NRC·CNRC

NRC Publications Archive
Archives des publications du CNRC

Critical role of EEG signals in assessment of sex-specific insights in neurological diagnostics via machine learning approach

Darvishi-Bayazi, Mohammad-Javad; Ghaemi, Mohammad Sajjad; Rish, Irina; Faubert, Jocelyn

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. / La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.

For the publisher's version, please access the DOI link below./ Pour consulter la version de l'éditeur, utilisez le lien DOI ci-dessous.

Publisher's version / Version de l'éditeur:
https://doi.org/10.1038/s41598-025-30848-y
Scientific Reports, 16, 1, pp. 1-12, 2025-12-10

NRC Publications Archive Record / Notice des Archives des publications du CNRC :
https://nrc-publications.canada.ca/eng/view/object/?id=a264eee2-ca92-44c5-bc1e-bc7da199dfcd
https://publications-cnrc.canada.ca/fra/voir/objet/?id=a264eee2-ca92-44c5-bc1e-bc7da199dfcd

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at
https://nrc-publications.canada.ca/eng/copyright
READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site
https://publications-cnrc.canada.ca/fra/droits
LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.

National Research Council Canada | Conseil national de recherches Canada

Canada

scientific reports

Check for updates

OPEN

Critical role of EEG signals in assessment of sex-specific insights in neurological diagnostics via machine learning approach

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

Early detection and diagnosis of neurological pathology are essential for timely treatment and intervention. While deep learning has shown promise in analyzing brain imaging data, the influence of sex-specific patterns in electroencephalogram (EEG) signals remains underexplored. In this study, we investigated the detectability and impact of biological sex in EEG data using Artificial Intelligence (AI) methods, with a focus on both biological sex classification and its confounding effects in pathological EEG diagnosis. We employed a lightweight yet effective convolutional neural network and evaluated its performance across three diverse EEG datasets (TUEG, TUAB, and NMT), including both healthy and pathological subjects. Our evaluation leveraged datasets from various sources and participant groups, featuring distribution shifts. Our model achieved balanced accuracy ranging from 65% to 80% in detecting biological sex from EEG signals, demonstrating the robustness and cross-dataset transferability of sex-related neural patterns. While the AI models demonstrated accurate biological sex detection on datasets without fine-tuning, their performance declined with significant distribution shifts. Furthermore, we explored the relationship between biological sex and pathology by visualizing salient features for target detection across distinct subgroups. Our findings revealed unprecedented insights into the negligible role of sex-specific patterns in pathology detection despite the presence of prominent and consistent patterns within each biological sex group. These findings are critical for advancing the development of more robust and unbiased AI models in disease prediction, as well as for informing treatment paradigms.

Keywords Electroencephalogram (EEG), Deep Learning, Model Robustness, Bias and Fairness in AI, Biological Sex Differences, Pathology Detection

Sex-related biases in medical and mental health research and practice demand urgent attention. This imperative is underscored by the US Food and Drug Administration's suspension of ten prescription drugs, with eight of these medications presenting disproportionately higher risks for women. A clear bias favouring males across various research stages contributes to this issue[1]. This pattern highlights the critical need for sex-specific considerations in drug development, from preclinical studies to clinical trials and therapeutic applications. A pervasive bias favouring male subjects across various research stages contributes substantially to this problem. The implications of this bias extend beyond pharmacology, affecting diverse areas of medical research. For instance, biological sex has been identified as a crucial biological variable in dementia research[2]. Similarly, in the field of neurodevelopmental disorders, the diagnosis of Autism Spectrum Disorder is reported to be up to four times more frequent in males than in females, raising questions about potential sex-based differences in presentation or diagnostic criteria[3]. These examples highlight the importance of recognizing biological sex as a fundamental biological variable in primary and preclinical research. Therefore, identifying sex-related biases is essential for ensuring research outcomes' accuracy, reproducibility, and clinical relevance, ultimately leading to more effective and equitable diagnostic and treatment strategies.

Understanding the complex interplay between brain function and biological sex is fundamental to advancing our knowledge of mental health[4]. EEG signals capture the brain's electrical activity and provide a unique window

[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. ✉email: mohammad.bayazi@mila.quebec; mohammad.javad.darvishi.bayazi@umontreal.ca

into sex-related neural patterns. When combined with large-scale datasets, machine learning techniques offer a powerful approach to deciphering these intricate neurological phenomena. This research holds significant promise for personalized medicine and the development of tailored mental health interventions. By leveraging EEG analysis and machine learning algorithms on extensive datasets, we may enhance our ability to detect, diagnose, and treat various neurological and psychiatric disorders at earlier stages[5].

The role of patient biological sex in neurological disorders remains an understudied area, with sex-based differences often relegated to the status of confounding variables rather than potential drivers of pathophysiology[6,7]. However, recent investigations have begun to challenge this paradigm. For instance[8], reported that event-related potential markers in Autism Spectrum Disorder may not be confounded by biological sex differences, suggesting a more nuanced relationship. Conversely[9], demonstrated sex-specific variations in functional brain connectivity patterns among individuals with depression. These contrasting findings underscore the complexity of sex-based influences on neurological conditions. Our study aims to further elucidate this relationship by investigating whether patient biological sex significantly impacts the detection and characterization of neurological pathologies.

While previous studies in the field primarily focused on differences in brain size and static features[10–12], a comprehensive research argued that most structural differences are minor[13] when normalized to whole brain volume[14]. To address this gap, we propose leveraging EEG, providing insights into brain dynamics and activity patterns. As EEG has a better temporal resolution than functional magnetic resonance imaging (fMRI), it can complement the findings for the fMRI study of biological sex differences[15]. However, a significant challenge in utilizing EEG data is its intrinsic noise. Despite the promising potential of brain imaging and machine learning in mental health research to classify sex-specific markers, a significant hurdle arises from often small and limited datasets used in these studies. For example[16], evaluated deep learning classifiers on a small number of participants with Major Depressive Disorder (MDD), and[17,18] applied them to a mid-size dataset on healthy participants (see Table 1 for a comparison). Relying on insufficient sample sizes can lead to incomplete and biased conclusions[13], hampering 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 necessitate comprehensive datasets.

Additionally, machine learning algorithms applied to EEG data face significant challenges that can impede their generalization performance. Two primary issues are distribution shifts and artifacts in EEG signals. Distribution shift occurs when the test data's statistical properties diverge from the training data's, potentially compromising the model's accuracy on unseen samples[19]. This phenomenon is particularly relevant in EEG analysis, where inter-subject and inter-session variability can lead to substantial differences between training and deployment conditions[20]. Concurrently, EEG signals are susceptible to various artifacts and noise sources that can distort the underlying neural activity. These unwanted variations may arise from diverse factors, including suboptimal electrode placement, cardiac electrical activity, muscle contractions, ocular movements, and environmental electromagnetic interference[21]. Such artifacts can significantly influence model performance by introducing spurious patterns or obscuring relevant neurophysiological features[17]. These artifacts and noise sources pose a considerable challenge for machine learning models in extracting meaningful features and patterns from EEG data. Consequently, these factors can substantially impact the quality and reliability of the derived insights, necessitating robust preprocessing techniques and adaptive learning algorithms to mitigate their effects[22]. Addressing these challenges is crucial for advancing the field of EEG-based machine learning, as it would enhance the generalizability and reliability of models across diverse experimental settings and subject populations.

