Healthcare Bias in AI: A Systematic Literature Review

Andrada-Mihaela-Nicoleta Moldovan[1][a], Andreea Vescan[1][b] and Crina Grosan[2][c]
[1]Babeş-Bolyai University, Faculty of Mathematics and Computer Science, Computer Science Department, Cluj-Napoca, Romania
[2]Applied Technologies for Clinical Care, Kings College London, London, U.K.

Keywords: Systematic Literature Review, Algorithmic Bias, Fairness, Health Disparities.

Abstract: The adoption of Artificial Intelligence (AI) in healthcare is transforming the field by enhancing patient care, advancing diagnostic precision, and optimizing clinical flows. Despite its promise, algorithmic bias remains a pressing challenge, raising critical concerns about fairness, equity, and the reliability of AI systems in diverse healthcare settings. This Systematic Literature Review (SLR) investigates how bias manifests across the AI lifecycle—spanning data collection, model training, and real-world application and examines its implications for healthcare outcomes. By rigorously analyzing peer-reviewed studies based on inclusion and exclusion criteria, this review identifies the populations most impacted by bias and explores the diversity of existing mitigation strategies, fairness metrics, and ethical frameworks. Our findings reveal persistent gaps in addressing health inequities and underscore the need for targeted interventions to ensure AI systems serve as tools for equitable and ethical care. This work aims to guide future research and inform policy development, in order to prioritize both technological progress and social responsibility in healthcare AI.

1 INTRODUCTION

Artificial Intelligence (AI) is transforming healthcare by improving diagnostic accuracy, personalizing treatment, and optimizing patient outcomes. Bias in medical environments is defined by Panch et al. (Panch et al., 2019) as "the instances when the application of an algorithm compounds existing inequities in socioeconomic status, race, ethnic background, religion, gender, disability or sexual orientation to amplify them and adversely impact inequities in health systems". Although it does not always happen, bias usually can lead to discrimination.

In recent years, the academic community has given the problem of algorithmic bias in Machine Learning (ML) systems (van Assen et al., 2024) used in the healthcare industry a lot of attention. Given their significant effects on healthcare outcomes and decision-making processes, researchers have directed more of their attention to comprehending and resolving possible biases that may exist in these systems.

In an Action Plan published in January 2021 (Clark et al., 2023), the Food and Drug Administration (FDA) highlighted the importance of detecting and mitigating bias in machine learning-based medical systems. The WHO (World Health Organization) Guidance on Ethics and Governance of AI for Health (Guidance, 2021) also acknowledges the possibility of prejudice being ingrained in AI technology. In October 2023, WHO adopted a plan of guiding principles which are: autonomy, safety, transparency, responsibility, equity, and sustainability (Bouderhem, 2024).

Burema et al. (Burema et al., 2023) have compiled a collection of situations to highlight the prevalence of ethical issues in medical systems. Some incidents that occurred from algorithmic bias and discrimination within the healthcare sector include a case that involved the distribution of care work (i.e., the number of hours a caregiver should spend with their patient) (Lecher, 2020), and another that reffers to the distribution of Covid-19 vaccinations (Wiggers, 2020a).

Researchers raise awareness of racial algorithmic bias in medical settings (Ledford, 2019), (Jain et al., 2023). Black patients were shown to have inferior diagnostic accuracy when using neural network algorithms trained to categorize skin lesions (Kamulegeya et al., 2023). It has also been discovered that health sensors exhibit racial bias (Sjoding et al., 2020).

Covid-19 prediction models with flaws (Thomp-

[a] https://orcid.org/0009-0006-6289-9869
[b] https://orcid.org/0000-0002-9049-5726
[c] https://orcid.org/0000-0003-1049-2136

Moldovan, A.-M.-N., Vescan, A. and Grosan, C.
Healthcare Bias in AI: A Systematic Literature Review.
DOI: 10.5220/0013480300003928
In *Proceedings of the 20th International Conference on Evaluation of Novel Approaches to Software Engineering (ENASE 2025)*, pages 835-842
ISBN: 978-989-758-742-9; ISSN: 2184-4895

son, 2020), as well as electronic symptom checks (Fraser et al., 2018) reveal problems with the AI accuracy. When assessing kidney function, prejudice or discrimination against specific populations was discovered in AI (Simonite, 2020). Even Google's lack of openness in their AI for breast cancer prediction was questioned by experts (Wiggers, 2020b).

