Dance of the Neurons: Unraveling Sex from Brain Signals

Mohammad-Javad Darvishi-Bayazi[1,2,3,*], Mohammad Sajjad Ghaemi[4], Jocelyn Faubert[2,3] and Irina Rish[1,2]

[1]Mila - Québec AI Institute, Montréal, QC, Canada
[2]Université de Montréal, Montréal, QC, Canada
[3]Faubert Lab, Montréal, QC, Canada
[4]National Research Council Canada, Toronto, ON, Canada

Abstract

Previous studies have shown that machine learning can predict biological sex from EEG data with high accuracy. However, the validity and generalizability of these findings across different datasets and tasks still need to be clarified. In this paper, we investigated the robustness and transferability of sex-related patterns in EEG data using a Convolutional neural network (CNN) trained on several corpora of EEG recordings ranging from 221 to 12,000 participants from healthy and diseased subjects. We evaluated the CNN on datasets from various sources and groups, with varying degrees of shift in their distributions. We found that CNNs can detect sex from EEG data accurately on datasets without fine-tuning or adaptation when the shift is low. However, performance drops where the shift is drastic. These results suggest that sex-related patterns in EEG data are robust and transferable across diverse datasets and relevant tasks. We discuss the implications of these findings for EEG analysis, machine learning applications, and best practices to avoid sex biases that enhance personalized mental health interventions.

Keywords

EEG, Sex Prediction, Artificial Neural Network, Machine Learning, Robustness, Transfer Learning, Mental Health

1. Introduction

Addressing sex biases in medicine and mental health is vital, as exemplified by the US Food and Drug Administration's suspension of ten prescription drugs, eight of which posed higher health risks in women. The root cause of this issue lies in a discernible bias towards males in various research stages [1]. Therefore, recognizing sex as a crucial biological variable in primary and preclinical research ensures accurate and replicable results.

Understanding the complex interplay between brain function and sex is vital to advancing mental health comprehension [2]. Electroencephalogram (EEG) signals, reflecting brain electrical activity, offer a unique avenue for exploring sex-related neural patterns. Combined with large datasets, machine learning has become a powerful tool in deciphering intricate neurological phenomena. This research endeavour holds significant implications for personalized medicine and mental health interventions, offering the potential to enhance early detection, diagnosis, and treatment of disorders [3]. The intersection of EEG analysis, machine learning, and large datasets opens new frontiers in mental health research, promising more precise and practical approaches to promoting mental well-being.

Most of the studies in the field have primarily focused on differences in brain size and static features [4, 5, 6, 7], ignoring the dynamic aspects of brain function. To address this gap, we propose using EEG, which provides insights into brain dynamics and activity patterns. However, one major challenge in utilizing EEG data is its inherent noise. To overcome this issue, we suggest employing a large number of samples to increase the signal-to-noise ratio, thus enhancing the reliability and accuracy of the findings. By incorporating EEG data into the analysis, we can better understand the brain's dynamic processes and their relationship to sex differences and behaviour.

Despite the promising potential of using brain imaging and machine learning in mental health research to classify sex-specific markers, a significant challenge arises from the often small and limited datasets employed in these studies. For instance, Bučková et al. [8] evaluated deep learning classifiers on a small number of participants with Major Depressive Disorder (MDD) and Jochmann et al. [9], Van Putten et al. [10] applied on a mid-size dataset on healthy participants(see Table 1 for a comparison). The reliance on insufficient sample sizes can lead to incomplete and biased conclusions [4], hindering the generalizability and reliability of findings. This issue is particularly critical in understanding the intricate connections between brain function and mental health, where individual variations and complexities require comprehensive datasets.

