On the Bias, Fairness and Bias Mitigation of a Wearable-based Artificial Intelligence Model for Detecting Freezing-of-Gait

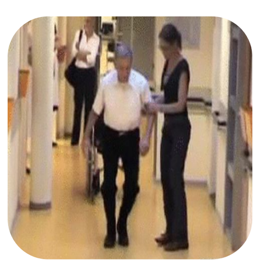
Timothy Odonga[1], Christine D. Esper[2], Stewart A. Factor[2], J. Lucas McKay[1,2,3], Hyeokhyen Kwon[1,3]

[1] Department of Biomedical Informatics, Emory University School of Medicine
[2] Department of Neurology, Emory University School of Medicine
[3] Wallace H.Coulter Department of Biomedical Engineering, Emory University and Georgia Institute of Technology

INTRODUCTION

- **Freezing-of-Gait (FOG)** is a debilitating Parkinson's disease (PD) symptom related to **increase** in **fall risk** and **injuries** [1].
- Despite progress in wearable and machine learning (ML) research for FOG detection, studies assessing **bias** and **fairness** of FOG detection models are **sparse**.
- Our study evaluated the bias and fairness of a **state-of-art** FOG detection model using standard **group fairness metrics** from four FOG wearable sensor datasets **and** evaluated the effectiveness of **multi-site transfer learning** as a **bias mitigation approach**.

CLINICAL SIGNIFICANCE

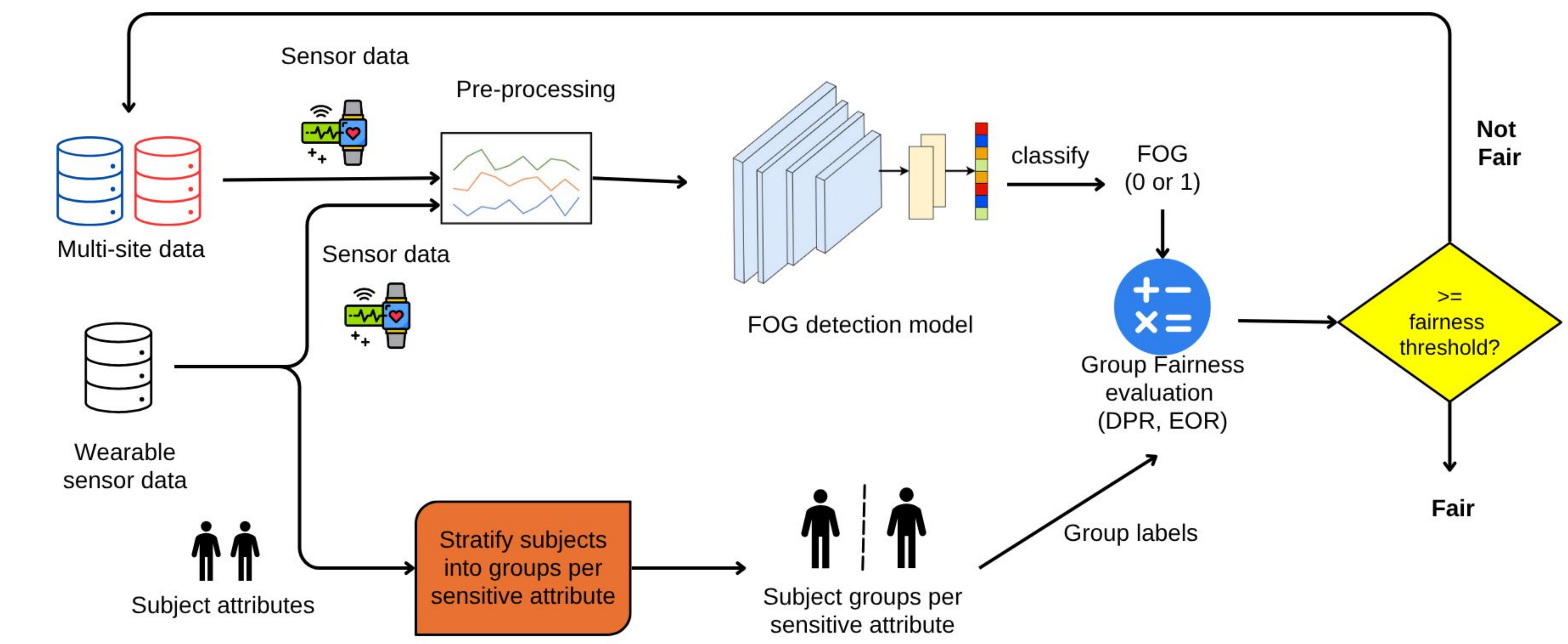


- Existing research has highlighted biases in PD diagnosis and care such as **delayed diagnosis** in **women** and significant **underdiagnosis** in **black patients** [2,3].
- Low-cost wearable-based FOG detection systems can **expand PD care access** through **continuous monitoring**, particularly for **underserved populations**, enable **personalized intervention plans**, and support development of **new therapeutic approaches** in clinical trials [4].

