IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# Addressing Bias In Machine Learning Algorithms Used For Ai In Healthcare

**Prajakta Kamble**,
Computer SCience,
SCMIRT, Bavdhan,
Pune, India.

**Manjusha Nagpure**,
Computer SCience,
SCMIRT, Bavdhan,
Pune, India.

Guided by,
**Dr. Archana Wafgaonkar**,
Assistant Professor,
SIBMT, Bavdhan
Pune India.

Guided by,
**Dr.Deepak Singh**,
Vice Principal,
SCIMRT, Bavdhan,
Pune, India.

**Abstract**—The research paper under the title "Addressing Bias in Machine Learning Algorithms Used for AI in Healthcare" focuses on the problem of bias within/across machine learning (ML) algorithms applied to the sphere of healthcare. Incorporating artificial intelligence as a tool in diagnosing diseases and tailoring treatment for patients ratchets up the risk of the negative impacts of bias from these models, to women, coloured people, and other minorities.

This paper also divides the major biases existing in the application of healthcare algorithms into selection bias, measurement bias and model bias. Such biases stem from using sketches and unbalanced data that are resulting in discriminated health care, wrong diagnosis, and late treatment for fragile groups of individuals. For example, several research works have indicated that issues of racism in algorithms mean that Black patients are less likely to be prompted to join high risk care programs than the white patient.

To minimize such biases' effects, the authors recommend the following strategies; data variety through augmentation, algorithmic fairness, monitoring the AI models' operations, and making the algorithms more open. It also includes the ethical issues being concerned with whether it is equitable, for patient and public benefit, appropriate and transparent and in compliance with the necessary regulations to apply AI in the healthcare sector.

According to the paper, there is a need to solve the problem of bias in the ML algorithm if the general healthcare system is to be made fair. When being fair, transparent, and representative the utilization of AI systems enhances patient experiences, increases general population trust in the AI, and is also in compliance with ethical and/or legal procedures.

**Keywords—**Bias in Machine Learning (ML), Healthcare AI, Selection bias, Measurement bias, Model bias, Algorithmic fairness, Data augmentation, Ethical AI in healthcare, Health disparities, Racial bias in AI, Misdiagnosis, Transparency in AI, AI regulation in healthcare, Equity in healthcare algorithms, Auditing ML models

## I. Introduction

In contemporary times, the inclusion of artificial intelligence (AI) in the health care system has the capacity to increase the quality of patient engagement as it improves the accuracy of diagnosis, health prescriptive measures and tailoring precise treatments to patients. The subfield of A I known as machine learning (ML) allows the processing of large amounts of data including medical data and the extraction of useful patterns which can aid in clinical decisions. However, as such technologies gain root in healthcare, the issues of equity and the inherent bias of some AI systems have gained prominence. As it relates to the use of ML models, there are disadvantages as biases influence who gets quality healthcare and who does not within the level of health systems affecting especially women of color, elders and racial minorities.

Systemic bias in AI used in health is attributable to the use of ML model datasets as data from which to make health decisions and these datasets will be representative of bias in health care in terms of access, treatment, and outcomes. For instance, datasets may capture a particular demographic in high or low fashion, resulting in unfair outcomes and advice. Other types of biases include sampling bias, caused by the method of data collection, algorithm bias or inherent bias in the model result interpretation. These biases in the context of healthcare – where the decision impacts a patient's health in one way or the other can lead to different negative health related effects, a patient may be diagnosed wrongly, his or her treatment delayed or even denied on the basis of such bias.

Bias in artificial intelligence systems not only presents a problem of technology as a problem to be solved, but it is also a question of principle. Aside from healthcare institutions, the providers of AI must guarantee application fairness in that the ML algorithms are made, trained, and tested for people of any color, ethnicity, or sex. The focus of this paper is on the sources of bias in healthcare AI, its effects on patients' outcomes, and ways to reduce bias in order to make AI useful in creating a fairer and more efficient health care system.

