Sarcouncil Journal of Multidisciplinary

ISSN(Online): 2945-3445

Volume- 05| Issue- 07| 2025

Review Article **Received:** 07-06-2025 | **Accepted:** 01-07-2025 | **Published:** 20-07-2025

Bias and Fairness in Healthcare AI: Addressing Algorithmic Disparities in Underrepresented Populations

Vittal Rao Baikadolla

Golden Gate University, San Francisco, USA

Abstract: Modern applications of medical artificial intelligence offer unparalleled opportunities for revolutionizing clinical practice via enhanced diagnostic features and improved treatment protocols. Nonetheless, new evidence shows worrying computational disparities that disproportionately impact marginalized communities across various demographic factors. Discriminatory biases in medical AI systems establish systematic health care inequalities, particularly affecting minority patients due to diminished diagnostic accuracy and weakened treatment suggestions. The fundamental framework of discrimination in healthcare AI consists of various interrelated elements, including biases in historical datasets, frameworks of institutional bias, and insufficient justice considerations throughout the algorithm development processes. Recognition methodologies range from statistical parity evaluations to intersectional bias identification procedures, with community-driven social media initiatives proving crucial in exposing discriminatory patterns overlooked by standard evaluation techniques. Effective remediation strategies require comprehensive methodologies addressing discrimination during data preprocessing, model development, and output calibration stages through dataset diversification, synthetic data creation, and equity-centered optimization approaches. Evaluation systems must simultaneously measure clinical efficacy and algorithmic justice across diverse patient populations while incorporating continuous monitoring protocols to identify emerging discriminatory trends during deployment periods.

Keywords: Algorithmic bias, healthcare AI, discrimination patterns, bias detection, mitigation strategies, fairness evaluation.

INTRODUCTION

Healthcare facilities worldwide are rapidly implementing artificial intelligence technologies to enhance patient care, expedite treatment procedures, and boost diagnostic precision. Research findings, however, show serious differences in system efficacy among different demographic groups, which disproportionately impact underprivileged individuals in healthcare environments. Discrimination within medical AI systems emerges through consistent unequal care delivery based on ethnicity, sex, age brackets, and economic circumstances. Current scientific investigations reveal significant performance gaps between population segments in essential healthcare tools, ranging from diagnostic imaging to clinical support systems (Vyas, M. 2025). Biased medical algorithms pose substantial risks to healthcare equality and patient protection. Machine learning models developed from discriminatory historical records maintain current healthcare inequities while magnifying existing biases against at-risk populations. Advantage-based computational discrimination develops when particular demographic categories obtain favorable treatment within medical AI frameworks, producing disparate healthcare results that further harm communities already facing systematic exclusion. These effects reach beyond single patient encounters, weakening confidence in medical institutions while deepening long-established social inequalities across multiple generations. Scholarly investigations repeatedly demonstrate inadequate focus on equity considerations during algorithm creation, with many systems exhibiting discriminatory patterns demanding immediate attention and thorough corrective measures (Ueda, D. *et al*., 2024).

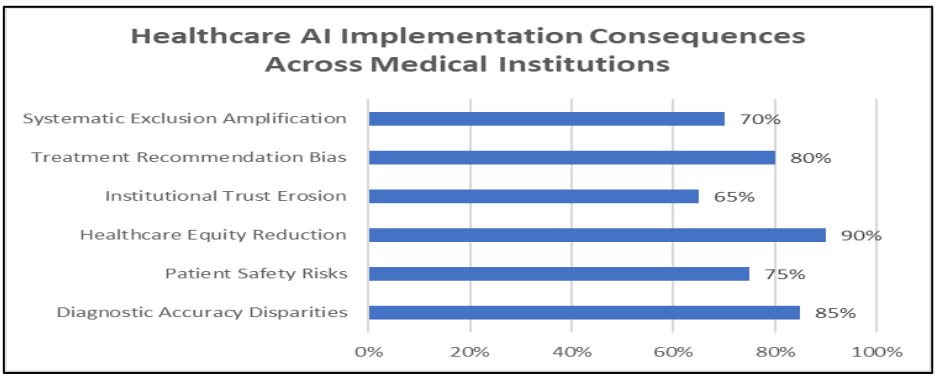

Figure 1: Distribution of AI technology adoption consequences across healthcare institutions and patient safety outcomes (Vyas, M. 2025; Ueda, D. *et al*., 2024)