To overcome the limitations inherent in structural brain imaging and small sample sizes, we propose leveraging large-scale datasets to enhance the signal-to-noise ratio, thereby improving the reliability and accuracy of findings. By integrating EEG data into our analysis, we aim to elucidate the brain's dynamic processes and associations with biological sex differences and behavioural outcomes. In this study, we employed machine learning algorithms on varying-sized datasets, encompassing both Non-Pathological and Pathological populations. Additionally, we investigated classifier performance under distribution shifts on unseen data to study model generalization and robustness. Ultimately, we explored features that were important for the models of biological sex and pathology detection in different subgroups, contributing to more robust and applicable insights for targeted and personalized mental health interventions.

Recent advances in neuro-AI have begun to shift toward the development of general-purpose foundation models trained on large-scale EEG datasets[23,24]. These models aim to capture transferable neural representations

Study	#Participants	#Rec	Conditions	Dataset
[18]	1308 (1000, 308)	1308	All Normal	In Lab
[16]	144	144	MDD	In Lab
[17]	1282 (1140, 142)	1282	Only Normal Split	Pub-TUAB
Ours	2417	2417	Normal/Abnormal	Pub-NMT
Ours	2329	2978	Normal/Abnormal	Pub-TUAB
Ours	14987	69000	Unlabeled	Pub-TUEG

Table 1. A comparison of previous studies on EEG biological 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.

that can support a range of downstream tasks, from cognitive state decoding to clinical prediction. However, as these models grow in size and complexity, they inherit the limitations and biases of the data they are trained on. In the context of medical AI, this raises urgent concerns about the representativeness and fairness of training datasets, particularly when key demographic attributes (such as biological sex) are imbalanced or under-annotated[25]. EEG datasets often suffer from such imbalances, yet their impact remains underexplored at scale[26]. Our study addresses this gap by systematically analyzing sex detectability and its influence on downstream EEG-based classification tasks across three diverse datasets. By doing so, we contribute to the foundational understanding necessary for developing robust, fair, and clinically meaningful neuro-AI models in the age of large-scale, multimodal brain modeling.

In more detail, we investigate sex-specific patterns in EEG signals using Artificial Neural Networks (ANN) approaches, specifically dominant Convolutional Neural Networks (CNNs), to understand their implications for neurological diagnostics. We evaluate biological Sex Detectability (SD) across three large-scale datasets: the Temple University Hospital EEG Corpus (TUEG), Temple University Hospital Abnormal EEG Corpus (TUAB), and NUST-MH-TUKL EEG (NMT) dataset, encompassing both healthy participants and those with various pathological conditions. Our methodology evaluates model performance using Balanced Accuracy (BAcc) metrics to account for dataset imbalances. To interpret the neural network's decision-making process, we utilize Amplitude Gradients Analysis (AGA) for feature visualization across different frequency bands. The following sections present our experimental results, demonstrating biological sex detectability in EEG signals, followed by investigations into the impact of biological sex imbalances on pathology detection, feature importance analysis, and comprehensive discussions of our findings. Detailed methodological approaches are provided in the subsequent Materials and Methods section.

Results

One of the objectives of this study is to investigate whether the sex of the subjects is detectable from their scalp EEG recordings, which are functional brain imaging. 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.

This section presents our comprehensive analysis of sex-specific patterns in EEG signals across multiple datasets and experimental conditions. We begin by examining the fundamental question of biological sex detectability in EEG signals across different populations, followed by an evaluation of model generalization capabilities on unseen data through zero-shot performance assessments. Subsequently, we investigate the potential impact of biological sex imbalances on pathology detection accuracy, and conclude with a detailed feature importance analysis to understand the neural mechanisms underlying our findings.

Biological sex detectability in EEG

To explore the detectability of biological sex in EEG signals, we conducted experiments using diverse datasets, including the TUEG EEG dataset, recognized as the current most extensive open-source corpus of EEG data. Our analysis encompassed both Normal and Abnormal populations. For classification, we employed a simple, yet effective shallow CNN, known for achieving competitive accuracy in predicting pathology from EEG data, as demonstrated in previous studies[26,27].

The outcomes of our experiments are summarized in Figure 1 and Table 2, showcasing the BAcc of the CNN classifier across each dataset. The figure illustrates BAcc when training and testing the model on the same dataset, representing an in-distribution scenario. Additionally, it displays the BAcc of a model trained on one dataset and tested on another, reflecting out-of-distribution performance. The results suggest that subjects' biological sex is discernible from their EEG recordings in all populations, yielding accuracy rates ranging from 65% to 80%. Furthermore, the findings indicate slightly superior biological sex detection performance for the Normal population compared to the Abnormal population. Notably, the TUEG dataset lacks pathology labels (Normal/Abnormal), and as a result, results for Normal and Abnormal participants are not presented. Importantly, the results also show a clear drop in performance under distribution shift (i.e., when the model is evaluated on a dataset different from the one it was trained on), highlighting the challenge of generalizing sex-specific EEG features across heterogeneous data sources.

Table 2 compares our models and a previous study focusing on biological sex detection in the TUAB dataset. Notably, the TUAB dataset, with a moderate sample size compared to the two other datasets in our study, serves as the benchmark. The outcomes reveal that ShallowNet outperforms the previous model, particularly excelling in the in-distribution scenario.

Performance on unseen data (zero-shot)

We conducted zero-shot performance assessments across various datasets to evaluate the model's generalization to unseen data. 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. 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.

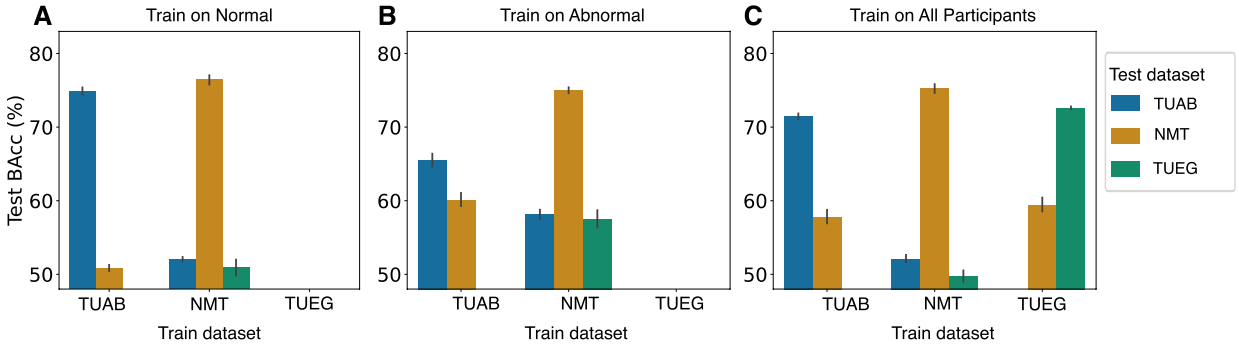

Fig. 1. Biological sex classification performance across three populations using EEG signals: (**A**) Normal, (**B**) Abnormal, and (**C**) All participants. Bars represent balanced accuracy (BAcc), and error bars indicate the standard error across ten random seeds. The TUEG dataset does not include pathology labels; therefore, results are only available for the combined population (**C**). Across datasets, BAcc ranged from 65% to 80%, with slightly higher performance observed in the Normal population compared to the Abnormal population. A notable drop in accuracy is observed when models are tested out-of-distribution, reflecting the challenge of generalizing across heterogeneous EEG datasets.

