Addressing Racial Bias in Cardiovascular Disease Risk Prediction with Fair Data Augmentation

1st Hannah Cui
i.equalcare
https://iequalcare.org
hannah.cui6@gmail.com

2nd Yatong Han
The Chinese University of Hong Kong (Shenzhen)
Shenzhen, China
hanyatong@cuhk.edu.cn

***Abstract*—In the past decade, the application of machine learning approaches to clinical tasks such as risk prediction and diagnosis of diseases has offered several potential benefits for both patients and medical professionals. However, in the development of these machine learning models, a factor is often overlooked: algorithms can be unfairly biased against certain populations, presenting disparate performances for different groups, defined by protected attributes such as race/ethnicity, sex or gender, and socioeconomic status, that are often insufficiently represented in datasets. Implementation of these biased models for clinical use could possibly worsen healthcare inequality. In this paper, we investigate the fairness of health-related ML research by using health survey data from the Centers for Disease Control and Prevention (CDC) to predict cardiovascular disease risk. We use two fairness metrics for the CDC data, and analyze the prediction model performance for racial minority groups to evaluate potential racial bias due to imbalanced or unrepresentative data. Then, we utilize Fair Mixup, a data augmentation method, to enhance fairness measures between racial groups and achieve improvements in predictive accuracy.**

***Index Terms*—Machine learning, Fairness, Data Augmentation, Cardiovascular Disease Risk Prediction**

I. Introduction

Artificial intelligence (AI) has shown increasing potential for use in healthcare, particularly in the applications of machine-learning (ML) algorithms for clinical risk prediction and diagnosis [1]. ML-based models can recognize correlations and patterns within medical datasets and determine the risk of patients developing conditions such as sepsis [2], diabetic retinopathy [3], or myocardial infarction [4], allowing clinicians to identify high risk patients and implement preventative treatments [5]. However, researchers have expressed concerns regarding racial bias in ML algorithms intended for use in healthcare [6], [7]. Algorithms are biased when they perform differently for one group compared to others, producing less accurate or even discriminatory results. Racially biased algorithms are often the result of biased training datasets [8], which are unrepresentative and imbalanced for specific racial groups due to discrimination, disparities in data collection, or limited participation in studies or surveys [9]. Several studies show that models have lower prediction accuracy for racial minorities when trained on noninclusive data [10]–[12]. This discrepancy may occur since algorithms become biased for features specific to the large majority group in datasets to achieve better overall performance [13]. Clinical implementation of racially biased algorithms for risk prediction may only exacerbate existing healthcare inequity and prevent fair prioritization of high-risk patients for medical intervention.

In this work, we focus on cardiovascular diseases (CVDs) as a specific case study. CVDs are a group of heart and blood vessel disorders that include strokes, atherosclerosis, aneurysms, and ischemic or coronary heart diseases (CHDs), which directly leads to conditions such as angina and myocardial infarction [14]. There were over 500 million cases of CVDs and 18 million CVD-related deaths worldwide in 2019, with coronary heart disease being the main cause of mortality, accounting for 49.2% of deaths [15]. Using datasets recording risk factors (age, smoking, hypertension, diabetes, obesity, dyslipidemia, etc. [16]) and disease occurrence, ML models can be developed to calculate the risk of patients developing certain CVDs. However, existing CVD clinical prediction models often fail to report or account for potentially imbalanced data and ensure unbiased performance of models across racial groups [17]–[19]. This is a significant issue for CVD prediction in particular since disease incidence and prominent risk factors can vary across racial-ethnic backgrounds [20]. For example, cardiovascular risk studies conducted in the US found that Black (African-American, Afro-Carribean, African) women developed CVD at a much earlier age than non-Hispanic white women [21], which may be the result of earlier onset and high prevalence of hypertension within this group [22]. Another study in the UK found that South East Asians had dyslipidemia at a lower body mass index (BMI) compared to other groups, and exhibited significant levels of inflammatory high-density lipoprotein (HDL) cholestrol [23]. It is important to achieve fair CVD risk prediction for all racial groups in diverse populations, as earlier implementation of treatment can prevent occurrence of deadly cardiovascular conditions such as myocardial infarction or stroke. We aim to enhance the influence of racial equity on the prediction task and boost the accuracy of CVD risk prediction by fine-tuning two widely used group fairness metrics: Demographic Parity (DP) and Equalized Odds (EO). A common way to achieve this goal is through utilization of data augmentation methods. In [29], the authors propose Fair Mixup, a data augmentation method that constructs interpolated samples, generating a path of distributions that connect sensitive groups and regularize the smoothness of transitions along the path to improve the

generalization of group fairness metrics. We follow this work to regularize neural networks for CVD risk prediction by interpolating data distributions between sensitive groups.

