Proceedings of the
10th Graz Brain-Computer Interface Conference 2026
10.3217/978-3-99161-093-9-034

Is User-Independent MI-BCI Performance Influenced by Gender Proportion in Training Users?

Simon Kojima[1], Fabien Lotte[1]

[1]Inria Centre at the University of Bordeaux, Talence, France

E-mail: simon.kojima@inria.fr

ABSTRACT: In motor-imagery brain–computer interfaces (MI-BCI) research, how user gender influences BCI classification performance and ERD/S responses is not yet fully understood. Thus, investigating the impact of gender on BCI systems is an important step toward realizing BCIs that are accessible to all users. In this study, we examined how varying the proportion of female users in the training data in cross-user MI-BCI affects BCI classification performance. The results (on 139 users from 2 data bases) revealed positive correlations (0.33–0.71) between the proportion of female users in the training data and BCI performance, and showed that including more female users in the training data was associated with performance improvements of up to 5.5%. These findings indicate that machine-learning models in MI-BCI may be influenced by gender-related biases and highlight the importance of considering gender distribution in the design of future BCI research.

INTRODUCTION

Motor Imagery (MI)-based brain-computer interfaces (BCIs) offer intuitive control and a wide range of potential applications; however, it is well known that approximately 15–30% of users fail to achieve sufficient control accuracy, a phenomenon referred to as BCI inefficiency [1]. Although various factors have been investigated, the fundamental causes of this issue remain unclear, and no definitive solution has been established. MI-BCIs operate by detecting, using machine learning, the power decreases and increases in the mu (8–13 Hz) and beta (15–30 Hz) bands, known as event-related desynchronization/synchronization (ERD/S), elicited by MI of body-part movements.

Multiple factors affecting BCI performance and ERD/S strength have been reported, particularly biological sex or gender-related factors, which have been shown to influence ERD amplitude and BCI performance. Randolph et al. reported that ERD amplitudes were larger in females [2]. Similar tendencies have been reported in [3, 4], where higher MI-BCI performance in females was also observed. The higher MI-BCI performance in females was observed in other studies [5, 6]. Furthermore, an fMRI study on post-stroke patients reported greater activation in the sensorimotor cortex in females compared with males [7]. Additionally, it has been suggested that the interaction between the genders of the experimenter and the user may influence MI-BCI performance [8]. Another study reported slightly better, though not statistically significant, MI-BCI performance in female users, and also showed that features learned by deep learning models can be used to classify user gender [9]. In contrast, [10–12] reported no significant gender differences. Thus, there is currently no clear consensus on how gender affects MI-BCI performance. Nevertheless, beyond BCI research, it is well known that various machine-learning algorithms may embed gender-related biases [13].Therefore, a careful investigation of gender effects is essential for realizing BCIs that are accessible to all users.

Previous studies mainly compared male vs. female performance but did not examine how the composition of the training data affects performance. In this study, we investigated an aspect that has not been sufficiently examined so far, namely how imbalances in the number of male and female samples in the training data - for cross-user MI-BCIs - affect the BCI performance of male and female users. In addition, by quantifying the variability of ERD/S responses under these conditions, we also examined potential gender differences in the ERD/S variability to interpret the performance difference.

MATERIALS AND METHODS

Dataset Description: We used two open datasets, (A) `Dreyer2023` [14] and (B) `Lee2019` [15], which were accessed via the Python package MOABB [16]. For both datasets, only EEG channels were used, and participant sex (male/female) was obtained from the dataset metadata and is referred to as gender in the remainder of the paper.

(A) `Dreyer2023`: It contains EEG recordings from 87 users performing left- and right-hand motor imagery (MI) tasks [14]. For each user, a single-session EEG recording consisting of six runs of MI tasks was provided. The first and second runs constituted the acquisition phase, during which data were collected for classifier training, whereas runs 3 to 6 comprised the online phase, during which users received real-time feedback from the BCI. In each run, 20 trials were performed per class, i.e., left- and right-hand MI. Among the 87 users, two (4 and 59) were excluded from further analysis due to excessive noise in the recordings (4) and missing runs (59), leaving 85 users

 10.3217/978-3-99161-093-9-034

(41 females and 44 males) for analysis. All 27 available EEG channels were used (cf. [14]). The original sampling frequency was 512 Hz.

