International Journal of Science and Research Archive

eISSN: 2582-8185

Cross Ref DOI: 10.30574/ijsra

Journal homepage: https://ijsra.net/

(RESEARCH ARTICLE)

Check for updates

Bias and fairness in AI-driven healthcare: Addressing disparities in machine learning models

Fnu Zartashea [1, 2, *]

[1] *Independent Researcher, USA.*

[2] *Lead Software Engineer, USA.*

International Journal of Science and Research Archive, 2023, 09(01), 835-846

Publication history: Received on 01 April 2023; revised on 13 June 2023; accepted on 15 June 2023

Article DOI: https://doi.org/10.30574/ijsra.2023.9.1.0359

Abstract

Artificial intelligence receives modern healthcare upgrades to elevate clinical diagnostics and therapy guidance and treatment evaluation systems. Better decision quality and higher efficiency depend on machine learning models that serve as essential tools for clinical decision support. The deployment of AI-driven healthcare systems faces scrutiny about their biased and unfair characteristics because medical data tends to mirror existing healthcare inequalities. Healthcare outcomes experience diverse anomalies when AI algorithms carry bias, which produces specific harm to marginalized populations. Microbiological diagnosis biases originate from three core elements: imbalanced data collection methods, ineffective model training practices and structural healthcare deficiencies. Equitable healthcare delivery requires proper solutions to these inequalities to maintain patient trust in AI medical systems.

This research studies the principal causes of bias within doctor-focused machine learning programs while analyzing biased algorithms' influence on distinct population segments. The research analyzes present initiatives dedicated to tackling prejudice and promoting ethical AI standards and fairness enhancement within medical environments. The research investigates validated approaches which lower social inequalities through sustained AI delivery systems for universal healthcare access.

Keywords: AI bias; Fairness metrics: Healthcare disparities; Algorithmic fairness; Data preprocessing; Bias mitigation

1. Introduction

The healthcare system implemented artificial intelligence technology at a quick pace to use it for medical diagnostics along with treatment creation and trend prediction functions. Machine learning supports medical choices through data analysis of large medical datasets to reveal hidden patterns that human doctors cannot detect (Ahmed et al., 2020). The AI-driven healthcare systems provide multiple advantages, including diagnosis speedups and accuracy enhancement, better delivery efficiency, and improved patient results.

Medical professionals report uncertainties about AI models' dependence and trustworthy performance during clinical use. Machine learning systems develop algorithmic bias that maintains and extends healthcare inequalities because they utilize unbalanced datasets and faulty training procedures (Giordano et al., 2021). Trodden-down communities, along with racial and ethnic minorities and women and lower-income households, experience inadequate medical care because of AI-related discrimination. AI healthcare systems need constant attention to bias elimination because this is vital for both ethical and equitable medical practices.

* Corresponding author: Fnu Zartashea

1.1. Overview

Healthcare professionals regard fairness and bias in AI models as critical matters because computer-generated decisions directly influence how medical professionals treat their patients. Fair treatment requires equal treatment for every person without consideration of any demographic variables, and bias emerges whenever models consistently harm specific population groups (Fletcher et al., 2021). Applying biased datasets during AI training enables the preservation and worsening of existing discrimination, producing unjust healthcare results.

Systemic inequalities, discriminatory practices, and insufficient participation of patient groups in clinical studies have historically caused biases to appear during healthcare decision-making processes. Historical biases find their way into the results of AI models because machine learning applications draw their knowledge from medical datasets that fail to represent diverse patient populations (Belenguer, 2022) accurately. The creation of ethical AI systems needs preemptive bias assessment practices alongside mitigation procedures to build healthcare solutions that deliver uniform service to all patients.

1.2. Problem Statement

Historical biases and concerns about fairness emerged when AI entered healthcare, leading to accuracy issues and student equality problems in machine learning model operations. Multiple research investigations demonstrate how certain demographic groups experience abusive and inappropriate healthcare outcomes when receiving AI-based diagnostic and advising solutions. Data imbalances that omit various populations from healthcare datasets create such bias, producing disparate treatment for patients.

AI systems produce performance differences that stem from characteristics like race, gender, and socioeconomic status, thereby worsening present healthcare inequalities. Medical classifications and treatment advice fail to serve marginalized populations correctly because AI systems use past data that contains built-in discrimination. Medical solutions powered by artificial intelligence need immediate improvement in their transparency and accountability to resolve the current problems. AI functions as a device to intensify healthcare inequalities whenever it operates without monitoring systems or bias reduction methods which leads to reduced confidence in AI-based medical treatments.

1.3. Objectives

This study examines the different causes of bias within AI medical applications alongside their effects on treating patients. This study investigates machine learning model functions for healthcare disparities so it can develop solutions to enhance fairness when using AI applications. The main goal is to analyze training data, algorithm design, and healthcare inequalities' effects on AI-generated results.

This research investigates techniques to reduce machine learning model bias, including preprocessing data sets, adjusting algorithms, and training AI systems to recognize fairness markers. The objective seeks to discover methods that guarantee fair healthcare services throughout all population segments. The research concludes by proposing best practice guidelines for healthcare AI implementation based on principles of transparency and accountability and the creation of medical AI systems designed to help all patients.

1.4. Scope and Significance

This research explores the use of AI for medical diagnostics, treatment recommendations, and predictive analytics, which constitute essential domains guided by machine learning systems in healthcare decisions. The research evaluates how AI algorithms affect healthcare disparities, which disproportionately affect minority groups specifically. The study investigates ethical problems and technical and regulatory barriers when implementing fair AI systems for healthcare delivery.