To tackle these challenges, we leveraged machine learning techniques on functional brain imaging data, specifically EEG, across diverse datasets encompassing

Machine Learning for Cognitive and Mental Health Workshop (ML4CMH), AAAI 2024, Vancouver, BC, Canada
*Corresponding author.
mohammad.bayazi@mila.quebec (M. Darvishi-Bayazi)
https://www.linkedin.com/in/mjdarvishi/ (M. Darvishi-Bayazi)
0000-0002-3251-8491 (M. Darvishi-Bayazi)

Table 1
A comparison of previous studies on EEG sex detection. The table shows the name of the study, the dataset used, the number of participants and recordings in the dataset and in (train, test) splits, participants' conditions, and the data availability.

Study	# of Participants	# of Recordings	Conditions	Dataset
Van Putten et al. [10]	1308 (1000, 308)	1308	All Normal	In Lab
Bučková et al. [8]	144	144	MDD	In Lab
Jochmann et al. [9]	1282 (1140, 142)	1282	Only Normal Split	TUAB
Ours	2417	2417	Normal/Abnormal	Public-NMT
Ours	2329	2978	Normal/Abnormal	Public-TUAB
Ours	14987	69000	Unlabeled	Public-TUEG

varying sample sizes and populations, including both normal and abnormal [1] populations. Also, we examined the performance of classifiers under the distribution shift on unseen data. Ultimately, our findings contribute to more robust and applicable insights for targeted and personalized mental health interventions.

2. Material and Methods

2.1. Datasets

We used three publicly available EEG datasets with different sample sizes and conditions to investigate the effect of sex on EEG signals. The datasets are:

NMT (NUST-MH-TUKL EEG): This dataset contains $2,417$ recordings from healthy and pathological subjects, with a total duration of 625 hours. The recordings are labelled as normal or abnormal by qualified neurologists and also include demographic information, such as sex and age [11].

TUAB (Temple University Hospital Abnormal EEG Corpus): This dataset is a subset of the TUEG corpus that contains $1,985$ recordings from $1,652$ subjects, with a total duration of 453 hours. The recordings are labelled as normal or abnormal by qualified neurologists [12, 13, 14].

TUEG (Temple University Hospital EEG Corpus): This dataset is a large open-source corpus of EEG data, containing over $69,000$ recordings from $14,987$ subjects, with a total duration of $27,062$ hours. The recordings are de-identified and annotated with clinical information, such as age and sex [13, 14].

We utilize patient sex information, encoded as 0 or 1 as our neural network target. We focus on sex instead of gender due to the dataset's clinical origin, assuming patients' records reflected assigned birth sex rather than self-identified gender. We applied several preprocessing steps to the EEG data, including 21 common channels, which were selected across the datasets. We used Artifact Subspace Reconstruction (ASR) [15] to remove artifacts from the EEG. We z-scored the EEG signals to each channel's statistics. We used predefined test sets to report the accuracy of our models and 15% of the training splits for model selection.

[1]The term "Normal/Abnormal" is used in original datasets to describe EEG recordings that contain pathological features, such as epileptic spikes, periodic discharges, or other abnormal patterns. It does not imply any value judgment or stigma but rather reflects the quality of the EEG signal. In this paper, we adhered to the same terminology.

2.2. Model

We used ShallowNet [16] as our model for all experiments, as it is a simple and efficient Convolutional Neural network (CNN) that can perform well on various EEG tasks. ShallowNet consists of only one convolutional layer followed by a fully connected layer, which reduces the number of parameters and the computational cost compared to deeper networks. We implemented ShallowNet in BrainDecode [17] and trained it using the AdamW optimizer with a learning rate of 0.000625, weight decay of 0, drop probability of 0.5, and a batch size of 64. We used binary cross-entropy as the loss function and balanced accuracy (BAC) as the evaluation metric.

2.3. Training and Evaluation

The training and evaluation of the model were conducted with the primary objective of classifying sex from EEG signals. Our focus extended beyond the training dataset to include a comprehensive analysis of model performance on both the test split of the training dataset and other unseen datasets. The overarching goal was to assess the *Sex Detectability* (SD) from EEG signals and evaluate the model's robustness to distribution shifts in unseen data. We investigated detectability and transferability under various conditions to investigate the model's capabilities. Specifically, we explored the model's performance when trained and tested on subsets of the data, considering scenarios where only Normal participants were included, only Abnormal participants were included, or when the entire dataset was utilized. This approach allowed us to understand how well the model generalizes across different participant profiles.