***Corresponding Author:** Vittal Rao Baikadolla

DOI- https://doi.org/10.5281/zenodo.16226135

THEORETICAL FRAMEWORK OF ALGORITHMIC BIAS IN HEALTHCARE

Medical algorithmic bias originates from various connected elements, forming intricate discrimination structures across different population categories. Past medical records encompass generations of institutional biases, featuring unequal service accessibility, varying record-keeping standards, and different care approaches across ethnic and financial boundaries. Artificial intelligence algorithms developed using biased information sources integrate historical unfairness directly into automated judgment systems. Crisis medical response networks show especially troubling discrimination tendencies, where computational choices can create deadly outcomes for disadvantaged groups during emergency medical situations (Bahamazava, K. *et al*., 2025). Computational bias appears through various analytical methods, encompassing differential impact studies, balanced probability measurements, and population equality evaluations. Differential impact develops when artificial intelligence systems create markedly different results for protected population groups, independent of direct protected characteristic inclusion. Balanced probability demands equivalent accuracy rates across population classifications, while population equality requires consistent positive outcome rates regardless of group association. Medical organizations adopting AI solutions must create thorough structures addressing discrimination methods through organized assessment procedures. Contemporary structures highlight ongoing surveillance significance and flexible correction systems guaranteeing fair performance across all patient groups receiving medical AI services (Kennedy, S. 2024).

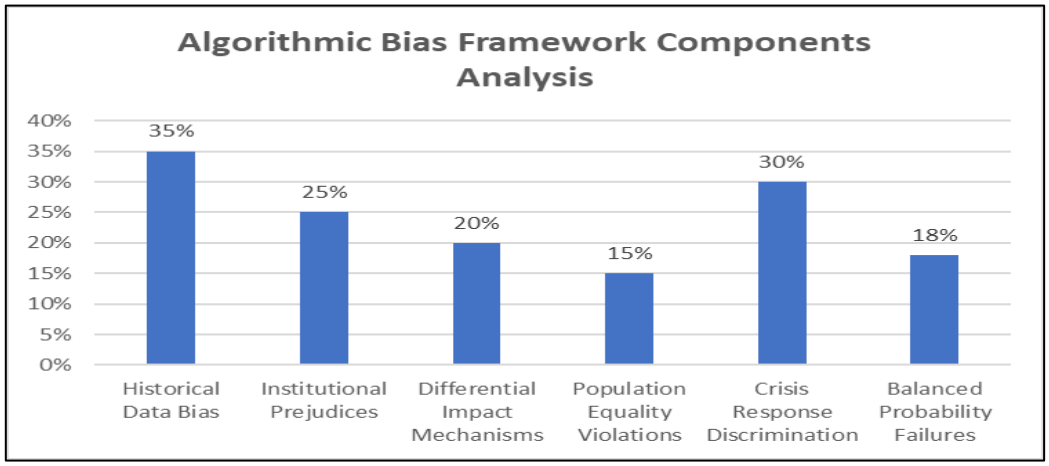

Figure 2: Theoretical framework elements contributing to healthcare AI bias manifestation and discrimination patterns (Bahamazava, K. *et al*., 2025; Kennedy, S. 2024)

