Blind Deconvolution of Task Paradigm Timecourses from fMRI Data using Recurrent Neural Networks

Mengozzi Federico, Pucci Valentina, Silvagni Leonardo, *Neuro-X Section, EPFL*
CS-433, ML4Science, 19.12.2024
Project Supervisor: Flavia Petruso, MIP Lab

***Abstract*—This study aims to estimate brain activity timecourses during task paradigms using fMRI data from the Human Connectome Project (HCP) [1]. The hypotesis behind is that neuronal firing patterns correspond to voxel activations in fMRI scans. We frame brain activity estimation as a blind deconvolution problem. Machine learning techniques, specifically Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs), are used to model the temporal dynamics of brain activity.**

I. Introduction

The human brain is a complex network of interconnected regions responsible for various cognitive and motor tasks. Functional magnetic resonance imaging (fMRI) allows researchers to observe these activities by measuring changes in blood oxygenation levels (BOLD signals), which indirectly reflect neuronal firing patterns. The Hemodynamic Response Function (HRF), denoted as $h(t)$, models the physiological processes in the brain that occur in response to neural activation. This includes the delivery of oxygen-rich blood to active brain regions, which is essential for neuronal metabolism. The BOLD signal observed in fMRI can be mathematically expressed as a convolution of the HRF with the neural activity of a voxel $a(t)$, combined with model and measurement noise $\varepsilon(t)$:

$$\text{BOLD}(t) = (h * a)(t) + \varepsilon(t)$$

This formula encapsulates the underlying idea that neuronal firing patterns induce metabolic demands, leading to a vascular response that affects the BOLD signal in functional neuroimaging data.

The primary goal of this study is to estimate the deconvolved brain activity timecourses during specific task paradigms using machine learning techniques, specifically Recurrent Neural Networks (RNNs). These models are well-suited for capturing the dynamic nature of brain activity over time. Additionally, Convolutional Neural Networks (CNNs) are considered due to their ability to process spatially structured data, such as the BOLD signal. This approach adopts a blind deconvolution framework, aiming to estimate brain activity without prior knowledge of the HRF.

II. Data Analysis

A. Dataset Description

The Human Connectome Project (HCP) [1] provides a comprehensive and high-resolution dataset aimed at understanding the human brain's functional and structural connectivity.

In this study, we focus on task-based fMRI data from the HCP. During the fMRI sessions, 100 participants engaged in controlled cognitive tasks that activate various brain regions. These tasks span a broad range of cognitive functions, including working memory, language processing, motor function, emotion processing, gambling, and social cognition. Each task follows a specific experimental paradigm with different onset times, durations, and *conditions*. Here, a condition refers to a specific and independent subtask, characterized by its own onset time. An example for the MOTOR task paradigm is presented in Figure 1.

B. Task Selection

For trainig our model, we decided to focus on the `MOTOR` task. This task was chosen because it engages well-defined and distinct regions of the brain, such as the primary motor cortex, making it an ideal candidate for studying localized task-related activations. We then tested our model on the other tasks in order to see its generalization power. The `MOTOR` task has timeseries of 284 timepoints comprehensive of 20 sequential conditions (that will define 20 independent regressors).

C. Dataset Preprocessing

fMRI images are represented in 4D format, where each 3D voxel has a corresponding timecourse across the acquisition period. The dataset provided by MIPlab has already undergone essential preprocessing steps, including motion correction, slice timing correction, and registration to MNI space, a common reference space where subjects ae mapped to allow comparisons between voxels.

Since neuronal cell bodies are predominantly located in the brain's grey matter, we refined the dataset by selecting voxels that overlap with a grey matter mask. This mask, provided in the dataset, was further cleaned to remove noisy borders, ensuring that the analysis focuses on the most relevant brain regions.

