Challenges of Trustable AI and Added-Value on Health
B. Séroussi et al. (Eds.)

doi:10.3233/SHTI220468

Preprocessing to Address Bias in Healthcare Data

Emel SEKER[a,b,1], John R. TALBURT[a], Melody L. GREER[b]
[a] *University of Arkansas at Little Rock, USA*
[b] *University of Arkansas for Medical Sciences, USA*

Abstract. Multimorbidity, having a diagnosis of two or more chronic conditions, increases as people age. It is a predictor used in clinical decision-making, but underdiagnosis in underserved populations produces bias in the data that support algorithms used in the healthcare processes. Artificial intelligence (AI) systems could produce inaccurate predictions if patients have multiple unknown conditions. Rural patients are more likely to be underserved and also more likely to have multiple chronic conditions. In this study, data collected during the course of care in a centrally located academic hospital, multimorbidity decreased with rurality. This decrease suggests a bias against rural patients for algorithms that rely on diagnosis information to calculate risk. To test preprocessing to address bias in healthcare data, we measured the amount of discrimination in favor of metropolitan patients in the classification of multimorbidity. We built a model using the biased data to test optimum classification performance. A new unbiased training data set and model were created and tested against unaltered validation data. The new model's classification performance on unaltered data did not diverge significantly from the performance of the initial optimal model trained on the biased data suggesting that bias can be removed with preprocessing.

Keywords. Data Bias, Artificial Intelligence, Underserved populations

1. Introduction

A rapid shift to data-centric processes punctuates recent history. Yet, healthcare has lagged behind other sectors like banking and retail, where machine learning and artificial intelligence were incorporated into the workflow decades ago. However, artificial intelligence is being incorporated into healthcare and healthcare-related settings such as insurance, population health, EHRs, disease screening, and clinical decision support systems (CDS) [1,2]. These computations are powered by data collected in the course of care. AI and real-world evidence can add value to patient care. Scrutiny will be required, however, as these tools are added into the healthcare process because our world is rife with examples of bias, and these are unfortunately captured in our data. We may be unknowingly propagating or even amplifying bias [3]. When an algorithmic prediction is based on biased information, it will result in biased predictions.

Comorbidity indices are common clinical data consumers used for risk adjustment based on patient characteristics [4]. Multimorbidity, having a diagnosis of two or more chronic diseases, is a risk factor for adverse clinical outcomes. Multi morbidity is known

[1] Corresponding Author, Emel SEKER, College of Medicine, Department of Biomedical Informatics, University of Arkansas for Medical Sciences, United States; E-mail: ESeker@uams.edu.

to have wide-ranging consequences and associations with poor outcomes, including decreased quality of life, psychological distress, more extended hospital stays, more postoperative complications, and a higher cost of care, ultimately resulting in higher mortality. Data from the 2018 National Health Interview Survey (NHIS) reported that 27.2% of US adults had multiple chronic conditions [5]. Clinical decision support systems need to incorporate complex information into risk analysis [2], but it isn't equally available for all populations. For example, when underserved Arkansas' rural communities were studied for diabetic neuropathy status, it was found that 79% had not been diagnosed with DPN (Diabetic Peripheral Neuropathy) among the patients with peripheral neuropathy symptoms [6]. Rural patients are, however, more likely to be underserved and also more likely to have multiple chronic conditions [7].

In this study, we examine bias in EHR data. We (1) measure the amount of discrimination, (2) de-bias the data by re-balancing class labels, and then (3) compare the pre-and post-processed modeling results.

2. Methods

An integrated data set was generated by appending zip code level data to 19,367 EHR records of patients with chronic diseases (asthma, diabetes, heart disease, congestive heart failure, coronary artery disease, heart attack, stroke) from the University of Arkansas for Medical Sciences Clinical Data Warehouse (AR-CDR) [8].

Patients are stratified by risk in order to receive care at the appropriate level of need and to produce the most optimal outcome for the patient. We modeled a simplified risk predictor to study how preprocessing data to remove bias impacts predictions. Because patients are evaluated for risk using primary vital signs (temperature, respiration, and heart rate), these were used as baseline clinical features along with demographic features (i.e., race, gender, geographic location).

The outcome variable was the class label *multimorbidity*. Urban, rural residence was designated as the sensitive attribute because multimorbidity decreased with rurality, as shown in Figure 1. This decrease suggests a bias against rural patients for algorithms that rely on diagnosis information to calculate risk. Geographic location data was appended in the form of Rural-Urban Commuting Area (RUCA) codes. Rural-Urban Commuting Area codes are indicators of the population level of a patient's geographic home location, which are generated using the United States Census Bureau data [9].

RUCA codes range from 1 to 9, indicating progressively more rural areas. Because the codes that are 6 or less indicate metropolitan areas and those equal to 7 or above indicate rural areas, they were binned into urban and rural categories accordingly.

