EEG and augment it with a machine learning model capable of predicting subcortical BOLD signals from the EEG data alone. The majority of studies attempted to establish the similarity of EEG and fMRI based findings about the underlying brain activity by directly correlating the two measures. To align the EEG and BOLD signals a convolutional transformation is applied to EEG-based source power profiles, e.g. [9], [7]. An interesting heuristic was described in [5] proposing a non-linear model stating that the increase in BOLD signal is correlated with the broadening (expansion) of the EEG signal spectrum. Most such studies were focused on the cortical BOLD activity [1] and the recovery of subcortical BOLD (sBOLD) signal from EEG data remains elusive, see, however [16] where the striatum BOLD activity is predicted from EEG data in a task-based setting.

To fill this technological gap we are presenting a novel AI-powered EEG-based fMRI digital twin technology capable of simultaneous recovery of BOLD signals of several subcortical structures solely based on the ecologically recorded EEG data. We capitalize on **the recent advances in deep contrastive learning and build individualized models capable of translating non-invasively recorded EEG into the BOLD signal of functionally distinct cortical and subcortical areas.** Using concurrently recorded EEG-fMRI data we train a deep neural network with convolutional and transformer layers on top of **the interpretable feature extractor layers [10, 11]**

* Corresponding author, email: ossadtchi@gmail.com

to capture the intricate hidden relations between head surface EEG and subcortical BOLD signals measured with fMRI. Once the model is built **it is used to predict regional BOLD activity from EEG alone** therefore bypassing the use of the fMRI equipment when operating in the inference mode. Figure 1 outlines the principles of the developed technology. To the best of our knowledge, **this is the first-ever demonstration of successful and simultaneous recovery of the BOLD signals reflecting the hemodynamics of multiple subcortical nuclei and the hippocampus**. Simultaneous recovery of BOLD activity of several brain regions allows us **to avoid the potentially bogus results explained by the common-mode signal**. In addition to removing the global signal during BOLD preprocessing, we focus on the variations of the regional BOLD signals with respect to their mean activity computed over 14 selected regions.

Additionally, **the interpretability of the front-end layers of our network** allows for the discovery of key EEG correlates pivotal for solving the EEG to BOLD decoding task and answering the questions regarding the basic principles underlying the fundamental neurophysiological mechanisms that link electrical activity of the brain to the associated hemodynamics and match the obtained answers against the existing theories.

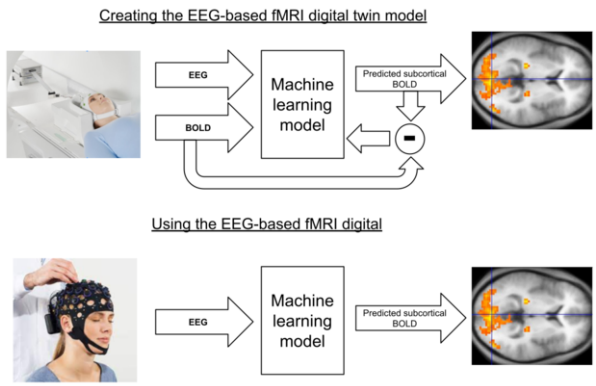

Figure 1. EEG-based fMRI digital twin principle. Concurrently recorded EEG-fMRI data are used to train a subject-specific ML model to predict regional BOLD signals from the multichannel EEG data. Once the model is built, EEG alone passed through the model predicts brain hemodynamics of both cortical and subcortical regions.

2 Dataset

In this demonstration we used a publicly available EEG/fMRI dataset [18]: multichannel EEG and fMRI-measured BOLD signals were concurrently recorded. To reduce the amount of gradient artifacts in the EEG data, the interference was picked up via Carbon Wire Loop (CWL) and adaptively subtracted from the EEG signals. Resting-state data was collected from 8 subjects, each recording lasting approximately 4.5 minutes. The EEG was recorded using a 30-channel MR-compatible electrode cap with a native sampling frequency of 5000 Hz during the fMRI scan. Signals were corrected for gradient artifacts and then downsampled to the sampling frequency of 1000 Hz by the authors of this dataset. The volumes of BOLD signal with dimensions [61 × 72 × 61] and resolution of 3×3×3 mm were acquired at TR 2000 ms, please refer for the detailed description to [18]. The obtained data was normalized to the MNI space, slice-timing corrected, and smoothed with a spatial Gaussian filter (5-mm kernel). Confound regression with respect to the motion and global signals including those from the white matter as regressors was used to remove from BOLD activity the signals of non-neuronal origin. Subsequently, we extracted BOLD activity of bilaterally symmetric regions of interest (ROI) using Harvard-Oxford structural atlas [8]. We focused on 14 ROIs - 12 subcortical regions and 2 hippocampi, see Figure 3. We focus on the CWL dataset because, to the best of our knowledge, it is the only public dataset using the advanced CWL technology to clean the EEG data from the gradient artifacts and containing the data recorded with the helium pump switched off. Additionally, to avoid the influence of the common mode signal on the obtained performance we computed the average region of interest (ROI) activity BOLD signal and subtracted it from the regional ROI BOLD time series. Our goal was then to recover this average (common mode) signal and the differential fluctuations around it characterizing the specific activation of the individual ROIs.