We conducted a Systematic Literature Review (SLR) in accordance with the guidelines proposed by Kitchenham and Charters (Kitchenham and Charters, 2007). The SLR assesses and synthesizes the state-of-the-art concerning healthcare bias in Artificial Intelligence (AI) systems. The focus is on existing literature regarding various aspects related to healthcare bias, specifically concerning algorithm bias.

The contributions in this paper are the following:

- Conducting an SLR regarding healthcare bias in AI systems that included a set of 97 articles from six database publication sources;
- Providing answers to five research questions regarding algorithm bias in AI systems, considering various aspects such as: algorithmic bias categories, sources of bias in healthcare algorithms, generated risks due to lack of algorithmic fairness, bias mitigation strategies, and metrics for identification of fairness in healthcare algorithms.
- Gaps, challenges, and open issues are discussed, with proposed opportunities/recommendations.

The remainder of this paper is structured as follows: Section 2 covers related work, Section 3 outlines the methodology, Section 4 details the review process, Section 5 presents the results, Section 6 discusses the findings, Section 7 addresses threats to validity, and Section 8 concludes the paper.

2 RELATED WORK

Yfantidou et al. (Yfantidou et al., 2023) presents a comprehensive study on bias in personal informatics, which includes the identification of bias types and the proposal of guidelines for mitigating discrimination. However, their research is specifically centered on personal informatics, focusing exclusively on systems implemented in devices (watches, wearables) rather than addressing bias within the healthcare domain.

While existing surveys and reviews (Mienye et al., 2024), (Kumar et al., 2024) define bias types and sources, propose frameworks, or offer critical perspectives, our SLR takes a complementary approach. Instead of focusing on fixed statements about bias types and sources, we aim to consolidate diverse perspectives from the literature, drawing parallels and highlighting variations in how bias is understood. Our goal is not to establish conclusions, but to present a comprehensive overview of existing knowledge, serving as a resource for researchers in AI healthcare.

There are qualitative surveys and systematic reviews on AI ethics in healthcare (Singh et al., 2023), (Williamson and Prybutok, 2024), however, our approach focuses strictly on algorithmic fairness.

3 METHODOLOGY

This section contains the details of the performed research, research questions, and protocol definition.

3.1 Review Need Identification

The literature on AI bias in healthcare is divided between technical studies on algorithm improvement and social science research on ethical and cultural implications, however, a combined approach is needed. Addressing both technical and societal biases in AI will lead to more effective and equitable healthcare solutions (Belenguer, 2022).

3.2 Research Questions Definition

We outline the research questions in our investigation:

RQ1: Which are the main algorithmic bias categories in healthcare?

RQ2: What are the most common sources of bias in healthcare algorithms?

RQ3: What risks are generated by the lack of algorithmic fairness in healthcare and who is most likely to be affected?

RQ4: Which bias mitigation strategies are proposed to reduce algorithmic disparities in AI for healthcare?

RQ5: How is bias assessed and which metrics have been proposed for identifying fairness in healthcare algorithms?

3.3 Protocol Definition

The steps of the conducted SLR protocol are the following: the search and selection process with the included and excluded criteria; the data extraction strategy; the synthesis of the extracted data with the responses to the research questions; and the discussion of the results, identified gaps and open issues, opportunities and recommendations. The next sections provide detailed descriptions of each step of the protocol.

4 CONDUCTING THE SLR

The SLR related activities are provided next, including the selection process that contains the database search and the specification of the selection criteria, followed by the data extraction.