To enhance the interpretation of the data, a spatial Gaussian smoothing kernel with a full width at half maximum (FWHM) of 10 mm was applied. This smoothing step reduces high-frequency noise, enhances the signal-to-noise ratio, and improves data reliability by averaging the fMRI signal across neighboring voxels. This also mitigates scat-

tered voxel activations while preserving significant neural activity patterns.

D. General Linear Model

The General Linear Model (GLM) is commonly used to identify voxel activations in fMRI images by quantifying how experimental conditions influence the observed fMRI signal, the BOLD signal. The beta coefficients are key to this process; they represent the relationship between the predictors (experimental conditions) and the target variable (the observed BOLD signal). These predictors are included in a design matrix ($X(t)$), each convolved with the HRF to account for the timing of the neural activation. The GLM posits a linear relationship between the observed BOLD signal and the predictors in the design matrix at the voxel level:

$$\text{BOLD}(t) = X(t)\beta + \epsilon(t) \tag{1}$$

where: $BOLD$ is the observed fMRI time-series signal at a voxel, X is the design matrix containing convolved predictors, β are the unknown coefficients (beta values) to be estimated, ϵ represents the residual error or noise.

The beta coefficients are estimated by minimizing the difference between the predicted signal ($X\beta$) and the observed signal (BOLD) using least-squares regression:

$$\hat{\beta} = (X^T X)^{-1} X^T \text{BOLD} \tag{2}$$

E. Estimation of the deconvolved signal

The justification for using beta coefficients from the GLM to achieve the study's objectives is based on the assumption that the brain functions as a linear time-invariant (LTI) system and that the convolution is a linear operation, meaning it does not alter the beta coefficients estimated from the General Linear Model (GLM). Therefore, the beta coefficients obtained from a convolved model (the design matrix contains convolved regressors) remain the same as those from a *deconvolved model*. The deconvolved model refers to a version without explicitly modeling the convolution with the HRF, as the exact shape of this signal is different for every participants. Given the LTI assumption (II-E), the process to estimate the deconvolved signal involves the following steps:

1) Estimate the beta coefficients by fitting the GLM (2)
2) Compute the deconvolved signal as a weighted sum of the non-convolved regressors, $r_i(t)$. Here, each regressor corresponds to an experimental condition modeled as a binary block for its onset time. The regressors are independent between each other.

$$Deconvolved\_signal(t) = \sum_{i=1}^{N} \beta_i \cdot r_i(t)$$

Here, N represents the total number of regressors. An example of BOLD and deconvolved signal after normalization can be seen in figure 2.

The model will receive as input a dataset consisting of time series of BOLD signals, with the corresponding deconvolved signals serving as the targets.

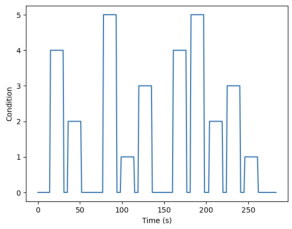

Figure 1: Illustration of a MOTOR task experiment. Conditions against onset time.

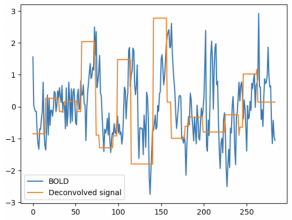

Figure 2: Example of normalized BOLD and deconvolved signal from the GLM of a MOTOR task.

F. Voxel Selection

To ensure the inclusion of meaningful brain activity while minimizing noise and artifacts, we applied a two-step voxel selection process. Given that true brain activation typically manifests as a region-wide phenomenon rather than isolated single-voxel activations, we retained only voxels belonging to clusters of at least 10 contiguous neighbors. This criterion helps exclude scattered noise or spurious activations that are biologically unlikely. The F-statistic maps, derived from the general linear model (GLM), measures the proportion of the BOLD signal variation that can be attributed to the task-specific predictors versus random noise. To control for multiple comparisons and reduce the risk of false positives, these maps were thresholded using a false discovery rate (FDR) of 0.05. This ensures that only regions with statistically significant activation are retained, focusing on the most task-relevant brain regions. From these maps, we retained the top 2% of active voxels (98th percentile), resulting in approximately 1,950 voxels per participant. An example of the refined F-statistic map, is presented in Figure 4.