(B) Lee2019: It contains EEG recordings from 54 users (23 females and 31 males) performing left- and right-hand motor imagery (MI) tasks [15]. For each user, two-session EEG recordings consisting of two runs (a training run and a test run) of MI tasks were provided. As the test run does not include true class labels (see the original paper for details [15]), only the training run was used in this study. Each run consisted of 100 trials (50 trials per class). All 62 available EEG channels were used (cf. [15]). The original sampling frequency was 1000 Hz.

Classification: In this study, the cross-user classification task was evaluated for each test user across varying proportions of female users in the training data.

For each dataset, one user was selected as the test user U_t, and the training data were constructed by selecting N_t users from the remaining users such that the proportion of female users in the training data was $R_f \in \{0.0, 0.25, 0.50, 0.75, 1.0\}$. Here, N_t was set to 40 for Dreyer2023 and 20 for Lee2019. The training and test data were generated as follows.

For each condition defined by the test user U_t and the female ratio R_f, the following steps were performed:

1. From the users excluding U_t, randomly select N_t users as the proportion of female users equals R_f.
2. Apply preprocessing (described later) to the data of each selected user and compute the covariance matrix for each trial. After applying Ledoit-Wolf regularization [17] to the covariance matrices, Riemannian Procrustes Analysis (RPA) [18] was applied to reduce between-user variability. In this study, only the recentering and rescaling steps of RPA were used, normalizing the data such that the geometric mean became the identity matrix and the dispersion around it was equal to 1. The rotation step of RPA was not applied.
3. Concatenate the data from all selected users to form the training dataset.
4. Apply the same preprocessing to the data of the test user U_t, compute the covariance matrices, and apply RPA. To simulate a realistic BCI usage scenario, RPA was applied sequentially; i.e., for the n-th trial, RPA was estimated and applied using only the data from the first n trials.
5. Train the classifier using the training data and classify the data of the test user. As the classifier, Riemannian Tangent Space + Logistic Regression (TSLR) [19] was used. For Logistic Regression, the lbfgs solver with an L2 penalty ($C = 1.0$) and $max\_iter = 100$ was employed.
6. The above procedure was repeated three times per test user, and the results were averaged.

Each user in the dataset was selected once as the test user, and the above procedure was performed for all conditions of R_f (the proportion of female users in the training data) for each test user.

The data preprocessing was performed as follows. First, a fourth-order Butterworth filter was applied bidirectionally in the 8–30 Hz band, and the MI task interval was extracted. The extraction interval was 0.5–4.5 s for Dreyer2023 and 0.5–3.5 s for Lee2019, post MI instruction cue. The data were then resampled to 128 Hz. Since Lee2019 contains data from two sessions, preprocessing and RPA were applied independently to each session, after which the two sessions were concatenated to form the data for each user. Classification performance was evaluated using accuracy. No artifact removal was applied to either the classification or the variability calculation.

Variability: For each test user, we examined (i) how variable the training data were with respect to that user (test-referenced variability) and (ii) how much variability existed within each gender group (gender-wise variability). Variability was evaluated for both the temporal waveform and the spatial distribution. To assess variability, we used the class-wise mean ERD/S pattern of each user. In this study, variability was uniformly defined as the average distance between each user's ERD/S pattern and a reference pattern M. The definition of the reference pattern M was adapted depending on the analysis setting. Specifically, for test-referenced variability, the ERD/S pattern of the test user served as the reference pattern, whereas for gender-wise variability, the representative pattern for each gender group was used.

First, preprocessing was applied to each user's data. A fourth-order Butterworth filter in the 1–45 Hz band was applied bidirectionally, and time–frequency representations were estimated from 1 to 45 Hz with a resolution of 1 Hz using the the Discrete Prolate Spheroidal Sequences (DPSS) multitaper method [20], followed by decimation by a factor of 2. ERD/S was then computed using the interval -2.0–0.0 s as the baseline. Subsequently, ERD/S was averaged across trials for each class and user, and the resulting class-wise mean ERD/S pattern was used as the representative data for each user.