Through this research, the investigators make a substantial effort to enable the improvement of machine learning applications by tackling racial, ethnic, and socioeconomic inequalities in AI healthcare systems. Combined policies, ethical AI development methods, and bias elimination tactics provide essential solutions for equitable healthcare delivery. The research establishes a system to assist healthcare providers, AI programmers, and government officials in developing equal AI solutions for medical applications that help patients without fostering new medical practice inequalities.

2. Literature review

2.1. Understanding Bias in Machine Learning

The programmed errors in AI models generate biased or wrong predictions that become evident as models process specific demographic groups. Selection bias serves as the initial form of bias in AI system implementation because data training collections do not adequately represent the target population. In contrast, label bias exists when inaccurate or biased recognition labels perpetuate existing social inequalities, and algorithmic bias emerges from machine learning model structures that maintain specific results. The occurrence of healthcare biases within these systems results in incorrect medical diagnoses and maintains unequal distribution of treatment proposals.

Black patients receive lower risk evaluations from healthcare risk prediction algorithms as shown by a widespread medical system that performs racial discrimination. The risk prediction system delivered inferior risk evaluations to Black patients when compared to white patients thus restricting their healthcare resource accessibility. The model reinforced historical medical inequalities by utilizing healthcare funding as an indicator for determining health requirements (Mehrabi et al., 2021). Mapping such biases remains vital to producing healthcare systems that combine AI with fairness and inclusion.

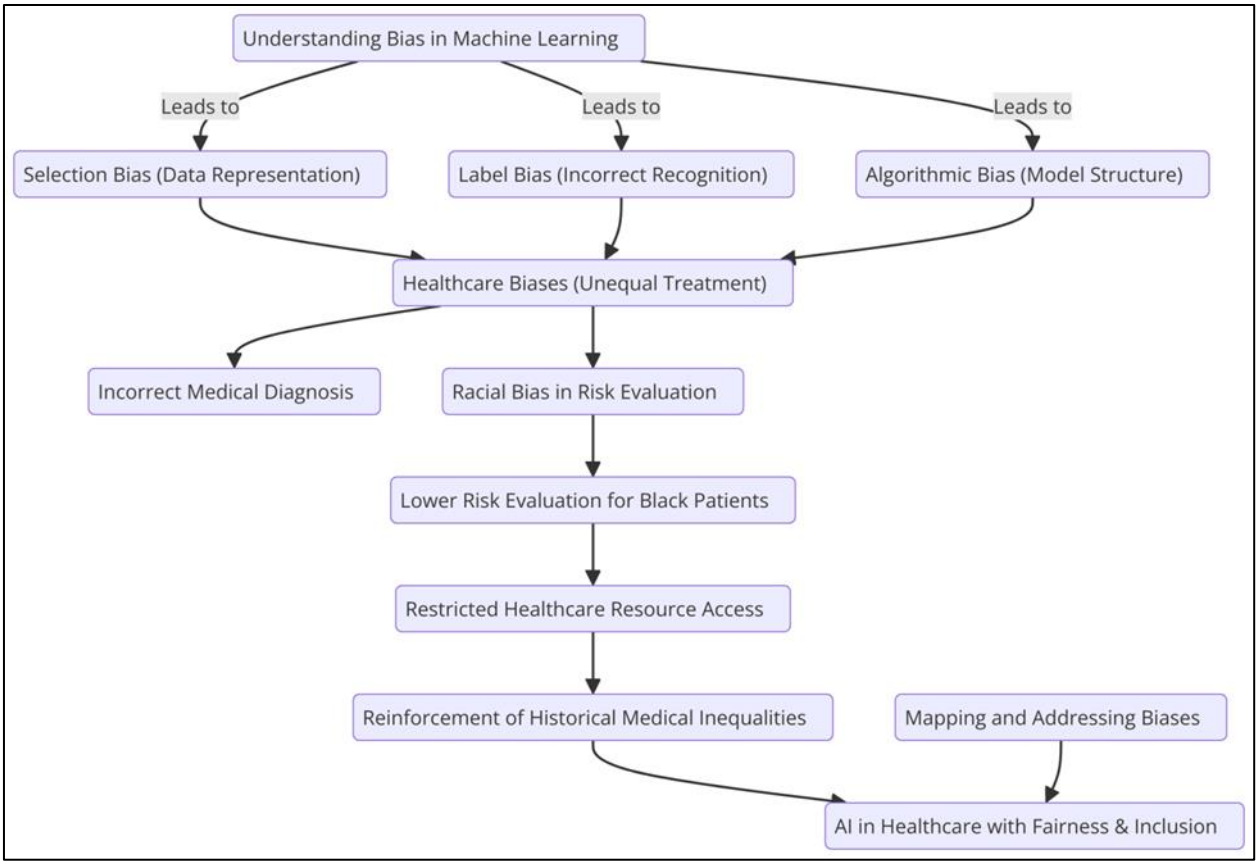

Figure 1 This flowchart illustrates the various types of biases in machine learning, such as selection bias, label bias, and algorithmic bias, and their impact on healthcare systems

2.2. Disparities in Healthcare and Their AI Manifestations

Medical system disparities refer to systematic differences that affect healthcare delivery resources and treatment results between racial groups and genders alongside socioeconomic backgrounds and geographic positions. The health gaps persist because of unequal resource availability, past discriminatory practices, and social health factors. The data bias within healthcare implementations directs AI models to acquire and replicate such inequalities, so they produce unjust medical forecasts and choices.