Method	TUAB
Simple CNN on clean data[17]	74.00±02.00
ShallowNet on clean data (Ours)	74.88±01.63

Table 2. Comparison of BAcc between previous work on the TUAB dataset and ours. Values show mean±SD over 10 randomly initialized models.

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

Biological sex imbalance's impact on EEG pathology detection

As we see in the previous sections (SD and Zero-Shot), biological 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 biological sex on pathology detection from EEG signals using the NMT Scalp EEG Dataset. The NMT dataset has a significant biological sex imbalance, as male samples are two times more frequent than female samples in the dataset. This raises the question of whether biological sex imbalance and the SD in EEG signals can affect the performance of the pathology detection models.

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

Figure 6 shows how the biological sex imbalance in the NMT dataset does not affect the pathology detection performance. We conducted an Independent Samples T-Test to assess the significance of differences between male and female groups for pathology detection. The results indicate a non-significant difference ($t(38) = -0.047$, $p = 0.962$) in test accuracy between the two groups. Therefore, based on our analysis, we did not find a significant distinction in pathology detection performance between male and female subjects. 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 biological sex imbalance in the NMT dataset does not hurt the pathology detection quality.

Feature importance

Figure 2 shows the gradient amplitude analysis of Female/Male classifiers of EEG signals across different frequency bands for Normal and Abnormal subjects. The colour gradient represents the gradient amplitude of the EEG signals. The Amplitude Gradients Analysis (AGA) reveals distinct patterns of feature importance across different conditions and frequency bands. The network strongly prefers features in the theta, alpha, and beta bands, as evidenced by the pronounced gradient values in these regions. This suggests that the network relies

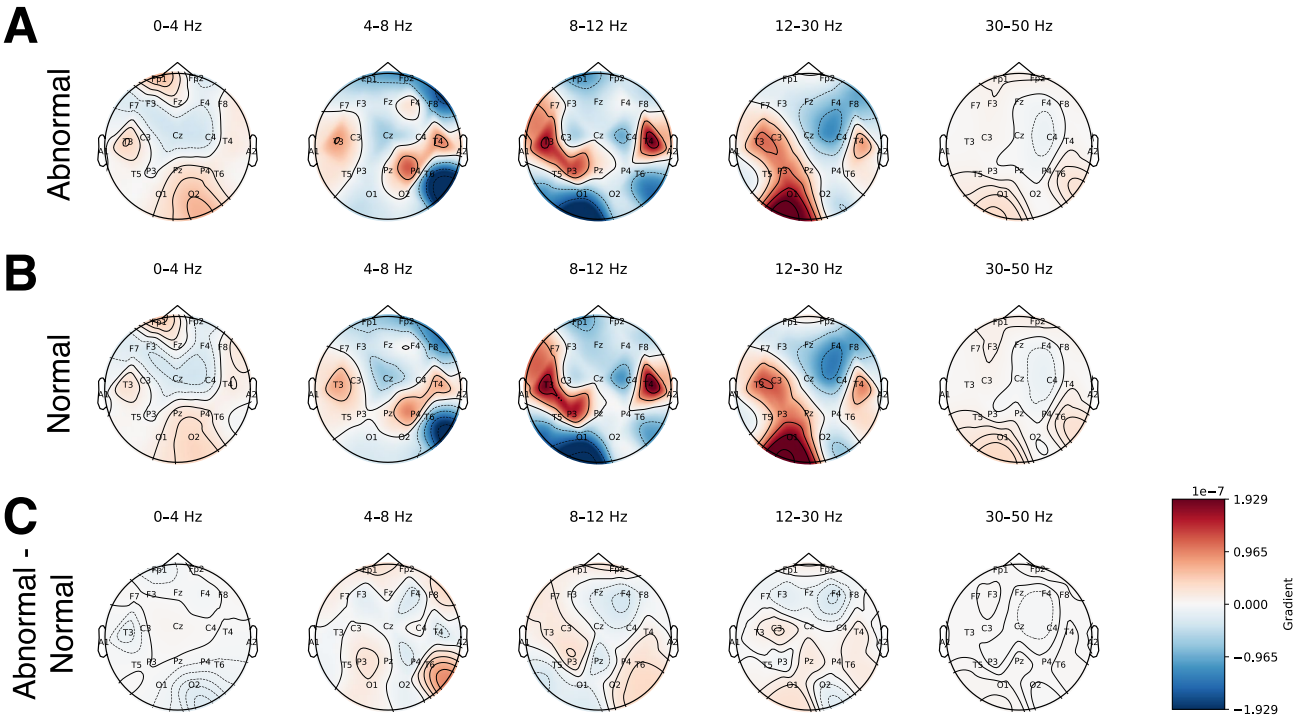

Fig. 2. AGA of different frequency bands of biological sex classifiers on NMT dataset. (**A**) Abnormal, (**B**) Normal, and (**C**) The difference between Abnormal and Normal. Red indicates a stronger relation with the female class, while blue indicates a stronger relation with the male class.

heavily on these frequencies when classifying between female and male classes. Figure 2C shows the difference between the Abnormal and Normal groups; these networks show a noticeable difference in these bands.

In addition to exploring the patterns of feature importance in the biological sex classifiers, our focus extended to discerning how specific frequencies in different brain areas contribute to the detection of pathology, distinguishing between Abnormal and Normal conditions, and whether female and male groups utilize different features. To achieve this, we employed AGA on these classifiers. The results were further stratified into distinct groups based on sex, specifically the Female and Male groups. This gender-based categorization allowed us to investigate potential variations in frequency contributions between sexes. Additionally, we calculated the difference in the AGA results between the Female and Male groups, providing insights into the gender-specific nuances of Abnormal condition detection. This approach not only broadened our understanding of the classifiers' behaviour but also unveiled gender-specific nuances in the frequency dynamics of brain areas implicated in the pathology detection task.

Figure 3 reveals that pathological conditions distinctly influence brain activity patterns across all frequency bands, particularly in the lower range (0–12 Hz). This underscores the classifiers' heightened sensitivity to anomalies in these lower frequency bands when distinguishing between Abnormal and Normal states, emphasizing their relevance in pathology detection. Interestingly, in Fig. 3C, notable differences emerge in the features employed by the pathology classifiers in Female and Male subjects.