Bias in Machine Learning (ML), Healthcare AI, Selection bias, Measurement bias, Model bias, Algorithmic fairness, Data augmentation, Ethical AI in healthcare, Health disparities, Racial bias in AI, Misdiagnosis, Transparency in AI, AI regulation in healthcare, Equity in healthcare algorithms, Auditing ML models

## II. Types of Bias in Healthcare Algorithms

Categories of bias in ML into several types:

1)**selection bias**, whereby the data on which the model is trained is not census;

2) the **measurement bias** from the improper or inaccurate recording of variable measurements.

3)**model bias** by design which means that bias inherent in the technique leads to variations in outcomes for different kinds of patients.

Several types of bias present different effects on the precision and non-privilege of healthcare predictions while scholars of applied methods attempted to eliminate them through various datasets and progressive approaches to training models.

## III. Real-World Impact of Biased Algorithms

Healthcare algorithms contain race bias, meaning that Black patients were the least likely to be recommended for high-risk care programs than our White counterparts despite the fact that we have similar healthcare needs. They said the bias could be traced to the fact that, relying on health costs as a measure of health needs, the algorithm was discriminating against rural citizens. This example explains why unequal healthcare experience due to colonialism and race reproduces itself through the ML model and discriminating against patients.

Likewise, inadequate representation in the training data set is another problem posed by ML. In some cases, both male and female are given different predictions depending on the algorithm, although they were tested on nearly identical data; similarly, racial minorities and elderly people are too given lower probability estimates because datasets of clinical algorithms contain less of them. The implications include the underrepresentation in clinical trials leading to wrong diagnosis or treatment regimens.

## IV. Consequences of Bias

The presence of bias in ML models can have severe implications:

• Misdiagnosis or Delayed Treatment: Vested models could underestimate risks among specific populations, thus delaying or incorrectly diagnosing them.

• Inequality in Care: Unsupported minorities may also suffer from disparities in the kind of attention they receive concerning their health.

• Loss of Trust in Technology: If healthcare systems are seen to be aligned to some earths more than others, this may bring dis-trust to health care solutions that are powered by ML.

### Advantages

1. Improved Patient Outcomes
2. Increased Fairness and Equity
3. Enhanced Trust in ML Systems
4. Regulatory Compliance
5. Broader Applicability of Models
6. Ethical Deployment

### Disadvantages

1. Increased Complexity in Model Development
2. Data Limitations
3. Higher Costs
4. Balancing Fairness and Accuracy
5. Constant Monitoring and Auditing
6. Lack of Standardization

## V. Objectives

1. To ensure fairness in patient treatment
2. To improve healthcare outcomes for underrepresented groups
3. To increase the accuracy and reliability of ML models
4. To promote the ethical use of technology
5. To comply with regulatory standards
6. To build trust in AI and ML healthcare solutions
7. To foster transparency and accountability
8. To promote data diversity and representation

9. To support the generalizability of ML models
10. To drive continuous improvement and monitoring

## VI. Strategies for Mitigating Bias

It has become possible to identify as many as there are guidelines or suggestions on how to minimize or completely do away with bias in healthcare related ML algorithms. There one is a data augmentation approach in which researchers intentionally include a more satisfactory diverse data sample. This helps to eliminate cases where the model has been trained on some specific kind of patients and their record, and therefore it shall be biased in predicting such records.

Another impressive technique is Algorithmic fairness, for instance re-weighting and changing model objectives in order to narrow the gap between disparate demography groups. Every one of the used techniques has the goal to minimize the between population variation and bring the settlements most favorable for all the samples of the data.

One other worthwhile best practice that needs to be discussed in this regard is the practice of auditing the ML models as frequently as is possible to establish the presence of bias. In auditing, the model developed is used on different types of data and results are compared based on the demographic factor. Therefore, when evaluating the model's outcomes, clinicians can determine the sources of new biases and restructure the model in order to make it more fair.

## VII. Challenges in Addressing Bias

In recent studies, the effectiveness of the mentioned methods has been closely examined. These methods are still not free from shortcomings. The main challenge is the problem of getting hold of good datasets, which are also similar to the target population. The data collected in the healthcare setting is possibly inconsistent, incomplete, or even only available for a defined population or region. Additionally, patient data is often private and shared, for instance HIPAA and GDPR prevent causality in certain data thus lack of diversity in data would cause bias.