III. Models and Methods

A. Dataset splitting

We considered 60 subjects from the HCP dataset, using stratified splitting to maintain class distribution: 60% for training, 20% for validation and 20% for testing.

B. Dataset Augmentation

We analyzed the correlation between timeseries of voxels from the same subject and observed substantial correlation in certain subjects. This high correlation can be partly attributed to the 10mm FWHM spatial smoothing applied to the fMRI data, which tends to blur the signal across neighboring voxels, causing them to exhibit similar activation patterns. In order to mitigate this problem and obtain a more generalizable model, the application of dataset augmentation strategies was a crucial aspect of our study. We applied several augmentation techniques to both the training and validation sets, targeting both the input data and their corresponding targets. Considering the assumed LTI properties of the *BOLD* process that generates the signal, we decided on opting for the following augmentation strategies:

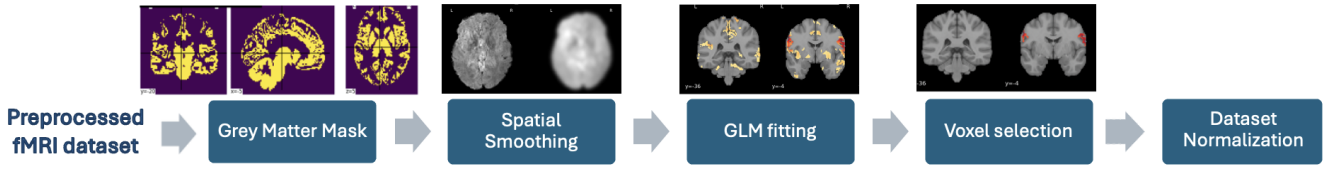

Figure 3: The preprocessing pipeline used in this study. The input is the original fMRI dataset from the HCP. The output consists of the BOLD timeseries, serving as the input for the model, and the deconvolved signal which is use as target.

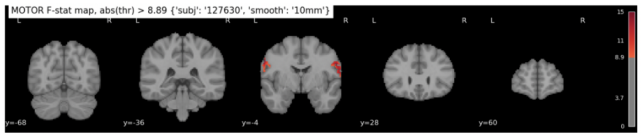

Figure 4: F-statistic map after voxel selection. The colormap on the right indicates the scale of activation, with red-colored voxels representing higher levels of activity. The regions highlighted correspond to the motor cortex of the brain.

1) Random Temporal Circular Shift: Each voxel's time-series was shifted by a random value between -100 and +100 time points. This created disalignment between the time series helps to prevent the model from memorizing sequential patterns across different voxels.

2) Random Amplitude Scaling: A random constant scaling factor between 0.5 and 2 was applied to each voxel's timeseries and corresponding target, changing its amplitude.

3) Random Gaussian Noise Addition: Random Gaussian noise with a standard deviation ranging from 5% to 25% of the maximum amplitude of each timeseries was added to it to enhance the diversity of the data.

4) Random Temporal Stretching: 60% of the time series of a subject were horizontally stretched by a random factor between 0.8 and 1.2 and subsequently resampled. This technique simulates variations in task duration and aids in generalization.

C. Model selection

The baseline of our model can be a simple LSTM layer or the Pure convolutional, which test results on MOTOR task are showed in Table I. Our final and most performant model 5 consists of a Long Short-Term Memory (LSTM) network with two layers, each having a hidden size of 64. This LSTM outputs a kernel of size 40, to be convolved with the original signal, feeding into another three-layer LSTM network, which estimates the block-like beta values of the input BOLD signal. We developed this model because the BOLD signal inherently requires preprocessing steps such as denoising and deconvolution with the hemodynamic response function (HRF) in order to reveal the underlying block-like structure of voxel activation, which our model is designed to estimate effectively. The intuition behind predicting a convolutional kernel is rooted in its ability to model a linear transformation through convolution. This process transforms the original signal into its deconvolved counterpart. Subsequent recurrent layers refine this deconvolved signal, progressively shaping it into a more block-like structure. We preferred LSTM over traditional RNNs or GRUs because of its superior ability to handle long-term dependencies, which are critical for modeling the BOLD signal's temporal structure. In the table I is shown the comparison between different models analyzed. The different types of loss are described in paragraph III-D