The measure of bias against rural patients was measured as shown in Eq. (1) where: D is discrimination or bias, s is the sensitive attribute (rural residence condition), $\bar{s}$ is the sensitive (metropolitan residence condition). Resulting in D = 0.063, meaning that the data is biased in favor of metropolitan patients, 6.3% are assigned a more complicated multimorbid status than the rural patients.

$$D = \frac{\bar{s} \wedge multimorbidity}{\bar{s}} - \frac{s \wedge multimorbidity}{s} \tag{1}$$

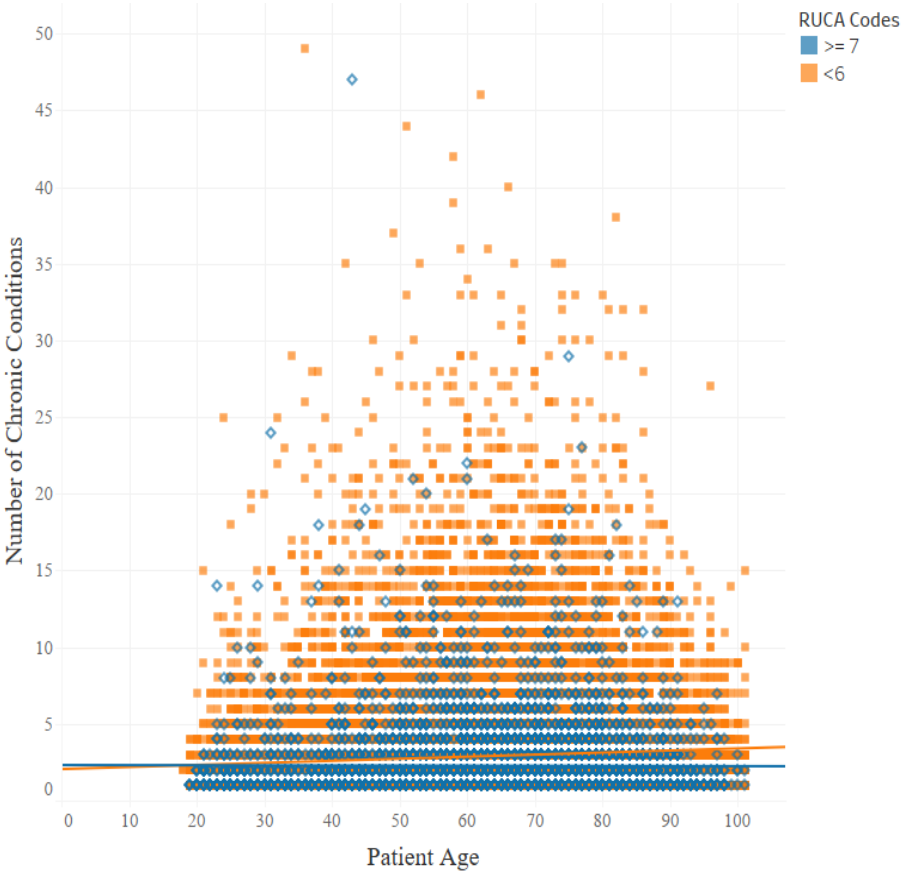

Figure 1. Trendlines show a 6.3% bias in multimorbidity increase with age for urban but not rural patients.

The data was then separated into a 60%/40% training and validation set, and we built a model using the biased data to test optimum classification performance. To address the bias in the data, the class labels of multimorbidity and no-multimorbidity were changed for *M* ranked patient rows in the training set as shown in Eq. (2). Ranking was done using logistic regression to determine the probability of *multimorbidity* associated with each patient. Urban patients with the lowest probability of being classified as multimorbid were 'demoted' to *no-multimorbidity*, and rural patients with the highest probability of being classified as multimorbid were 'promoted' to *multimorbidity*. To maintain the balance between the two classes, promotion and demotion were done at the same time for each of 66 rows, as shown in Table 1. [10]

$$M = \frac{(s \; x \; \bar{s} \wedge multimorbidity) - (\bar{s} \; x \; s \wedge multimorbidity)}{s + \bar{s}} \tag{2}$$

A logistic regression model was then learned using the debiased training data and tested on the unaltered validation set. A logistic regression model was also learned using the unaltered training data with the sensitive attribute removed for comparison. We compared the results of these three methods in model classification performance.

Table 1. The data was debiased by rebalancing the multimorbidity class labels for M rows.

		Debiased Data - Number of Chronic Conditions	
Original Data - Number of Chronic Conditions		*no* Multimorbidity	Multimorbidity
	no Multimorbidity	5840	66
	Multimorbidity	66	13395

3. Results

We built a model using the biased data to test optimum classification performance. A new unbiased training data set and model were then created and tested against unaltered validation data. The new model's classification performance on unaltered data did not diverge significantly from the performance of the initial optimal model trained on the biased data suggesting that bias can be removed with preprocessing.