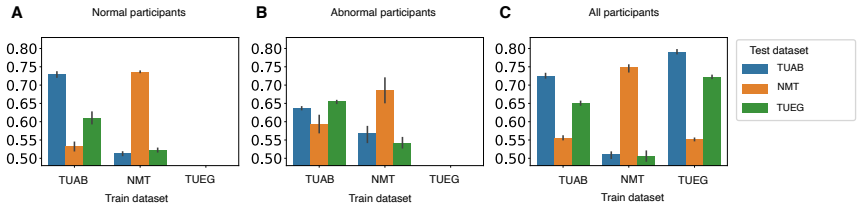

Figure 1: SD from EEG in three populations: A) Normal B) Abnormal C) All participants. Error bars show the standard error of BAC across three random seeds. It is worth noting that the TUEG dataset does not have pathology labels. Therefore, the results for the TUEG dataset are not available in A and B, and we only visualize the results for all participants in C

Furthermore, we conducted experiments to understand the impact of SD on pathology detection. To achieve this, we trained the model on the NMT dataset, which features imbalances in different aspects. Our analysis focused on different subgroups within the dataset, including Male Normal, Female Normal, Male Abnormal, and Female Abnormal participants. We aimed to elucidate any potential associations between SD and pathology detection by examining the model's performance on these subgroups. We ran each experiment with three random seeds. All error bars show the standard error of the metrics of the three seeds.

3. Results

One of the objectives of this study is to investigate if the biological sex of the subjects is detectable from their scalp EEG recordings. This question is relevant for understanding the sex-specific differences in brain activity and their implications for the diagnosis and treatment of various neurological and psychiatric disorders. Moreover, this question is also essential for evaluating the potential biases and limitations of machine learning classifiers trained on EEG data.

3.1. *Sex Detectability* (SD) from EEG

To address SD from EEG, we experimented with several datasets of different sizes and compositions, including the TUEG EEG dataset, the world's most extensive open-source corpus of EEG data. We also considered the normal and abnormal populations of the subjects. We used a shallow and deep convolutional neural network (CNN) as our classifier, Previous studies have shown that this CNN can achieve competitive accuracy with larger models in predicting pathology from EEG data [18, 19].

The results of our experiments are summarized in Figure 1 and Table 2, which show the BAC of the CNN classifier for each dataset. The figure shows the BAC when we train the model on a dataset and test on its own test split, which is in distribution. It also shows the BAC of a model trained on a dataset and tested on another dataset, which is out of distribution performance. The results indicate that the sex of the subjects is detectable from their EEG recordings in all of the distribution scenarios, with accuracy ranging from 60% to 80%. The results also show that the sex detection performance is slightly higher for the normal population than for the abnormal population. It is worth noting that the TUEG dataset does not have pathology (Normal/Abnormal) labels, therefore, we do not show the results for normal and abnormal participants.

3.2. Performance on Unseen Data (Zero-Shot)

To evaluate the model's adaptability to unseen data, we conducted zero-shot performance assessments across various datasets. Zero-shot performance means that the model can predict the class of a sample from an unseen dataset without having seen any examples from that dataset during training. Notably, the highest accuracy of 79.11% was achieved when training on TUEG and testing on TUAB EEG datasets. Conversely, the lowest accuracies were observed when the model was evaluated across the TUH datasets and the NMT Scalp EEG Dataset.

Our investigation extended beyond the original training and testing datasets to explore out-of-distribution accuracy, mainly focusing on the abnormal population. Strikingly, the model exhibited higher accuracy in out-of-distribution scenarios when dealing with the abnormal population. To comprehensively gauge the generalization of learned features, each model was tested on other datasets to evaluate zero-shot performance. This analysis provided insights into how well the model leverages learned features when confronted with entirely new data.