3 Methods

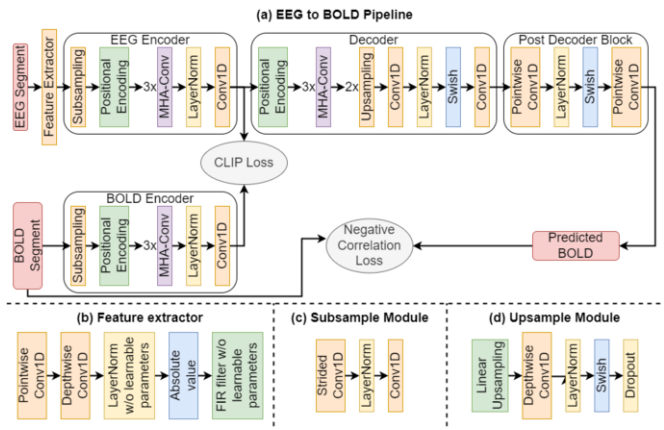

Figure 2. Deep learning model for the EEG to BOLD prediction task. The complete architecture is presented in (a). Custom modules are shown in detail in their own boxes (b), (c), (d). Note: during inference we switch off the BOLD Encoder block to prevent data leak. It is used only for training to reduce the domain gap between EEG and BOLD embeddings.

The neural network architecture diagram is presented in Figure 2 (a). It consists of four major components: an interpretable Feature Extractor, an EEG Encoder, a Decoder, and a Post Decoder Block. We also utilize a BOLD Encoder during training for contrastive learning.

The Feature Extractor shown in Figure 2 (b) is based on the factorized and trainable spatial and temporal filters [10]. It applies spatial and then temporal filtering to a raw EEG segment, with filters implemented via pointwise and depthwise one-dimensional convolutions, respectively. The filtered signal is transformed to absolute values and then convolved with Hamming windows for temporal smoothing. Overall, it extracts the envelopes of activity of the specific and relevant to the decoding task neuronal populations. Subsequent interpretation of the Feature Extractor weights allows for a query into the geometric location and dynamics of the neuronal populations relevant to the specific decoding task [11].

A series of resultant features is further fed to the EEG Encoder, which maps it to a high-dimensional embedding space. The stem of an Encoder is a stack of multiple hybrid blocks, each combining multi-head self-attention and convolutional layers (MHA-Conv blocks similar to the Conformer blocks [3]). This core subpart is preceded by a small block that reduces sampling rate via a single-strided

convolution and applies a 1D convolution in order to match the dimension of the embedding space (see Figure 2 (c)).

The stem of the Decoder consists of the same building blocks as the EEG Encoder. In order to match the shapes of the predicted and target BOLD signals, the stem is followed by a small block of temporal upsampling and spatial reduction, achieved via linear interpolation followed by convolutional layers (see Figure 2 (d)).

The Post Decoder Block aims to smooth Decoder's output signal and/or to allow some additional flexibility for temporal structure. In our experiments, we used two point-wise convolutions with a layer normalization and non-linearity in between as shown in Figure 2 (a). A multichannel output of this module represents the predicted BOLD activity of 14 + 1 subcortical ROIs.

In addition to the primary architectural components described above and designed specifically for the EEG to BOLD prediction task we introduced the BOLD Encoder. It mapped raw interpolated time series of a BOLD signal into the same embedding space as the EEG Encoder does for the EEG segments. The goal of having this extra component is to turn the embedding space into a multimodal one and to facilitate learning of more informative as well as generalizable embeddings for the EEG. Except for the dimensionality of an input we used the BOLD Encoder of the same architecture as the EEG Encoder which is shown in Figure 2 (a). During the actual inference, the BOLD Encoder is not used since the BOLD signal is unknown and has to be predicted.

To train the model we utilized the loss function that consists of multiple components. The first component is dedicated to measuring the quality of predicting BOLD signal itself. This component is a negative correlation value between a true BOLD signal and a BOLD signal predicted from EEG by the model (output of the Post Decoder Block). The second component we used is the CLIP contrastive loss [14] aimed at reducing the modality gap by bringing embeddings of EEG and BOLD closer. To compute it we take an output of the EEG Encoder and of the BOLD Encoder. Then, we calculate the CLIP loss between these embeddings. Both of these components and their usage are shown in Figure 2 (a). Our final loss function is a weighted sum of these two components: CLIP and negative correlation.