D. Loss Function

Early versions of our models seemed to prefer the prediction of smooth or constant signals, as shown in Figure 6. Particular care was therefore posed in choosing appropriate loss functions. In order to promote a block design, a variation [2] of the *Total Variation (TV)* was implemented. Furthermore, to penalize a constant prediction which is an optimum for the *TV* term, a loss term that penalizes non-unitary variance was introduced *(UV)*. These losses were added with a weight λ to the main data fidelity term *MAE*. Here is the formula for the final combined objective:

$$L_y(f(x)) = L_{\text{MAE}}(f(x), y) + \lambda(L_{\text{TV}}(f(x)) + L_{\text{UV}}(f(x))) \tag{3}$$

where:

$$L_{\text{MAE}}(f(x), y) = |f(x) - y|$$
$$L_{\text{TV}}(f(x)) = \sqrt{2\alpha\varepsilon}|f(x)_{n+1} - f(x)_n|e^{-\alpha(f(x)_{n+1} - f(x)_n)^2}$$
$$L_{\text{UV}}(f(x)) = (\sigma^2_{f(x)} - 1)^2 : y \text{ has unitary variance}$$
$$x = \text{Input of the model, } f(x) = \text{NN prediction}$$
$$y = \text{True deconvolved signal}$$

In the equation above, $\sigma^2_{f(x)}$ is the prediction's variance and $f(x)_{n+1} - f(x)_n$ is the first discrete derivative. The hyperparameters λ, α were pre-tuned on the expected loss of smooth and constant signals derived from our training set, in particular λ was chosen to be the inverse of the maximum expected loss. The variation of the *TV* is not a convex function, for this reason the term λ was introduced as a quadratic function of the training epoch, to reduce the risks of converging to a bad local minima.

E. Model Parameters

As an early stopping criterion, the model was stopped if the validation loss did not improve by at least 2% within the last 15 epochs. A learning rate scheduler with a patience of 3 was used, reducing the learning rate by a factor of 10 if no improvement was observed within the patience window. The ADAM optimizer was chosen for its robustness and ability to partially handle the non-convex nature of the

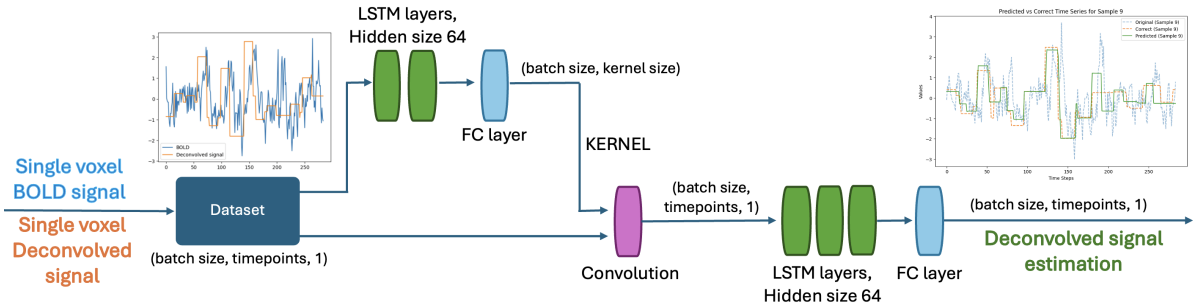

Figure 5: Diagram of the full modeling pipeline. The pipeline begins with BOLD timeseries and estimated deconvolved signal of the selected voxels. The model includes an initial two-layer LSTM generating a kernel to convolve with the original signal, and a final three-layer LSTM network to predict block-like beta values representing voxel activations.