This paper consists of three main sections. First, we review related works in the study of machine learning fairness and the use of data augmentation for neural networks. Then, we discuss our dataset, risk feature selection, and network architecture, as well as our measurements for group fairness. Lastly, we detail our experiment procedure and analyze the experimental results, which show that we can obtain a $2\%-3\%$ relative improvement in accuracy while maintaining fairness.

II. Related work

A. Machine Learning Fairness

With the increasing prevalence of Machine Learning (ML) predictive systems in decisions that wield substantial influence over the lives of individuals—such as university admissions, job recruitment, child custody, and criminal risk assessment—fairness has emerged as a paramount requirement [24]. Gajane and Pechenizkiy, in their work [25], delve into the formalization of significant fairness concepts, including statistical parity, equal opportunity, and individual fairness. They also explore the impact of these concepts on distributive justice based on social science literature. Furthermore, they elucidate two other fairness concepts extensively studied in social science: resource equality and capability equality. However, these concepts lack mathematical formalization. In a separate study [26], the author categorizes fairness notions into four distinct classes: statistical tests, absolute measures, conditional measures, and structural measures. Statistical tests merely signify the absence or presence of discrimination. Absolute and conditional measures both gauge the extent of discrimination, while conditional measures also consider reasonable explanations for the observed discrimination.

B. Data Augmentation and Regularization

In [27], the authors suggest that when sufficient implicit regularization is provided, not only are contributions to weight decay and lost generalization redundant, but such techniques can significantly degrade performance if hyperparameters are not carefully tuned to the architecture and dataset. In contrast, data augmentation systematically provides large generalization gains and does not require retuning of hyperparameters. In [28], regularization is adopted as a fundamental technique used to enhance the generalization performance of a model by constraining its complexity. Deep Neural Networks (DNNs) often suffer from overfitting to the training data, and rely heavily on regularizers such as data augmentation (DA) or weight decay through the minimization of structural risk (i.e., cross-validation) to adjust hyperparameters. In [29], the authors propose the Fair Mixup method to generate a path of distributions that connect sensitive groups and regularize the smoothness of transitions among the path, improving the generalization of group fairness metrics.

TABLE I
Features Selected for Risk Prediction

Features from BRFSS 2015 Data	
Demographics	Health
Age (AGEG5YR)	Hypertension (RFHYPE5)
Sex (SEX)	High Cholestrol (TOLDHI2)
Education (EDUCA)	Recent Cholestrol Check (CHOLCHK)
Income (INCOME2)	Previous Stroke (CVDSTRK3)
BMI (BMI5)	Difficulty Walking (DIFFWALK)
Health Plan (HLTHPLN1)	Physical Health (PHYSHLTH)
Medical Costs (MEDCOST)	General Health (GENHLTH)
Smoking Status (SMOKE100)	Mental Health (MENTHLTH)
Diet (FRTLT1, VEGLT1)	Physical Activity (TOTINDA)
Heavy Drinker (RFDRHV5)	