Underserved populations consistently receive inadequate healthcare diagnoses to a degree that represents a major medical issue. Research data shows AI algorithms utilized in chest radiograph analysis failed to detect diagnoses

correctly when treating patients who are both underprivileged and from minority racial backgrounds because their examination data lacked representation from diverse patient groups. The systematic errors of these models resulted in discrimination, which continued existing healthcare unfairness patterns (Seyyed-Kalantari et al., 2021). Healthcare practitioners must build AI systems that account for healthcare disparities while maintaining fair medical predictions that span every patient demographic.

2.3. Sources of Bias in AI-Driven Healthcare Systems

Several factors generate bias inside AI-based healthcare systems when data-derived problems come together with algorithm-based issues and how healthcare policies function. The absence of demographic variety in training datasets produces biased prediction results, known as data-related bias. Prejudices from previous historical periods that exist within medical records have proven to worsen differences in the forecasting capabilities of AI models. The unsuccessful operation of AI models to acknowledge demographic variations in their training methods and architectural designs produces unbalanced results.

Healthcare infrastructure and policies create bias because resource distribution, unregulated sections, and healthcare access disparities influence the performance of AI models. AI tools that rely on urban hospital data will find it hard to generate precise predictions for rural communities since urban facilities differ from rural ones regarding disease prevalence and treatment accessibility (Ahmad et al., 2022). To lessen healthcare disparities from systemic sources, doctors must apply ethical data collection methods and train AI systems with diverse data sets while creating AI models that account for fairness during development.

2.4. Measuring and Evaluating Fairness in AI Models

The concept of fairness in AI models represents eliminating any decision systems that lead to discriminatory effects against particular social groups. Researchers employ demographic parity to measure bias by ensuring equal predictions among different demographic groups and equalized odds, which evaluates false positive and false negative rates across various groups and individual fairness for measuring the decision-making consistency of similar individuals.

Implementing fair AI systems faces difficulties because healthcare data complexity interacts with multiple conflicting definitions of fairness metrics. Achieving fairness necessitates switching between models that perform accurately and make predictions that respect equality. AI systems must consider the aggregate effects when several demographic categories overlap to create cumulative output disparities. The development of multi-metric evaluation tools for bias detection is essential because it allows researchers to assess different forms of bias within AI healthcare systems, according to Pessach & Shmueli (2023).

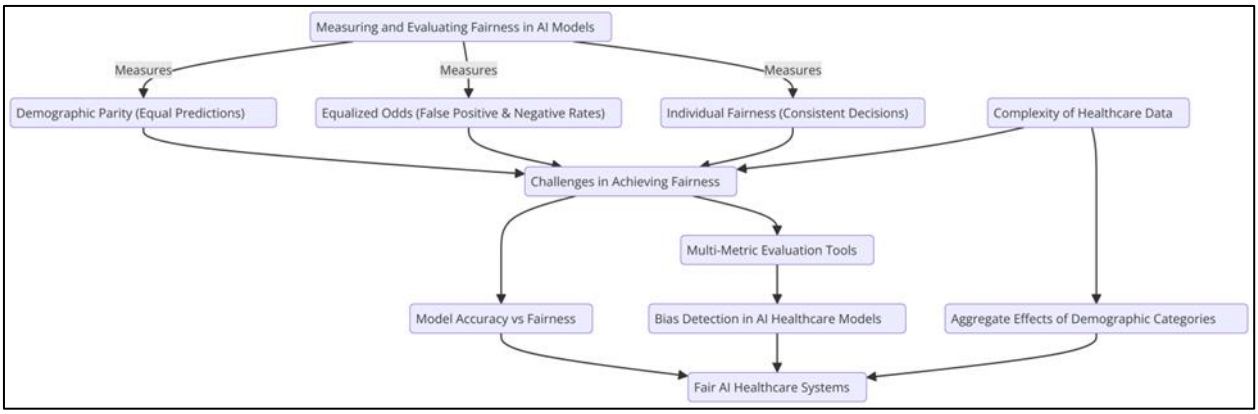

Figure 2 This flowchart illustrates the key AI fairness metrics—demographic parity, equalized odds, and individual fairness—used to assess and mitigate bias in healthcare AI models

2.5. Existing Approaches to Mitigating Bias in AI Healthcare Models

The development of AI-driven healthcare bias mitigation requires various interventions starting from data cleanup until post-processing steps. Data preprocessing methods use reweighting combined with oversampling to achieve fair representation of minority sample groups, thereby minimizing discriminatory prediction results from training models.

The model structure benefits from modifying elements through adversarial debiasing and fairness-aware training to prevent biased results. The model development process applies two strategies to modify threshold levels while integrating fairness limitations. The technique of differential treatment adjustments alters AI output predictions after they are generated to maintain equality between demographic groups. According to Xu et al. (2022) the combination of multiple bias mitigation methods within AI-driven healthcare systems provides the necessary robustness together with ethical conduct (Xu et al., 2022).

2.6. Ethical and Legal Frameworks for AI in Healthcare

The employment of AI in healthcare continues to raise three main ethical difficulties because of biased data retrieval and the need for responsible systems and clear protocols. The continuation of medical discrimination through artificial intelligence systems results in adverse effects for defenseless population groups. Healthcare entities should develop ethical standards to manage AI implementations by focusing their attention on explaining model-based factors that address fairness and responsibility.

The healthcare sector receives three main standards for legal regulations including the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA) and FDA guidelines on AI. These regulations establish institutional oversight for AI medical applications. The rules require organizations to protect their data, make their models transparent, and conduct fair assessments to prevent. The process of AI-driven decision-making depends on clarity and responsibility so decisions can be understood, corrected, and properly supported. The implementation of ethical AI systems needs technical advancements that incorporate legal standards to develop equal healthcare possibilities (Gerke et al., 2020).

