Potential Biases in Using Machine Learning For Healthcare Applications

Arun Pa Thiagarajan [1]

[1]Indian Institute of Technology, Madras

Abstract

Big data and Machine Learning (ML) systems find numerous applications in the healthcare sector like improving patient care, predicting risk scores etc. One of the main challenges in applying ML techniques for healthcare applications is detecting and mitigating bias. Bias stems from experiments which do not consider complete factors about the data generating process or choices made in the predictive algorithm, thereby resulting in solution techniques which are discriminative to marginalized groups or not achieving the intended usage goal. In this work, we identify potential causes of bias in healthcare applications and discuss directions on detecting and mitigating bias which can ultimately help in creating an equitable healthcare system.

Introduction

Machine Learning (ML) techniques are increasingly used in healthcare for various purposes like medical diagnosis, predicting healthcare costs etc. It is imperative that an algorithm deployed in the healthcare sector makes fair decisions. But machine learning algorithms are vulnerable to bias and systematic-errors. ML-based systems have been shown to demonstrate bias in various applications like machine translation [5], image classification [19] etc. Socially, bias in an algorithm can be defined as an algorithm being unfair to a sub-group or (un)privileged population and statistically, bias can be defined as a deviation from the true distribution which an estimator is trying to estimate [15].

Broadly, [16] identifies two sources of bias in a predictive system: bias arising from training data and bias arising from algorithms. Evidence suggests that healthcare systems also demonstrate these biases [8]. The presence of bias in healthcare systems increases societal disparities between (un)privileged groups of population and creates allocation harms. Further, when the biased machine learning models influence clinical practices, it creates an implicit feedback-loop which can aggravate the issues caused by existing bias in the system. In the following sections, we highlight potential causes of biases in healthcare system and discuss directions to mitigate them.

Bias due to data

Machine Learning algorithms are data-driven. They use training data to learn about the patterns in a sample and generalize it for the unseen population [13]. Thus, the quality of the predictions are tightly coupled to the quality of training data. When the underlying training data is biased, the algorithms trained on it will learn the biases present in training data along with the data patterns and reflect the same in their predictions.

Representation Bias

Most samples of data used in machine learning systems are drawn entirely from western, educated, industrialized, rich and democratic societies [9], thereby not representative of the human population as a whole. When the sample is skewed over a particular subgroup or when it does not cover the target population adequately, the arising bias is known as representation bias and the patterns learnt using such data will produce skewed outcomes. The lack/presence of too few/many healthy individuals in the dataset can also cause representative bias [6].

In [23], the authors found that an image classification model to detect pneumonia from chest X-rays trained on one sample failed to generalize to an another sample. Similarly, [20] discovered that pulse oximeters which measure blood oxygen saturation level by sending infrared light through the skin are racially biased because they were calibrated using white population. To assess the generalizability of the predictive model, [14] tested the model on patient samples from different populations, thereby building a model robust to representational bias.

Observational Error Bias

Observational error biases are seldom discussed when studying bias in machine learning algorithms. Observational error occurs when the measured values of a quantity differs from its true value due to calibration error or measuring device inaccuracies [1]. They manifest in predictive models as observational bias. For example, an inaccurate measure of blood cholesterol level due to faulty measuring equipment can cause observational error bias in the predictive model.

Missing Variable Bias

Missing variable bias are bias which occur due to loss of information caused by missing data points or other relevant variables [11] in the training data. As healthcare system generates lot of data [12], it is easy to miss out some important and relevant features for the problem in hand. To compare treatments or to find effect of an intervention, a healthcare researcher must take into account many different variables like treatments administered, comorbidities etc. Often, these variables are hard to collect and sometimes they remain missing, resulting in missing variable bias [22]. Another source of missing variables in healthcare records is the different level of access, practice or recording by patients and physicians. For example, healthcare records from a hospital can miss a diagnostic result as the hospital may be lacking in the laboratory equipment required to make the diagnosis [7].

In [4], a predictive model learns that having asthma as a precondition lowers the risk of dying from pneumonia. Analysis by the authors revealed that patients who had a history of asthma who also had pneumonia were directly admitted to ICU units and received aggressive care and hence they had lower risk of dying. If an additional variable to account for the level of care had been included, the model may instead have found that having asthma increases the risk of death. Hence, it is important to know about the missing variables like confounding factors and incorporate them into the predictive model.

Bias Due to Algorithm

Bias due to algorithms are bias caused by the algorithm itself and not by the training data [2]. Algorithms can produce biased outcomes due to choices made during training and other design choices of the algorithm. In the following subsection, we discuss two such biases which can occur in healthcare systems - measurement bias and learning bias.

Measurement Bias

Measurement bias are biases which are due to how we choose, collect or compute variables and labels used in predictive models [21]. In predictive models, it is common to use proxy variables as targets when ideal targets are not directly measurable or unavailable, like *creditworthiness* of a loan applicant, *risk score* of patients. The use of proxy turns problematic when the proxy is not a proper measure of the ideal target. In [17], the authors showed that when *future healthcare costs* is used as a proxy to predict future healthcare needs of a patient, Black patients assigned the same level of predicted risks by the algorithm were more sicker than White patients. The root cause of the bias is the wrong design choice of the target variable - *future healthcare costs* in the algorithm design. Since the black people faced more barriers to access healthcare historically, less money was spent on their healthcare needs and they generated lower healthcare costs even though they were more sick than white patients. If the design had focused on the right target variable which in this case is the illness of a patient, the predictive model would have been more equitable.