III. Methods

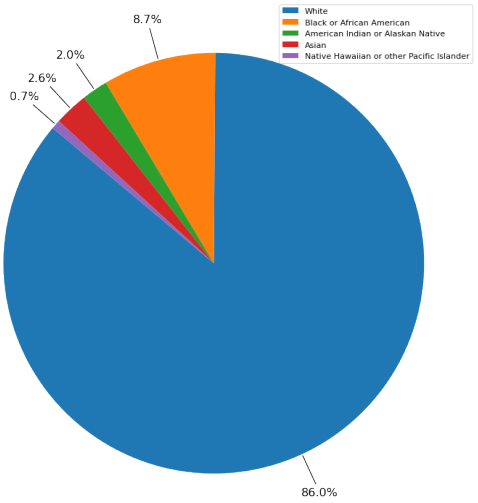
A. Dataset and Feature Selection

We used the 2015 dataset from the Behavioral Risk Factor Surveillance System (BRFSS), which collects health-related information via an annual telephone-based survey conducted by the Centers for Disease Control and Prevention (CDC). Responses were collected by interview from over 400,000 US residents in all 50 states and 3 territories (Guam, Puerto Rico, and US Virgin Islands) [30]. Respondents were asked questions regarding their demographics, mental and physical health, medical conditions, access to healthcare, and lifestyle. Fig. 1 shows the racial distribution of respondents in the dataset before data processing. More than half of the respondents were Non-Hispanic whites, while Asian, and American Indian/Alaskan Native, and Native Hawaiian/Pacific Islander individuals were the most underrepresented groups, collectively making up less than 6% of the data, indicating that the original dataset is significantly biased in terms of race.

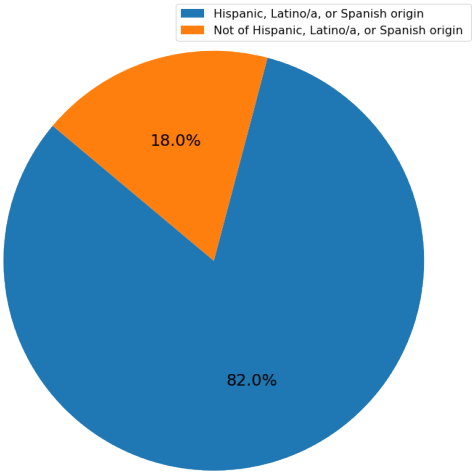
We primarily focused on predicting risk of coronary heart disease and one of its commonly associated conditions, Myocardial Infarction (MI), as CHD is the most prominent cardiovascular disease [15]. We chose features from the BRFSS dataset based on general risk factors for CVDs, including diabetes, hypertension, high cholestrol, BMI, diet, smoking habits, and low physical activity, as well as basic demographics including age, healthcare access, education, and sex. Table I shows the features selected for use in our CVD risk prediction model.

B. Network Architecture

We treated cardiovascular disease risk prediction as a standard fair binary classification. We referenced the CDC BRFSS codebook to select specific risk features as inputs $X \in \mathcal{X} \subset \mathbb{R}^d$ (RFHYPE5 represents high blood pressure, BMI5 represents body mass index, etc.) and CHD/MI occurrence as truth labels $Y \in \mathcal{Y} = \{0, 1\}$. Race was used as the sensitive attribute $A \in \{0, 1\}$. For example, if we want to adjust the racial fairness for black people, we set samples from black respondents in the dataset to 1 and all other samples to 0. Prediction label $\hat{Y} \in [0, 1]$ is predicted by model $f : \mathbb{R}^d \to [0, 1]$. The network architecture is a neural network with 4 layers: Fully Connected (FC) Layer



(a) Percentages of different racial groups in the dataset before data processing.



(b) The ratio of Hispanic to non-Hispanic individuals in the dataset before data processing.

Fig. 1. The two race classification settings used in the experiment.

(input size, 200) ⇒ activation layer (RELU) ⇒ FC(200,400)⇒ RELU ⇒ FC(400,400) ⇒ RELU ⇒ FC(400,200) ⇒ RELU ⇒ FC(200,1). Fairness metrics are added to model losses as regularization constraints.

C. Group Fairness Metrics

Various metrics have been proposed to measure the fairness of machine learning models. Our interest in CVD risk prediction is group fairness, where samples are grouped according to a particular sensitive attribute, and statistics about model predictions are calculated for each group and compared across groups. While different kinds of group fairness measurements have been proposed, we will focus on the two most commonly used ones listed below.

Demographic Parity (DP): This requires the predictions $\hat{Y}$ to be independent of the sensitive attribute A, i.e., $P(\hat{Y}|A = 0) = P(\hat{Y}|A = 1)$.

Equality of Opportunity (EO): This metric takes into consideration that different groups could have different distributions in terms of label Y. EO requires $\hat{Y}$ and A to be conditionally independent with respect to Y, i.e., $P(\hat{Y}|A = 1, Y = y) = P(\hat{Y}|A = 0, Y = y)$ for $y \in \{0, 1\}$.