Hyper-parameters selection

In addition to ShallowNet, we included EEGNet[28], ShallowNet[29], Deep4Net[29], and Temporal Convolutional Network (TCN-EEG)[30] in our experiments to evaluate the model-independence of biological sex and pathology classification in EEG data. These models differ in depth, architecture, and inductive biases, yet they consistently achieved above-chance performance across datasets. This consistency supports our central claim that EEG-based classification of biological sex and pathology is a robust signal, not tied to a specific model architecture. Among these, ShallowNet was chosen as the primary model for its simplicity, interpretability, and competitive performance, which aligns well with the goals of our study.

In our hyperparameter search for the neural network models, we consider a search space for learning rate, weight decay, dropout probability, and data augmentation. The hyperparameters are explored on all four well-established models. The Table 3 displays the search space for hyperparameters.

We adapt Randaugment[31] to randomly select two transformations from a predefined pool in each epoch for data augmentation. Inspired by prior work on augmenting EEG data[32], we employ four augmentation methods, each with a probability of 0.4 during training: **SignFlip**: Randomly flips the sign of the EEG signals, simulating a change in electrode polarity. **ChannelsDropout**: Randomly drops out some channels of the EEG signals, simulating a loss of contact or electrode malfunction, with a dropout probability of 0.2. **FrequencyShift**:

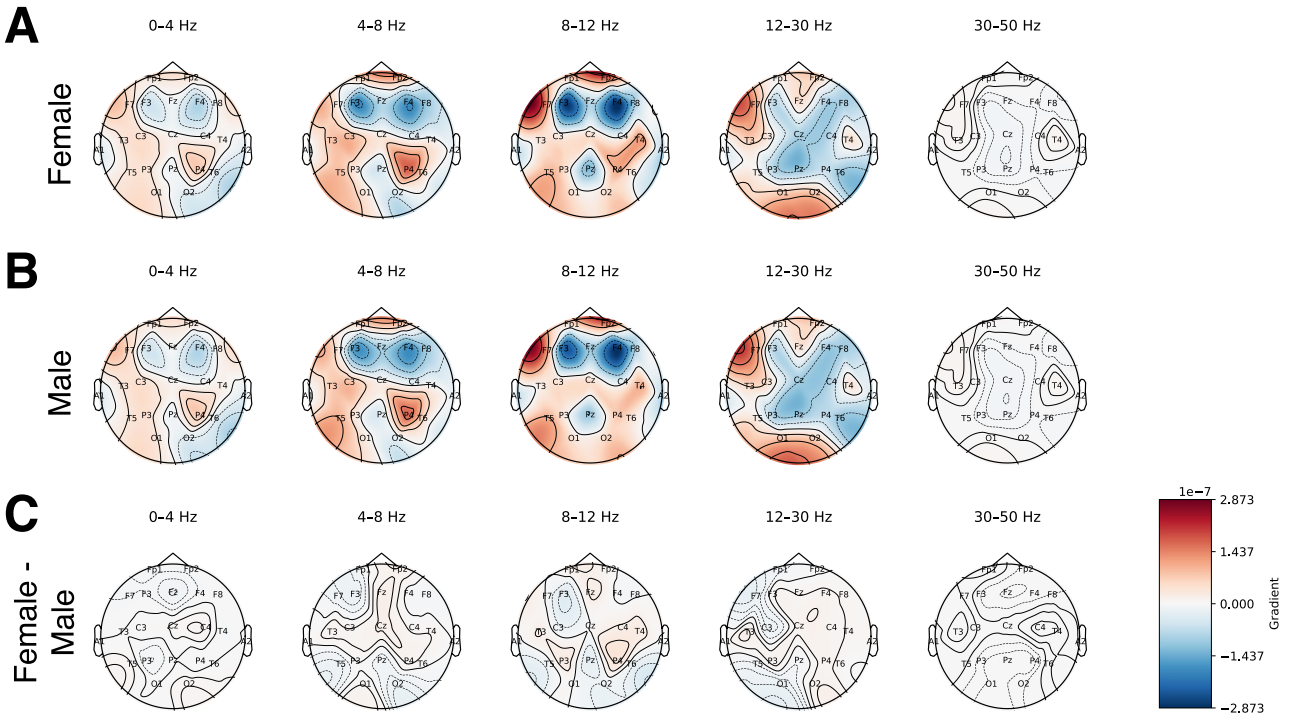

Fig. 3. AGA of different frequency bands of Pathology classifiers on NMT dataset. (**A**) Female (**B**) Male (**C**) The difference between the Female and Male classes. Red indicates a stronger relation with the female class, while blue indicates a stronger relation with the male class.

Parameter	Model	Drop prob	lr	Weight decay	Augmentation
Distribution	Categorical	Uniform	log uniform	log uniform	Categorical
Values	ShallowNet, Deep4Net EEGNet, TCN	[0–1]	[1e-6 - 1e-3]	[1e-7 - 1e-5]	[True False]
Selected HP	ShallowNet	0.5	0.000625	0.0	False

Table 3. Hyper-Parameters (HP) search space and the HPs selected for experiments in the paper.

Model	TUEG	TUAB	NMT
ShallowNet	67.55 ± 5.20	69.76 ± 5.41	60.51 ± 3.16
Deep4Net	59.57 ± 8.56	60.95 ± 9.40	54.14 ± 4.48
TCN	59.04 ± 7.17	61.20 ± 7.99	54.54 ± 4.71
EEGNet	59.68 ± 5.24	62.16 ± 4.63	54.85 ± 5.31

Table 4. Balanced Accuracy (BAcc %) across datasets and models for biological sex classification. Values are reported as mean ± standard deviation.

Randomly shifts the frequency spectrum of the EEG signals, simulating a change in the sampling rate or frequency drift, with a maximum frequency shift of 2 Hz. **SmoothTimeMask**: Randomly masks some time segments of the EEG signals with a smooth transition, simulating a temporary occlusion or signal distortion, with a mask length of 600 samples. Additionally, we include two other methods, **BandstopFilter** and **ChannelsShuffle**, with the same probability and aimed at simulating noise reduction or electrode placement changes.

Performance comparison across models. Table 4 and Fig. 4 summarize the classification performance of four neural models across three EEG datasets. ShallowNet consistently achieved the highest mean balanced accuracy (BAcc), particularly on TUAB and TUEG, while also exhibiting lower standard deviation compared to the other models. This stable and superior performance supports our choice of ShallowNet as the primary model for the main experiments. Although Deep4Net, TCN, and EEGNet achieved above-chance results, their