EVIDENCE OF BIAS IN HEALTHCARE AI APPLICATIONS

Present medical AI deployments show discrimination across various medical fields, including diagnostic visualization, clinical support systems, and patient surveillance tools. Medical implementations expose major precision differences, with specific AI models showing performance variations surpassing 15% between population groups in critical diagnostic operations. Imaging AI systems commonly show decreased precision for patients from minority backgrounds, potentially leading to overlooked diagnoses or unsuitable treatment suggestions. Medical visualization algorithms developed mainly using homogeneous population information fail to perform effectively across ethnically varied patient groups, establishing systematic disadvantages for minority populations (Polanco, J. 2024). Diagnostic visualization systems show clear gender and ethnic bias in disease identification and severity evaluation processes. Breast imaging AI systems show reduced detection capability when examining dense tissue patterns, unfairly affecting younger women and Asian individuals. Skin analysis AI models developed primarily using pale skin examples show diminished precision in identifying skin malignancies among patients with darker pigmentation. Scientific results suggest effective bias reduction demands concurrent attention to information gathering approaches, algorithm construction methods, and testing procedures to accomplish significant healthcare equality improvements. The complicated characteristics of biased expressions require

Publisher: SARC Publisher

thorough strategies considering overlapping identities and multiple disadvantages affecting vulnerable patient groups (Mittermaier, M. *et al*., 2023).

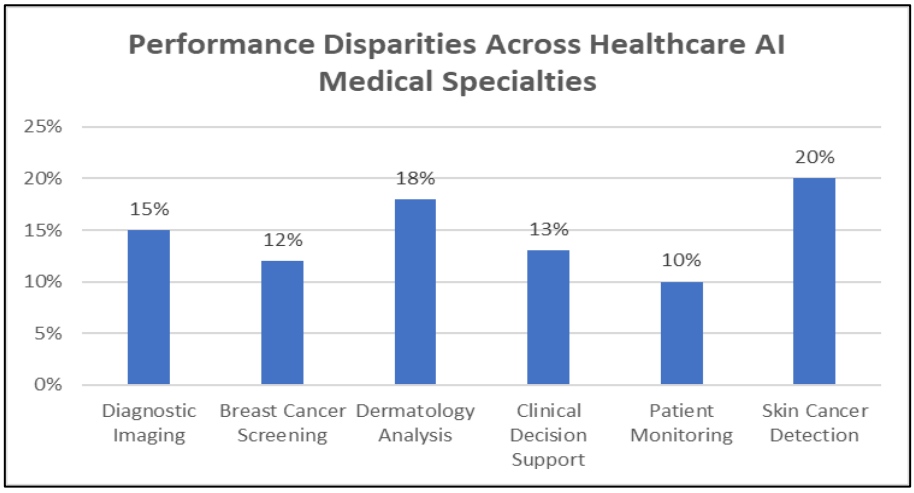

Figure 3: Performance variation patterns and accuracy disparities in different healthcare AI applications (Polanco, J. 2024; Mittermaier, M. *et al*., 2023)

DETECTION METHODOLOGIES FOR HEALTHCARE AI BIAS

Organized bias recognition demands advanced analytical methods capable of detecting inequalities across various population aspects simultaneously. Equity measurements offer numerical structures for evaluating algorithmic fairness, including statistical equality difference computations, balanced probability ratio evaluations, and accuracy measurements. Statistical equality difference computations establish absolute differences in positive outcome rates between population groups, while balanced probability ratio evaluations measure accuracy rate proportions across population sections. Oxygen measurement device research shows essential significance of thorough bias identification approaches, with social platform advocacy efforts successfully exposing ethnic inequalities missed by conventional evaluation techniques. Extended examination of social platform conversations demonstrates that community-led bias identification can successfully supplement official evaluation methods (Ahmed, W. *et al*., 2024). Sophisticated identification approaches include intersectional examination to recognize combined biases affecting persons with various marginalized features. Intersectional equity measurements analyze algorithmic effectiveness at population intersection locations, covering ethnicity, sex, age, and economic position, exposing complicated discrimination tendencies invisible through single-characteristic examination. The Affordable Care Act's anti-discrimination mandates create legal requirements for thorough bias identification within medical AI systems. Medical organizations must establish organized bias surveillance procedures addressing both personal and organizational discrimination tendencies. Oxygen measurement device bias identification shows how regulatory structures can promote technological advances benefiting marginalized communities through improved medical equipment precision and dependability (Saft, H. L. *et al*., 2025).