Test-referenced variability: First, we evaluated how much the ERD/S patterns of the N_t training users selected for a given test user U_t varied around U_t. This metric quantifies how varied the training users' ERD/S patterns are from the test user's pattern.

For temporal variability, the corresponding channel for each class (C4 for left-hand imagery and C3 for right-hand imagery) was extracted and their ERD/S averaged over the 8–30 Hz band. This resulted in a temporal vector $x_t \in \mathbb{R}^T$ for each user, where T denotes the number of time samples.

The temporal variability v_t was quantified using DTW (Dynamic Time Warping), a distance measure based on nonlinear alignment between two time series [21]. It was defined using the reference pattern M, given by the temporal waveform of the test user ($M = x_t^{c,U_t}$), as follows:

$$v_t = \frac{1}{N_c}\frac{1}{N_t}\sum_{c=1}^{N_c}\sum_{i=1}^{N_t} \mathrm{DTW}\left(x_t^{c,i}, M\right), \quad M = x_t^{c,U_t} \tag{1}$$

10.3217/978-3-99161-093-9-034

Here, N_c denotes the number of classes, c denotes the class (left-hand and right-hand imagery), and i indexes the training users.

Next, spatial variability was evaluated. All EEG channels were used, and ERD/S data were averaged over the 8–30 Hz band and subsequently over time. This resulted in a spatial vector $x_s \in \mathbb{R}^C$ for each user, where C is the number of EEG channels.

The spatial variability v_s was defined using the reference pattern M as the spatial pattern of the test user, $M = x_s^{c,U_t}$, as follows:

$$v_s = \frac{1}{N_c}\frac{1}{N_t}\sum_{c=1}^{N_c}\sum_{i=1}^{N_t} \arccos\left(\frac{x_s^{c,i}\cdot M}{\|x_s^{c,i}\|\,\|M\|}\right), \quad M = x_s^{c,U_t} \tag{2}$$

Here, $d = \arccos((w_1 \cdot w_2)/(\|w_1\|\|w_2\|))$ compute the angular distance between two vectors (w_1 and w_2).

These two metrics represent the variability around the test user U_t in terms of the temporal waveform and spatial distribution, respectively.

Gender-wise variability: Next, for each dataset, we evaluated the variability within the male group, within the female group, and within the entire user population. Here, the reference pattern M was defined as the representative ERD/S pattern of each group.

For the temporal waveform, the reference pattern M in Eq. (1) was defined as the Euclidean mean over the set of users belonging to the group $U \in \{U_{\text{male}}, U_{\text{female}}, U_{\text{all}}\}$:

$$M = \frac{1}{N_U}\sum_{i\in U} x_t^{c,U} \tag{3}$$

where N_U is the number of users in the given group U.

For the spatial distribution, the reference pattern M in Eq. (2) was defined as the Fréchet mean on the hypersphere for the set of spatial vectors:

$$M = \underset{\|m\|_2=1}{\operatorname{argmin}} \sum_{i\in U} \arccos\left(\frac{x_s^{c,i}\cdot m}{\|x_s^{c,i}\|\,\|m\|}\right) \tag{4}$$

Statistical Analysis: In the correlation analysis, we evaluated the relationships between the proportion of female users in the training data, R_f, and each metric: machine-learning classification accuracy and test-referenced variability metrics. Specifically, for each test user, repeated-measures data obtained under different R_f conditions were used to analyze the correlations between R_f and (i) classification accuracy, (ii) temporal waveform variability, and (iii) spatial distribution variability, using repeated measures correlation. In this study, a total of 12 correlation analyses were conducted (2 datasets $\times$ 2 genders $\times$ 3 metrics: accuracy, temporal variability, and spatial variability). Therefore, multiple comparison correction using the Benjamini–Hochberg (BH) procedure was applied to the resulting p-values, with a significance level of 0.05.