Learning Bias

The bias arising out of modeling choices like architecture, hyper-parameters, optimizer, and objective function is called as learning bias [10]. For example, it is well-known that a model which overfits to capture the regularities in the training data fails to generalize well to unseen data.

Conclusion

Given the overwhelming potential impact of AI applications in healthcare and the impact on human welfare which it creates, the cost of deploying biased predictive models is very high. In this work, we explored different types of bias which can creep into the healthcare system. We further showed examples of bias and potential harm which they can cause. One way to realize the promise of machine learning in the healthcare system is to improve the quality of data but often getting quality data is hard. To this end, in the future work, we aim to study different techniques by which bias in a machine learning models can be detected and resolved. We believe that building more explainable machine learning models and imbibing ideas from causality theory can play a big role in improving challenges presented by bias in the healthcare applications of machine learning. For example, causal analysis can help in controlling for confounding factors, bias due to representation [3], problems arising from missing data [18] etc.

References

[1] Alaa Althubaiti. Information bias in health research: definition, pitfalls, and adjustment methods. *Journal of multidisciplinary healthcare*, 9:211, 2016.

[2] Ricardo Baeza-Yates. Bias on the web. *Communications of the ACM*, 61(6):54–61, 2018.

[3] Elias Bareinboim and Judea Pearl. Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences*, 113(27):7345–7352, 2016.

[4] Rich Caruana, Yin Lou, Johannes Gehrke, Paul Koch, Marc Sturm, and Noemie Elhadad. Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1721–1730, 2015.

[5] Joel Escudé Font and Marta R. Costa-jussà. Equalizing gender biases in neural machine translation with word embeddings techniques, 2019.

[6] Anna Fry, Thomas J Littlejohns, Cathie Sudlow, Nicola Doherty, Ligia Adamska, Tim Sprosen, Rory Collins, and Naomi E Allen. Comparison of sociodemographic and health-related characteristics of uk biobank participants with those of the general population. *American journal of epidemiology*, 186(9):1026–1034, 2017.

[7] Marzyeh Ghassemi, Tristan Naumann, Peter Schulam, Andrew L Beam, Irene Y Chen, and Rajesh Ranganath. A review of challenges and opportunities in machine learning for health. *AMIA Summits on Translational Science Proceedings*, 2020:191, 2020.

[8] Marzyeh Ghassemi and Elaine Okanyene Nsoesie. In medicine, how do we machine learn anything real? *Patterns*, 3(1):100392, 2022.

[9] Joseph Henrich, Steven J. Heine, and Ara Norenzayan. The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2-3):61–83, 2010.

[10] Sara Hooker. Moving beyond "algorithmic bias is a data problem". *Patterns*, 2(4):100241, 2021.

[11] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)*, 54(6):1–35, 2021.

[12] LB Minor. Harnessing the power of data in health. *Stanford Med. Heal. Trends Rep.*, 2017.

[13] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

[14] Zaid Nabulsi, Andrew Sellergren, Shahar Jamshy, Charles Lau, Edward Santos, Atilla P Kiraly, Wenxing Ye, Jie Yang, Rory Pilgrim, Sahar Kazemzadeh, et al. Deep learning for distinguishing normal versus abnormal chest radiographs and generalization to two unseen diseases tuberculosis and covid-19. *Scientific reports*, 11(1):1–15, 2021.

[15] Natalia Norori, Qiyang Hu, Florence Marcelle Aellen, Francesca Dalia Faraci, and Athina Tzovara. Addressing bias in big data and ai for health care: A call for open science. *Patterns*, 2(10):100347, 2021.

[16] Ziad Obermeyer, Rebecca Nissan, Michael Stern, Stephanie Eaneff, Emily Joy Bembeneck, and Sendhil Mullainathan. Algorithmic bias playbook. *Center for Applied AI at Chicago Booth*, 2021.

[17] Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464):447–453, 2019.

[18] Judea Pearl. Theoretical impediments to machine learning with seven sparks from the causal revolution. *arXiv preprint arXiv:1801.04016*, 2018.

[19] Shreya Shankar, Yoni Halpern, Eric Breck, James Atwood, Jimbo Wilson, and D. Sculley. No classification without representation: Assessing geodiversity issues in open data sets for the developing world, 2017.

[20] Michael W Sjoding, Robert P Dickson, Theodore J Iwashyna, Steven E Gay, and Thomas S Valley. Racial bias in pulse oximetry measurement. *New England Journal of Medicine*, 383(25):2477–2478, 2020.

[21] Harini Suresh and John Guttag. A framework for understanding sources of harm throughout the machine learning life cycle. In *Equity and Access in Algorithms, Mechanisms, and Optimization*, pages 1–9. 2021.

[22] David Thesmar, David Sraer, Lisa Pinheiro, Nick Dadson, Razvan Veliche, and Paul Greenberg. Combining the power of artificial intelligence with the richness of healthcare claims data: Opportunities and challenges. *PharmacoEconomics*, 37(6):745–752, 2019.

[23] John R Zech, Marcus A Badgeley, Manway Liu, Anthony B Costa, Joseph J Titano, and Eric Karl Oermann. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. *PLoS medicine*, 15(11):e1002683, 2018.

DAI 2022

arunppsg@gmail.com