How Aging Impacts Bias in Healthcare

AI ALCHEMISTS-Mohamed Arshad Peermohamed, Brandon Huang, Rafael Sanchez

DuckAI

What are we doing?

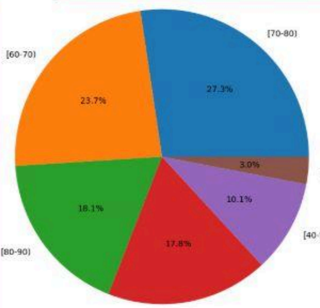
Exploring how **age impacts fairness** in AI-driven hospital readmission predictions

Analyzing results from both **Decision Trees** and **Neural Networks** to detect patterns of bias

Key focus areas:

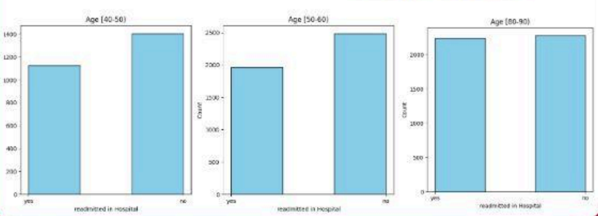
- Are models treating age groups equally?
- Do prediction rates vary unfairly across age ranges?
- What do fairness metrics (Demographic Parity, Equal Opportunity, Equalized Odds) reveal?

Data Distribution

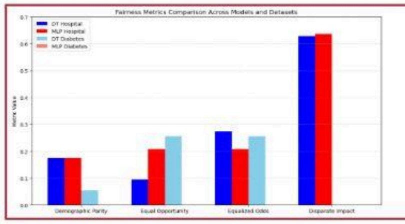


Age Distribution Insights:

- Majority of respondents are from older age groups
- Younger demographics are notably underrepresented
- This age imbalance may introduce a generational bias in the machine learning models used to predict the readmissions
- Ran against statistical learning models like decision trees vs deep learning models like MLP, we found decreasing performance with age!
- Reference datasets: Hospital Readmissions Datasets



Bias Metrics



Fairness Snapshot:

Demographic Parity = balanced positive predictions across age

Equal Opportunity = fair treatment across age groups

Disparate Impact

Fairness Insights:

Selection Rate:

DT and MLP show similar range in hospital data

DT dominates in diabetes, MLP close to zero across all groups

Recall (Equal Opportunity):

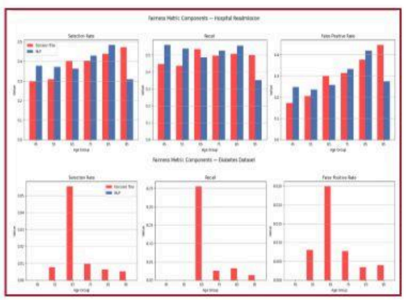
MLP hits higher recall in most hospital age groups

MLP recall flatlines in diabetes — bias likely from imbalance

False Positive Rate (Equalized Odds):

Hospital: DT rises with age, MLP more stable

Diabetes: DT shoots up for a single group — unstable



Visual Takeaways

= **Decision Tree**

= **MLP**

False Positive Rate (Hospital): Bars increase with age — higher error rates for older groups in DT

Selection & Recall (Hospital): Smaller gaps between bars in MLP → more consistent fairness across ages

Selection & Recall (Diabetes - MLP): Bars remain flat near zero — consistent but signals poor group attention

Recall (Diabetes - DT): One bar spikes up while others stay low — strong bias or imbalance

Each bar = one age group's value — big jumps = more disparity, flat = less variation

Our Conclusions

What Are We Seeing?

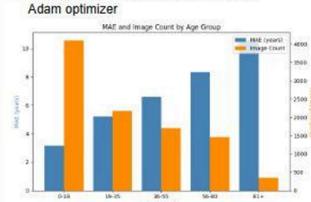
- Age impacts accuracy — performance drops as age increases
- Data ≠ performance — older groups have more data but worse outcomes
- Decision Tree predicts across age groups, but not always fairly
- MLP balances results better in hospital data, but fails in diabetes
- Deep Learning skips positive predictions for diabetes — zero recall across groups
- Model struggles most with oldest age groups — complexity + data imbalance
- Fairness ≠ metric scores — real-world usefulness matters more
- AI in healthcare must treat all ages with care, not just accuracy

Bias in Images

Investigating bias of CNN on UTKFace images dataset.

Model Architecture

- 3 convolutional layers (32→64→128), ReLU, max-pool
- Dense layer with 256 units + dropout
- Single output neuron for age prediction
- Trained 30 epochs using MSE loss and Adam optimizer

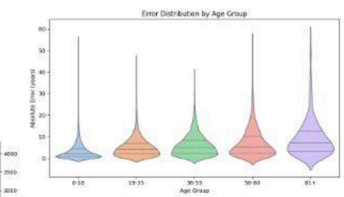


Insight – There is a strong relationship between dataset size and MAE

- MAE ↑ with age: 3.2 yrs (0–18) → 10.6 yrs (81+)
- Fewer samples → higher error ($r \approx -0.95$)
- Older groups show most bias

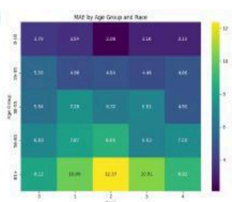
Takeaways:

- Bias increases with age due to underrepresentation of older faces.
- MAE by age and race reveals some compounding bias



Error Ranges Widen in Older Groups

- Younger groups: tight, symmetric error distribution
- Older groups: wider "tails" → more large mistakes (10+ yrs)
  Median error also increases across age bins
- Highlights inconsistency and underperformance in older predictions



- MAE increases with age across all racial groups
- Some racial groups (e.g. 2, 4) show steeper error growth
- Differences between races are small in younger ages
- **Insight** – Suggests compounding effects of age and representation gaps

STEVENS
INSTITUTE OF TECHNOLOGY