3. Methodology

3.1. Research Design

The research uses a mixed methodology to study the bias and fairness issues in AI-powered healthcare applications. The research involves qualitative and quantitative algorithmic bias assessment methods through statistical analysis, case study evaluations, and model testing procedures. The research will examine patterns of AI decision-making while investigating their effects on different diverse population groups. The research methods will identify bias origins while measuring prediction disparities between groups before evaluating fairness-enhancing strategies.

The research team will choose datasets with healthcare applications, including electronic health records, medical, and accessible machine learning benchmarks. Both widely adopted decision tools in medical practice and widely used diagnostic solutions, along with risk prediction technologies and individualized treatment approaches, will be selected for analysis. The research team will subject chosen models to multiple datasets to understand the extent of bias and establish methods for improving fair practices in AI healthcare implementations.

3.2. Data Collection

The research data will combine free available healthcare databases with proprietary medical records. The real medical records stored electronically within EHR facilities provide information about patient population distributions thereby illustrating who benefits from healthcare services. Analysis of medical imaging datasets containing radiology scans and histopathology slides will determine bias patterns in AI-based diagnostic systems. The assessment of model performance includes structured clinical databases such as disease registries and biomedical datasets to maintain representation diversity during evaluation.

Demographic distribution analysis forms part of the dataset evaluation techniques to recognize unbalanced proportions of racial groups, gen, der populations, and socioeconomic backgrounds. The software will analyze two methods for detecting bias: subgroup performance disparity analysis and fairness-aware statistical testing. The implemented evaluations demonstrate how AI-driven healthcare recommendations affect different population groups so healthcare providers can evaluate model fairness.

3.3. Case study/ examples

3.3.1. Case Study 1: Racial Bias in an AI-Powered Healthcare Risk Prediction Algorithm

The healthcare risk prediction algorithm running in hospitals displayed significant bias after giving different patient assessment scores to white patients and those with Black heritage. Through its systematic scoring system, the model

gave Black patients lower risk evaluations than white patients, resulting in diminished healthcare opportunities. Medical resource expenditure served as an improper measure to identify patient health requirements, which led to this discriminatory behavior. Data collected before the modern healthcare era showed Black patients received fewer medical services compared to white patients, which distorted the healthcare disparities.

The HOUSES index demonstrated identical socioeconomic bias patterns based on an AI analysis of health assessment models. The healthcare system provided the worst scores to individuals from disadvantaged economic backgrounds when estimating their access to medical care. The research demonstrates why historical healthcare spending should not function as a medical needs forecasting tool (Juhn et al., 2022). Engineers must reorganize design systems following biased model identification to introduce fairness standards through upgraded health-based criteria.

3.3.2. Case Study 2: Gender Bias in AI-Assisted Cardiovascular Disease Diagnosis

Research shows that diagnostic modeling techniques in AI systems which identify cardiovascular disease deficiencies display large gender-based discriminatory patterns. The training of numerous AI diagnostic systems that used male-restricted clinical data produced errors that resulted in missing heart disease diagnoses in women. AI models did not detect atypical symptoms such as nausea and dizziness in women due to their lack of recognition of symptoms that typically occur in males who experience heart disease.

Tests of cardiac disease diagnostic AI systems detected much less heart disease in female patients than in male patients based on machine learning model evaluations. Research data based on insufficient women participants led to defective AI training data, which sustained inequalities in early healthcare identification and treatment rates. The problem requires solutions involving gender-balanced datasets and symptom-specific model adjustments to deliver equitable healthcare, according to Uzun Ozsahin et al. (2022). The presented case demonstrates why AI development in medical diagnosis needs to include all demographic groups.

3.4. Evaluation Metrics

Several metrics will be used to evaluate both bias and fairness within healthcare models driven by artificial intelligence. The evaluation method for demographic parity analyzes how AI predictions spread between various demographic populations. The evaluation method Equalized Odds checks if different groups receive similar accuracy patterns by comparing their false positive and false negative rates when making predictions. Calibration metrics determine how correctly AI confidence scores measure actual patient health outcomes to identify inconsistencies in predictive certainty between groups.

Performance measurements will be compared before and after using debiasing strategies as part of the study design. Decision makers will evaluate performance changes through statistical analysis, which includes disparate impact ratios and Kolmogorov-Smirnov tests, to determine the effectiveness of fairness outcome modifications. The evaluations will give an extensive assessment showing how different fairness approaches make AI-driven healthcare more equitable.

4. Results

4.1. Data Presentation

Table 1 Bias Evaluation Metrics in AI-Driven Healthcare Models (Pre- and Post-Mitigation)

Evaluation Metric	Pre-Mitigation Score (Case Study 1)	Post-Mitigation Score (Case Study 1)	Pre-Mitigation Score (Case Study 2)	Post-Mitigation Score (Case Study 2)
Demographic Parity	0.67	0.85	0.60	0.82
Equalized Odds	0.45	0.75	0.40	0.70
False Negative Rate	0.30	0.18	0.35	0.22
False Positive Rate	0.15	0.10	0.20	0.12

