Task 3: Ethics in personalized medicine

Dataset: https://www.cancer.gov/tcga

Potential Biases in the Cancer Genomic Atlas(TCGA) Dataset.

- Bias in a dataset refers to a systematic error or distortion that causes the data to misrepresent the real world scenarios which could result in flawed AI/ML model outcomes.

1. **Demographic Underrepresentation Bias**

    - TCGA primarily includes data from White patients with very limited samples from Black, Asian, Hispanic and other races. Models created using this data will be efficient and effective for White people and may perform poorly for the other races.
    - Gender imbalance in certain cancers as prostate cancer studies focus more on male and breast cancer studies skew female. This can affect treatment recommendations.

2. **Data Collection Bias**

    - TCGA samples are mostly taken from high income counties with advanced healthcare systems, missing genetic prevalent in low-resourced regions.
    - Heterogeneity may not be fully captured leading to basic or oversimplified AI prediction.

3. **Algorithmic Bias**

    - If AI models are trained on data that is not diverse, they may prioritize treatments effectively only for the majority group worsening cancer disparities outcome.
    - Socioeconomic factors are less likely to be included, limiting real world applicability.

Fairness Strategies for Mitigating Bias

1. **Diverse and Representative Data**

    - Expand TCGA with global datasets to include more ethnic groups

- Add more samples of underrepresented groups when training models to balance class distribution.

2. **Bias Aware AI Development**

- Adversarial debiasing: Train models to ignore demographic factors while maintaining accuracy.
- Fairness metrics: Use statistical parity, equalized odds or demographic parity to evaluate model fairness.

Conclusion

AI has great potential in precision oncology, but biases in TCGA can lead to inequitable treatment recommendations. By improving diversity ,applying fairness-aware algorithms and enforcing regulatory standards we can develop AI systems that benefit all cancer patients regardless of ethnicity or background.