Furthermore, equality and precision are workable options in any set up; nonetheless, numerous challenges arise in the middle. When working with Kleinberg et al. (2016) to improve the fairness that is attached to models, it was seen that this process slightly reduced the general accuracy because the data set that was original was imbalanced heavily. It is this balance between these two objectives that is perhaps a major area of current research.

## VIII. Case Study

A good example of bias in healthcare ML is the racial bias established in diagnosis models. An analysis revealed that machines developed to identify which patients are likely to need additional health care are treating black patients unfairly by giving them less attention than white patients. This was especially the case since the algorithm used historical data of healthcare spending, which retained previous inequalities in service utilization.

Solution of Addressing Bias:

1. AI bias in health care: a primary acquisition where machine learning algorithms demonstrate bias such that patients are discriminated against and other unfair treatments rendered on them. To remedy this problem, the following related strategies are considered crucial:.

2. The first is data accumulation, in a broad spectrum and across the population. This can be achieved by use of surveys from various demographic facets such as ages, gender, ethnicity, and economic standings. When a dataset is designed, therefore, some effort could be made to ensure that they are fair and represent the population of the interest so as to avoid.bias.

3. After the data has been collected, bias detection and analysis become a must. Machine learning models can be built and evaluated using statistical methodologies for comparing impacts of models across different groups of data. In order to detect and mitigate any discrepancies within model performance throughout and after a deployment, constant monitoring of the model is required.

4. The application of algorithmic fairness can be further improved and expanded allowing for the refinement of certain elements. This involves setting of models of fairness during training in order to reduce the amount of disparity that may be observed, use of adversarial training whereby models will be built to be more resistant to bias.

5. Another area which is also significant is the transparency of model development. It becomes equally important to build AI models that have explainability, so decision-makers ranging from healthcare givers to patients are able to comprehend the decision-making process. Thus, the involvement of multiple stakeholders during the development process can reveal hidden biases and create confidence in the discussed technology.

6. This can only be done if a system of auditing and a process of evaluation of models post-deployment for fairness are put in place. The creation of feedback mechanisms also helps healthcare professionals and the patient to raise concerns regarding bias to see adjustments.

7. There has to be a general and application-specific ethical normative framework and corresponding regulation to reduce biases and unseen unfairness in healthcare AI. Adherence to the regulations provided by such bodies as the FDA and WHO helps to achieve compliance to their standards by the developed systems.

8. One can learn both the identification of bias risk factors and measures necessary to prevent its implementation in the model's equation by increasing awareness and conducting training among data scientists and other healthcare employees. It can also inform patients through public participating and campaigning on the impacts of biases in healthcare innovation thus improving and developing an informed consumer.

9. Last but not the least; such hybrid approaches like the composited models where several algorithms are merged in order to reduce the competitive model's bias can be beneficial.

10. These areas can be addressed systematically to create a playbook for how the healthcare sector can build a better machine learning algorithm, for creating less bias and more benefit in patient care, while developing more trust in AI solutions. This paper concludes that sustained practice, involvement of an assortment of stakeholders, coupled with adherence to ethical principles form the bedrock of fashioning fair technologies in healthcare.

## IX. Ethical Considerations

The issue of bias states that with Machine Learning algorithms, it's not only a technical issue but also an ethical one. London (2019) argues that pre-existing algorithms are ethically unjust due to the nazism issue as it violates the justice principle that requires everyone should be treated fairly in all settings including the healthcare facilities. The ethical problems are therefore compounded by the fact that many ML models are opaque, and it may be hard to determine how the algorithm came to the conclusions that it has, let alone unbiased itself.

In the same paper, Char et al. (2018) also affirmed why there is the need for the implementation of interpretable and explainable ML in healthcare. Through interventions to make decision-making of algorithms more explainable, the healthcare givers and patients are in a position to understand and question regulators and the process that may have skewed outcomes due to biasing. This is important for the provision of trust in AI-based health care solutions.

