Experimental setup for models

Add appropriate range for each feature

Performance for final fold value

* Before bias mitigation

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Metric | SVM | KNN | Random Forest | AdaBoost |
| Recall/ Sensitivity |  |  |  |  |
| Precision |  |  |  |  |
| F1-score |  |  |  |  |
| Demographic parity |  |  |  |  |
| Equal odds/Opportunity |  |  |  |  |

* After bias mitigation

Images and Graphs

Stratified

* Fairness model plot
* Stratified accuracy, recall
* Combine sensitive features across F1 Score
* Final statistics being used

Considerations

Avoid data leakage by keeping development and test set separate

Remove RestingBP outlier of 0.

Abstract

Heart Disease is a leading cause of death and morbidity globally, early detection is a good start to preventing this. Diagnostic technology has been increasingly used, Healthcare Artificial Inteligence models being a tool.

Trustworthiness of these tols are important, as users must understand how reliable, reproducible, and fair the diagnosis are.

Using a heart failure prediction dataset publicly available on kaggle, this project trains 5 models and evaluates them on fairness, performance and explainability. Adaboost, Random Forest, Support Vector Matrix, K-Nearest Neighbour, and a linear Neural Network are mdels with varying complexities and implementation making the comparison fair. This dataset included sensitive demographic features, age and sex, which shall be stratified.

The best performance by traditional metrics

Introduction

Related Works

Methodology

Results

Analysis

Conclusion and Future considerations

Although this dataset was created less than 3 years ago, the data collation and observations are over 2 decades old. Reliability of this with a change of times would need to be addressed.