We trained our model using ≈34 minutes of data in total from 8 subjects. 7 subjects were only used for training. The final subject's data was split the following way: 68.7 initial seconds for training; the next 60 seconds for validation; and the final 150 seconds for testing.

4 Results

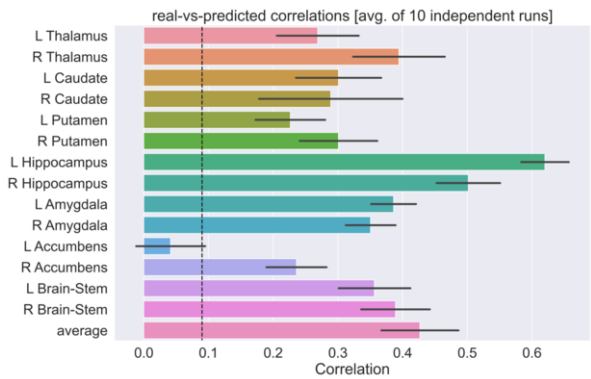

Figure 3. The obtained correlation coefficient values along with the standard deviation for 14 deep brain ROIs and the average common mode signal.

As a meaningful measure of accuracy, we used the Pearson correlation coefficient between the actual and the EEG-derived BOLD activity. The correlation coefficients and the standard deviations for 14 deep brain ROIs and the average common mode signal are shown in Figure 3. We can observe that the hemodynamic activity of most deep brain regions can be recovered from EEG with accuracy that significantly exceeds the chance level marked with the vertical dashed line. We calculated the chance level as the performance of the surrogate model with the same architecture but trained using temporally unrelated EEG and fMRI segments. Such pairs used from testing and training comprised two different non-overlapping sets. Importantly, despite the fact that we focused on the more difficult task of predicting relative BOLD activity, for most of the subcortical structures we observed significantly higher decoding accuracy values as compared to those reported in the state-of-the-art study [16] for the striatum within a task-based setting.

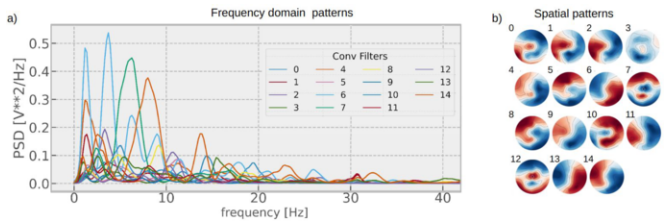

Figure 4. a) Temporal (shown in frequency domain) and b) spatial patterns derived from feature extractor's weights according to [10] and corresponding to the sources that appeared pivotal to the decoding task.

In Figure 4 we show the results of interpretation of the feature extractor's weights for each of the 15 branches. The analysis of the frequency (a) and spatial color-coded (b) patterns of each network's branch reveals pairs of nearly identical up to the sign (blue vs. red) topographies (e.g. 0-10, 6-14, 7-12, etc.) with the corresponding distinct spectral curves having peaks over adjacent frequency bands. This can be interpreted as the attempt of the network to measure the "spectral expansion" in the activity of specific neuronal sources. This is in agreement with the spectrum broadening hypothesis as a correlate of the BOLD increase described in [5]. Judging by the topographies, the electrical activity of both superficial somatosensory [13] and deeply located sources appear pivotal to the subcortical BOLD decoding task which is in line with the known presence of a strong network of structural and functional links between several subcortical structures and cortical sensory-motor areas.

5 Conclusion

Our AI-powered EEG-based fMRI digital twin solution for the first time yields simultaneous and consistent recovery of hemodynamics in multiple subcortical structures based solely on the EEG data in the task-free resting state setting. Our model uses only 10 seconds of the most recent EEG time series data to estimate the corresponding BOLD activity (that is typically available delayed by 5-7 seconds later from the fMRI scanner). Therefore, the proposed EEG-based sBOLD signal recovery furnishes latency-free access to the hemodynamic activity of subcortical structures. Our results open up opportunities for ecologically valid and accessible exploration of deep brain activity. This will foster discoveries in the field of affective neuroscience and neurointerfaces. The presented technology will become instrumental in developing novel diagnostic solutions and unraveling mechanisms of a range of neuropsychiatric disorders.

Acknowledgements

The article was prepared within the framework of the Basic Research Program at HSE University.

References

[1] R. Abreu, A. Leal, and P. Figueiredo. Eeg-informed fmri: A review of data analysis methods. *Frontiers in Human Neuroscience*, 12, 2018. ISSN 1662-5161. doi: 10.3389/fnhum.2018.00029. URL https://www.frontiersin.org/articles/10.3389/fnhum.2018.00029.

[2] H. J. Groenewegen et al. The basal ganglia and motor control. *Neural plasticity*, 10(1-2):107–120, 2003.

