



Ilya Zakharov<sup>1</sup>, Alexey Shovkun<sup>1</sup>, Andrey Kiryasov<sup>1</sup>, Polina Pshonkovskaya<sup>1</sup>, Timofey Adamovich<sup>1</sup>, Andrey Zhdanov<sup>1</sup>, Mariam Khayretdinova<sup>1\*</sup>

\*E-mail: [mariam@brainify.ai](mailto:mariam@brainify.ai)

<sup>1</sup>Brainify.AI, New York City, US

## Introduction

Current research on sex-related electrical signatures of the brain shows that some of these features are more common in females and others are more common in males. Overall, sex-related brain variance is better described as a continuous rather than a binary variable. Moreover, fMRI studies have found the mosaic “male” and “female” zones (Joel et al., 2015), and the distribution of such zones can be unique for a person. The “brain sex” phenotype may act as a biomarker to mark certain mental health disorders (Phillips et al., 2019).

## Objectives

1. To investigate the accuracy of “brain sex” prediction by utilizing machine learning algorithms
2. To examine what would be considered the most important features for “brain sex” prediction
3. To examine individual differences in sex-related EEG features (“mosaic” distribution of qEEG for a person)

## Methods

**Dataset:** An initial TD-Brain dataset (van Dijk et al., 2022) consisted of 1,274 patients (620 females), aged  $38.67 \pm 19.21$  (range 5–88) years, with a total of 1,346 EEG sessions.

### qEEG features:

We calculated 4298 features from 2 min EEG recordings with eyes-opened and eyes-closed conditions. We used

- EEGlib package (L. Cabañero-Gomez et al., 2017) focused primarily on calculation of complexity and connectivity features
- Neural\_PY (Toole and Boylan, 2017) focused primarily statistical properties of EEG signals
- Phase synchronization (wPLI) calculated with MNE-Python (Gramfort et al, 2017)

### ML model development:

We used the tree-based LightGBM model (Ke et al., 2017) to analyze the accuracy of brain sex prediction. To avoid overfitting or selection bias, 5-fold cross-subject cross-validation was used.

To find the best hyperparameters for the model we used Scikit-Optimize library (Head et al., 2018).

The hyperparameters for the best model were:  
**learning\_rate:** 0.09, **num\_leaves:** 83, **max\_bin:** 179, **min\_child\_samples:** 83, **subsample:** 0.5, **colsample\_bytree:** 0.37

## Results: Sex prediction

- balanced accuracy (whole set of features): 73.7% (AUC = 0.808, f1 score = 0.76)
- balanced accuracy (top 40 features): 72% (AUC = 0.77, f1 score = 0.74)