optimization problem. A grid search was conducted for hyperparameters such as batch size, convolutional kernel size, and LSTM hidden size within a linear range of 5 to 150, while loss hyperparameters λ and α were searched in a log-space between 10^{-3} and 10^{3}. The final chosen hyperparameters are: *batch size*: 16, *kernel size*: 40, *hidden size*: 64, λ: 2.61, α: 8. In addition to these explicitly optimized hyperparameters, the model comprises 136,809 learnable parameters distributed across the LSTM and Fully Connected layers.

IV. Results

Our network was trained only on MOTOR tasks but tested on all the others II which have different onset times and experiment duration, the loss is higher but still a reasonable value. Additionally, it is important to highlight that visually the deconvolved signal estimated seems reasonable 8 on the tests on new tasks.

Model	MAE	Custom Loss
BiLSTM‡	0.2857 ± 0.0085	0.4423 ± 0.0106
ConvLSTM‡	0.6341 ± 0.0146	1.3610 ± 0.0333
LSTM_attention‡	0.6635 ± 0.0142	1.5037 ± 0.0192
Conv‡	0.4141 ± 0.0100	0.6661 ± 0.0285
BiLSTMConvLSTM†	0.3593 ± 0.0099	$\mathbf{0.0785 \pm 0.0156}$
BiLSTMConvLSTM‡	0.3037 ± 0.0091	0.4361 ± 0.0204
LSTM_1L‡	0.6772 ± 0.0142	1.4783 ± 0.0244
LSTM_3L‡	0.6657 ± 0.0160	1.4918 ± 0.0289
LSTMConvLSTM†	0.3511 ± 0.0100	$\mathbf{0.0790 \pm 0.0134}$
LSTMConvLSTM‡	$\mathbf{0.2771 \pm 0.0086}$	0.4097 ± 0.0085

Table I: Models test performance metrics on MOTOR task, training with: MAE‡, custom loss†

Task	MAE	Custom Loss
Relational	0.5019 ± 0.0106	0.1421 ± 0.0121
Emotion	0.4347 ± 0.0147	0.1339 ± 0.0205
Working memory	0.5284 ± 0.0123	0.0660 ± 0.0192
Language	0.4861 ± 0.0134	0.0915 ± 0.0105
Gambling	0.5520 ± 0.0133	0.2288 ± 0.0250

Table II: Model Performance Metrics, model LSTMConvLSTM trained on MOTOR tasks with Custom Loss and tested on the other tasks

V. Conclusions

Our final model, trained exclusively on MOTOR tasks, demonstrated ability to generalize to other tasks. The results

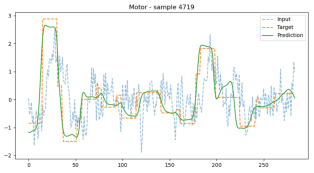

Figure 6: Estimation of the deconvolved signal of the LSTMConvLSTM model trained on MOTOR task with the MAE Loss, tested on a MOTOR task.

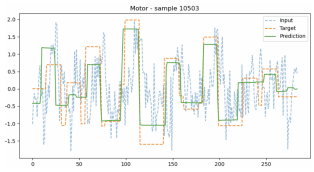

Figure 7: Estimation of the deconvolved signal of the LSTMConvLSTM model trained on MOTOR task with the Custom Loss, tested on a MOTOR task.