In reality, it is quite restrictive to have the equality constraint, so we typically consider relaxed metrics:

$$\Delta DP(f) = |\mathbb{E}_{x\sim P_0} f(x) - \mathbb{E}_{x\sim P_1} f(x)| \tag{1}$$

$$\Delta EO(f) = \sum_{y\in\{0,1\}} |\mathbb{E}_{x\sim P_0} f(x) - \mathbb{E}_{x\sim P_1} f(x)| \tag{2}$$

where $P_a = P(\cdot|A = a)$.

The simplest way of optimizing the above problem is to regard it as a penalized optimization problem, and use Gap Regularization to regularize the fairness measurement:

$$\min_f \mathbb{E}_{(x,y)\sim P}[l(f(x), y)] + \lambda \Delta DP(f), \tag{3}$$

$$\min_f \mathbb{E}_{(x,y)\sim P}[l(f(x), y)] + \lambda \Delta EO(f), \tag{4}$$

where l is the classification loss. In [29], the authors show that small training values of $\Delta P(f)$ do not necessarily generalize well during evaluation. To improve generalizability, they introduce a data augmentation strategy via a dynamic form of group fairness metrics.

D. Dynamic Formulation of Fairness: Paths Between Groups

We follow [29], which considers a dynamic metric that measures the change of $\hat{Y}$ while transitioning gradually from P_0 to P_1. This dynamic formulation bridges two groups with an interpolator $T(x_0, x_1, t)$, which generates interpolated samples between x_0 and x_1 based on step t. Let $T : \mathcal{X}^2 \times [0, 1] \to \mathcal{X}$ be a function that is continuously differentiable for t w.r.t such that $T(x_0, x_1, 0) = x_0$ and $T(x_0, x_1, 1) = x_1$. For any differentiable function f, we have

$$\begin{aligned} &\Delta\mathrm{DP}(f) = \\ &\left| \int_0^1 \frac{d}{dt} \int f(\underbrace{T(x_0, x_1, t)}_{\text{interpolation}}) dP_0(x_0) dP_1(x_1) dt \right| =: \\ &\left| \int_0^1 \frac{d}{dt} \mu_f(t) dt \right|, \end{aligned} \tag{5}$$

$$\Delta\mathrm{EO}(f) = \sum_{y\in\{0,1\}} \left| \int_0^1 \frac{d}{dt} \int f(T(x_0, x_1, t)) dP_0^y(x_0) dP_1^y(x_1) dt \right|. \tag{6}$$

where we define $\mu_f(t) = \mathbb{E}_{x_0\sim P_0, x_1\sim P_1} f(T(x_0, x_1, t))$ *, the expected output of f with respect to* $T(x_0, x_1, t)$.

The interpolator T adopts the standard mixup by setting the linear interpolation in the input space: $T(x_0, x_1, t) =$

$tx_0 + (1-t)x_1$. The resulting smoothness regularizer has the following closed-form expression:

$$R^{DP}_{mixup}(f) = \int_0^1 \left| \int \langle \bigtriangledown_x f(T), x_0 - x_1 \rangle dP_0(x_0) dP_1(x_1) \right| dt. \quad (7)$$

$$R^{\text{EO}}_{\text{mixup}}(f) = \sum_{y \in \{0,1\}} \left| \int_0^1 \langle \bigtriangledown_x f(T), x_0 - x_1 \rangle dP^y_0(x_0) dP^y_1(x_1) \right| dt. \quad (8)$$

The regularizer can be easily optimized by computing the Jacobian of f on mixup samples. Here, the methods from [29] regularizes the expected inner product between the Jacobian on mixup samples and the difference $x_0 - x_1$. We can reformulate Gap Regularization (Equations 3 and 4) using Equations 7 and 8, i.e. enforce fairness via dynamic data augmentation of paths between groups:

$$(\text{FairMixup}) : \min_f \mathbb{E}_{(x,y)\sim P}[l(f(x), y)] + \lambda R_{\text{mixup}}(f). \quad (9)$$

IV. Experiment

A. Data Pre-processing

For a machine learning model, we hope that the labels of the training samples input for the model are as balanced as possible. Only in this way can good prediction performance be obtained. Therefore, we make the dataset balanced in a 50-50 split of non-CHD and CHD individuals. There are a sufficient number of samples to do this randomly and yield predictive results. There are 398,881 (reporting no MI or CHD) + 38,633 (reporting having MI or CHD) such that we can make a new 50-50 binary dataset of 23154 heart disease risk samples, and 23154 samples without heart disease.