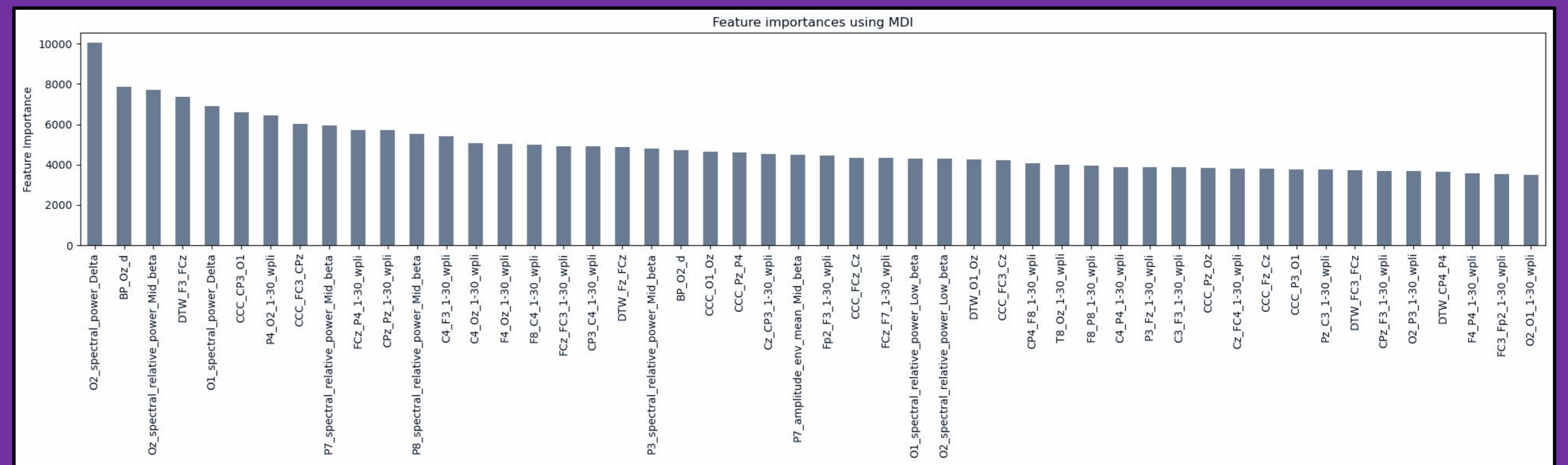


Figure 1. represents feature explanation “by gain” for the Light GBM model. The gain implies the relative contribution of the corresponding feature to the model calculated by taking each feature contribution for each tree in the model