[3] A. Gulati, J. Qin, C.-C. Chiu, N. Parmar, Y. Zhang, J. Yu, W. Han, S. Wang, Z. Zhang, Y. Wu, and R. Pang. Conformer: Convolution-augmented transformer for speech recognition. *arXiv:2005.08100*, 2020.

[4] K. Janacsek, T. M. Evans, M. Kiss, L. Shah, H. Blumenfeld, and M. T. Ullman. Subcortical cognition: The fruit below the rind. *Annual Review of Neuroscience*, 45:361–386, 2022.

[5] J. Kilner, J. Mattout, R. Henson, and K. Friston. Hemodynamic correlates of eeg: A heuristic. *NeuroImage*, 28(1):280–286, 2005. ISSN 1053-8119. doi: https://doi.org/10.1016/j.neuroimage.2005.06.008. URL https://www.sciencedirect.com/science/article/pii/S1053811905004167.

[6] D. Koshiyama, M. Fukunaga, N. Okada, F. Yamashita, H. Yamamori, Y. Yasuda, M. Fujimoto, K. Ohi, H. Fujino, Y. Watanabe, et al. Role of subcortical structures on cognitive and social function in schizophrenia. *Scientific reports*, 8(1):1183, 2018.

[7] H. Laufs, A. Kleinschmidt, A. Beyerle, E. Eger, A. Salek-Haddadi, C. Preibisch, and K. Krakow. Eeg-correlated fmri of human alpha activity. *Neuroimage*, 19(4):1463–1476, 2003.

[8] N. Makris, J. M. Goldstein, D. Kennedy, S. M. Hodge, V. S. Caviness, S. V. Faraone, M. T. Tsuang, and L. J. Seidman. Decreased volume of left and total anterior insular lobule in schizophrenia. *Schizophrenia research*, 83(2-3):155–171, 2006.

[9] E. Martınez-Montes, P. A. Valdés-Sosa, F. Miwakeichi, R. I. Goldman, and M. S. Cohen. Concurrent eeg/fmri analysis by multiway partial least squares. *NeuroImage*, 22(3):1023–1034, 2004.

[10] A. Petrosyan, M. Sinkin, M. Lebedev, and A. Ossadtchi. Decoding and interpreting cortical signals with a compact convolutional neural network. *Journal of Neural Engineering*, 18(2):026019, mar 2021. doi: 10.1088/1741-2552/abe20e. URL https://dx.doi.org/10.1088/1741-2552/abe20e.

[11] A. Petrosyan, A. Voskoboinikov, D. Sukhinin, A. Makarova, A. Skalnaya, N. Arkhipova, M. Sinkin, and A. Ossadtchi. Speech decoding from a small set of spatially segregated minimally invasive intracranial eeg electrodes with a compact and interpretable neural network. *Journal of Neural Engineering*, 19(6):066016, nov 2022. doi: 10.1088/1741-2552/aca1e1. URL https://dx.doi.org/10.1088/1741-2552/aca1e1.

[12] J. E. Pierce and J. Péron. The basal ganglia and the cerebellum in human emotion. *Social cognitive and affective neuroscience*, 15(5):599–613, 2020.

[13] D. Purves, G. J. Augustine, D. Fitzpatrick, L. C. Katz, A.-S. LaMantia, J. O. McNamara, and S. M. Williams, editors. *Neuroscience*. Sinauer Associates, Sunderland (MA), 2nd edition, 2001.

[14] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever. Learning transferable visual models from natural language supervision. *arXiv:2103.00020*, 2021.

[15] S. Salloway and J. Cummings. Subcortical structures and neuropsychiatric illness. *The Neuroscientist*, 2(1):66–75, 1996.

[16] N. Singer, G. Poker, N. Dunsky-Moran, S. Nemni, S. Reznik 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*, 276:120183, 2023. ISSN 1053-8119. doi: https://doi.org/10.1016/j.neuroimage.2023.120183. URL https://www.sciencedirect.com/science/article/pii/S1053811923003348.

[17] Y. Tian, D. S. Margulies, M. Breakspear, and A. Zalesky. Topographic organization of the human subcortex unveiled with functional connectivity gradients. *Nature neuroscience*, 23(11):1421–1432, 2020.

[18] J. van der Meer, A. Pampel, E. van Someren, J. Ramautar, Y. van der Werf, G. Gomez-Herrero, J. Lepsien, L. Hellrung, H. Hinrichs, H. Möller, and M. Walter. "eyes open – eyes closed" eeg/fmri data set including dedicated "carbon wire loop" motion detection channels. *Data in Brief*, 7:990–994, 2016. ISSN 2352-3409. doi: https://doi.org/10.1016/j.dib.2016.03.001. URL https://www.sciencedirect.com/science/article/pii/S2352340916301056.