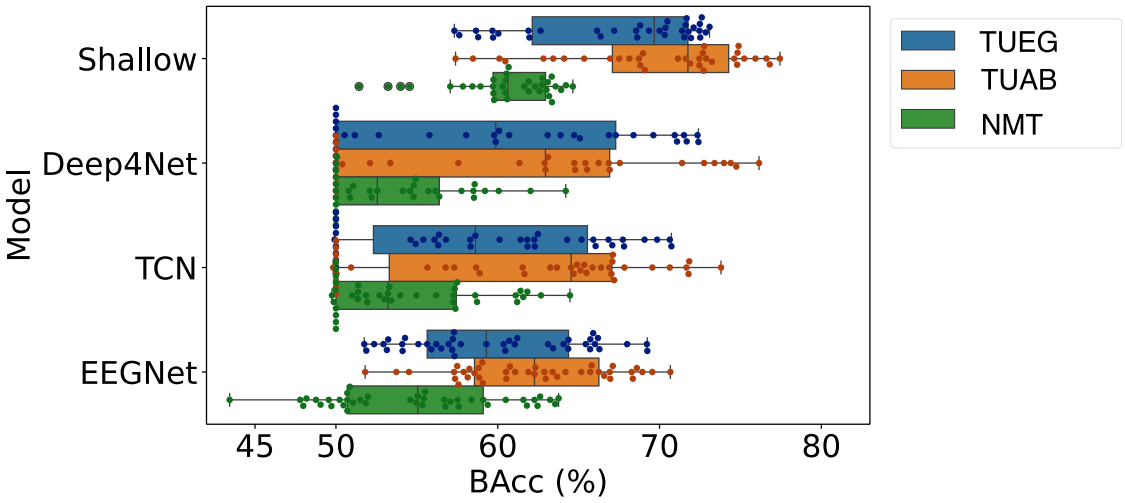

Fig. 4. Hyperparameter search results. BAcc for all models.

comparatively lower mean scores and higher variability suggest less consistent performance across datasets. These findings confirm that sex-related signals in EEG can be robustly extracted using diverse neural network architectures, reinforcing the generalizability of our results. The specific hyperparameter settings used for training ShallowNet are detailed in Table 3. Notably, our experiments showed no significant improvement when using data augmentation; therefore, we conducted all main experiments using ShallowNet without augmentation.

Discussion and conclusion

Historically documented biological sex differences in EEG patterns and the successful application of machine learning for automatic biological sex detection suggest that sex-related patterns can act as confounders in machine learning-based EEG assessments[16,17]. In our investigation of 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, biological sex does not function as a confounder due to an equal distribution of pathological participants in the male/female splits. However, as demonstrated in the SD section, we show that biological sex remains detectable. Consequently, acknowledging biological 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 biological sex differences is that, despite observable behavioural distinctions between men and women, differences in brain structure and function are minimal and inconsistent when controlling for individual brain size and accounting for inadequate sample sizes[13]. 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.

Our experiments on SD, particularly ShallowNet, demonstrate superior performance in biological sex detection on the TUAB dataset in both in-distribution and zero-shot scenarios. Notably, ShallowNet outperforms previous models by a margin in the zero-shot scenario, showcasing its robust generalization capabilities. The substantial improvement can be attributed to utilizing the TUEG dataset, which offers several advantages: a considerably larger sample size, seven times more unique participants, and a data distribution closely aligned with TUAB. This enrichment in the training data contributes to a noteworthy enhancement in ShallowNet's performance on the TUAB dataset, achieving an improvement. Importantly, our focus is not solely on surpassing previous benchmarks; we believe a thorough exploration of architecture and hyperparameter settings could further elevate the model's performance, providing avenues for future refinement and optimization of the biological sex detection task. However, performance is weaker than that of pathology detection[26].

Furthermore, in a related experiment[33], the investigation provided additional insights into the relationship between biological sex prediction and participants' diagnoses. Noteworthy findings revealed significant distinctions in only one condition, namely Parkinson's Disease. However, the authors acknowledged that this condition comprises one of the most minor groups, and its influence on the overall accuracy of the models is likely inconsequential. One potential explanation for the absence of differences could be the variations in prevalence between male and female subtypes within disorders. Specifically, Alzheimer's disease dementia tends to be more prevalent in females, while males face a higher risk of developing vascular dementia[2].

Frequency bands are widely recognized as critical features in quantitative EEG analysis. Despite their prominence, the significance of these features in biological sex detection remains unclear[34]. Some studies assert that brain rhythms exhibit sex-specific patterns[18,35–37], while others argue that none of the traditional frequency bands play a particularly crucial role in biological sex detection[17]. Several studies[18,37] demonstrated that a primary distinguishing characteristic lies in the beta activity and its spatial distribution; our results show a similar pattern. However, our results show that the theta and alpha bands also contribute to biological sex classification. Moreover, features in the beta band are similar but are different in the theta and alpha bands

between Abnormal and Normal groups. This might indicate that the beta band is more robust. In pathology detection, Fig. 3 in comparison to Fig. 2 highlights the importance of low-frequency bands over higher bands in pathology detection, with a notable difference between Female and Male groups in 3C, the distinct visual representation emphasizes the need to consider biological sex in pathology detection applications. These findings underscore gender-specific nuances in feature utilization, highlighting the importance of a nuanced understanding of classifier dynamics within different demographic groups and enhancing our comprehension of pathology classification intricacies.

By comparing Fig. 2 and Supplementary Figure 1, it becomes evident that our model exhibits consistency in the utilization of specific channels within frequency bands for biological sex classification across the TUAB and NMT datasets. The similarities in the AGA patterns for biological sex classification in the two datasets (TUAB and NMT) suggest that certain frequency-related features play a central role in the model's decision-making process for sex-related distinctions. Conversely, notable differences emerge when examining pathology classification in Fig. 3 and Supplementary Figure 2. The network appears to leverage distinct features and channels within frequency bands when discerning pathology, highlighting dataset-specific nuances in the neural network's learning patterns. This observation underscores the model's adaptability, indicating its ability to tailor feature utilization based on each dataset's specific characteristics and complexities. Such insights from the AGA visualizations contribute to a deeper understanding of the neural network's behaviour and its capacity to extract relevant information for biological sex and pathology classification across diverse datasets.

Brain connectivity and topography research has yielded diverse and sometimes conflicting perspectives, providing a rich field for future investigations. A seminal study by[38], involving 949 youths, revealed distinct patterns in supratentorial connections between males and females. Their findings suggest 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. Specifically, they observed stronger intrahemispheric connections in males and stronger interhemispheric connections in females. Advancements in neuroimaging techniques have further expanded our understanding of brain topography[17]. demonstrated the significance of EEG topographies in biological sex detection, revealing that even with disrupted waveforms, biological sex could be accurately identified. This finding highlights the potential of EEG topographies as a robust biomarker for sex-specific brain characteristics. However, the field is not without controversy and methodological challenges. As such[16], observed that the incorporation of multivariate classification models did not consistently improve performance in brain signal analysis. This finding underscores the need for careful consideration of analytical methods in brain connectivity research. Moreover[13], presents a critical perspective on the longstanding belief in sex-based brain lateralization. Despite decades of research examining biological sex effects on lateralized brain function, they argue that no substantial evidence supports the widely held belief that male brains are significantly more lateralized than female brains. This challenge to established notions highlights the importance of rigorous, unbiased research. 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 potential direction for future studies could be to examine which connections trained neural networks prefer when classifying brain signals, potentially revealing new insights into the functional significance of specific connectivity patterns.

In conclusion, our comprehensive training and evaluation process demonstrated the model's efficacy in classifying biological sex from EEG signals. We rigorously assessed its generalization to unseen data, analyzed detectability and transferability across varied conditions, and explored its utility for pathology detection in a heterogeneous and imbalanced dataset. These analyses provide a nuanced understanding of the model's strengths and its potential clinical applications. Our findings contribute to the broader effort to characterize brain connectivity and topography through neural signal processing.