For the diabetes records, we need to combine the pre-diabetics with the diabetics, and non-diabetics, or remove them entirely. We refer to the previous work [33], and add prediabetic individuals to the diabetic group, as it indicates potential progression towards diabetes and entails similar health risks [31]. Ultimately, we make a new 50-50 binary dataset of 33,998 diabetic risk individuals to 33,998 randomly selected non-diabetics. The respondent distribution of various racial groups in the processed datasets for CHD and diabetes is shown in Table II.

B. Experiment Setup

We use 3 methods for experiments: 1) FairMixup data augmentation method (Equation 9), 2) GapReg basic fairness regularization method (Equations 3,4), and 3) Without Fairness Regularization (WFR). We retrain each model 3 times and recorded the mean accuracy and fairness measurements. The dataset is randomly split into a training, validation, and testing sets with partitions of 60%, 20%, and 20%, respectively. The models are then selected according to the performance on the validation set.

TABLE II
Distribution of underrepresented races after data processing

Disease: MICHD		
Race	sensitive attribute $A = 1$	sensitive attribute $A = 0$
Black	3575	42733
Asian	815	45493
American Indian	941	45367
Native Hawaiian	275	46033
Hispanic	2118	44190
Disease: DIABETES		
Black	6623	61373
Asian	1443	66553
American Indian	1470	66526
Native Hawaiian	482	67514
Hispanic	3836	64160

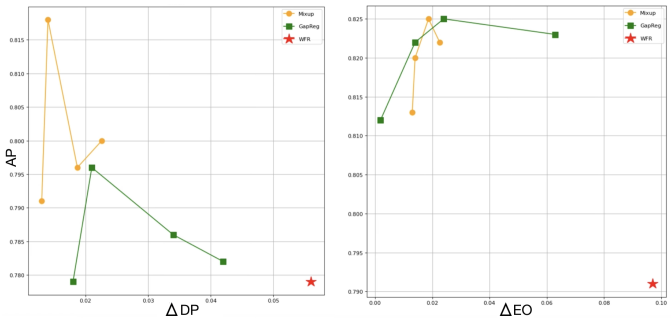


Fig. 2. Race: Black, Changes in the average precision (AP) observed with changes in fairness metrics (ΔDP and ΔEO).

C. Results

Fig. 2 to Fig. 6 show changes in Average Precision (AP) of the model across different fairness metrics DP and EO with application of the Fair Mixup and GapReg methods for the underrepresented racial groups in the dataset (Black, American Indian/Alaskan Native, Asian, Native Hawaiian/Pacific Islander, and Hispanic). The Fair Mixup data augmentation method could almost consistently achieve a higher AP for every racial minority group while minimizing ΔDP compared to the GapReg and WFR methods. An exception is the performance of the model on the Asian group, shown in Figure 4 (a), where the AP achieved is still higher than the baseline methods, while the trade-off with ΔDP is slightly worse compared to results for other races.

For the EO fairness metric, the Fair Mixup method also performs better than the other methods, as it achieves a high AP at relatively low ΔEO for most racial groups in comparison to performance from direct regularization of the EO measurement with GapReg. This indicates that Fair Mixup not only improves generalization of fairness constraints, but overall accuracy of the model for predictions in underrepresented groups as well.

We also perform additional tests of our model for diabetes risk prediction using the mixup method and the WFR method. Individuals with diabetes have a high risk of developing CVD, as many of the same major risk factors are relevant for both conditions (obesity, dyslipidemia, hypertension, smoking, etc.) [32]. This also suggests that diabetes risk prediction algorithms

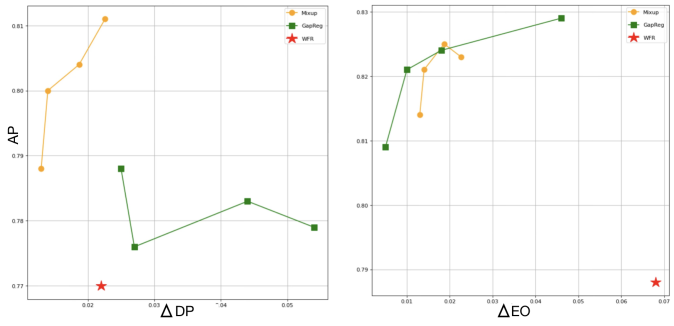


Fig. 3. Race: American Indian, Changes in the average precision (AP) observed with changes in fairness metrics (ΔDP and ΔEO).

