 DOI: 10.3217/978-3-99161-050-2-104

Can sex be decoded from MI features in deep learning based BCIs?: an exploratory analysis

B. J. Zorzet[1]*, V. Peterson[2], D. H. Milone[1] and R. Echeveste[1]

[1] *Research Institute for Signals, Systems and Computational Intelligence sinc(i), FICH–UNL/CONICET, Santa Fe, Argentina.;* [2] *Instituto de Matemática Aplicada del Litoral, IMAL, UNL-CONICET, Santa Fe, Argentina*

*Ciudad Universitaria UNL, Ruta Nacional Nº 168, km 472.4, FICH, 4to Piso (3000) Santa Fe – Argentina. E-mail: bzorzet@sinc.unl.edu.ar

Introduction: Deep Learning (DL) methods in Motor Imagery Brain-Computer Interfaces (MI-BCI) are promising models for capturing complex EEG representations. Nevertheless, biases based on attributes such as sex and age can, although inadvertent, be introduced [1]. Previous studies have explored decoding age and sex through various techniques [2, 3, 4, 5], but it remains unexplored whether sex-related information is present in the features of DL models trained for MI tasks. This study aims to examine if features in the intermediate layers of a MI-BCI DL model include such information.

Material, Methods and Results: Two datasets were used, referred here to as Cho 2017 [6] and Lee 2019 [7]. Preprocessing included band-pass filtering, re-referencing to Fz, and downsampling to 128 Hz. A 2-second EEG window (0.5–2.5 s post-MI cue) was extracted for channels C3, C4, and Cz. The EEGNet [8] was used as DL model. It was trained using Leave-One-Subject-Out cross-validation balanced by sex. Per each leave-out subject, 100 models were trained (20 data splits and 5 model initializations). We first verified that models were successfully trained for the MI task for both male and female subjects (Table 1, top). A consistent trend favoring female performance is observed, but no significant sex differences were found. Next, we assessed the presence of sex-related information in the learned features. To do this the MI models were modified by freezing all layers except the final classification layer, which was re-trained for sex classification. Results reveal that sex-related information is present for male subjects (significantly above chance level classification), while not for female subjects (Table 1, bottom).

Conclusion: Our results show that sex-related information in DL trained for an unrelated MI-BCI task is present in the features of the model. At the same time, we observe a slight (though n.s.) trend toward better performance for women in MI tasks in both datasets. In contrast, sex can be better detected from model activations in the case of male subjects, suggesting that spurious correlations with sex are detrimental to BCI performance. Our work thus raises a concern: if demographic information like sex are still present in features extracted by DL models (like EEGNet) trained with balanced by sex data, and this correlates with model performance, more complex architectures may amplify biases. Further assessment of these potential issues is crucial for ensuring fair, transparent, and robust DL systems for MI-BCIs.

Dataset	Global Accuracy	Female	Male
Motor Imagery Classification			
Cho 2017	0.709 ± 0.131 ***	0.755 ± 0.132 ***	0.682 ± 0.124 ***
Lee 2019	0.710 ± 0.115 ***	0.738 ± 0.111 ***	0.690 ± 0.120 ***
Sex Classification			
Cho 2017	0.591 ± 0.276 *	0.446 ± 0.309 n.s.	0.674 ± 0.248 *
Lee 2019	0.550 ± 0.258 n.s.	0.470 ± 0.297 n.s.	0.609 ± 0.254 n.s.

Table 1: Global accuracy values and sex-specific performances for each dataset for Motor Imagery (MI) and sex classification. A Wilcoon test was applied to each accuracy greater than chance level (Wilcoxon test). *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$.

References:

[1] Ricci Lara MA, Echeveste R, Ferrante E. Addressing fairness in artificial intelligence for medical imaging. *Nature Communications*, 13(1):4581, 2022.

[2] Kaur B, Singh D, Roy PP. Age and gender classification using brain–computer interface. *Neural Computing and Applications*, 31(10):5887–5900, 2019.

[3] Wang P, Hu J. A hybrid model for EEG-based gender recognition. *Cognitive Neurodynamics*, 13(6):541–554, 2019.

[4] Kaushik P, Gupta A, Roy P, Dogra D. EEG-based age and gender prediction using deep BLSTM-LSTM network model. *IEEE Sensors Journal*, 18(1):1–1, 2018. doi:10.1109/JSEN.2018.2885582.

[5] Van Putten MJAM, Olbrich S, Arns M. Predicting sex from brain rhythms with deep learning. *Scientific Reports*, 8(1):3069, 2018.

[6] Cho H, Ahn M, Ahn S, Kwon M, Jun SC. EEG datasets for motor imagery brain–computer interface. *GigaScience*, 6(7):gix034, 2017.

[7] Lee MH, Kwon OY, Kim YJ, Kim HK, Lee YE, Williamson J, et al. EEG dataset and OpenBMI toolbox for three BCI paradigms: An investigation into BCI illiteracy. *GigaScience*, 8(5):giz002, 2019.

[8] Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP, Lance BJ. EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. *Journal of Neural Engineering*, 15(5):056013, 2018.

 Published by Verlag der Technischen Universität Graz