In table 2, we compare our models with previous work on sex detection on the TUAB dataset, which has a moderate sample size compared to two other datasets in our study. The results show that ShallowNet archives comparable results in the distribution scenario and outperforms by a high margin on the zero-shot scenario. The reason for this improvement might be that the TUEG dataset has seven times more unique participants, and the data

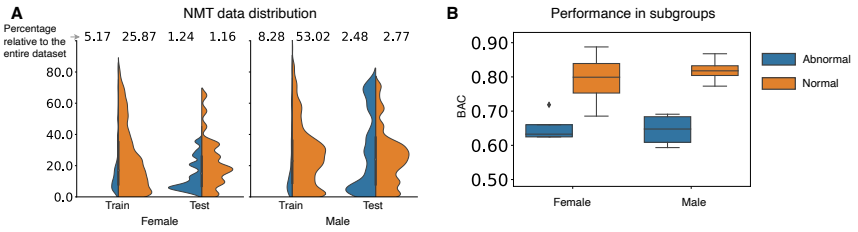

Figure 2: Sex imbalances in the NMT dataset and its effect on pathology detection: A) Data distribution of the NMT dataset, the number of male samples is two times higher than that of females. B) Accuracies of subgroups. The difference in the number of samples does not affect the pathology detection.

Table 2
Comparison of BAC (%) between previous work on TUAB dataset and ours

Method	TUAB
Clean [9]	74.00±02.00
Clean (Ours)	72.89±01.21
Zero-Shot (Pre-trained-Ours)	**79.11±01.47**

distribution is close to TUAB. Therefore, it improves the performance of the TUAB dataset by 7%.

These results demonstrate that the biological sex of the subjects is a significant factor that machine learning methods could capture. However, these results also may imply that the sex of the subjects should be taken into account when developing and evaluating machine learning classifiers for EEG pathology detection, as the sex distribution of the training and testing data may affect the generalization and robustness of the models.

3.3. Sex Imbalance's Impact on EEG Pathology Detection

As we see in the previous sections (SD and Zero-Shot), sex is detectable from the EEG signals and is an important biological factor that can influence human brain activity and behaviour. Therefore, considering it in the analysis is essential, especially when the datasets are imbalanced. In this section, we aim to investigate the effect of sex on pathology detection from EEG signals using the NMT Scalp EEG Dataset. The NMT dataset has a significant sex imbalance, as men are two times more frequent than women in the dataset. This raises the question of whether the sex imbalance and the SD from the EEG signals can affect the performance of the pathology detection models.

To address this question, we conducted several experiments using different deep-learning architectures. We first verified that sex is detectable from the EEG signals using a simple convolutional neural network (CNN) that achieved a good accuracy on the sex classification task on several EEG datasets with different sample sizes (see Table 2 and Figure 1). We then evaluated the pathology detection models on the NMT dataset for different subgroups.

Figure 2 shows how the sex imbalance in the NMT dataset does not affect the pathology detection performance. Although the NMT dataset has twice as many male samples as female samples, as shown in panel A. However, this does not lead to a significant difference in the accuracy of the pathology detection models for the male and female subgroups, as shown in panel B. This suggests that the sex imbalance in the NMT dataset does not hurt the pathology detection quality.

One possible reason for this finding is that the NMT dataset has a balanced ratio of normal and abnormal samples in each sex subgroup, as shown in Figure 2A. This means that the models can learn the features related to the pathology, not the sex, of the subjects. Therefore, even though the sex is detectable from the EEG signals, it does not interfere with the pathology detection task.

4. Discussion and conclusion

Historically documented sex differences in EEG patterns and the successful application of machine learning for automatic sex detection suggest that sex-related patterns can act as confounders in machine learning-based EEG assessments [8, 9]. In our experimentation on potential confounding factors within the NMT dataset, we explored a scenario involving an imbalance in male and female participants. Our findings indicate that, in this dataset, sex does not function as a confounder due to an equal distribution of abnormal participants in the male/female splits. However, as demonstrated in the SD section, we reveal that sex remains detectable. Consequently, acknowledging sex as a factor is essential for precision medicine in mental health.