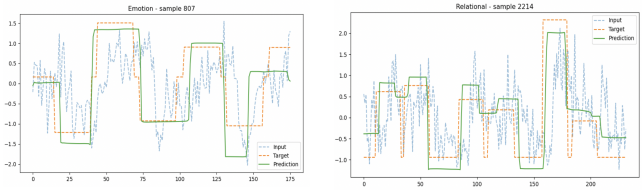

Figure 8: Estimation of deconvolved signal of the model LSTMConvLSTM trained on MOTOR tasks with the Custom Loss, tested on a EMOTION task(left) and on RELATIONAL task(right). The green line is the estimated deconvolved signal.

suggest that the proposed approach is effective in estimate the deconvolved brain activity. Nevertheless, several promising future directions could enhance this work. One possibility is exploring leaner network designs, which could result in more efficient models. Another avenue is developing an input-dependent kernel to dynamically deconvolve the signal based on input characteristics. Incorporating spatial data to account for variations in hemodynamic BOLD functions across different brain regions is another important consideration. Finally, testing the model on resting-state fMRIs, where ground truth for brain activation is unavailable, and comparing it with other model-based approaches, would evaluate its robustness in more complex and less controlled scenarios. These directions represent valuable opportunities to extend the current work and deepen our understanding of brain dynamics through task-based and resting-state fMRI analysis.

VI. Ethical Risk Assessment - Situation A

A key ethical risk identified in this project is the potential for bias in predictive models generated from task-based fMRI data. This risk arises from the possibility of misrepresenting minorities, leading to biased predictions or misclassifications when applied to certain groups of individuals. Such biases could result in unfair or discriminatory outcomes, especially in clinical or research contexts.

A. Impact on Stakeholders

The primary stakeholders impacted by this risk include participants from the Human Connectome Project (HCP), researchers, and healthcare professionals. Participants may be affected if their data is used to draw inaccurate conclusions about brain activity or cognitive abilities, potentially influencing their treatment or participation in future studies. Researchers and healthcare professionals may inadvertently rely on biased models, leading to erroneous interpretations or decisions. Moreover vulnerable groups, such as individuals with mental health conditions, are at risk of being unfairly stigmatized or misdiagnosed if biased models are used in clinical settings.

B. Significance of the Risk

The likelihood of this risk occurring is moderate, as machine learning models can unknowingly capture biases in the data, particularly when the dataset lacks sufficient diversity or when certain demographic groups are underrepresented. The only way to ensure a low ethical risk is to pay strong attention while recording data, but a small bias will be always present when talking about human health conditions.

C. Risk Mitigation and Evaluation

To mitigate this risk, we considered only anonymous fMRI data to protect participant's privacy and reduce the possibility of biased predictions based on sensitive personal information. Additionally, training the model on augmented data tries to break possible biases in the original data of the patients. Despite these efforts, challenges remain to fully ensure fairness in our machine learning models, as biases can still emerge in subtle ways.

In conclusion, while the ethical risk of bias was addressed by implementing several mitigation strategies, ensuring complete fairness and avoiding unintended consequences remains a challenge.

References

[1] H. C. Project, "The human connectome project," 2024, accessed: 2024-12-12. [Online]. Available: http://www.humanconnectomeproject.org

[2] G. L. Zeng, "Better than the total variation regularization," *Int J Biomed Res Pract*, vol. 4, no. 2, Jun. 2024.

[3] F. I. Karahanoğlu, C. Caballero-Gaudes, F. Lazeyras, and D. Van De Ville, "Total activation: fmri deconvolution through spatio-temporal regularization," *Neuroimage*, vol. 73, pp. 121–134, 2013.

[4] K. C. Chuang, S. Ramakrishnapillai, K. Kirby, A. W. Van Gemmert, L. Bazzano, and O. T. Carmichael, "Joint estimation of neural events and hemodynamic response functions from task fmri via convolutional neural networks," in *International Workshop on Machine Learning in Clinical Neuroimaging*. Springer Nature Switzerland, 2023, pp. 67–78, https://inria.hal.science/hal-02085810v2/document.

[5] J. A. Livezey and J. I. Glaser, "Deep learning approaches for neural decoding: from cnns to lstms and spikes to fmri," *arXiv preprint arXiv:2005.09687*, 2020, https://arxiv.org/pdf/2005.09687.

[6] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[7] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019, pp. 8024–8035. [Online]. Available: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf