Final Project Draft Report

College of Professional Studies, Northeastern University

EAI 6400 – Data Governance and Responsible AI

Professor Mimoza Dimodugno

12/10/2024

Title:

The Impact of Bias in AI Technology: Ethical Challenges and Analysis in Diabetes Prediction Using Facial

Recognition

Group Members:

Harsh Bharatbhai Anadkat

Abstract

This study examined the use of machine learning approaches to the prediction of type 2 diabetes mellitus (T2DM) using face recognition technology (FRT). We used 18 facial morphological measures and 81 critical points to assess the faces of 2,574 patients (1,590 with T2DM and 984 without). Using least absolute shrinkage and selection operator (LASSO) regression, two prediction models were created. The area under the curves (AUCs) for the "lambda.min" and "lambda.1se" models were 0.799 and 0.776, respectively. The most important facial characteristic for prediction was found to be the breadth of the jaw. The work raises important ethical questions about algorithmic fairness and dataset bias in AI-driven healthcare, even as it shows promise for non-invasive T2DM screening.

Introduction

Substantial progress has been made in the application of artificial intelligence (AI) in healthcare, namely in the prediction of diabetes. However, there are serious ethical issues with the use of face recognition technology (FRT) in medical diagnoses. In addition to addressing the important problems of bias in AI technology and its ethical ramifications, this work attempts to investigate the possibility of FRT in T2DM prediction. With 12.8% of persons with diabetes and 35.2% in pre-diabetic stages1, the incidence of diabetes in China is concerning. Early detection is important yet difficult, which highlights the necessity for creative prediction techniques. This work creates a unique method for T2DM prediction using face characteristics by fusing the theory of traditional Chinese medicine (TCM) with contemporary machine learning techniques.

Analysis

A Synopsis of the Dataset

768 samples with 8 characteristics and a binary outcome variable indicating the existence of diabetes make up the dataset. Age, BMI, and glucose are important characteristics. A preliminary examination identifies several difficulties:

Potential Bias in characteristics: Missing values may be indicated as zeros in characteristics such as BMI and glucose, which would disproportionately impact model predictions and data quality.

Outcome Imbalance: With around 35% of the instances being positive, the dataset exhibits an imbalance that might skew model performance.

The consequences of Feature Bias

Glucose and Insulin Levels: If these important indicators are missing or have zero values, the results may be skewed, especially for underprivileged groups that might not have as much access to healthcare.

BMI Representation: The disparity in BMI between age groups emphasizes how crucial it is to take demographic subtleties into account when designing models.

Prototype Implementation

To forecast diabetes outcomes, a fairness-aware categorization algorithm was created. The modelemploys strategies like stratified sampling to solve class imbalance and preprocessing processes to deal with missing variables. Important measures like recall, accuracy, and precision were assessed, with an emphasis on fair performance across demographic groups**.**

Method

We collected facial images from 2,574 subjects at Lanzhou Second People's Hospital between December 2017 and September 2021. The study adhered to the Declaration of Helsinki and received ethical approval. Facial images were analyzed using 81 key points and 18 morphological parameters. LASSO regression was employed to construct two prediction models: "lambda.min" and "lambda.1se".

Results

The "lambda.min" and "lambda.1se" models achieved AUCs of 0.799 and 0.776, respectively. The width of the jaw was identified as the most significant facial feature for T2DM prediction. Both models demonstrated good predictive efficiency and reproducibility in the validation set, with AUCs of 0.695 and 0.682.

Discussion

This study presents a novel approach to T2DM prediction by integrating TCM facial inspection theory with machine learning. The models' performance suggests potential for clinical application in non-invasive T2DM screening. However, the use of FRT in healthcare raises critical ethical concerns:

1. Dataset Bias: The study's dataset may not represent diverse ethnic groups, potentially leading to biased predictions for underrepresented populations.

2. Algorithmic Fairness: The models may struggle with features that vary across populations, such as skin pigmentation and facial structure, potentially resulting in disparate performance across demographic groups.

3. Health Inequities: Biases in the system could reinforce existing healthcare disparities by misdiagnosing or over-testing certain populations.

To address these ethical challenges, future research should focus on:

1. Developing diverse and representative datasets

2. Implementing advanced bias mitigation techniques

3. Combining FRT results with other diagnostic factors to improve reliability

4. Establishing robust ethical and governance frameworks for AI in healthcare

Conclusion

While this study demonstrates the potential of FRT in T2DM prediction, it also underscores the critical need to address bias and ethical concerns in AI-driven healthcare. Future development of such technologies must prioritize fairness, transparency, and equitable performance across all demographic groups to ensure responsible and beneficial implementation in clinical settings.

References

https://www.healthcareproducts.org/wp-content/uploads/2023/02/Final-Bias-in-Artificial-Intelligence-11.27.22.pdf

https://pmc.ncbi.nlm.nih.gov/articles/PMC9577725/

https://www.researchgate.net/publication/361606995_Diabetes_Analysis_with_a_Dataset_Using_Machine_Learning

Ahmed, F. & Farid, F. (2018). Applying internet of things and machine-learning for personalizedhealthcare: Issues and challenges. 2018 International Conference on Machine Learning and Data Engineering (iCMLDE). 19-21. doi: 10.1109/iCMLDE.2018.00014 Reddy, S., Allan, S.,

Coghlan, S., & Cooper, P. A governance model for the application of AI in health care. Journal ofthe American Medical Informatics Association, 0(0), 2019, 1–7. doi: 10.1093/jamia/ocz192

Chaves, L., & Marques, G. (2021). Data Mining Techniques for Early Diagnosis of Diabetes: A Comparative Study. Applied Sciences, 11(5), 2218.

Fitriyani, N., Syafrudin, M., Alfian, G., and Rhee, J. (2019). Development of Disease PredictionModel Based on Ensemble Learning Approach for Diabetes and Hypertension. IEEE Access. 7. 144777 - 144789. 10.1109/ACCESS.2019.2945129.

Komi, M., Jun Li, Yongxin Zhai, Xianguo Zhang, 2017. Application of data mining methods indiabetes prediction. Presented at the 2017 2nd International Conference on Image, Vision and Computing (ICIVC), IEEE, Chengdu, China, pp. 1006–1010. https://doi.org/10.1109/ICIVC.2017.7984706

Samant, P., Agarwal, R., 2018a. Machine learning techniques for medical diagnosis of diabetesusing iris images. Computer Methods and Programs in Biomedicine 157, 121–128. https://doi.org/10.1016/j.cmpb.2018.01.004