In addition, to examine whether each metric changed significantly with variations in the female ratio R_f, the following analysis was conducted. First, for the total of 12 conditions (2 datasets $\times$ 2 genders $\times$ 3 metrics), repeated-measures data for each test user were analyzed using a repeated-measures ANOVA with the female ratio R_f as a within-subject factor. This allowed us to test whether there was a significant main effect of R_f on each metric. The resulting 12 p-values were corrected for multiple comparisons using the Benjamini–Hochberg (BH) procedure, and statistical significance was determined with a significance level of $\alpha = 0.05$.

For the conditions in which a significant main effect was observed in the ANOVA, post hoc comparisons were performed for all pairwise combinations of the five levels of R_f (10 pairs in total) using Wilcoxon signed-rank tests. The p-values obtained from these comparisons (10 values in total) were also corrected for multiple comparisons using the BH procedure.

Table 1: The results of the post hoc comparisons using the Wilcoxon signed-rank test are presented. A and B denote the levels of the female ratio R_f that were compared. p denotes the p-values corrected using the Benjamini–Hochberg procedure. With a significance level of 0.05, results showing significant differences are indicated by bold p-values. G., TempVar and SpatVar denote gender, temporal variability and spatial variability, respectively.

Dataset	Modality	G.	A	B	p
`Dreyer2023`	Accuracy	F	0.0	0.5	7.7×10^{-2}
			0.5	1.0	8.7×10^{-2}
			0.0	1.0	5.5×10^{-2}
	TempVar	F	0.0	0.5	$\mathbf{2.2 \times 10^{-4}}$
			0.5	1.0	$\mathbf{6.9 \times 10^{-3}}$
			0.0	1.0	$\mathbf{4.5 \times 10^{-11}}$
		M	0.0	0.5	$\mathbf{7.8 \times 10^{-6}}$
			0.5	1.0	$\mathbf{6.7 \times 10^{-3}}$
			0.0	1.0	$\mathbf{2.3 \times 10^{-12}}$
	SpatVar	F	0.0	0.5	$\mathbf{1.9 \times 10^{-6}}$
			0.5	1.0	$\mathbf{8.4 \times 10^{-8}}$
			0.0	1.0	$\mathbf{1.1 \times 10^{-7}}$
		M	0.0	0.5	$\mathbf{1.3 \times 10^{-4}}$
			0.5	1.0	$\mathbf{4.9 \times 10^{-5}}$
			0.0	1.0	$\mathbf{4.1 \times 10^{-5}}$
`Lee2019`	Accuracy	F	0.0	0.5	$\mathbf{8.3 \times 10^{-3}}$
			0.5	1.0	$\mathbf{4.1 \times 10^{-4}}$
			0.0	1.0	$\mathbf{6.0 \times 10^{-5}}$
		M	0.0	0.5	$\mathbf{2.7 \times 10^{-3}}$
			0.5	1.0	$\mathbf{7.7 \times 10^{-3}}$
			0.0	1.0	$\mathbf{7.8 \times 10^{-4}}$

RESULTS

Fig. 1 shows bar plots of the results for classification accuracy and each test-referenced variability metric across different proportions of female users in the training data R_f, separately for male and female test users. The R values in the figure denote the correlation coefficients obtained by repeated measures correlation.