## X. Future Scope

1. Fairness in this case can be expanded to encompass the creation of more complex fairness algorithms.
2. Ethical Dilemmas in Artificial Intelligence: The Integration of Ethical Framework into AI Model
3. Organizational Relations Among Professional Fields
4. In this step, diversification and broad representations of datasets across populations will have to increase.

## XI. Conclusion

Therefore, identifying bias that exists in machine learning algorithms in health care is an essential point towards attaining fairness in patient's treatment. It is valuable to construct source models for making specific health care forecasts that will work with both impartial and prejudiced information and work only on the fair data by implementing the following suite of solutions: using a diverse range of data sources, configuring and detecting both corpus- and algorithm-shift prejudices, and employing fairness-promoting algorithms. Concerning ethics and policy in regard to such technologies, it reduces the possibility that the same will violate set health standards or deny clients their right to privacy.

Besides, for the model to attract the end-users, methods that enhance explainability of the model and feedback loop will be employed with concern to healthcare professionals. With the practice of this approach, a repeated implementation of auditing will be made possible such that models and initiatives can undergo test runs where diverse data are needed and where performance is tested for enhanced and improved real world applications of healthcare ML. Therefore, I will remark that with the help of these strategies, the industry of healthcare will be able to improve its experience for subjugated groups and come closer to achieving equity for AI and machine learning to the same degree. Both these joint commitments toward establishing enhanced quality of care and increased trust of the public in the increasing application of technologies in the healthcare sector will go hand in hand.

## Acknowledgments

This research work was carried out by **Prajakta Kamble** and **Manjusha Nagpure**, department of computer science, SCMIRT, Bavdhan, Pune. The authors would like to place on record their appreciation to **Dr. Archana Wafgaonkar**, Assistant Professor at SIBMT, Bavdhan Pune, and **Dr. Deepak Singh**, Vice-Principal of SCMIRT Bavdhan Pune, for their administration support throughout the duration of this study. I must express my gratitude for their valuable contribution to the finish of this work.

## References

[1] Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys (CSUR), 54(6), Article No. 115, 1-35. https://doi.org/10.1145/3457607

[2] Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447-453. https://doi.org/10.1126/science.aax2342

[3] Gianfrancesco, M. A., Tamang, S., Yazdany, J., & Schmajuk, G. (2018). Potential biases in machine learning algorithms using electronic health record data. JAMA Internal Medicine, 178(11), 1544-1547. https://doi.org/10.1001/jamainternmed.2018.3763

[4] Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., Qiu, Y., Wang, J., Liu, Y., Wei, Y., Xia, J., Yu, T., Zhang, X., & Zhang, L. (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study. The Lancet, 395(10223), 507-513. https://doi.org/10.1016/S0140-6736(20)30211-7

[5] Zemel, R., Wu, Y., Swersky, K., Pitassi, T., & Dwork, C. (2013). Learning fair representations. Proceedings of the 30th International Conference on Machine Learning, PMLR, 28(3), 325-333.

[6] Vayena, E., Blasimme, A., & Cohen, I. G. (2018). Machine learning in medicine: Addressing ethical challenges. PLOS Medicine, 15(11), e1002689. https://doi.org/10.1371/journal.pmed.1002689

[7] Wiens, J., Saria, S., Sendak, M., Ghassemi, M., Liu, V. X., Doshi-Velez, F., Jung, K., Heller, K., Kale, D., Saeed, M., Ossorio, P. N., Thadaney-Israni, S., & Goldenberg, A. (2019). Do no harm: A roadmap for responsible machine learning for health care. Nature Medicine, 25(9), 1345-1354. https://doi.org/10.1038/s41591-019-0548-6

[8] Kleinberg, J., Ludwig, J., Mullainathan, S., & Rambachan, A. (2018). Algorithmic fairness. AEA Papers and Proceedings, 108, 22–27. https://doi.org/10.1257/pandp.20181018

[9] London, A. J. (2019). Ethical challenges in the use of machine learning in health care. Journal of Medical Ethics, 45(2), 131-134. https://doi.org/10.1136/medethics-2018-105141

[10] Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing machine learning in health care — addressing ethical challenges. New England Journal of Medicine, 378(11), 981–983. https://doi.org/10.1056/NEJMp1714229