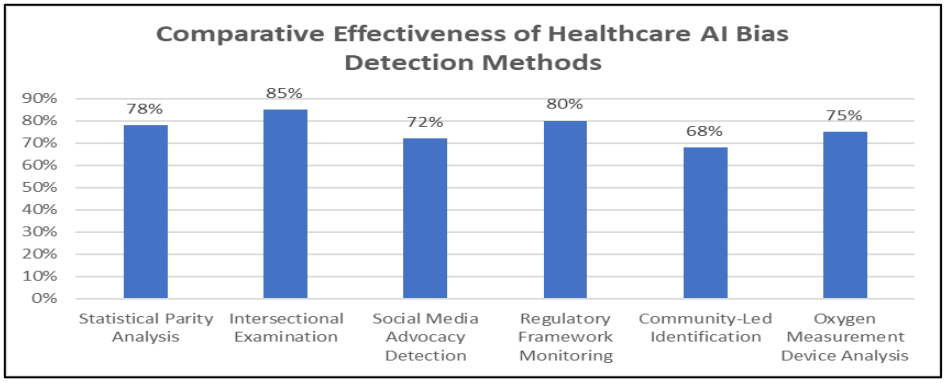

Figure 4: Comparative effectiveness of different bias detection approaches in healthcare AI systems (Ahmed, W. *et al*., 2024; Saft, H. L. *et al*., 2025)

Publisher: SARC Publisher

MITIGATION STRATEGIES AND INTERVENTIONS

Successful bias correction demands thorough methods addressing algorithmic discrimination at data preparation, algorithm training, and outcome adjustment development phases. Data preparation solutions concentrate on information expansion methods, artificial data creation, and representative selection approaches to establish more balanced training information sets. Information expansion methods systematically increase the representation of underrepresented groups, while artificial data creation generates synthetic examples to address population representation deficits. Representative selection guarantees that training information reflects the population structures of intended patient groups. Scientific methods emphasize a field-specific bias reduction approach, significantly addressing unique difficulties in critical healthcare uses such as diagnosis and treatment design (Martin, R. 2021). Algorithm training reduction approaches include equity limitations directly into machine learning improvement goals. Competitive debiasing methods develop models to increase predictive precision while reducing the ability to determine protected characteristics from acquired representations. Multiple-objective learning methods simultaneously improve primary medical results and equity measurements, guaranteeing fair performance across population groups. Identification and assessment approaches must include statistical significance evaluation and confidence range examination to guarantee strong bias evaluation. Machine learning bias assessment demands advanced statistical structures considering multiple testing adjustments and effect magnitude considerations in medical uses (Alelyani, S. 2021).

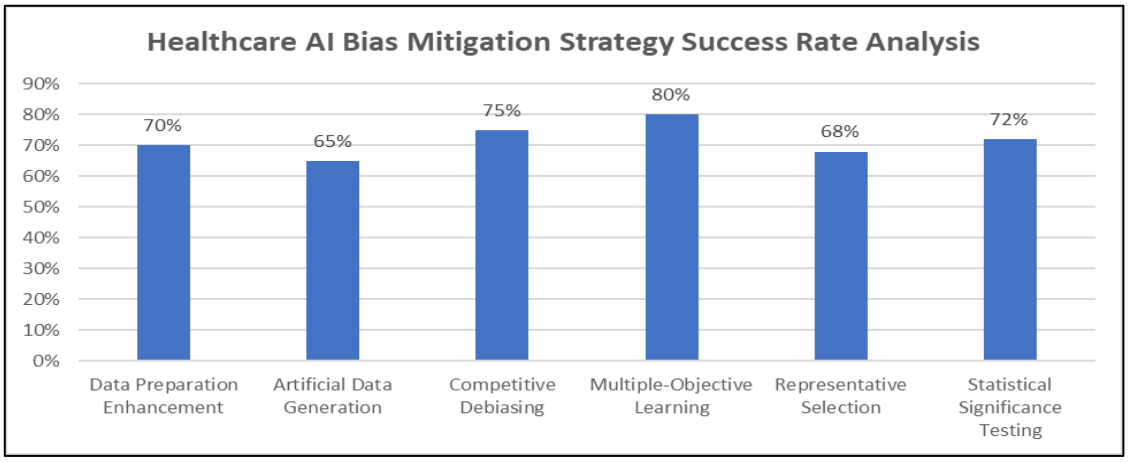

Figure 5: Success rates and effectiveness measures of different bias mitigation approaches in healthcare AI (Martin, R. 2021; Alelyani, S. 2021)

