'This work was supported by the Engineering and Physical Sciences Research Council [Grant EP/X031012/1]'

Assessing Accuracy and Fairness of Dimensionality Reduction Techniques in Health Data Analysis

Adedayo Omoniyi | Supervisor: Dr. Carlos Vlad González, School of Computing, Newcastle University |

linkedin.com/in/adedayo-omoniyi

Introduction / Aim

- Health data are often complex and high-dimensional, requiring dimensionality reduction (DR) to simplify analysis.
- But an important question arises: are these methods fair to everyone?
- Fairness means ensuring results are not biased against groups defined by age, gender, education, or ethnicity.
- Without fairness, predictions may be reliable for some groups but not others. For example, a diabetes risk model that works well for men but poorly for women.
- In this work, fairness refers to the extent to which the embedding represents different attributes (e.g., ethnicity, education) equally well. This is measured using Mean Embedding Discrepancy (MED): the smaller the MED, the fairer the representation.
- The idea is: if the visualisation is fair, the position of a data point in the visualisation should not reveal its sensitive attribute.
- This research examines the fairness and performance of dimensionality reduction techniques, with the goal of informing the design of healthcare AI tools that are accurate, equitable, trustworthy, and beneficial for all.

Methods

- **Dataset**
  NHANES (National Health and Nutrition Examination Survey), including demographic, body measure, and clinical features.
- **Dimensionality Reduction (DR)**
  Applied t-SNE, UMAP, Fair-NeRV, and Fair-t-NeRV to simplify high-dimensional data.
- **Fairness Evaluation**
  Checked whether DR methods preserved fair representation across sensitive groups (age, gender, education, ethnicity).
- **Performance Assessment**
  Evaluated the quality and fairness of embeddings from dimensionality reduction (t-SNE, UMAP) and fairness-aware methods (Fair-NeRV, Fair-t-NeRV) using k-NN F1-scores across multiple neighborhood sizes (k = 1, 3, 5, 10, 20) and mean embedding discrepancy (MED).

Fairness Metric Equations

1. Mean Embedding Discrepancy (MED) - Fairness
MED = (1 / |S|) Σ (|| μ_s − μ ||²) , for s ∈ S
S = set of sensitive groups (e.g. gender, ethnicity)
μ_s = mean embedding of group s
μ = global mean embedding
👉 Smaller MED = fairer embedding.

2. F1 Score at k (F1-k) - Accuracy
F1-k = (2 · Precision-k · Recall-k) / (Precision-k + Recall-k)
Measures how well true neighbours are retrieved
👉 Higher F1-k = more accurate embedding.

3. Accuracy versus Fairness Trade-off - Balancing loss
Minimize [Accuracy Loss + λ · Fairness Loss]
Accuracy Loss = preserves data structure
Fairness Loss = penalises bias (MED)
λ = balance parameter
👉 Shows the trade-off between accuracy and fairness.

Dimensionality Reduction Performance on Diabetes and Marital Status Attributes

Sensitive attribute		t-SNE		UMAP		Fair-NeRV		Fair-t-NeRV	
	K	$f1_k$	MED	$f1_k$	MED	$f1_k$	MED	$f1_k$	MED
Diabetes	10	0.9000	44.5646	0.8934	4.5838	0.8640	0.1421	0.695	0.8444
Marital Status	20	0.5622	16.8702	0.5568	1.2890	0.5608	0.0936	0.302	0.7032

Diabetes: t-SNE has the highest accuracy (F1 = 0.90) but is the least fair (MED = 44.56). UMAP improves fairness (MED = 4.58) with minimal loss in accuracy. Fair-NeRV and Fair-t-NeRV achieve the fairest embeddings (MED = 0.14, 0.84), though with slightly lower F1.

Marital Status: t-SNE again shows high F1 (0.56) but poor fairness (MED = 16.87). UMAP reduces bias (MED = 1.29) with similar accuracy. Fair-NeRV and Fair-t-NeRV produce very fair embeddings (MED = 0.09, 0.70) at some cost to accuracy.

These suggest that Fair-aware embeddings drastically reduce bias while slightly trading off predictive performance.

