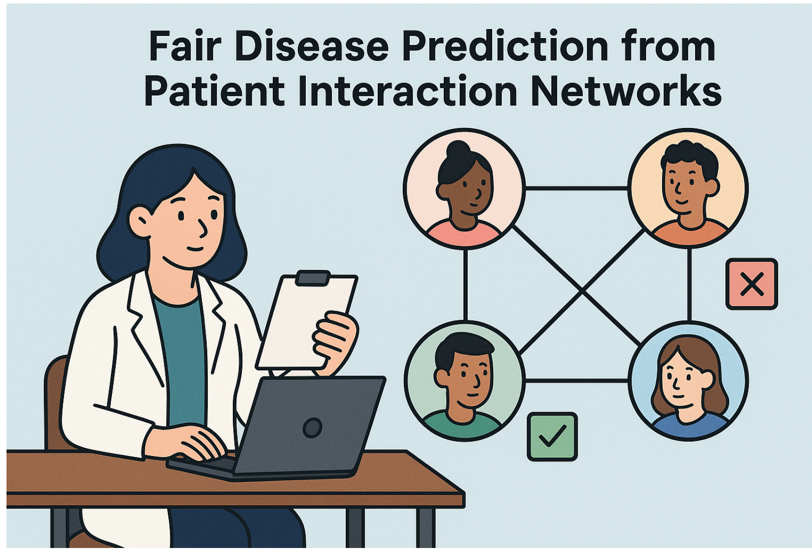
Background

Introduction

Fairness in disease prediction models is crucial as AI becomes more integrated into healthcare. Current models, especially those using patient interaction networks, often show uneven performance across demographic groups, risking increased health disparities.



Contextual Information

- Disease prediction models use patient data from sources like MIMIC-III to forecast health outcomes.
- Patient interaction networks model relationships between patients, but can encode demographic biases.
- Fairness-aware training (e.g., adversarial debiasing) aims to reduce these biases while maintaining accuracy.