For `Dreyer2023`, a positive correlation between R_f

 10.3217/978-3-99161-093-9-034

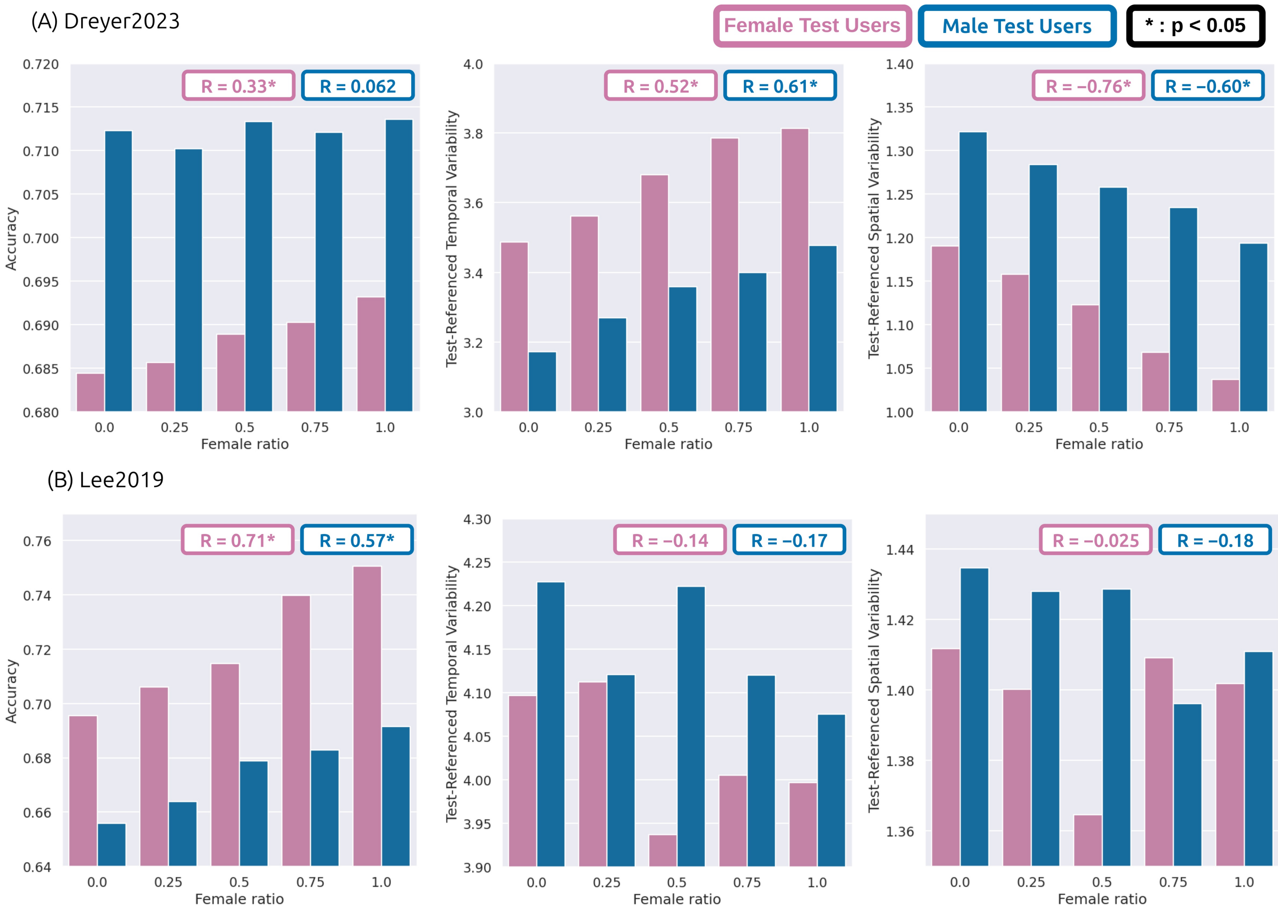

Figure 1: Classification accuracy as well as test-referenced temporal and spatial variability are shown for different values of the female ratio R_f in the training data. The correlation coefficients R obtained by repeated measures correlation are also indicated in the figure. An asterisk ($*$) attached to a correlation coefficient denotes that the correlation was significant at $\alpha = 0.05$ after applying the Benjamini–Hochberg correction to the obtained p-values. Note that the scale of the y-axis differs across the plots.

and classification accuracy was observed for female test users. For temporal variability, moderate positive correlations between R_f and variability were observed for both male and female test users. In contrast, for spatial variability, negative correlations between R_f and variability were observed for both genders.

For `Lee2019`, regarding classification accuracy, positive correlations between R_f and accuracy were observed for both male and female test users. On the other hand, for both temporal variability and spatial variability, no significant correlations were found for either gender.