Future work should focus on resolving remaining inconsistencies in the literature, refining methodological approaches, and leveraging emerging technologies to further explore the complex relationships between brain structure, function, and individual differences. As large-scale and general-purpose foundation models for neuroimaging data continue to emerge[24], it is crucial to address dataset imbalance, particularly in sensitive applications such as medical classification. The development and deployment of such models require careful attention to demographic and clinical representation to avoid encoding or amplifying biases. This consideration is especially important when subtle physiological differences, such as those related to biological sex or pathology, may influence clinical decisions or scientific conclusions. A particularly promising direction involves investigating the preferential connections utilized by the trained neural network during EEG classification. This approach could reveal the most salient features underlying sex-based differences in brain activity and enhance our understanding of the neurophysiological mechanisms that distinguish individuals. Such insights may ultimately support the development of more personalized and targeted applications in neuroscience and clinical practice.

Materials and methods

This section provides a comprehensive overview of our experimental methodology and analytical approaches. We first describe the three large-scale EEG datasets utilized in this study, along with our preprocessing pipeline for signal preparation and artifact removal. Following this, we detail our training and evaluation procedures, including model selection and performance metrics. An overview of the complete experimental pipeline is provided in Fig. 5, offering a high-level summary of the data, models, tasks, and evaluation procedures described in detail below. We then outline our specific experimental design for investigating biological sex detectability and pathology detection across different population subgroups. Finally, we present our visualization techniques for interpreting neural network decision-making processes through gradient-based analysis methods.

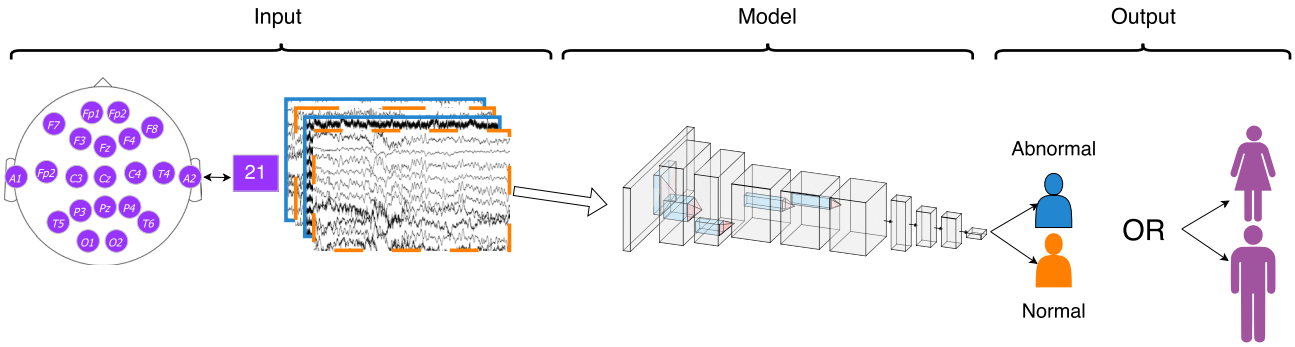

Fig. 5. Overview of the experimental pipeline. The figure summarizes the core components of our methodology, including the EEG datasets used, model architecture, and the two primary tasks: biological sex classification and pathology detection.

Datasets and preprocessing techniques

We analyzed three publicly available EEG datasets, each characterized by distinct sample sizes and conditions, to explore the impact of biological sex and pathology on EEG signals. The utilized datasets are as follows:

TUEG (Temple University Hospital EEG Corpus): This extensive open-source EEG data corpus encompasses over 69, 000 recordings from 14, 987 subjects, with a cumulative duration of 27, 062 hours. The recordings are de-identified and annotated with clinical information, including age and biological sex[39,40]. Table 1 provides an overview of these datasets, presenting comprehensive information about each dataset.

TUAB (Temple University Hospital Abnormal EEG Corpus): A subset of the TUEG corpus, this dataset consists of 1, 985 recordings from 1, 652 subjects, totalling 453 hours. Expert neurologists have labelled the recordings as normal or abnormal, and demographic information such as biological sex and age is provided[39–41]. It is important to note the overlap of participants between the TUAB and TUEG datasets; therefore, we do not present cross-dataset results between the two.

NMT (NUST-MH-TUKL EEG): Comprising 2, 417 recordings from Normal and Abnormal subjects, this dataset spans a total duration of 625 hours. Expert neurologists have labelled the recordings as either normal or abnormal (The term "Normal/Abnormal" is originally used by the datasets to describe EEG recordings with Abnormal features. This term does not imply judgment but reflects the condition of the EEG signal). Demographic information, including biological sex and age, is also included[42].

Data Preprocessing: Patient biological sex or pathology information, encoded as 0 or 1, served as our neural network target. We focused on biological sex rather than gender due to the dataset's clinical origin, assuming records reflected assigned birth sex. Preprocessing steps included selecting 21 common channels across datasets, cropping between 1–20 minutes, resampling to 100 Hz, Artifact Subspace Reconstruction (ASR)[43] for artifact removal, re-referencing to average and z-scoring EEG signals to each channel's statistics. *Data Splitting:* Predefined test sets were used to report model accuracy, with 15% of training splits reserved for model selection.

Training and evaluation

We used ShallowNet[29] as our model for all experiments. Given its relative computational simplicity and fewer non-linearities, ShallowNet emerges as a strategic choice for our experiments (see Sect. "Hyper-parameters selection" for more details). ShallowNet's architecture, comprising only one convolutional layer followed by a fully connected layer, mitigates the computational cost associated with deeper networks, making it particularly appealing for our study. The streamlined design of ShallowNet also offers an advantage in model explainability. The simplicity of the architecture implies that explainability methods are likely to provide more precise insights into what the model has captured, especially concerning biological sex differences. In our experiments, we implemented ShallowNet using BrainDecode[44] and trained it with the AdamW optimizer, utilizing a learning rate of 0.000625, weight decay of 0, drop probability of 0.5, and a batch size of 64. The training process comprised 35 epochs, and model selection was based on performance using the BAcc metric on the validation set. The BAcc[45], calculated as the arithmetic mean of sensitivity and specificity, offers a more reliable performance assessment, particularly in the context of imbalanced data similar to our datasets. It is strongly advised as a robust evaluation metric for applications involving brain decoding with imbalanced data[46]. Notably, it is equivalent to standard accuracy in scenarios with balanced data.