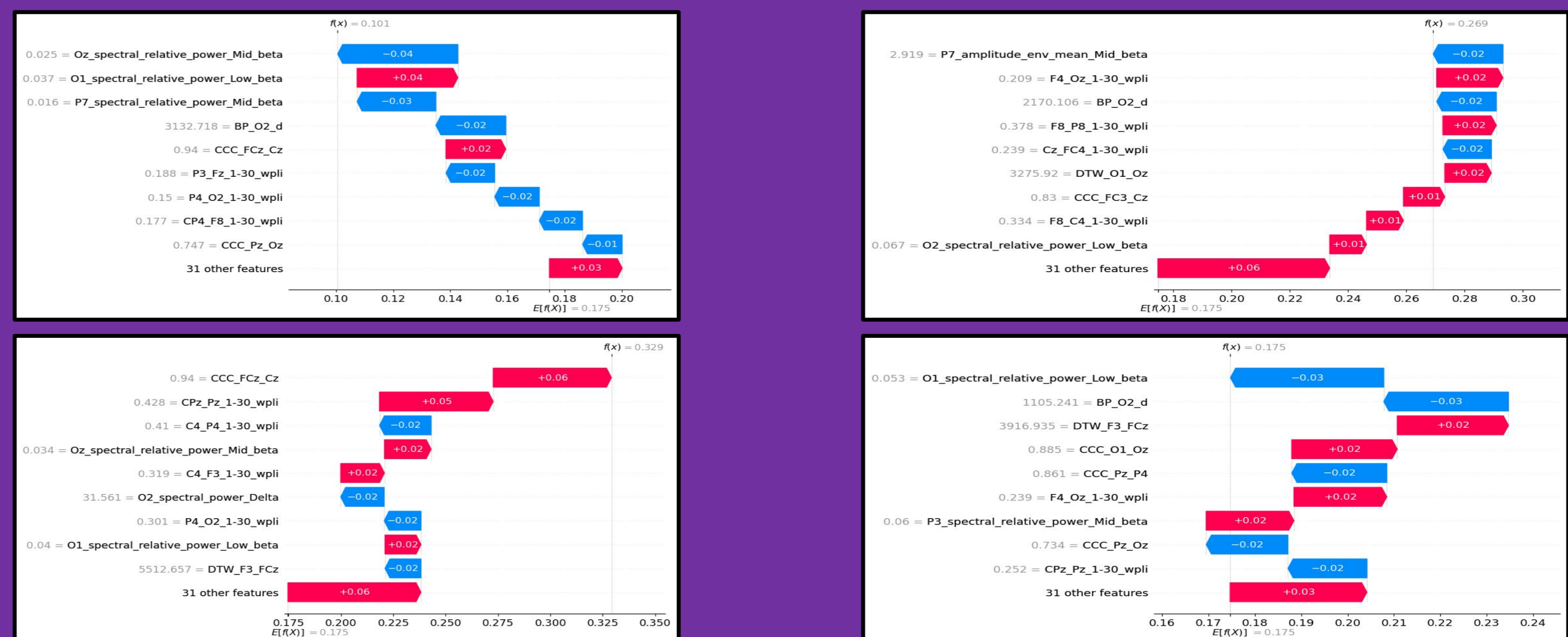


Figure 2. shows example of the weights of the different features for the sex prediction of 4 different people,  $f(x)$  - The base logit value for the prediction for the given person. Calculated with SHAP (SHapley Additive exPlanation) library, based on the a game theoretic approach to explain the output of machine learning model

### Top features explanation:

- BP\_band - absolute spectral power for the given band (a - alpha, b - beta, t - theta, d - delta)
- Spectral relative power - spectral power in the frequency band, normalised to total spectral power EEG power
- CCC - cross correlation coefficient between the data in c1 and the data in c2.
- DTW - the dynamic time warping algorithm between the data of the given channels. DTW It is a method to calculate the optimal matching between two time series.
- 1-30\_wpli - weighted Phase Lag Index (Vinck et al, 2011), calculated between c1 and c2 in the 1-30 Hz frequency range

## Discussion

The feature importance analysis shows that the most important features are related to cross-electrode connectivity (primarily, based on wPLI synchronization) in fronto-central areas of the brain. Results are in line with the Zhang et al. (2018) study, where 87% prediction accuracy was achieved from fMRI data with functional connectivity features within the default mode, fronto-parietal and sensorimotor networks contributed most to the “brain sex” prediction. The individual patterns of sex-related qEEG patterns are unique for an individuals.

## Conclusion

Accurate prediction of the person “brain sex” is feasible from qEEG data with high accuracy. The individual EEG “brain sex” phenotype score can be in future investigated as a biomarker for various mental health disorders.

### REFERENCES:

1. Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A. Engemann, Daniel Strohmeier, Christian Brodbeck, Roman Goj, Mainak Jas, Teon Brooks, Lauri Parkkonen, and Matti S. Hämäläinen. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7(267):1–13, 2013. doi:10.3389/fnins.2013.00267.
2. Cabanero-Gomez L., Hervas, R., Gonzalez, I., & Rodriguez-Benitez, L. (2021). eeglib: a Python module for EEG feature extraction. *SoftwareX*, 15, 100745.
3. Head, T., MechCoder, G. L., & Sheherbati, I. (2018). scikit-optimizers/scikit-optimizers: v0.5.2. Version v0.5
4. Joel, D., Berman, Z., Tavor, I., Wexler, N., Gaber, O., Stein, Y., ... & Assaf, Y. (2015). Sex beyond the genitalia: The human brain mosaic. *Proceedings of the National Academy of Sciences*, 112(50), 15468–15473.
5. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30, 3146–3154.
6. Phillips, O. R., Onopa, A. K., Hsu, V., Ollila, H. M., Hillary, R. P., Hallmayer, J., ... & Singh, M. K. (2019). Beyond a binary classification of sex: An examination of brain sex differentiation, psychopathology, and genotype. *Journal of the American Academy of Child & Adolescent Psychiatry*, 58(8), 787–798.
7. Toole, J. M., & Boylan, G. B. (2017). NEURAL: quantitative features for newborn EEG using Matlab. *arXiv preprint arXiv:1704.05694*.
8. Vinck, M., Oostenveld, R., Van Wingerden, M., Battaglia, F., & Pennartz, C. M. (2011). An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. *Neuroimage*, 55(4), 1548–1565.
9. Zhang, C., Dougherty, C. C., Baum, S. A., White, T., & Michael, A. M. (2018). Functional connectivity predicts gender: Evidence for gender differences in resting brain connectivity. *Human brain mapping*, 39(4), 1765–1776.