A repeated-measures ANOVA was conducted to examine the main effects of the female ratio R_f on each metric. For `Dreyer2023`, significant main effects were observed for classification accuracy (female test users; $p = 1.5 \times 10^{-2}$), temporal variability (female; $p = 6.0 \times 10^{-11}$, male; $p = 1.2 \times 10^{-7}$), and spatial variability (female; $p = 1.2 \times 10^{-11}$, male; $p = 2.7 \times 10^{-6}$). In contrast, for `Lee2019`, significant main effects were observed only for classification accuracy (female; $p = 2.4 \times 10^{-8}$, male; $p = 3.1 \times 10^{-6}$).

In the post hoc comparisons, among the five levels of R_f, the results for the three pairs 0.0 vs. 0.5, 0.5 vs. 1.0, and 0.0 vs. 1.0, out of ten pairs, are reported in Table 1. Significant differences were found in all conditions except for the classification accuracy in `Dreyer2023` when female users were used as the test users. Additionally, significant differences were found in 0.75 vs. 1.0 in `Lee2019` regardless the gender of the test user (female; 4.8×10^{-2} and male; 4.3×10^{-2}).

In `Lee2019`, the highest classification accuracy was obtained when the model was trained using only female data ($R_f = 1.0$), regardless of the gender of the test user. Even when male users were used as the test users, the highest accuracy was achieved when training was performed using only female data. When the female user was the test user, the accuracy was improved by 5.5% when $R_f = 1.0$ against $R_f = 0.0$.

Fig. 2 shows the gender-wise variability for each dataset, and the distances between the centroids of the female and male groups. In `Dreyer2023`, males exhibited lower variability in temporal variability, whereas females exhibited lower variability in spatial variability. In contrast, in `Lee2019`, females tended to show lower variability for both temporal and spatial metrics.

DISCUSSION

10.3217/978-3-99161-093-9-034

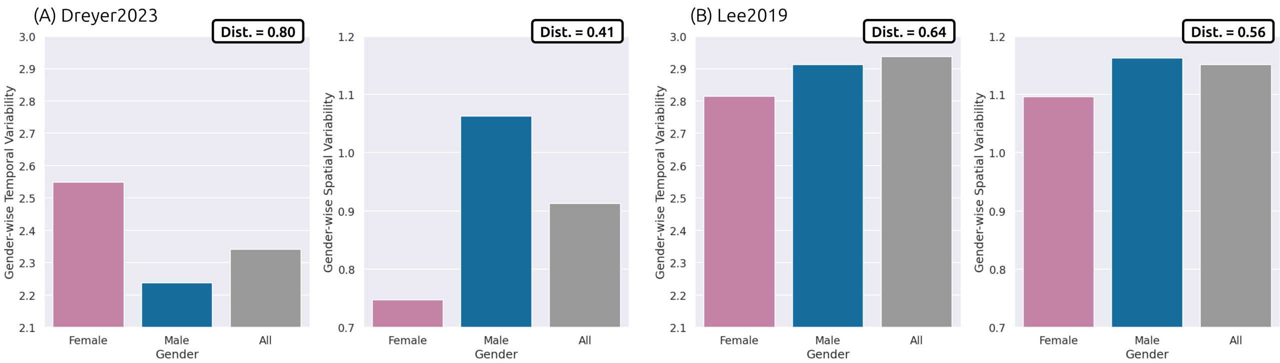

Figure 2: Results of the gender-wise variability analysis. Temporal and spatial variability are shown for the male group, the female group, and all users. The brown box in the upper-right corner of each subfigure shows the distance between the centroids of the female and male groups (DTW for temporal variability and angular distance for spatial variability). The y-axis scales are identical across datasets for each metric.

First, this study revealed that, in cross-user MI-BCI classification, bias may arise depending on the gender ratio of users included in the training data. Both the correlation analysis between the female ratio R_f and BCI performance (Fig. 1) and the differences in BCI performance across R_f conditions within each setting (Table 1) showed significant correlations (3 out of 4 conditions) and significant differences (2 out of 4 conditions). These findings provide important evidence that the gender distribution can influence performance in the design and experimentation of cross-user MI-BCI systems.