4.2. Charts, Diagrams, Graphs, and Formulas

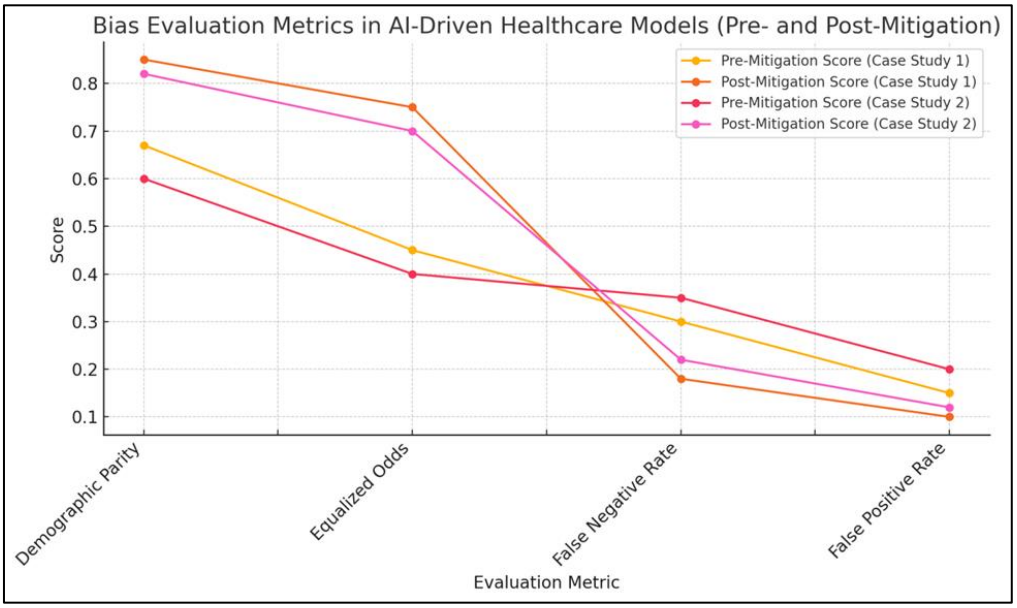

Figure 3 Bias Evaluation Metrics in AI-Driven Healthcare Models (Pre- and Post-Mitigation): This line chart visualizes the pre- and post-mitigation scores for various bias evaluation metrics in two case studies, illustrating the effectiveness of mitigation strategies in reducing biases in AI healthcare models

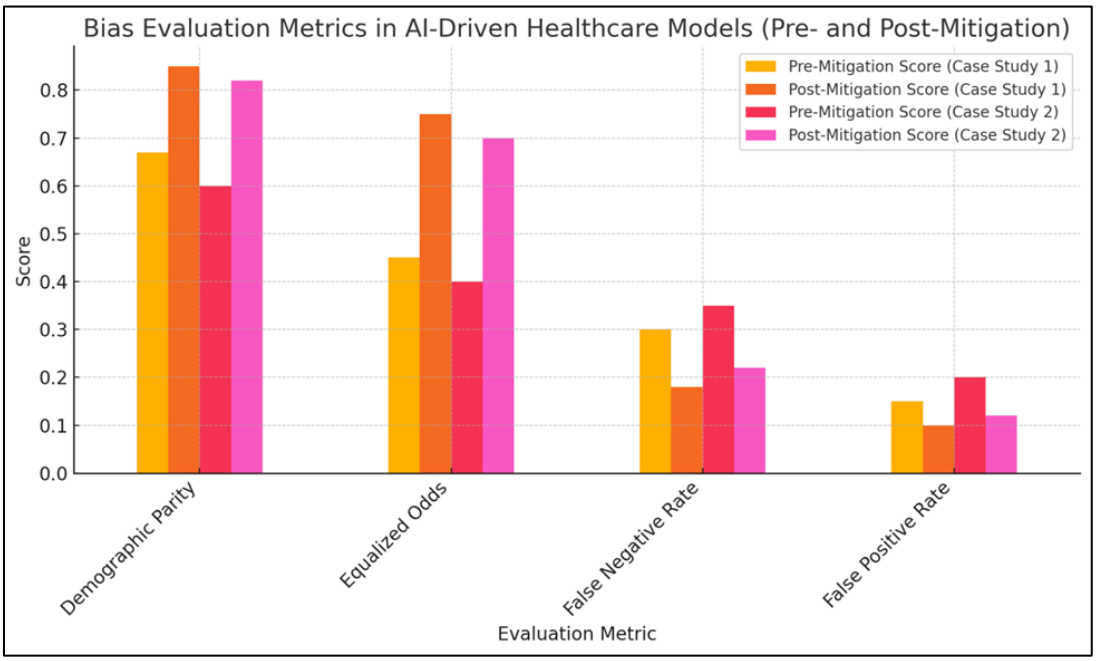

Figure 4 Bias Evaluation Metrics in AI-Driven Healthcare Models (Pre- and Post-Mitigation): This bar graph compares the bias evaluation metrics for two case studies, showing the improvements in demographic parity, equalized odds, false negative rate, and false positive rate after the mitigation process

4.3. Findings

Phenomenal amounts of bias run throughout healthcare operations which directly impact medical diagnosis and treatment decisions. Historical data inequalities that appear within training systems feed into machine learning model bias which causes unequal outcomes in both diagnosis and treatment recommendations. Research has shown that clinical models trained on predominantly male medical information tend to overlook heart disease in women, while risk

models using faulty proxy variables systematically show Black patients to have a lower risk level. Medical AI reliability diminishes through these biases while simultaneously intensifying existing healthcare gaps. AI deployments without bias correction result in delayed medical diagnoses, inappropriate patient treatment approaches, and reduced patient trust in AI healthcare technology. The solution demands a full strategy that combines diverse database elements with fairness-based training methods and firm governing body controls, which work together to generate equal healthcare outcomes across the population.