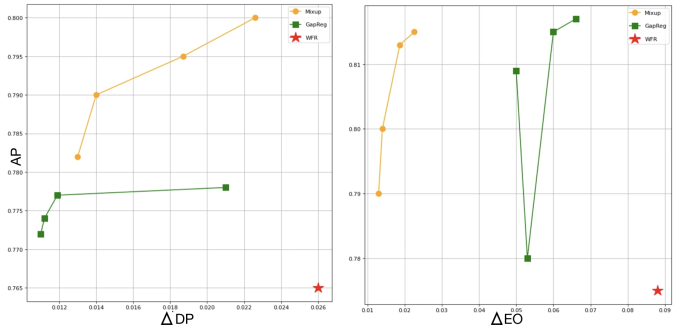


Fig. 4. Race: Asian, Changes in the average precision (AP) observed with changes in fairness metrics (ΔDP and ΔEO).

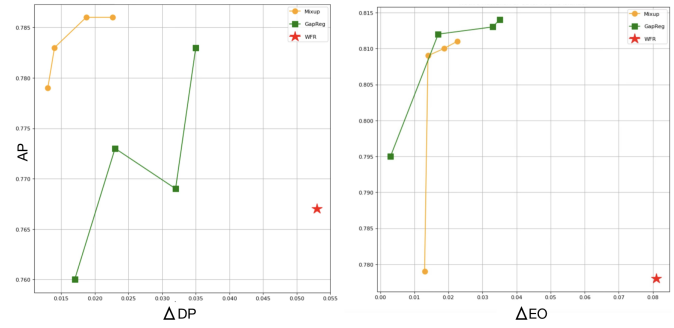


Fig. 5. Race: Native Hawaiian, Changes in the average precision (AP) observed with changes in fairness metrics (ΔDP and ΔEO).

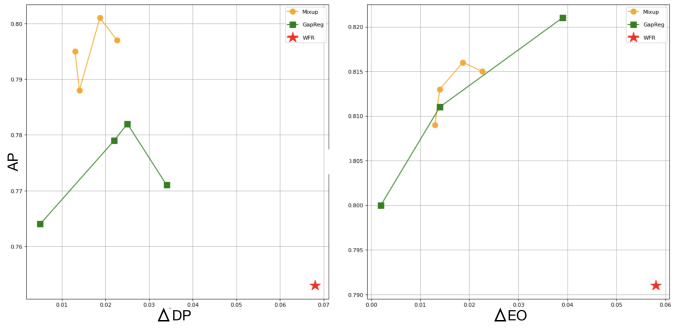


Fig. 6. Race: Hispanic, Changes in the average precision (AP) observed with changes in fairness metrics (ΔDP and ΔEO).

could face similar issues with bias when trained on imbalanced and unrepresentative datasets. The results for these experiments are shown in Table III. For almost all underrepresented races, the mixup method achieves comparably high AP at lower ΔDP and ΔEO compared to the WFR method. There is slightly less AP with the mixup method for the Native Hawaiian and Hispanic groups, while for the Black, Asian, and American Indian groups, there is a general improvement in AP for both group fairness metrics. Overall, the results are similar to those from the CVD risk prediction experiments, as fair mixup achieves a better trade-off between ΔDP or ΔEO and AP compared to baseline methods.

TABLE III
RESULTS FOR DIABETES PREDICTION

Disease: DIABETES				
Method:mixup				
Race	AP(DP)	ΔDP	AP(EO)	ΔEO
Black	0.787	0.017	0.787	0.001
Asian	0.784	0.015	0.788	0.001
American Indian	0.781	0.021	0.787	0.001
Native Hawaiian	0.776	0.017	0.762	0.001
Hispanic	0.795	0.019	0.782	0.001
Method:WFR				
Black	0.778	0.029	0.778	0.102
Asian	0.764	0.020	0.782	0.745
American Indian	0.770	0.039	0.776	0.086
Native Hawaiian	0.769	0.037	0.773	0.067
Hispanic	0.769	0.047	0.783	0.088

