Disparity Correction in Machine Learning for Healthcare

Contact

Dr Vlad González-Zelaya: carlos.gonzalez@newcastle.ac.uk

Research Project

Machine learning (ML) plays an increasingly pivotal role in disease and ailment detection within healthcare systems. However, these mechanisms are susceptible to biases that can exacerbate disparities in diagnostic accuracy and treatment recommendations across demographic groups. This project seeks to advance responsible AI practices by addressing algorithmic fairness, privacy, and transparency in ML systems for disease detection.

ML models in healthcare often rely on imbalanced datasets, where certain demographic groups are overrepresented while others are underrepresented. Other sources of disparity include feature selection, and algorithmic design choices. These issues can lead to models that perform well for some groups but poorly for others, perpetuating disparities in diagnostic accuracy and treatment outcomes. Addressing these challenges requires a comprehensive approach to ensure that ML mechanisms are fair, transparent, and privacy-preserving, fostering trust and equity in healthcare.

The primary objectives of this project are to:

1. Develop ML algorithms for healthcare that incorporate fairness constraints to address data imbalances and minimise bias across demographic groups.
2. Investigate the impact of feature selection and algorithmic design choices on fairness and develop strategies to mitigate associated biases.
3. Design privacy-preserving data processing techniques to protect sensitive patient information while ensuring model effectiveness.
4. Establish frameworks for algorithmic transparency, enabling stakeholders to understand and evaluate model decision-making processes.
5. Evaluate the impact of these methods on reducing disparities in disease detection through simulation studies and real-world trials.

This project addresses the critical need for equitable and responsible AI applications in healthcare. The outcomes will:

- Enhance the fairness and accuracy of ML-driven disease detection systems across diverse populations.
- Address multiple sources of unfairness, including data imbalances, feature biases, and algorithmic limitations.
- Promote privacy-preserving practices, ensuring patient trust in data usage.
- Foster algorithmic transparency, allowing healthcare providers and patients to better understand diagnostic decisions.
- Advance the field of responsible AI in healthcare, contributing to both academic research and practical implementation.

Project Timeline

Phase	Timeline	Key Activities
Literature Review	Months 1-6	Review existing research on ML in disease detection, fairness in AI, and privacy-preserving methods.
Data Collection	Months 4-9	Acquire and process healthcare datasets, ensuring demographic diversity and addressing imbalances.
Algorithm Development	Months 7-15	Develop and validate ML models with fairness constraints and privacy-preserving mechanisms.
Framework Design	Months 16-20	Design frameworks for algorithmic transparency and stakeholder engagement.
Evaluation	Months 21-27	Conduct simulation studies and pilot tests in healthcare settings to measure disparity reduction.
Dissertation Writing	Months 28-36	Write and revise the PhD dissertation.

Supervision Environment

The project will be supervised within a multidisciplinary environment, leveraging expertise in machine learning, health informatics, and ethics in AI.
Dr Vlad GZ, is an expert in AI ethics, focusing on fairness and transparency. The secondary supervisor will be a professional in the healthcare sector, with knowledge in the application of ML mechanisms to address medical issues.
Newcastle University will provide access to state-of-the-art computational facilities, collaborations with healthcare providers, and opportunities for interdisciplinary research.

Applicant Skills / Background

The ideal candidate will possess:

- A strong academic background in computer science, data science, or a related field, with a focus on machine learning.
- Experience in programming languages such as Python or R and familiarity with ML frameworks like TensorFlow or PyTorch.
- A commitment to advancing health equity.
- Strong analytical and communication skills, with the ability to work collaboratively in a multidisciplinary setting.

References

Chen, Richard J., et al. "Algorithmic fairness in artificial intelligence for medicine and healthcare." Nature biomedical engineering 7.6 (2023): 719-742.
Friedler, Sorelle A., Carlos Scheidegger, and Suresh Venkatasubramanian. "The (im) possibility of fairness: Different value systems require different mechanisms for fair decision making." Communications of the ACM 64.4 (2021): 136-143.
González-Zelaya, Vladimiro, et al. "Preprocessing matters: Automated pipeline selection for fair classification." International Conference on Modelling Decisions for Artificial Intelligence. Cham: Springer Nature Switzerland, 2023.