Fairness-Aware Disease Risk Prediction using Machine Learning

Abstract

Machine learning models are increasingly adopted for disease risk prediction, yet their deployment in healthcare raises critical concerns regarding algorithmic bias and unequal performance across demographic groups. This project investigates fairness-aware machine learning techniques for disease risk prediction using structured clinical datasets. Baseline classification models were trained and evaluated across sensitive attributes such as gender and age. Fairness metrics including demographic parity difference and equalized odds were used to quantify disparities in model predictions. To address observed biases, fairness-aware interventions such as sample reweighting and decision threshold adjustment were applied. The study examines the trade-offs between predictive performance and fairness, emphasizing the importance of equitable model behavior in high-stakes healthcare applications.

Results

Baseline models achieved competitive predictive performance but exhibited measurable disparities across demographic groups, particularly in false positive and false negative rates. After applying fairness-aware techniques, disparities in demographic parity and equalized odds were reduced while maintaining comparable overall accuracy. Although slight reductions in performance were observed in some cases, the adjusted models demonstrated more balanced outcomes across groups. These results highlight the effectiveness of incorporating fairness constraints into healthcare prediction pipelines.

Conclusion

This project demonstrates that fairness-aware machine learning methods can meaningfully reduce demographic disparities in disease risk prediction without severely compromising predictive performance. The findings underscore the necessity of evaluating models beyond aggregate accuracy, particularly in healthcare contexts where biased predictions can have significant ethical and societal implications. Future work may explore casual fairness approaches and extend the analysis to more diverse datasets and sensitive attributes.