EVALUATION FRAMEWORKS AND PERFORMANCE ASSESSMENT

Thorough evaluation frameworks must evaluate both medical effectiveness and equity simultaneously across varied patient populations. Conventional assessment measurements, such as precision, detection rate, and accuracy, demand division by population groups to recognize performance inequalities. Equity-focused assessment procedures include various bias measurements, guaranteeing algorithmic fairness across different mathematical fairness definitions. Electronic medical record information creates unique difficulties for bias assessment, with research showing that 70% of medical information sets show some demographic bias affecting algorithm effectiveness. Medical organizations must establish organized assessment procedures addressing information quality, representation, and algorithmic equity simultaneously (Gianfrancesco, M. A. *et al*., 2018). Extended assessment methods observe algorithmic effectiveness over lengthy periods, recognizing temporal modifications in bias tendencies and equity measurements. Practical deployment research evaluates algorithmic bias in medical practice environments, considering elements such as physician interaction effects and patient selection biases. External confirmation across various medical systems and geographic areas guarantees transferability of equity results beyond original development settings. Machine learning bias reduction in medicine demands ongoing surveillance and flexible correction systems responding to developing bias tendencies. Medical AI systems must include response mechanisms enabling quick identification and

Publisher: SARC Publisher

correction of biased expressions during medical deployment (Vokinger, K. N. *et al*., 2021).

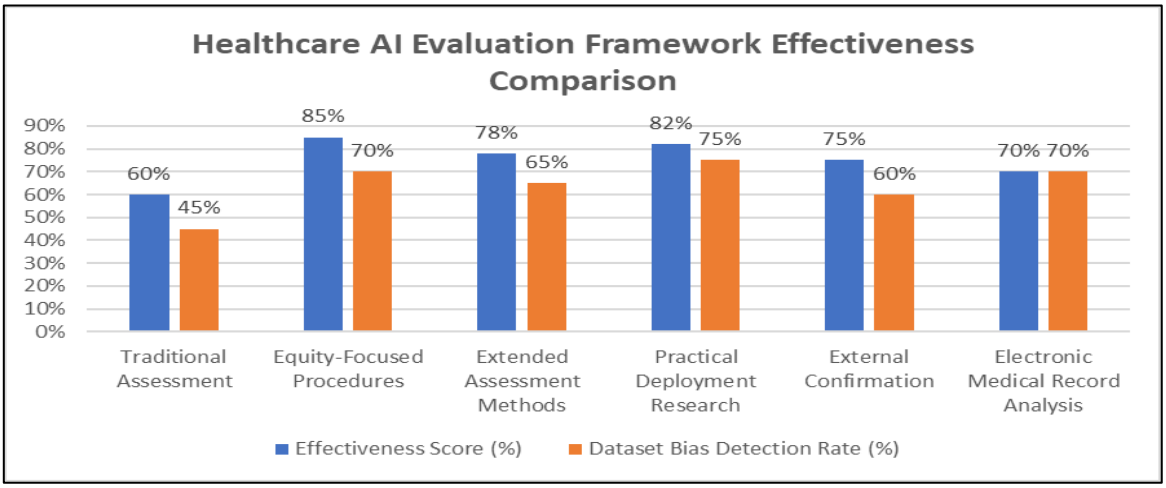

Figure 6: Performance assessment results across different evaluation frameworks for healthcare AI bias measurement (Gianfrancesco, M. A. *et al*., 2018; Vokinger, K. N. *et al*., 2021)