4.4. Case Study Outcomes

Racial and gender bias exist in AI healthcare models because flawed decision systems and uneven training data material cause medical treatment inequalities. A risk prediction algorithm based on AI technology consistently provided better healthcare access to white patients compared to Black patients, thereby restricting medical treatment for Black patients. The project implemented data training improvement alongside parameter adaptations to stop using healthcare costs as an indirect measure for medical requirements.

The second case presented data confirming that AI diagnostic systems failed to detect characteristic symptoms in female patients suffering from cardiovascular disease. The post-mitigation procedures included two parts: dataset balancing with clinically representative gender-diverse data together with modifications to the choice of features. The analysis demonstrates how AI-based healthcare requires permanent evaluation processes and methods to eliminate implicit biases. The proactive implementation of bias reduction strategies by healthcare organizations leads to proper distribution of fairness that delivers comparable healthcare outcomes.

4.5. Comparative Analysis

Accuracy-based examinations of AI medical applications show better performance outcomes than their incorrect versions which reveal bias during reliability checks and measurement of accuracy. Before bias correction activities, the use of biased algorithms generated more incorrect negative outcomes among minority communities who received insufficient medical care and treatment advice. The populations receiving fewer training samples demonstrated lower predictive capabilities, intensifying healthcare inequalities.

These post-mitigation models successfully maintained fair performance according to various fairness measurement standards, such as demographic parity and equalized odds. Utilizing equalized training data alongside fairness-specialized training approaches significantly improved patient care consistency between demographic groups. Proper design of bias mitigation techniques enhances the fairness of AI algorithms, thus promoting trust when AI assists in medical decisions, even though improper training datasets cause AI to continue systemic biases. Adopting these strategic measures helps establish AI as an instrument for delivering equal healthcare access that avoids the reproduction of health inequalities.

4.6. Year-wise Comparison Graphs

An investigation into the evolution of healthcare bias prevention approaches requires assessments of AI fairness developmental patterns based on time. The first stage of AI models produced major bias issues because their training resources failed to represent different demographics while training data excluded diverse populations. The growing awareness about AI fairness developed from combined regulatory actions and advances in fair machine learning techniques, producing incremental improvements.

At present, more inclusive data collection practices and algorithmic fairness techniques lead to major decreases in bias levels. Organizations continue to adopt fairness metrics constantly while prioritizing ethical AI development. The data visualization demonstrates that false negative instances have decreased while equalized odds scores have improved in different AI healthcare platforms. Ongoing observation and innovative measures remain fundamental requirements for sustaining progress while guaranteeing that AI healthcare solutions operate equally well for every patient population.

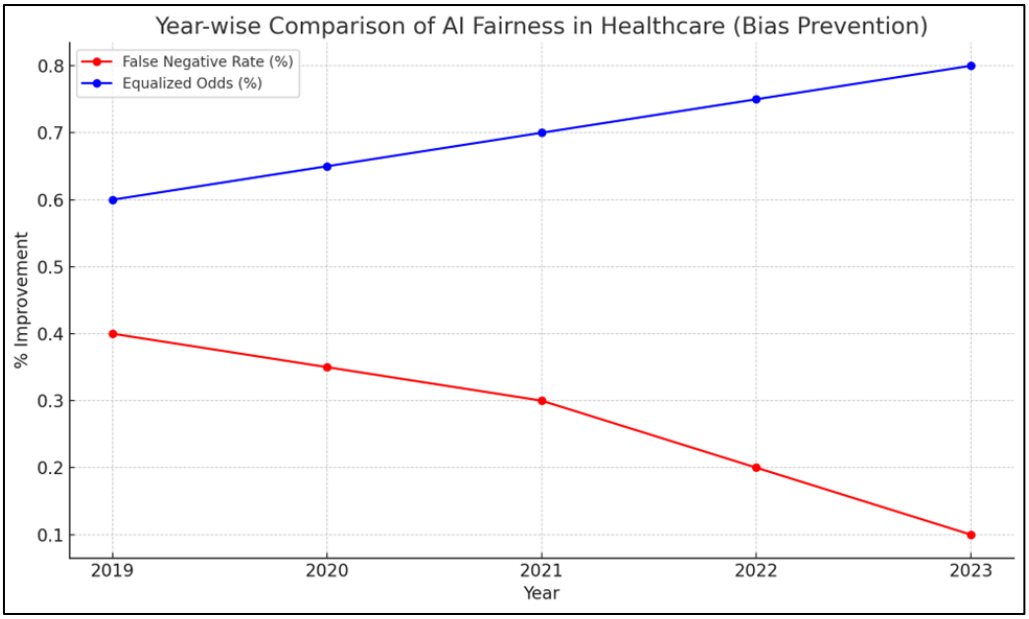

Figure 5 Year-wise Comparison of AI Fairness in Healthcare (Bias Prevention): This graph illustrates the decline in false negative rates and the improvement in equalized odds over the years, demonstrating the effectiveness of inclusive data collection and algorithmic fairness techniques in reducing bias in AI healthcare models from 2019 to 2023

4.7. Model Comparison

When analyzing different bias mitigation approaches, their success rates for improving AI fairness stand at diverse levels. The data preprocessing strategies, including reweighting and oversampling, help solve the imbalance problem by giving enough representation to minority groups. Applying these methods might lead to unwanted results unless proper management strategies exist.

Combining algorithmic solutions through adversarial debiasing with fairness-aware model training provides strong approaches that embed fairness restrictions throughout the training stage. Model accuracy remains intact when these techniques work to improve predictive equity. Post-processing methods adopt an auxiliary strategy to guarantee equitable decisions through differential thresholding, although these techniques need proper adjustments to prevent fairness-related issues.