V. Conclusion

We employ a data augmentation technique called Fair Mixup to enhance the accuracy of ML models for cardiovascular disease risk prediction across different racial minority groups in health datasets, with the goal of promoting unbiased clinical algorithms for diverse populations, so that high-risk patients requiring medical interventions or treatment can be equitably identified. Experimental results indicate that the method is effective, demonstrating a relative increase in accuracy of about 2%-3% for most underrepresented minorities in the dataset while maintaining fairness. Therefore, utilization of Fair Mixup or related data augmentation methods could be an effective option for reducing biases in clinical risk prediction, by improving the generalization of models to all groups in the population, even with unrepresentative training datasets. In our experiment, we examined various features that are significantly involved in CVD risk, including hypertension, high cholesterol, BMI, smoking habits, and diet. However, our current methods do not differentiate between the contributions of each feature to fairness and accuracy. An development that could be made in future experiments is the implementation of dynamic feature selection for fairness and accuracy during the model training process, which could further improve both aspects. Another potential area to consider in future research is feature selection using feature decomposition methods to improve the model's fairness and accuracy in regard to racial factors.

REFERENCES

[1] K. H. Goh et al., "Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare," Nature Communications, vol. 12, no. 1, 2021. doi:10.1038/s41467-021-20910-4

[2] K. H. Goh et al., "Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare," Nature Communications, vol. 12, no. 1, 2021. doi:10.1038/s41467-021-20910-4

[3] I. Rojek, M. Kozielski, J. Dorożyński, and D. Mikołajewski, "AI-Based Prediction of Myocardial Infarction Risk as an Element of Preventive Medicine," Applied Sciences, vol. 12, no. 19, p. 9596, Sep. 2022, doi: 10.3390/app12199596

[4] Y. Zhao et al., "Using machine learning techniques to develop risk prediction models for the risk of incident diabetic retinopathy among patients with type 2 diabetes mellitus: A cohort study," Frontiers in Endocrinology, vol. 13, 2022. doi:10.3389/fendo.2022.876559

[5] M. Javaid, A. Haleem, R. Pratap Singh, R. Suman, and S. Rab, "Significance of machine learning in Healthcare: Features, pillars and applications," International Journal of Intelligent Networks, vol. 3, pp. 58–73, 2022. doi:10.1016/j.ijin.2022.05.002

[6] N. Norori, Q. Hu, F. M. Aellen, F. D. Faraci, and A. Tzovara, "Addressing bias in big data and AI for health care: A call for open science," Patterns, vol. 2, no. 10, p. 100347, 2021. doi:10.1016/j.patter.2021.100347

[7] A. Jindal, "Misguided artificial intelligence: How racial bias is built into clinical models," Brown Hospital Medicine, vol. 2, no. 1, 2022. doi:10.56305/001c.38021

[8] [1] M. A. Gianfrancesco, S. Tamang, J. Yazdany, and G. Schmajuk, "Potential biases in machine learning algorithms using electronic health record data," JAMA Internal Medicine, vol. 178, no. 11, p. 1544, 2018. doi:10.1001/jamainternmed.2018.3763

[9] I. Y. Chen et al., "Ethical machine learning in Healthcare," Annual Review of Biomedical Data Science, vol. 4, no. 1, pp. 123–144, 2021. doi:10.1146/annurev-biodatasci-092820-114757

[10] M. Barton, M. Hamza, and B. Guevel, "Racial equity in healthcare machine learning: Illustrating bias in models with minimal bias mitigation," Cureus, 2023. doi:10.7759/cureus.35037

[11] L. H. Kamulegeya et al., Using artificial intelligence on dermatology conditions in Uganda: A case for diversity in training data sets for machine learning, 2019. doi:10.1101/826057

[12] P. Burlina, N. Joshi, W. Paul, K. D. Pacheco, and N. M. Bressler, "Addressing artificial intelligence bias in retinal diagnostics," Translational Vision Science and Technology, vol. 10, no. 2, p. 13, 2021. doi:10.1167/tvst.10.2.13

[13] E. Puyol-Antón et al., "Fairness in cardiac magnetic resonance imaging: Assessing sex and racial bias in deep learning-based segmentation," Frontiers in Cardiovascular Medicine, vol. 9, 2022. doi:10.3389/fcvm.2022.859310

[14] E. O. Lopez, B. D. Ballard, and A. Jan, "Cardiovascular disease - statpearls - NCBI bookshelf," National Library of Medicine, https://www.ncbi.nlm.nih.gov/books/NBK535419/ (accessed Aug. 28, 2023).