The initial bias or discrimination measurement was just over 6%, suggesting that patients from a metropolitan area had an advantage in that they would be evaluated at a higher risk stratification than their rural counterparts. This bias is unexpected because rural patients are more likely to have multiple chronic illnesses. Reclassifying ranked patients in a training set removed the discrimination and had a negligible impact on classification performance. Because removing sensitive attributes is also a potential solution, we compared these results with modeling using a training set with the sensitive attribute completely removed. This model was tested on unaltered validation data, and classification performance wasn't significantly different, as shown in Table 2.

Table 2. The AUC was measured and used to compare the classification performance of models built using data that was (1) unaltered, (2) debiased, and (3) had the sensitive attribute removed. The AUC was robust to debiasing techniques indicating that preprocessing can be applied without negatively impacting model performance in at least some cases.

	Unaltered	**Debiased**	**SA Removed**
Test	0.7033	0.7042	0.7031
Validation	0.7081	0.7088	0.7080

4. Discussion

Real-world clinical data is important for clinical decision-making and valuable when used within artificial intelligence algorithms. However, real-world data has everyday biases imprinted within it and can preserve and even amplify health disparities. Underserved rural populations are less likely to get needed healthcare due to distance, costs, and poor insurance coverage leading to underdiagnosis of illnesses even though they are more commonly affected by chronic conditions.

To study this bias problem within healthcare data, we have analyzed and preprocessed a real-world data set of patients with chronic conditions from geographically disparate locations. We have chosen to preprocess because it prepares the data set for any following modeling and does not need to be repeated for each new classifier. When correcting for bias it is essential to maintain the integrity of the data set for it to still be useful for prediction. We also tested the removal of the sensitive attribute

altogether. Each of these tests produces similar AUC results. Comparable AUC results indicate that classification bias can be removed while maintaining strong classification performance.

5. Conclusion

Our work has resulted in a desirable, in fact a necessary, outcome of stable prediction capability with reduced bias. Classification performance was not altered significantly in any of these cases, which suggests debiasing can be conducted without a drastic negative effect on predictive modeling. Although removing the sensitive attribute did not degrade classification performance, it is potentially detrimental because there are often multiple features associated with the sensitive attribute in question. This can continue to produce bias even in the absence of sensitive information.

Debiasing techniques may have unexpected downstream consequences that need to be evaluated. Further research is required in a broader range of institutions and sensitive attributes. We believe this is an essential first step in debiasing healthcare data used for algorithmic prediction and directly impacts patient health outcomes.

Funding Acknowledgment and Consent

Patients data used were obtained under IRB approval (IRB# 261145) at the UAMS.

References

[1] Makino M, Yoshimoto R, Ono M, Itoko T, Katsuki T, Koseki A, Kudo M, Haida K, Kuroda J, Yanagiya R, Saitoh E, Hoshinaga K, Yuzawa Y, Suzuki A. Artificial intelligence predicts the progression of diabetic kidney disease using big data machine learning. Sci Rep. 2019 Aug 14;9(1):11862. doi: 10.1038/s41598-019-48263-5. PMID: 31413285; PMCID: PMC6694113.

[2] Fraccaro P, Arguello Casteleiro M, Ainsworth J, Buchan I. Adoption of clinical decision support in multimorbidity: a systematic review. JMIR Med Inform. 2015 Jan 7;3(1):e4. doi: 10.2196/medinform.3503. PMID: 25785897; PMCID: PMC4318680.

[3] Seyyed-Kalantari L, Zhang H, McDermott MBA, Chen IY, Ghassemi M. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. Nat Med. 2021 Dec;27(12):2176-2182. doi: 10.1038/s41591-021-01595-0. Epub 2021 Dec 10. PMID: 34893776; PMCID: PMC8674135.

[4] Alonso-Morán E, Nuño-Solinis R, Onder G, Tonnara G. Multimorbidity in risk stratification tools to predict negative outcomes in adult population. Eur J Intern Med. 2015 Apr;26(3):182-9. doi: 10.1016/j.ejim.2015.02.010. Epub 2015 Mar 6. PMID: 25753935.

[5] Boersma P, Black LI, Ward BW. Prevalence of Multiple Chronic Conditions Among US Adults, 2018. Prev Chronic Dis. 2020 Sep 17;17:E106. doi: 10.5888/pcd17.200130. PMID: 32945769; PMCID: PMC7553211.

[6] Wang W, Balamurugan A, Biddle J, Rollins KM. Diabetic neuropathy status and the concerns in underserved rural communities: challenges and opportunities for diabetes educators. Diabetes Educ. 2011 Jul-Aug;37(4):536-48. doi: 10.1177/0145721711410717. PMID: 21750334.

[7] CDC, Rural Health, in: *Preventing Chronic Diseases and Promoting Health in Rural Communities*, 2019.

[8] T.R. Institute, Clinical Data Repository (AR-CDR), in, 2022.

[9] USDA Economic Research Service, Rural-Urban Commuting Area Codes, in: *Economic Research Service*, U.S. Department of Agriculture, 2020.

[10] F. Kamiran and T. Calders, Classifying without discriminating, in: *2009 2nd international conference on computer, control and communication*, IEEE, 2009, pp. 1-6.