Discussion and Conclusion

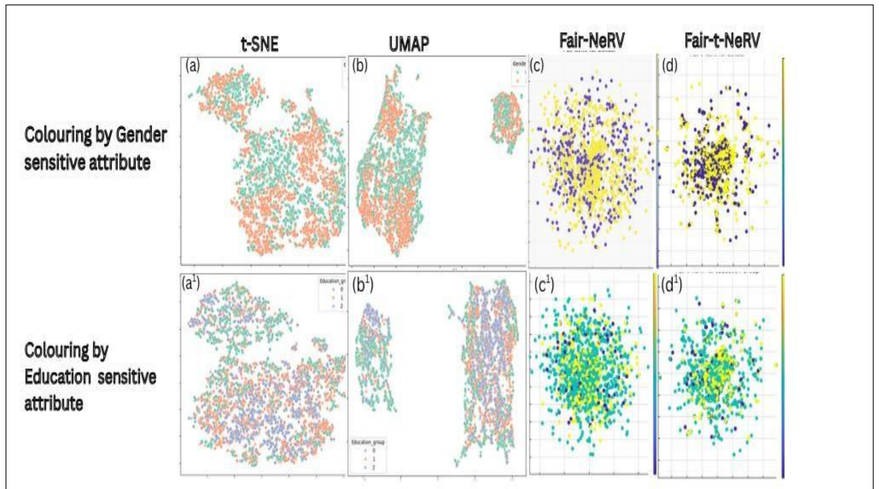
- **Key Observations**
  Standard DR methods (t-SNE, UMAP) achieve high predictive performance but exhibit high MED, showing sensitive-attribute clusters and potential bias in the embeddings.
  Fairness-aware DR methods (Fair-NeRV, Fair-t-NeRV) drastically reduce MED and distribute data more uniformly, mitigating sensitive-attribute influence.
- **Accuracy versus Fairness Tradeoff**
  Fair embeddings show slightly lower F1-score compared to standard DR, highlighting the tradeoff between predictive power and fairness.
- **Fairness Comparison**
  Fair-NeRV: Showed Lowest MED, most effective at reducing bias.
  Fair-t-NeRV: Balances fairness and predictive performance, with improving F1- score, providing the best balance between fairness and accuracy.
- **Conclusion**
  Incorporating fairness in dimensionality reduction promotes equitable data representation.
  It supports bias-aware analysis and fairer downstream AI tasks, making results more trustworthy and responsible.

Acknowledgments

Thanks to Dr. Vlad, HDR UK, Newcastle University, and NorthFutures for their guidance, support, and the internship opportunity.

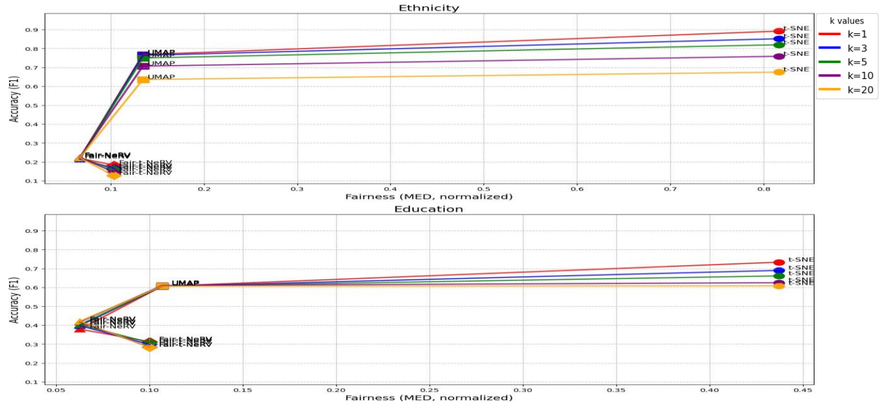
Results

Embedding of High-Dimensional NHANES Data



Embeddings by t-SNE (a, a^1) and UMAP (b, b^1) clearly reveal sensitive information through concentration of the minority and majority sensitive categories in Gender (male, green dots: female, red dots). Likewise, the embeddings for Education (low, green dots: middle, red dots: high, blue dots). In contrast, the Fair-NeRV (c, c^1) and Fair-t-NeRV (d, d^1) mixes sensitive categories better than t-SNE and UMAP, while still showing high dimensional data organisation well.

Accuracy versus Fairness Trade-off Across Embeddings and K (Nearest Neighbours) for Ethnicity and Education Sensitive Attributes



Fair-NeRV and Fair-t-NeRV achieved significantly lower MED, indicating higher fairness compared to t-SNE and UMAP for both ethnicity and education, while t-SNE and UMAP maintained higher predictive accuracy across k-values. This illustrates the trade-off between accuracy and fairness across different embeddings and neighbourhood sizes.