[15] G. A. Roth et al., "Global burden of cardiovascular diseases and risk factors, 1990–2019," Journal of the American College of Cardiology, vol. 76, no. 25, pp. 2982–3021, Dec. 2020. doi:10.1016/j.jacc.2020.11.010

[16] J. M. Torpy, "Coronary heart disease risk factors," JAMA, vol. 302, no. 21, p. 2388, 2009. doi:10.1001/jama.302.21.2388

[17] E. Dritsas, S. Alexiou, and K. Moustakas, "Cardiovascular disease risk prediction with supervised machine learning techniques," Proceedings of the 8th International Conference on Information and Communication Technologies for Ageing Well and e-Health, 2022. doi:10.5220/0011088300003188

[18] S. Hageman et al., "SCORE2 risk prediction algorithms: New models to estimate 10-year risk of cardiovascular disease in Europe," European Heart Journal, vol. 42, no. 25, pp. 2439–2454, 2021. doi:10.1093/eurheartj/ehab309

[19] S. Dalal et al., "Application of machine learning for cardiovascular disease risk prediction," Computational Intelligence and Neuroscience, vol. 2023, pp. 1–12, 2023. doi:10.1155/2023/9418666

[20] P. Mody, A. Gupta, B. Bikdeli, J. F. Lampropulos, and K. Dharmarajan, "Most important articles on cardiovascular disease among racial and ethnic minorities," Circulation: Cardiovascular Quality and Outcomes, vol. 5, no. 4, 2012. doi:10.1161/circoutcomes.112.967638

[21] L. S. Mehta et al., "Cardiovascular disease risk factors in women: The impact of Race and ethnicity: A scientific statement from the American Heart Association," Circulation, vol. 147, no. 19, pp. 1471–1487, 2023. doi:10.1161/cir.0000000000001139

[22] A. Reeves et al., "Study selection bias and racial or ethnic disparities in estimated age at onset of cardiometabolic disease among midlife women in the US," JAMA Network Open, vol. 5, no. 11, 2022. doi:10.1001/jamanetworkopen.2022.40665

[23] E. Dal Canto, B. Farukh, and L. Faconti, "Why are there ethnic differences in cardio-metabolic risk factors and cardiovascular diseases?," JRSM Cardiovascular Disease, vol. 7, p. 204800401881892, 2018. doi:10.1177/2048004018818923

[24] K. Makhlouf, S. Zhioua, and C. P. "On the applicability of machine learning fairness notions." ACM SIGKDD Explorations Newsletter 23.1 (2021): 14-23.

[25] P.Gajane . "On formalizing fairness in prediction with machine learning." arXiv preprint arXiv:1710.03184 (2017).

[26] Z. Indre. "A survey on measuring indirect discrimination in machine learning." arXiv preprint arXiv:1511.00148 (2015).

[27] A. Hernandez-Garcia, P. König. "Data augmentation instead of explicit regularization." arXiv preprint arXiv:1806.03852 (2018).

[28] R. Balestriero, L. Bottou, Y. LeCun. "The effects of regularization and data augmentation are class dependent." Advances in Neural Information Processing Systems 35 (2022): 37878-37891.

[29] C. Chuang, Y. Mroueh. "Fair Mixup: Fairness via Interpolation." International Conference on Learning Representations. 2021.

[30] "CDC-2015 BRFSS survey data and Documentation," Centers for Disease Control and Prevention, https://www.cdc.gov/brfss/annual_data/annual_2015.html (accessed Aug. 30, 2023).

[31] X. Cai et al., "Association between Prediabetes and risk of all cause mortality and cardiovascular disease: Updated meta-analysis," BMJ, p. m2297, 2020. doi:10.1136/bmj.m2297

[32] I. Martín-Timón, "Type 2 diabetes and cardiovascular disease: Have all risk factors the same strength?," World Journal of Diabetes, vol. 5, no. 4, p. 444, Aug. 2014. doi:10.4239/wjd.v5.i4.444

[33] https://www.kaggle.com/code/alexteboul/diabetes-health-indicators-dataset-notebook