4.1 Search and Selection Process

For our literature review, we followed the PRISMA 2020 statement (Page et al., 2021), which ensures a transparent, unbiased, and reproducible process. Figure 1 shows the filtering stages, resulting in the selection of 97 papers from the databases.

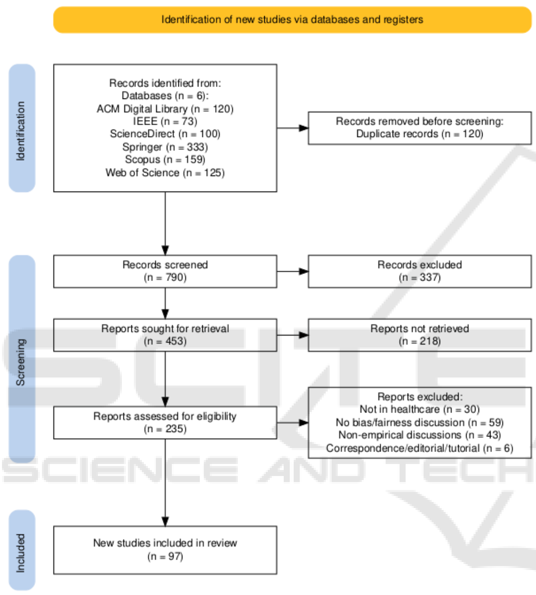

Figure 1: Diagram of the Scoping Review Flow in accordance with the PRISMA 2020 declaration (Haddaway et al., 2022).

4.1.1 Database Search

To ensure the inclusion of the most pertinent research on the topic (Kitchenham and Charters, 2007), a manual search was conducted across six prominent databases: ACM Digital Library, IEEE Xplore, ScienceDirect, Scopus, Springer, and Web of Science (WOS) [1]. Three researchers were assigned databases for the October 2024 search using "AI algorithmic bias in healthcare", applying filters for publication year or domain to manage excessive results and sorting by relevance.

4.1.2 Merging, and Duplicates and Impurity Removal

Upon completing the search of each database, we utilized Zotero software [2] to extract the BibTeX files containing the citations of all retrieved papers as the first filtering stage. These files were subsequently parsed for the automatic removal of duplicate entries.

4.1.3 Application of the Selection Criteria

The main objective was to identify a selection of approximately 100 papers that provided the most substantial insights into the topic of AI algorithmic bias in healthcare. These papers were assessed on whether they include answers to the research questions and on inclusion and exclusion criteria (Meline, 2006).

The inclusion and exclusion criteria that guided the evaluation of each article to determine its applicability to the research topics.

4.2 Data Extraction

The review protocol produced the initial findings (Table 1). After retrieving articles, removing duplicates, and excluding papers based on relevance and lack of access, 235 articles were re-evaluated for eligibility in a detailed third review and 97 papers were identified as relevant to AI algorithmic bias in healthcare.

Table 1: Table showing the search and review statistics across various sources.

Source	**DB search**	**Sought for retrieval**	**Assessed for eligibility**	**Selected**
ACM Digital-Library	120	91	15	10
IEEE	73	93	93	14
Science Direct	100	13	13	8
Springer	333	212	90	42
Scopus	159	25	11	9
WOS	125	19	13	14
Total	910	453	235	97

To filter out papers on unrelated topics, we manually assessed a set of keywords before identifying the main ideas, using keyword-based initial screening (Kitchenham and Charters, 2007). Our selection

[1] ACM: https://dl.acm.org, IEEE Xplore: https://ieeexplore.ieee.org, ScienceDirect: https://www.sciencedirect.com, Scopus: https://www.scopus.com, Springer: https://link.springer.com, WOS: https://www.webofscience.com

[2] Zotero: https://www.zotero.org

of terms included *AI*, *ML*, *healthcare*, *bias*, *fairness*, *correctness*, *metrics*, *discrimination*, *clinic* or *medical*. We recorded occurrences for each term found in a document, and documented the main ideas of the papers that scored positively on most terms.

Figure 2 shows the final selection of papers, with a six-fold increase in publications from 2022 to 2024.