The results suggested a tendency toward improved performance when a higher proportion of female user data was included in the training set. Significant positive correlations between R_f and BCI performance were observed for female users in `Dreyer2023` and for both male and female users in `Lee2019`. In particular, for female users in `Lee2019`, a strong correlation coefficient of 0.71 was obtained. In addition, the results in Table 1 indicate that, in `Lee2019`, training the classifier using only female users ($R_f = 1.0$) yielded significantly higher performance than the other R_f conditions. Moreover, for both male and female users in `Dreyer2023`, although no significant differences were observed, the highest average accuracy among all R_f conditions was also obtained when $R_f = 1.0$. As discussed in the Introduction, several studies [3–6] have reported that females tend to exhibit higher MI-BCI performance and larger ERD amplitudes; our findings further suggest the potential advantage of using more female data for training.

The results suggest that differences in ERD/S responses may exhibit a hierarchical scale structure. In Fig. 2, the within-group variability for both males and females was consistently larger than the centroid distance between the two gender groups ($v_f > d$ and $v_m > d$, where d denotes the centroid distance and v_f and v_m denote the variability of females and males, respectively). This implies that inter-individual differences exceed gender differences. Furthermore, the centroid distance between gender groups was larger than the difference in within-group variability ($d > |v_f - v_m|$) in all conditions. This indicates that the direct gender difference in mean responses may be larger than the difference in variability within each gender group. Taken together, these findings suggest a hierarchical scale structure of inter-individual differences > gender differences > differences in within-gender variability. In other words, although inter-individual differences are likely to have the largest impact on BCI performance, direct gender differences may exert a greater influence than within-gender variability differences. Therefore, when designing cross-users BCI systems that account for gender effects, ensuring robustness to differences in mean responses between genders may be more important than addressing differences in variability within each gender group. However, in this study, each variability metric was estimated as a single group-level value using all available data, and thus, statistical tests of gender differences in variability were not performed. Future work should further investigate gender-related differences in variability using resampling approaches such as bootstrap methods.

The performance changes with varying R_f in `Dreyer2023` may be partly explained by variability. For example, spatial variability showed a strong negative correlation with R_f, suggesting that lower spatial variability was associated with higher accuracy. In contrast, temporal variability exhibited a moderate positive correlation with R_f, implying that higher temporal variability may have had a relatively smaller impact on accuracy. These results suggest that the Riemannian TS+LR classifier used in this study may be sensitive to spatial variability while being relatively robust to temporal variability. However, a similar tendency was not observed in `Lee2019`, and further validation using a larger number of datasets will be necessary.

In this study, we used two datasets, `Dreyer2023` and `Lee2019`; however, these data were collected at Inria Bordeaux (Talence, France) and Korea University (Seoul, Korea), respectively, and may therefore reflect populations with different geographical backgrounds. Such geographical population differences could be one of the factors leading to the differing results observed between the

10.3217/978-3-99161-093-9-034

datasets. Future work should thus include validation using a larger number of datasets to enable evaluation based on geographically diverse populations.
In addition, the present analysis was conducted using only TS+LR, and different trends may emerge when other classifiers are employed. Therefore, validation using alternative classifiers is an important direction for future research.

CONCLUSION

In this study, we investigated how varying the gender ratio in the training data affects BCI performance in cross-user MI-BCI. The results showed that classification accuracy was significantly higher in three out of four conditions when the machine-learning model was trained using only female users. In addition, across the same three conditions, positive correlations (up to 0.71) were observed between the proportion of female users in the training data and BCI performance, suggesting that training the classifier with more female user data may be associated with improved BCI performance, even for male users.
Several previous studies have also suggested a female advantage in MI-BCI performance, and the present findings can be considered to be consistent with these observations. Moreover, since our results indicate that machine-learning models for MI-BCI may be subject to gender-related biases in the training data, they suggest that incorporating a gender-aware perspective will be important in future BCI research.