A key takeaway from an extensive review spanning three decades of research on human brain sex differences is that, despite evident behavioural distinctions between men and women, disparities in brain structure and function are minimal and inconsistent when adjusted for in-

dividual brain size and inefficient participant numbers [4]. In contrast, our study employs EEG, which has high temporal but low spatial resolution, to assess functional brain activity. Our findings reveal distinct patterns across datasets with varying subject numbers, highlighting the unique insights provided by EEG in uncovering differences.

Brain connectivity and topography research has yielded diverse perspectives, providing a rich field for future investigations. For instance, Ingalhalikar et al. [20] found that male brains exhibit enhanced connectivity between perception and coordinated action, while female brains are structured to facilitate communication between analytical and intuitive processing modes. Their study, involving 949 youths, demonstrated distinct patterns in supratentorial connections, with stronger intrahemispheric connections in males and stronger interhemispheric connections in females. Jochmann et al. [9] highlighted the significance of EEG topographies in sex detection, revealing that even with disrupted waveforms, the sex could be accurately identified. On the other hand, Bučková et al. [8] observed that the incorporation of multivariate classification models did not consistently improve performance. Also, Eliot et al. [4] argues that despite decades of examining sex effects on lateralized brain function, there is no substantial evidence supporting the widely held belief that male brains are significantly more lateralized than female brains. The diversity of findings in the literature underscores the complexity of brain connectivity and topography, making it an intriguing and promising avenue for future research. One could examine where the trained neural network looks when classifying brain signals.

Frequency bands are widely recognized as critical features in quantitative EEG analysis. Despite their prominence, the significance of these features in sex detection remains unclear. Some studies assert that brain rhythms exhibit sex-specific patterns [21, 10], while others argue that none of the traditional frequency bands play a particularly crucial role in sex detection [9]. A potential avenue for future research would be to explore and substantiate these claims using an extensive dataset, such as TUEG.

In summary, our training and evaluation process thoroughly explored the model's performance in classifying sex from EEG signals. We systematically assessed its ability to generalize to unseen data, examined detectability and generalization under varying conditions, and investigated potential implications for pathology detection using a diverse and imbalanced dataset. The results of these analyses contribute to a nuanced understanding of the model's capabilities and potential applications in clinical settings.

5. Acknowledgements

We extend our sincere appreciation to Mathilde Besson for their valuable comments, which greatly contributed to the refinement of this paper. This work was funded by Canada CIFAR AI Chair Program and from the Canada Excellence Research Chairs (CERC) program, National Research Council Canada, Natural Sciences and Engineering Research Council (NSERC-CAE-CRIAC-CARIQ, NSERC discovery grant RGPIN-2022-05122), Doctoral Research Microsoft Diversity Award (Microsoft-Mila), Faculty of medicine-UdeM, and Faculté des études supérieures et postdoctorales. We thank Compute Canada for providing computational resources.

References

[1] S. K. Lee, Sex as an important biological variable in biomedical research, BMB reports 51 (2018) 167.

[2] D. M. Christiansen, M. M. McCarthy, M. V. Seeman, Understanding the influences of sex and gender differences in mental disorders, Frontiers in Psychiatry 13 (2022) 984195.

[3] T. J. Sejnowski, P. S. Churchland, J. A. Movshon, Putting big data to good use in neuroscience, Nature neuroscience 17 (2014) 1440–1441.

[4] L. Eliot, A. Ahmed, H. Khan, J. Patel, Dump the "dimorphism": Comprehensive synthesis of human brain studies reveals few male-female differences beyond size, Neuroscience & Biobehavioral Reviews 125 (2021) 667–697.

[5] A. M. Chekroud, E. J. Ward, M. D. Rosenberg, A. J. Holmes, Patterns in the human brain mosaic discriminate males from females, Proceedings of the National Academy of Sciences 113 (2016) E1968–E1968.