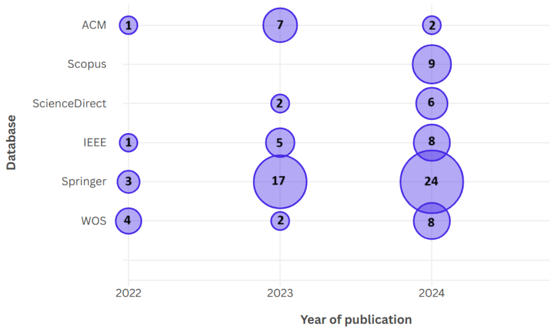

Figure 2: Final database selection: distribution by year of publication.

A replication package is available containing the list of all the selected articles from this study (Author(s), 2025). Further, in this paper, we will refer to the selected papers with S1 to S97.

5 RESULTS

The following section presents the findings for the key RQs, addressing the main categories of algorithmic bias in healthcare (RQ1), their sources (RQ2), the risks stemming from unfair algorithms (RQ3), mitigation strategies to reduce disparities (RQ4), and metrics to assess fairness (RQ5). We provide examples from the selected literature for all answers.

5.1 RQ1: Main Algorithmic Bias Categories in Healtcare

Seven distinct types of algorithmic bias were identified in medical settings (*S6*): *historical bias*, which mirrors existing societal prejudices against specific groups (*S15*, *S16*, *S79*); *representation bias*, arising from sampling methods that under-represent certain population segments (*S57*, *S64*, *S66*, *S74*, *S75*); *measurement bias*, resulting from poorly fitted ML models (*S78*); *aggregation bias*, occurring when universal models fail to account for variations among subgroups; *learning bias*, where modeling decisions exacerbate disparities in performance; *evaluation bias*, which emerges when benchmark datasets do not accurately reflect the intended target population; and *deployment biases* (*S25*) when a model is applied in ways that diverge from its intended purpose. Many researchers agree on similar types of bias (*S27*, *S36*, *S48*, *S49*, *S83*) but use varying terms: 'representation bias' is also called 'sample bias' or 'selection bias'; 'aggregation bias' overlaps with 'linking bias'; and 'deployment bias' is sometimes termed 'feedback bias' (*S23*).

Investigating the selected publications, we extracted information regarding only *mention* of the bias and/or containing *in depth* investigation regarding algorithmic bias. The algorithmic biases that are most mentioned in the articles are those related to *historical bias* and *representation/sampling/selection bias*, the same being also in the case of in-depth investigation as seen in Figure 3. When considering what database publications investigated in-depth the algorithmic biases, the data indicated that Springer and ACM have the largest number of such papers.

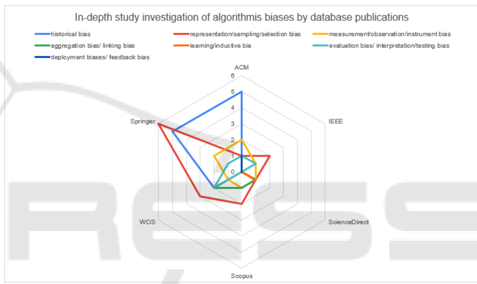

Figure 3: In-depth study investigation of algorithmic biases in database publications.

Other studies (*S46*, *S82*) categorize bias into: genetic variations, intra-observer labeling variability, data acquisition processes. When classified by cause, similar issues are stated as prevalence-, presentation-, and annotation-sourced disparities (*S70*).

Keeling (*S93*) outlines how typical clinical ML models are limited and task-specific, and that their biases are frequently caused by the under-representation of certain demographics in training data. On the other hand, whereas generalist models are wide-ranging, they also incorporate more intricate biases like stereotype associations. Ustymenko and Phadke (*S3*) introduce some guidelines that aim to be useful in a future framework addressing bias in LLMs for healthcare.