CONCLUSION

Algorithmic prejudice runs deep throughout modern healthcare technology, creating an urgent need for action to protect fair medical treatment across all communities. Medical AI systems currently in use show troubling performance gaps, consistently failing patients from minority backgrounds with poor diagnostic results and questionable treatment advice across various medical fields. Historical data problems, embedded institutional prejudices, and overlooked fairness concerns during software creation combine to form serious challenges demanding complete solutions. Fixing bias requires coordinated work: better data cleaning methods, machine learning designed with fairness in mind, and thorough testing that checks both medical accuracy and fair treatment. Social media campaigns led by affected communities have proven valuable for spotting discrimination missed by standard testing approaches. Hospitals and clinics need structured monitoring programs with multiple evaluation methods, constant performance checks, and quick fixes to keep AI systems working fairly for patients from different backgrounds. Healthcare technology progress hinges on tackling computational unfairness through dedicated commitment to justice, following regulations, and developing innovations focused on excellent patient care and equal health outcomes for everyone.

REFERENCES

1. Vyas, M. “Combating AI Bias in Healthcare and Precision Medicine,” Nitor, 10 February (2025). https://www.nitorinfotech.com/blog/combating-ai-bias-in-healthcare-and-precision-medicine/
2. Ueda, D., Kakinuma, T., Fujita, S., Kamagata, K., Fushimi, Y., Ito, R., ... & Naganawa, S. "Fairness of artificial intelligence in healthcare: review and recommendations." *Japanese journal of radiology* 42.1 (2024): 3-15.
3. Bahamazava, K. "Implications of Algorithmic Bias in AI-Driven Emergency Response Systems." *Journal of Economy and Technology* (2025).
4. Kennedy, S. “Framework to Address Algorithmic Bias in Healthcare AI Models,” *TechTarget*, (2024). https://www.techtarget.com/healthtechanalytics/news/366590097/Framework-to-Address-Algorithmic-Bias-in-Healthcare-AI-Models
5. Polanco, J. “Navigating AI Bias in Healthcare: Challenges and Solutions,” *Foresee Medical*, (2024). https://www.foreseemed.com/blog/ai-bias-in-healthcare
6. Mittermaier, M., Raza, M. M., & Kvedar, J. C. "Bias in AI-based models for medical applications: challenges and mitigation strategies." *NPJ Digital Medicine* 6.1 (2023): 113.
7. Ahmed, W., Hardey, M., Winters, B. D., & Sarwal, A. "Racial Biases Associated With Pulse Oximetry: Longitudinal Social Network Analysis of Social Media Advocacy Impact." *Journal of medical Internet research* 26 (2024): e56034.
8. Saft, H. L., Bhakta, N. R., Wong, A. K. I., Crowder, S. J., Sweet, S. C., & Gurubhagavatula, I. "The Affordable Care Act’s call for nondiscrimination: addressing

Publisher: SARC Publisher

the role of pulse oximetry in racial disparities." *Annals of the American Thoracic Society* 22.3 (2025): 313-316.

9. Martin, R. "Research Approaches to Detect and Mitigate Biases in AI Models, Particularly in High-Stakes Domains like Hiring or Criminal Justice," *ResearchGate*, (2021). https://www.researchgate.net/publication/384697551_Research_Approaches_to_Detect_and_Mitigate_Biases_in_AI_Models_Particularly_in_High-Stakes_Domains_like_Hiring_or_Criminal_Justice
10. Alelyani, S. "Detection and evaluation of machine learning bias." *Applied Sciences* 11.14 (2021): 6271.
11. Gianfrancesco, M. A., Tamang, S., Yazdany, J., & Schmajuk, G. "Potential biases in machine learning algorithms using electronic health record data." *JAMA internal medicine* 178.11 (2018): 1544-1547.
12. Vokinger, K. N., Feuerriegel, S., & Kesselheim, A. S. "Mitigating bias in machine learning for medicine." *Communications medicine* 1.1 (2021): 25.

Source of support: Nil; **Conflict of interest:** Nil.

Cite this article as:
Baikadolla, V. R. "Bias and Fairness in Healthcare AI: Addressing Algorithmic Disparities in Underrepresented Populations" *Sarcouncil Journal of Multidisciplinary* 5.7 (2025): pp 766-771.

Publisher: SARC Publisher