[6] F. Sepehrband, K. M. Lynch, R. P. Cabeen, C. Gonzalez-Zacarias, L. Zhao, M. D'Arcy, C. Kesselman, M. M. Herting, I. D. Dinov, A. W. Toga, et al., Neuroanatomical morphometric characterization of sex differences in youth using statistical learning, Neuroimage 172 (2018) 217–227.

[7] C. Sanchis-Segura, M. V. Ibañez-Gual, N. Aguirre, Á. J. Cruz-Gómez, C. Forn, Effects of different intracranial volume correction methods on univariate sex differences in grey matter volume and multivariate sex prediction, Scientific Reports 10 (2020) 12953.

[8] B. Bučková, M. Brunovskỳ, M. Bareš, J. Hlinka, Predicting sex from eeg: validity and generalizability of deep-learning-based interpretable classifier, Frontiers in Neuroscience 14 (2020) 589303.

[9] T. Jochmann, M. S. Seibel, E. Jochmann, S. Khan, M. S. Hämäläinen, J. Haueisen, Sex-related patterns

in the electroencephalogram and their relevance in machine learning classifiers, Human Brain Mapping 44 (2023) 4848–4858.
[10] M. J. Van Putten, S. Olbrich, M. Arns, Predicting sex from brain rhythms with deep learning, Scientific reports 8 (2018) 3069.
[11] H. A. Khan, R. Ul Ain, A. M. Kamboh, H. T. Butt, S. Shafait, W. Alamgir, D. Stricker, F. Shafait, The nmt scalp eeg dataset: an open-source annotated dataset of healthy and pathological eeg recordings for predictive modeling, Frontiers in neuroscience 15 (2022) 755817.
[12] S. López, I. Obeid, J. Picone, Automated interpretation of abnormal adult electroencephalograms, Ph.D. thesis, Temple University, 2017.
[13] N. Shawki, M. G. Shadin, T. Elseify, L. Jakielaszek, T. Farkas, Y. Persidsky, N. Jhala, I. Obeid, J. Picone, Correction to: The temple university hospital digital pathology corpus, in: Signal Processing in Medicine and Biology: Emerging Trends in Research and Applications, Springer, 2022, pp. C1–C1.
[14] I. Obeid, J. Picone, The temple university hospital eeg data corpus, Frontiers in neuroscience 10 (2016) 196.
[15] S. Blum, N. S. Jacobsen, M. G. Bleichner, S. Debener, A riemannian modification of artifact subspace reconstruction for eeg artifact handling, Frontiers in human neuroscience 13 (2019) 141.
[16] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, T. Ball, Deep learning with convolutional neural networks for eeg decoding and visualization, Human brain mapping 38 (2017) 5391–5420.
[17] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, T. Ball, Deep learning with convolutional neural networks for eeg decoding and visualization, Human Brain Mapping (2017). URL: http://dx.doi.org/10.1002/hbm.23730. doi:10.1002/hbm.23730.
[18] M.-J. Darvishi-Bayazi, M. S. Ghaemi, T. Lesort, M. R. Arefin, J. Faubert, I. Rish, Amplifying pathological detection in eeg signaling pathways through cross-dataset transfer learning, Computers in Biology and Medicine (2023) 107893.
[19] L. A. Gemein, R. T. Schirrmeister, P. Chrabąszcz, D. Wilson, J. Boedecker, A. Schulze-Bonhage, F. Hutter, T. Ball, Machine-learning-based diagnostics of eeg pathology, NeuroImage 220 (2020) 117021.
[20] M. Ingalhalikar, A. Smith, D. Parker, T. D. Satterthwaite, M. A. Elliott, K. Ruparel, H. Hakonarson, R. E. Gur, R. C. Gur, R. Verma, Sex differences in the structural connectome of the human brain, Proceedings of the National Academy of Sciences 111 (2014) 823–828.
[21] P. Kaushik, A. Gupta, P. P. Roy, D. P. Dogra, Eeg-based age and gender prediction using deep blstm-lstm network model, IEEE Sensors Journal 19 (2018) 2634–2641.