Model

The ShallowNet architecture consists of temporal and spatial convolutional layers, followed by a squaring non-linearity and mean pooling. Given an EEG input $X \in \mathbb{R}^{C \times T}$, where C is the number of channels and T is the number of time points, the model applies:

- A temporal convolution: $X' = \mathrm{Conv}_{\mathrm{temp}}(X)$
- A spatial convolution: $X'' = \mathrm{Conv}_{\mathrm{spat}}(X')$
- Element-wise squaring: $X''' = (X'')^2$
- Mean pooling over time and flattening

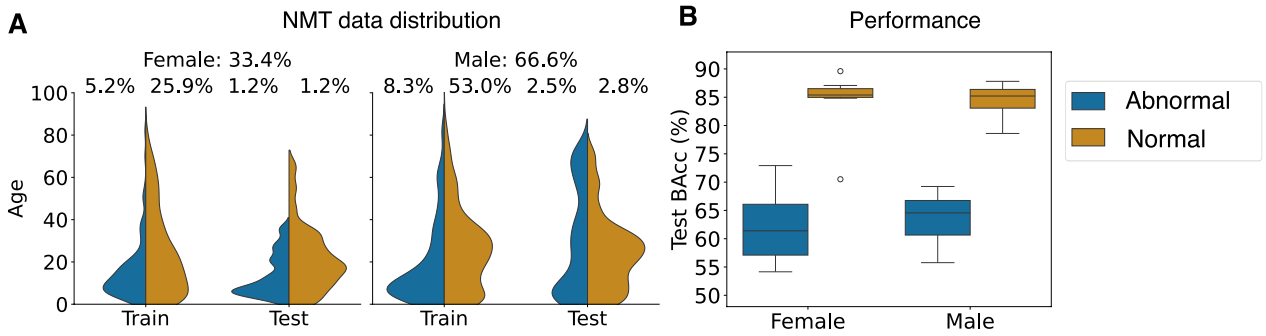

Fig. 6. Effect of biological sex imbalances on pathology detection in the NMT dataset: (**A**) Distribution of male and female samples in the NMT dataset, with the number of male samples being twice as high as that of females. (**B**) Performance (accuracies) of subgroups. The discrepancy in sample numbers does not impact pathology detection.

- A fully connected layer with softmax activation

For classification, the model uses the categorical cross-entropy loss:

$$\mathcal{L} = -\sum_{i=1}^{K} y_i \log(\hat{y}_i)$$

where y_i is the ground-truth label and $\hat{y}_i$ is the predicted probability for class *i*.

Experiment design

The training and evaluation of the model were conducted with the primary objective of classifying biological sex and pathology 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 *biological 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 examine 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.

To ensure robust training and evaluation of our models, we randomly sampled 2000 participants from the TUAB and NMT datasets under various conditions. This approach involved oversampling in scenarios with fewer than 2000 participants, such as the NMT Abnormal subset, and undersampling when the dataset exceeded 2000 participants, as observed in the case of the entire participant pool. This balanced sampling strategy aimed to mitigate potential biases from uneven dataset distributions. We selected 14,000 unique participants from the TUEG dataset for training, ensuring a diverse and representative training set for the neural network.

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 (see Fig. 6A). Our analysis focused on different subgroups within the dataset, including Male Normal, Female Normal, Male Abnormal, and Female Abnormal participants. We aimed to explore any potential associations between SD and pathology detection by examining the model's performance on these subgroups. We ran each experiment with ten random seeds for all experiments. All error bars show the standard error of the metrics of the ten seeds. We used JASP[47] for statistical analysis to conduct t-tests.

EEG signal visualization

We employed a visualization technique to interpret the deep neural network's decision-making process. Specifically, we utilized AGA to gain insights into the network's behaviour. To explain the internal mechanisms of the deep neural network, we conducted an AGA. This process involved computing the gradients of ten distinct models while classifying the target evaluation set derived from the TUAB dataset, focusing on the frequency domain. We examined the traditional EEG frequency bands: delta (0–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz), and gamma (30–50 Hz).

The obtained gradients were systematically grouped to discern how changes in specific frequency bands influenced feature importance. This grouping facilitated a nuanced understanding of the neural network's sensitivity to different aspects of the input signal. We plotted the resulting gradient values on a head diagram to present the spatial distribution of informative brain regions for the network's predictions. Each region was colour-coded or annotated based on its importance, visually representing the neural network's focus during the brain decoding task. The resulting patterns on the head diagram served as a key tool for interpreting the neural

network's behaviour. Regions with larger absolute gradient values indicated higher sensitivity to changes in the input signal concerning the output. This interpretation sheds light on areas crucial for the network's decision-making process and provides valuable insights into the brain decoding task.

Data availability

All data used in this study are publicly available and can be accessed through the following repositories: The Temple University Hospital EEG (TUEG and TUAB) datasets: https://isip.piconepress.com/projects/nedc/html/tuh_eeg. The NUST-MH-TUKL EEG (NMT) dataset: https://dll.seecs.nust.edu.pk/downloads/. These datasets are openly accessible, and we encourage readers to refer to the respective publications for detailed information regarding their structure and intended use. The source code used for all experiments, model training, and evaluation is available at our GitHub repository: https://github.com/JavadBayazi/AURA-EEG.

Received: 5 January 2025; Accepted: 26 November 2025
Published online: 10 December 2025