REFERENCES

[1] Vidaurre C, Blankertz B. Towards a Cure for BCI Illiteracy. Brain Topography. 2010;23(2):194–198.
[2] Randolph AB. Not All Created Equal: Individual-Technology Fit of Brain-Computer Interfaces. In: 2012 45th Hawaii International Conference on System Sciences. Jan. 2012, 572–578.
[3] Cantillo-Negrete J, Gutierrez-Martinez J, Carino-Escobar RI, Carrillo-Mora P, Elias-Vinas D. An approach to improve the performance of subject-independent BCIs-based on motor imagery allocating subjects by gender. BioMedical Engineering OnLine. 2014;13(1):158.
[4] Alimardani M, Gherman DE. Individual Differences in Motor Imagery BCIs: A Study of Gender, Mental States and Mu Suppression. In: International Winter Conference on Brain-Computer Interface. 2022, 1–7.
[5] Cantillo-Negrete J, Gutiérrez-Martínez J, Cariño-Escobar RI, Flores-Rodríguez TB, Elías-Viñas D. Time-frequency analysis of EEG signals from healthy subjects allocated by gender for a subject-independent BCI-based on motor imagery. In: International IEEE/EMBS Conference on Neural Engineering. 2013, 37–40.
[6] Wang P, Huang S, Jamil Z, Cheung VCK, Chan RH. Exploring Gender Differences in Motor Imagery EEG for Brain-Computer Interface Applications*. In: IEEE EMBS Special Topic Conference on Data Science and Engineering in Healthcare, Medicine and Biology. 2023, 15–16.
[7] Klick SM et al. Females versus males exhibit greater brain activation in sensorimotor regions during motor imagery after stroke. Neuroscience. 2025;582:104–113.
[8] Roc, Aline, Pillette, Léa, N'Kaoua, Bernard, Lotte, Fabien. Would Motor-Imagery based BCI user training benefit from more women experimenters? 2019.
[9] Zorzet B, Echeveste R, Milone D, Peterson V. Can sex be decoded from MI features in deep learning based BCIs?: An exploratory analysis. In: Int. BCI Meeting. Verlag der Technischen Universität Graz, 2025.
[10] Hagedorn LJ, Leeuwis N, Alimardani M. Prediction of Inefficient BCI Users Based on Cognitive Skills and Personality Traits. In: Neural Information Processing. 2021, 81–89.
[11] Jeunet C, N'Kaoua B, Subramanian S, Hachet M, Lotte F. Predicting Mental Imagery-Based BCI Performance from Personality, Cognitive Profile and Neurophysiological Patterns. PLOS ONE. 2015;10(12):e0143962.
[12] von Groll VG et al. Large scale investigation of the effect of gender on mu rhythm suppression in motor imagery brain-computer interfaces. Brain-Computer Interfaces. 2024;11(3):87–97.
[13] Nadeem A, Abedin B, Marjanovic O. Gender Bias in AI: A Review of Contributing Factors and Mitigating Strategies. ACIS 2020 Proceedings. 2020.
[14] Dreyer P, Roc A, Pillette L, Rimbert S, Lotte F. A large EEG database with users' profile information for motor imagery brain-computer interface research. Scientific Data. 2023;10(1):580.
[15] Lee MH et al. EEG dataset and OpenBMI toolbox for three BCI paradigms: An investigation into BCI illiteracy. GigaScience. 2019;8(5):giz002.
[16] Chevallier S et al. The largest EEG-based BCI reproducibility study for open science: The MOABB benchmark. Apr. 2024. arXiv: 2404.15319.
[17] Ledoit O, Wolf M. A well-conditioned estimator for large-dimensional covariance matrices. Journal of Multivariate Analysis. 2004;88(2):365–411.
[18] Rodrigues PLC, Jutten C, Congedo M. Riemannian Procrustes Analysis: Transfer Learning for Brain–Computer Interfaces. IEEE Transactions on Biomedical Engineering. 2019;66(8):2390–2401.
[19] Yger F, Berar M, Lotte F. Riemannian Approaches in Brain-Computer Interfaces: A Review. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2017;25(10):1753–1762.
[20] Slepian D. Prolate spheroidal wave functions, fourier analysis, and uncertainty — V: The discrete case. The Bell System Technical Journal. 1978;57(5):1371–1430.
[21] Sakoe H, Chiba S. Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans Acoust Speech Sig Proc. 1978;26(1):43–49.