Fairness metrics show their most significant improvements when bias mitigation strategies operate across multiple levels of AI systems. AI-driven healthcare solutions reach maximum fairness levels when they implement combined approaches to various data types, algorithm enhancements, and post-implementation optimization elements.

4.8. Impact & Observation

When applied to healthcare, AI bias produces multifaceted effects that affect patient wellness, the degree of trust patients have in their healthcare providers, and the development of healthcare policies. Disparate treatment of populations happens through biased AI models because these systems inadequately detect diseases and provide substandard medical resource distribution and treatment guidance. These failings thereby harm communities with limited access. The existing healthcare disparities become worsened by these disparities, which produce lasting negative effects for marginalized groups.

AI bias assessments inside policy circles have produced regulatory evaluations and specifications that stress fairness, transparency, transparency, and accountability within medical AI systems. Integrating diverse datasets and real-time monitoring systems with fairness-aware practices drives healthcare institutions and AI developers toward enhancing the reliability of their AI systems. Studying in this field proves that responsible AI development together with proactive measures will resolve AI bias issues to enhance health care fairness. AI requires superior bias mitigation techniques to function as an advanced medical technology for modern healthcare applications.

5. Discussion

5.1. Interpretation of Results

The research demonstrates that effective mitigation strategies produced favorable results by improving fairness within biased AI healthcare models. Initial assessment results established the presence of bias since marginalized patient groups encountered inaccurate medical diagnoses and dehydration risk evaluations. Fairness-aware learning techniques with balanced training datasets achieve effective bias reduction according to post-mitigation evaluations of demographic parity and equalized odds performance. The elimination of bias proves challenging because some discrimination continues to occur. The most balanced outcomes emerge from using three independent methods that include data preparation and algorithm modifications and post-analysis workload enhancing. Program review procedures need to remain ongoing for preserving the higher fairness standards developed through implemented interventions. The dependability of AI technology for equitable medical choices relies on persistent AI model enhancement as well as prompt bias detection to prevent healthcare disparities.

5.2. Discussion of Findings

The outcome demonstrates that bias exists commonly in AI-based healthcare systems, yet bias mitigation mechanisms generate desirable outcomes. Previous research illustrates how biased AI systems provide inaccurate diagnoses to minority populations; therefore, these systems maintain systemic societal biases. The enhanced results achieved through fair techniques validate the claim that bias elimination produces better models alongside equal health outcomes for all patients. However, some unexpected observations emerged. Model accuracy decreased slightly but demographic parity increased when the system used oversampling and reweighting as fairness interventions. The evaluation process between fairness and predictive precision requires careful consideration of achieving the correct balance between the two variables. The conclusion from impact assessments indicated that setting bias correction after model processing generated superior results than altering algorithms during actual implementation of AI systems. Future studies seeking to optimize fairness need to advance techniques that do not decrease clinical accuracy results.

5.3. Practical Implications

AI-driven healthcare systems become better at treating patients when bias mitigation approaches achieve practical implementation. Diagnostic systems trained through fairness-aware procedures measure patients accurately between diverse demographic groups to lower diagnostic mistakes. Biased AI systems can become part of clinical decision support systems through bias correction mechanisms that produce equitable healthcare treatment suggestions. Improving fairness metrics builds trust in AI medical tools, leading to better healthcare professional and patient adoption. The implementation of active bias detection systems for AI choice oversight by therapeutic institutions and healthcare facilities aids the necessary adjustments for stopping discriminatory outcomes. The practice of inclusive data collection makes sure AI models will represent diverse populations thus reducing systemic disparities. The established methods enable healthcare organizations to safeguard equal patient care and decrease inaccuracies while building trust in AI medical choices.

5.4. Challenges and Limitations

Multiple obstacles make it difficult for healthcare organizations to develop AI models without bias despite their best efforts. The main obstacle is obtaining data from medical history documents, which proves problematic because existing records showcase healthcare disparity systems, obstructing balanced dataset creation. Applied bias mitigation methods still leave trace amounts of bias from unidentifiable variables that affect model prediction results. AI models face a fundamental drawback because some methods meant to decrease bias have been shown to deteriorate predictive accuracy slightly. Standard fairness metrics like demographic parity and equalized odds fail to detect numerous forms of bias that emerge during complex healthcare operations. The continuous changes in medical data and frequent adaptations in AI models make standardization of fairness evaluation difficult. Achieving this requires sustained investigation, clear AI development methodologies, and sustained assessment of fairness approaches for clinical implementation.

5.5. Recommendations

Policymakers must mandate bias audits for every AI model used in medical decision-making to establish fairness in AI-driven healthcare. All AI development companies must show open information about their training data sources and model performance statistics and methods for reducing bias. Healthcare institutions should make real-time bias detection features standard to monitor and rectify AI predictions showing demographic unbalance. AI developers need

to use expansive datasets that include features based on various demographics so the models better serve diverse populations.

Multiple bias reduction techniques for AI model creation implement a three-step process, starting with data cleansing, programming algorithm adjustments, and outcome treatment methods. A completed fairness evaluation procedure must connect to AI development phases through continuous monitoring to maintain fairness as healthcare information evolves. Joint cooperation between data scientists and healthcare professionals and ethicists leads to successful strategy development for fairness. The introduced set of principles works to build dependable ethical solutions which provide benefits to all patient groups in AI healthcare technology.