METHODS

- Four publicly available wearable sensor datasets: **DeFOG** (N=38) from **home** environments and **tDCS FOG** (N=62), **Daphnet** (N=10), **De Souza *et al.*** (N=35) from **clinical** settings [5, 6, 7].
- Subject attributes evaluated included **demographic attributes** such as **Age** and **Sex** and **clinical variables** such as **FOG type** and **Disease Duration** (years) [5].
- For **FOG type**, freezing episodes (FOG=1) were classified as **tremulous** or **akinetic** after analyzing the total power in the **locomotion band** (0-3Hz) and **freeze band** (3-8 Hz) [8].
- Trained a **Deep Convolutional and LSTM** (**DeepConvLSTM**) model known for its effectiveness in capturing complex FOG movements [9]. **Random Forest** and a **transformer-based models** [10] were also trained for comparison.
- Group fairness metrics used in evaluation: **Demographic Parity Ratio** (**DPR**) and **Equalized Odds Ratio** (**EOR**) [11].
- Applied transfer learning approaches i.e. pretraining using multi-site data (**multi-site transfer**) and from large scale pretrained human activity recognition models [12,13] (**generic activity feature transfer**) as bias mitigation methods. Conventional bias mitigation approaches such as **adversarial debiasing** [14] were applied for comparison .

ML PIPELINE



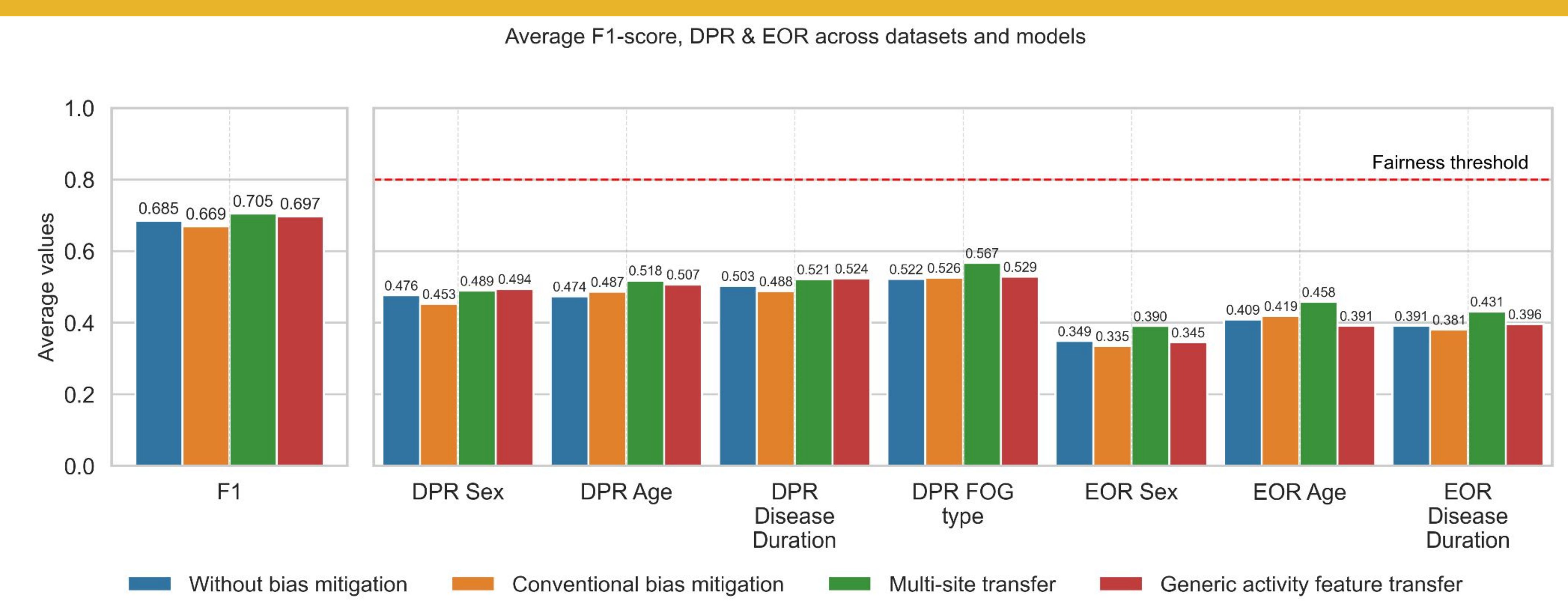
Following the standard human activity recognition pipeline [15], we applied the **sliding window** approach to segment the data into 3-second windows for DeFOG, tDCS FOG, and De Souza *et al.* and 4.5 second windows for Daphnet. For fairness analysis, self-reported sex, FOG type and median split groups for age (69: DeFOG/tDCS FOG/De Souza, 66:Daphnet) and disease duration (De Souza=7, tDCS FOG=9 , DeFOG=13, Daphnet=12.5 years) were used.