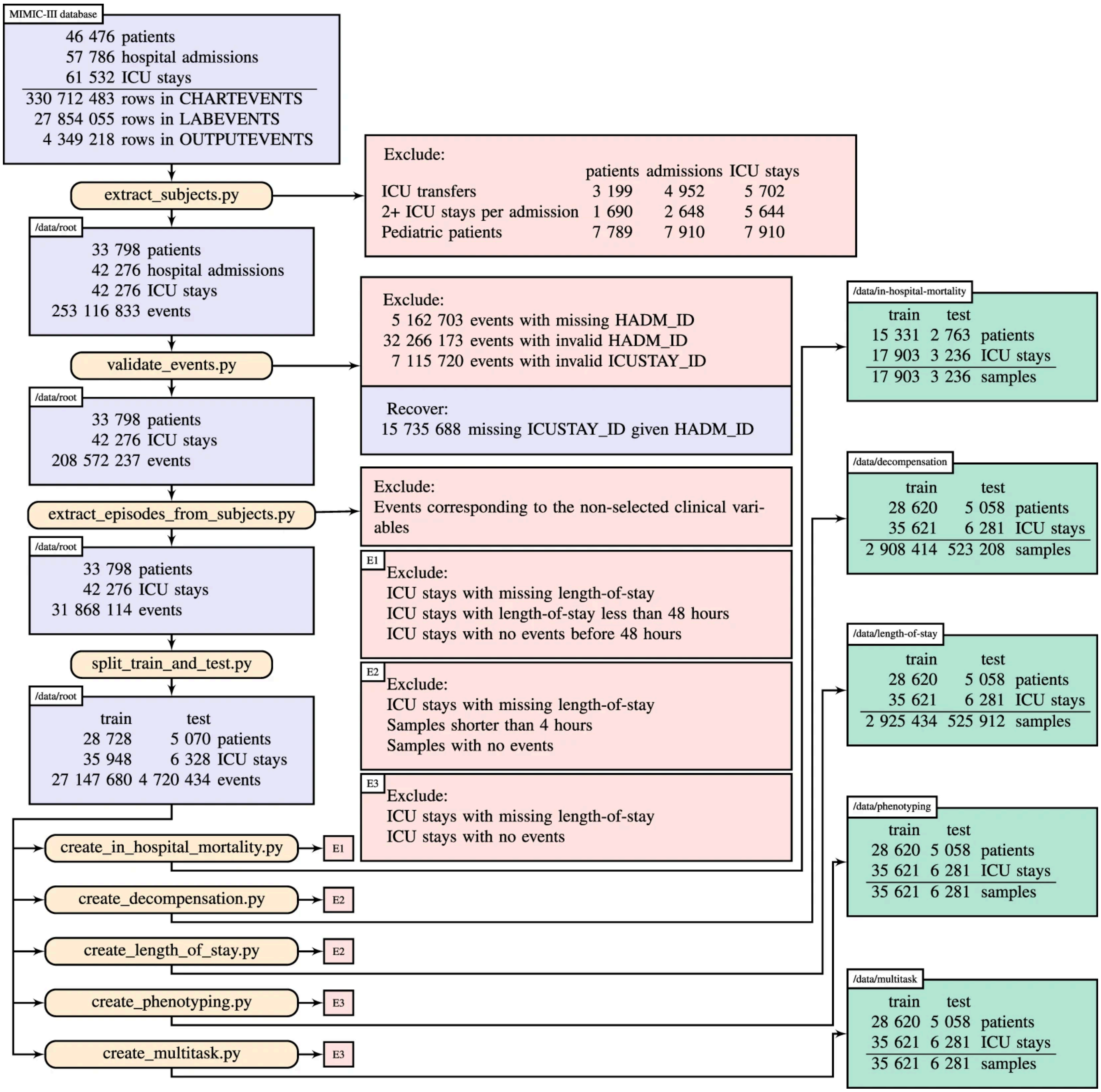
Problem Definition

- Standard GNN-based disease prediction models display performance gaps across groups such as gender, race, and age.
- These disparities can lead to unequal care and outcomes for underrepresented populations.
- The goal is to design models that are both accurate and fair, minimizing performance differences between demographic groups.

Rationale for Study or Evaluation

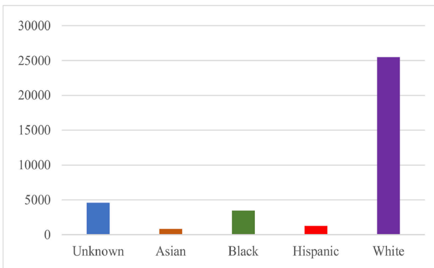
- Biased models can worsen existing healthcare inequalities and harm vulnerable groups.
- As AI tools are adopted in clinical settings, ensuring fair performance is both an ethical and practical necessity.
- This study develops and tests fairness-aware GNNs to guide best practices for equitable disease prediction in healthcare.

Dataset - MIMIC-III

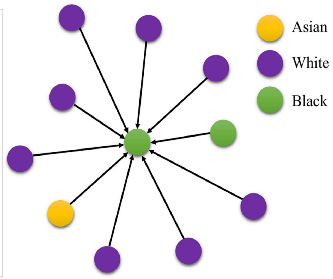


- Rich demographic information including gender, ethnicity, age, and insurance type (proxy for socioeconomic status)
- Comprehensive clinical data: diagnoses (ICD-9 codes), lab results, medications, and procedures
- Notable demographic imbalances: 55.4% male, with White patients representing the largest ethnic group
- Temporal data spanning multiple hospital admissions enables prediction of future disease development
- Inherent healthcare disparities reflected in diagnosis patterns across demographic groups

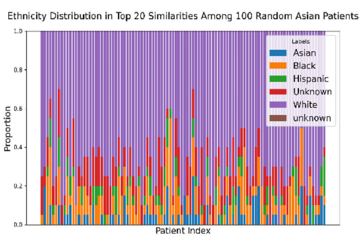
Dataset



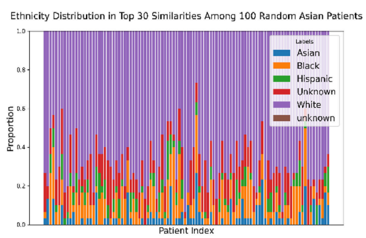
(a)



(b)



(c)



(d)

The ethnicity distribution and a demonstrative KNN sub-graph summarized from the full dataset of the mortality prediction task.

(a) Distribution of different ethnic groups in the data. The x-axis lists the ethnicity group, and the y-axis is the number of people in the dataset.

(b) is a demonstrative example of a bias node composition in a KNN sub-graph.