References

1. Lee, S. K. Sex as an important biological variable in biomedical research. *BMB reports* **51**, 167 (2018).
2. Podcasy, J. L. & Epperson, C. N. Considering sex and gender in alzheimer disease and other dementias. *Dialogues in clinical neuroscience* **18**, 437–446 (2016).
3. Williams, O. O., Coppolino, M. & Perreault, M. L. Sex differences in neuronal systems function and behaviour: beyond a single diagnosis in autism spectrum disorders. *Transl. Psychiatry* **11**, 625 (2021).
4. Christiansen, D. M., McCarthy, M. M. & Seeman, M. V. Understanding the influences of sex and gender differences in mental disorders. *Front. Psychiatry* **13**, 984195 (2022).
5. Sejnowski, T. J., Churchland, P. S. & Movshon, J. A. Putting big data to good use in neuroscience. *Nat. neuroscience* **17**, 1440–1441 (2014).
6. Baron-Cohen, S., Knickmeyer, R. C. & Belmonte, M. K. Sex differences in the brain: implications for explaining autism. *Sci.* **310**, 819–823 (2005).
7. Drapeau, A., Lesage, A. & Boyer, R. Is the statistical association between sex and the use of services for mental health reasons confounded or modified by social anchorage?. *The Can. J. Psychiatry* **50**, 599–604 (2005).
8. Mason, L. et al. Stratifying the autistic phenotype using electrophysiological indices of social perception. *Sci. Transl. Medicine* **14**, eabf8987 (2022).
9. Dorfschmidt, L. et al. Sexually divergent development of depression-related brain networks during healthy human adolescence. *Sci. Adv.* **8**, eabm7825 (2022).
10. Chekroud, A. M., Ward, E. J., Rosenberg, M. D. & Holmes, A. J. Patterns in the human brain mosaic discriminate males from females. *Proc. Natl. Acad. Sci.* **113**, E1968–E1968 (2016).
11. Sepehrband, F. et al. Neuroanatomical morphometric characterization of sex differences in youth using statistical learning. *Neuroimage* **172**, 217–227 (2018).
12. Sanchis-Segura, C., Ibañez-Gual, M. V., Aguirre, N., Cruz-Gómez, Á. J. & Forn, C. Effects of different intracranial volume correction methods on univariate sex differences in grey matter volume and multivariate sex prediction. *Sci. Reports* **10**, 12953 (2020).
13. Eliot, L., Ahmed, A., Khan, H. & Patel, J. Dump the "dimorphism": Comprehensive synthesis of human brain studies reveals few male-female differences beyond size. *Neurosci. & Biobehav. Rev.* **125**, 667–697 (2021).
14. Sanchis-Segura, C., Aguirre, N., Cruz-Gómez, Á. J., Félix, S. & Forn, C. Beyond, "sex prediction": Estimating and interpreting multivariate sex differences and similarities in the brain. *NeuroImage* **257**, 119343 (2022).
15. Ryali, S., Zhang, Y., de Los Angeles, C., Supekar, K. & Menon, V. Deep learning models reveal replicable, generalizable, and behaviorally relevant sex differences in human functional brain organization. *Proc. Natl. Acad. Sci.* **121**, e2310012121 (2024).
16. Bučková, B., Brunovskỳ, M., Bareš, M. & Hlinka, J. Predicting sex from eeg: validity and generalizability of deep-learning-based interpretable classifier. *Front. Neurosci.* **14**, 589303 (2020).
17. Jochmann, T. et al. Sex-related patterns in the electroencephalogram and their relevance in machine learning classifiers. *Hum. Brain Mapp.* **44**, 4848–4858 (2023).
18. Van Putten, M. J., Olbrich, S. & Arns, M. Predicting sex from brain rhythms with deep learning. *Sci. reports* **8**, 3069 (2018).
19. Gulrajani, I. & Lopez-Paz, D. In search of lost domain generalization. In *International Conference on Learning Representations*, 0 (2020).
20. Lan, Z., Sourina, O., Wang, L., Scherer, R. & Müller-Putz, G. R. Domain adaptation techniques for eeg-based emotion recognition: A comparative study on two public datasets. *IEEE Transactions on Cogn. Dev. Syst.* **11**, 85–94 (2018).
21. Roy, Y. et al. Deep learning-based electroencephalography analysis: a systematic review. *J. neural engineering* **16**, 051001 (2019).
22. Banville, H., Wood, S. U., Aimone, C., Engemann, D.-A. & Gramfort, A. Robust learning from corrupted eeg with dynamic spatial filtering. *NeuroImage* **251**, 118994 (2022).
23. Wang, J. et al. Eegmamba: An eeg foundation model with mamba. *Neural Networks* 107816 (2025).
24. Bayazi, M. J. D. et al. General-purpose brain foundation models for time-series neuroimaging data. In *NeurIPS Workshop on Time Series in the Age of Large Models* (2024).
25. Buslón, N., Cortés, A., Catuara-Solarz, S., Cirillo, D. & Rementeria, M. J. Raising awareness of sex and gender bias in artificial intelligence and health. *Front. Glob. Women's Heal.* **4**, 970312 (2023).
26. Darvishi-Bayazi, M.-J. et al. Amplifying pathological detection in eeg signaling pathways through cross-dataset transfer learning. *Comput. Biol. Medicine* **169**, 107893 (2024).
27. Gemein, L. A. et al. Machine-learning-based diagnostics of eeg pathology. *NeuroImage* **220**, 117021 (2020).
28. Lawhern, V. J. et al. Eegnet: a compact convolutional neural network for eeg-based brain-computer interfaces. *J. neural engineering* **15**, 056013 (2018).
29. Schirrmeister, R. T. et al. Deep learning with convolutional neural networks for eeg decoding and visualization. *Hum. brain mapping* **38**, 5391–5420 (2017).
30. Elsken, T., Metzen, J. H. & Hutter, F. Neural architecture search: A survey. *The J. Mach. Learn. Res.* **20**, 1997–2017 (2019).
31. Cubuk, E. D., Zoph, B., Shlens, J. & Le, Q. V. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops* (2020).
32. Rommel, C., Paillard, J., Moreau, T. & Gramfort, A. Data augmentation for learning predictive models on eeg: a systematic comparison. *J. Neural Eng.* **19**, 066020 (2022).
33. Khayretdinova, M. et al. Prediction of brain sex from eeg: using large-scale heterogeneous dataset for developing a highly accurate and interpretable ml model. *NeuroImage* **285**, 120495 (2024).

34. Chapman, R. et al. Sex differences in electrical activity of the brain during sleep: a systematic review of electroencephalographic findings across the human lifespan. *BioMedical Eng. OnLine* **24**, 33 (2025).
35. Dijk, D. J., Beersma, D. G. & Bloem, G. M. Sex differences in the sleep eeg of young adults: visual scoring and spectral analysis. *Sleep* **12**, 500–507 (1989).
36. Ujma, P. P. et al. Sleep eeg functional connectivity varies with age and sex, but not general intelligence. *Neurobiol. aging* **78**, 87–97 (2019).
37. Kaushik, P., Gupta, A., Roy, P. P. & Dogra, D. P. Eeg-based age and gender prediction using deep blstm-lstm network model. *IEEE Sensors J.* **19**, 2634–2641 (2018).
38. Ingalhalikar, M. et al. Sex differences in the structural connectome of the human brain. *Proc. Natl. Acad. Sci.* **111**, 823–828 (2014).
39. Shawki, N. et al. Correction to: The temple university hospital digital pathology corpus. In *Signal Processing in Medicine and Biology: Emerging Trends in Research and Applications*, C1–C1 (Springer, 2022).
40. Obeid, I. & Picone, J. The temple university hospital eeg data corpus. *Front. neuroscience* **10**, 196 (2016).
41. López, S., Obeid, I. & Picone, J. *Automated interpretation of abnormal adult electroencephalograms*. Ph.D. thesis, Temple University (2017).
42. Khan, H. A. et al. The nmt scalp eeg dataset: an open-source annotated dataset of healthy and pathological eeg recordings for predictive modeling. *Front. neuroscience* **15**, 755817 (2022).
43. Blum, S., Jacobsen, N. S., Bleichner, M. G. & Debener, S. A riemannian modification of artifact subspace reconstruction for eeg artifact handling. *Front. human neuroscience* **13**, 141 (2019).
44. Schirrmeister, R. T. et al. Deep learning with convolutional neural networks for eeg decoding and visualization. *Hum. Brain Mapp.* https://doi.org/10.1002/hbm.23730 (2017).
45. Brodersen, K. H., Ong, C. S., Stephan, K. E. & Buhmann, J. M. The balanced accuracy and its posterior distribution. In *2010 20th international conference on pattern recognition*, 3121–3124 (IEEE, 2010).
46. Thölke, P. et al. Class imbalance should not throw you off balance: Choosing the right classifiers and performance metrics for brain decoding with imbalanced data. *NeuroImage* **277**, 120253 (2023).
47. JASP Team. JASP (Version 0.18.3)[Computer software] (2024).

Acknowledgements

We sincerely appreciate Mathilde Besson for their valuable comments, which greatly contributed to the refinement of this paper. This work was funded by the Canada CIFAR AI Chair Program and 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. Additionally, we thank Compute Canada for providing computational resources.

Author contributions

M.J. D.B. conducted the experiment(s). M.J. D.B., M.S.Gh. analyzed the results. All authors contributed to conceiving the experiment and reviewed the manuscript.

Funding

This work was funded by the Canada CIFAR AI Chair Program and 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.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-30848-y.

Correspondence and requests for materials should be addressed to M.-J.D.-B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025