80% Applied the **four-fifths** rule (threshold = 0.8) as the **fairness threshold**, where DPR=1 & EOR=1 indicate perfect fairness.

We hypothesized that transfer learning across datasets exposes the model to **increased** and **more diverse** data thus improving its overall **model generalizability** and **reducing bias** with respect to the subject attributes in the target dataset.

With **multi-site transfer**, the model was pretrained on source/pretrain datasets and fine tuned on a target dataset with similar on-body sensor locations such as lower back and ankle.

RESULTS



- No single attribute was found to achieve the fairness threshold with the **sex** attributes showing the **lowest values** for DPR and EOR across all datasets. This was due to the **significant underrepresentation** of **females** in all four datasets.
- The transfer learning-based bias mitigation approaches consistently **increased** the DPR and EOR of the **sex** attribute.
- These transfer learning-based bias mitigation methods i.e multi-site transfer and generic activity feature transfer **increased** the **performance** (average change in F1-score of **+0.026** and **+0.018**, respectively) and **reduced** the **bias** (average change in DPR of **+0.027** and **+0.039**, respectively) across datasets and attributes.

CONCLUSION

- **No single attribute** is **fair** across both datasets. This implies the inconsistency of model performance across different subject groups (e.g., longer disease duration) can result in care disparities.
- Our results show empirically that transfer learning through multi-site datasets can **improve overall model performance** and **reduce biases**. Future work may include datasets with race attributes and conducting intersectional analysis to investigate how protected attributes interact to influence predictions.

REFERENCES

1. Nutt et al. (2011) "Freezing of gait: moving forward on a mysterious clinical phenomenon" The Lancet Neurology, vol. 10, no. 8, pp. 734-744.
2. Saunders-Pullman et al. (2011), "Diagnosis and Referral Delay in Women with Parkinson's disease," Gender medicine, vol. 8, no. 3, pp. 209-217.
3. Dahodwala et al. (2009) "Racial differences in the diagnosis of Parkinson's disease," Movement Disorders, vol. 24, no. 8, pp.1200-1205.
4. Mancini et al. (2019) "Clinical and methodological challenges for assessing freezing of gait: future perspectives," Movement Disorders, vol 34, no. 6, pp. 783 – 790.
5. Howard et al. (2023) "Parkinson's Freezing of Gait Prediction" https://kaggle.com/competitions/tlvmc-parkinsons-freezing-gait-prediction
6. Bachlin et al. (2010) "Wearable assistant for Parkinson's disease patients with freezing of gait symptoms"
7. De Souza et al. (2022) "A Public Data Set With Ground Reaction Forces of Human Balance in Individuals With Parkinson's Disease" Frontiers in Neuroscience
8. Moore, Steven T., et al. "Autonomous identification of freezing of gait in Parkinson's disease from lower-body segmental accelerometry." Journal of neuroengineering and rehabilitation.
9. Kwon et al. (2019) "Handling annotation uncertainty in human activity recognition," ACM ISWC, pp. 109-117.
10. Haresamudram et al. (2020) Masked reconstruction based self-supervision for human activity recognition. ACM ISWC, pp 45-49
11. Agarwal et al. (2018) "A Reductions Approach to Fair Classification," PMLR, pp. 60-69.
12. Yuan et al. (2024) "Self-supervised learning for human activity recognition using 700,000 person-days of wearable data" NPJ Digital Medicine
13. Ruan et al. (2025) "Foundation Models for Wearable Movement Data in Mental Health Research"
14. Zhang et al. (2018) "Mitigating unwanted biases with adversarial learning," AAAI/ACM Conference on AI, Ethics and Society, pp. 335-340.
15. Bulling et al. (2014) "A tutorial on human activity recognition using body worn inertial sensors" ACM Computing Surveys

ACKNOWLEDGEMENTS

This work was supported in part by grants from the Michael J. Fox foundation and the McCamish Parkinson's Disease Innovation Program center at Georgia Tech to J. Lucas McKay