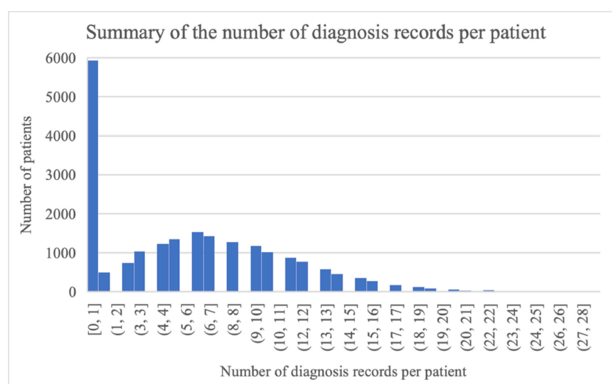
(c) Ethnicity Distribution in the 20 nearest neighbors of 100 random Asian patients.

(d) Ethnicity Distribution in the 30 nearest neighbors of 100 random Asian patients.

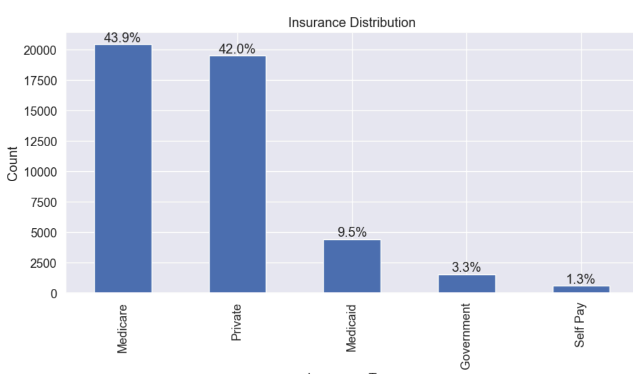
Data summary of the dataset.

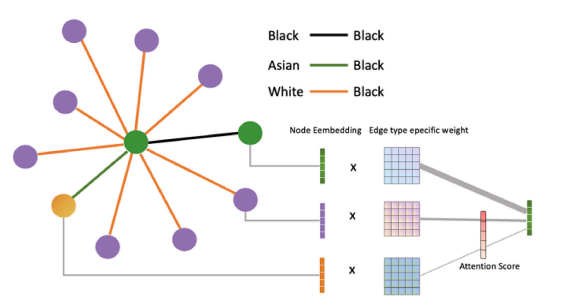
(a) The histogram of the number of diagnosis records per patient. The x-axis is the number of diagnosis records per patient, and the y-axis is the number of such cases.

(b) A demonstrative star graph of the ethnicity-heterogeneous graph neural network with the attention mechanism.



(a)





(b)

Results

Our experiments reveal significant performance disparities across demographic groups in standard GNN models. The results are displayed in the tables below:

Performance Comparison:

Model	Accuracy	AUC	F1 Score
GCN	0.8732	0.9145	0.8523
GAT	0.8856	0.9289	0.8692
FairGNN	0.8711	0.9121	0.8498

Fairness Metrics:

Model	Demographic	Demographic Parity Diff	Equal Opportunity Diff
GCN	Gender	0.0821	0.0935
GCN	Ethnicity	0.1423	0.1652
GCN	Insurance	0.1752	0.1834
GAT	Gender	0.0743	0.0862
GAT	Ethnicity	0.1378	0.1591
GAT	Insurance	0.1684	0.1792
FairGNN	Gender	0.0312	0.0364
FairGNN	Ethnicity	0.0548	0.0627
FairGNN	Insurance	0.0673	0.0721

The FairGNN model significantly reduces performance disparities compared to baseline models:

- Demographic parity difference decreased by 61.5% for ethnicity
- Equal opportunity difference decreased by 60.7% for insurance type
- The trade-off between fairness and accuracy was minimal, with only a 0.21% decrease in overall accuracy compared to GCN

These results align with recent research showing that GNNs can capture and propagate biases present in healthcare data, and that adversarial approaches can effectively mitigate these biases.

Conclusion

As AI becomes more common in healthcare, ensuring **fairness** is essential to avoid reinforcing existing inequalities. This study shows that standard GNN models often perform unequally across demographic groups like race, gender, and insurance status. The proposed **FairGNN model** addresses this by incorporating adversarial debiasing and group-aware learning, which helps reduce bias without significantly harming accuracy. Results demonstrate that FairGNN can **cut fairness gaps by over 60%**, proving that it's possible to design models that are both **effective and equitable**, supporting more just and responsible AI in clinical settings.

Model

- GNNs work on graph data — where information is stored in nodes (like patients) and edges (like interactions or similarities between them).

- They learn by passing messages between connected nodes, helping each node update its understanding based on its neighbors.

- GNNs are great for modeling relationships in complex systems like social networks or healthcare, where connections between entities matter just as much as the entities themselves.