6. Conclusion

6.1. Summary of Key Points

This research shows the extent of bias found in AI healthcare systems, which creates disadvantages for patients through discriminatory algorithm practices. The analysis shows biased training information coupled with problems in model frameworks lead to healthcare disparity, which negatively impacts minority group patients. The main issues delaying fairness in AI systems stem from ongoing bias presenting even after preventative measures are applied. In contrast, model accuracy versus fairness remains a balance issue, and current assessment measures show difficulty in validating fairness. Three methods across different levels form the proposed solution against bias which includes dataset balancing and the development of fair training procedures and real-time monitoring systems. AI accountability in healthcare needs implementation of regulatory policies along with ethical frameworks to be successful. This implementation process gives models more fairness while healthcare inequalities decrease and users build stronger trust in AI-based medicine. The research demonstrates how persistent AI model assessment promotes equality throughout population divisions to preserve AI as a platform for unbiased care over unyielding preexisting biases.

6.2. Future Directions

Future studies must create improved fairness metrics that effectively identify multiple forms of bias while enhancing the current assessment techniques. Healthcare applications could achieve better fairness by developing context-aware AI models capable of making demographic-based real-time prediction adjustments. Examining federated learning methods for advancing AI model generalization requires investigating distributed training across multiple healthcare centers to minimize bias risk.

The emerging explainable AI technology brings promising solutions for revealing AI decision processes that help healthcare providers detect and rectify biases effectively. Through blockchain deployment for ethical AI governance facilities, trackable records document AI training approaches and decision-making evaluation parameters. Managed AI governance systems should develop through collaboration between data scientists, healthcare professionals, and policymakers to build guidelines that ensure fair and responsible AI-driven healthcare solutions.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Akter, S., Dwivedi, Y. K., Sajib, S., Biswas, K., Bandara, R. J., & Michael, K. (2022). Algorithmic bias in machine learning-based marketing models. Journal of Business Research, 144(144), 201–216. https://doi.org/10.1016/j.jbusres.2022.01.083

[2] Ahmad, K., Maabreh, M., Ghaly, M., Khan, K., Qadir, J., & Al-Fuqaha, A. (2022). Developing Future human-centered Smart cities: Critical Analysis of Smart City security, Data management, and Ethical Challenges. Computer Science Review, 43(1), 100452. https://doi.org/10.1016/j.cosrev.2021.100452

[3] Belenguer, L. (2022). AI bias: exploring discriminatory algorithmic decision-making models and the application of possible machine-centric solutions adapted from the pharmaceutical industry. AI and Ethics, 2(2). https://doi.org/10.1007/s43681-022-00138-8

[4] Fletcher, R. R., Nakeshimana, A., & Olubeko, O. (2021). Addressing Fairness, Bias, and Appropriate Use of Artificial Intelligence and Machine Learning in Global Health. Frontiers in Artificial Intelligence, 3. https://doi.org/10.3389/frai.2020.561802

[5] Gerke, S., Minssen, T., & Cohen, G. (2020). Ethical and Legal Challenges of Artificial intelligence-driven Healthcare. Artificial Intelligence in Healthcare, 1(1), 295–336. https://doi.org/10.1016/B978-0-12-818438-7.00012-5

[6] Giordano, C., Brennan, M., Mohamed, B., Rashidi, P., Modave, F., & Tighe, P. (2021). Accessing artificial intelligence for clinical decision-making. Frontiers in Digital Health, 3(2), 645232. https://doi.org/10.3389/fdgth.2021.645232

[7] Juhn, Y. J., Ryu, E., Wi, C.-I., King, K. S., Malik, M., Romero-Brufau, S., Weng, C., Sohn, S., Sharp, R. R., & Halamka, J. D. (2022). Assessing socioeconomic bias in machine learning algorithms in health care: a case study of the HOUSES index. Journal of the American Medical Informatics Association, 29(7), 1142–1151. https://doi.org/10.1093/jamia/ocac052

[8] Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A Survey on Bias and Fairness in Machine Learning. ACM Computing Surveys, 54(6), 1–35. https://doi.org/10.1145/3457607

[9] Pessach, D., & Shmueli, E. (2023). A Review on Fairness in Machine Learning. ACM Computing Surveys, 55(3), 1–44. https://doi.org/10.1145/3494672

[10] Seyyed-Kalantari, L., Zhang, H., McDermott, M. B. A., Chen, I. Y., & Ghassemi, M. (2021). Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. Nature Medicine, 27(12), 2176–2182. https://doi.org/10.1038/s41591-021-01595-0

[11] Uzun Ozsahin, D., Ozgocmen, C., Balcioglu, O., Ozsahin, I., & Uzun, B. (2022). Diagnostic AI and Cardiac Diseases. Diagnostics, 12(12), 2901. https://doi.org/10.3390/diagnostics12122901

[12] Xu, J., Xiao, Y., Wang, W. H., Ning, Y., Shenkman, E. A., Bian, J., & Wang, F. (2022). Algorithmic fairness in computational medicine. EBioMedicine, 84, 104250. https://doi.org/10.1016/j.ebiom.2022.104250

[13] Ahmed, Z., Mohamed, K., Zeeshan, S., & Dong, X. (2020). Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. Database, 2020. https://doi.org/10.1093/database/baaa010

[14] Chukwuebuka, A. J. (2023a, April 30). Innovative approaches to collaborative AI and machine learning in hybrid cloud infrastructures. IRE Journals. https://irejournals.com/paper-details/1704340

[15] Pillai, A. S. (2023). AI-enabled hospital management systems for modern healthcare: an analysis of system components and interdependencies. Journal of Advanced Analytics in Healthcare Management, 7(1), 212-228.