5.2 RQ2: Sources of Bias in Healthcare Algorithms

Several sources of bias are identified in the investigated approaches, Figure 4 depicts the main ones from label bias and target specification bias to cognitive biases in decision-making and domain shifts during clinical use. Specific biases are related to data: data acquisition and health disparities. For each of such biases, several solutions were proposed from multitask learning (*S67*, *S76*) and feature-swapping augmentation to algorithmic fairness audit frameworks (*S80*). In the next paragraphs, we provide details about the sources of biases and some solutions.

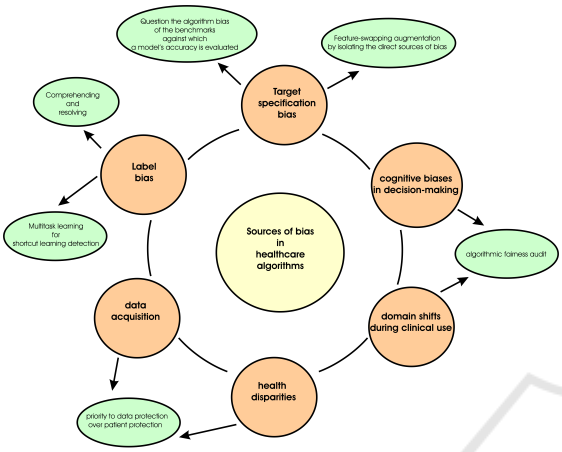

Figure 4: Sources of bias in healthcare algorithms.

The types and sources of bias are highly related. Sources of bias in Clinical Decision Support Systems (CDSS) were identified as cognitive biases in decision-making, in domain transitions during clinical usage, as health inequities, and during data gathering (*S33*).

According to Mhasawade et al. (*S2*), label bias in healthcare algorithms occurs when proxy labels, used instead of actual labels, vary in their connection to true health status across subgroups. Another similar approach employs multitask learning for the detection of shortcut learning (*S76*). Similarly, QP-Net uses multitask learning alongside a domain adaptation module (*S67*), aligning feature distributions across subgroups to improve less-frequent subgroup fitting.

Questioning the bias in benchmarks used to evaluate model accuracy is crucial. The authors (*S1*) argue that label-matching accuracy may bias foreign data and should be avoided in high-stakes medical contexts.

The debiased Survival Prediction solution proposed by Zhong et.al. (*S18*) reduces the effect of biases on model performance by isolating the direct source of bias. By detaching identification information, researchers have proven that disentanglement learning produces more equitable survival estimates free from population biases (*S18*). This approach was also empirically assessed in combination with federated learning for healthcare equity (*S46*, *S50*, *S69*).

5.3 RQ3: Generated Risks by the Lack of Algorithmic Fairness in Healthcare

The lack of algorithmic fairness generates several risks from inaccurate diagnoses and harmful stereotypes to the erosion of trust in the healthcare systems. In Figure 5 several elements as source of risks are identified: inequities, demographic diversity, gender bias, and lack of confidence in the healthcare system.

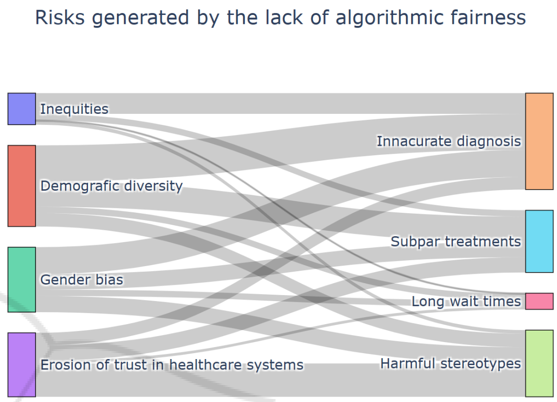

Figure 5: Risks generated by the lack of algorithmic fairness in healthcare.

Children and young people (CYP) under 18 are particularly underrepresented in research (*S42*). A study on Covid-19 diagnosis data (*S32*) revealed significant age disparities, with only 10% of data representing individuals under 20, while 50% pertained to those over 70. Another study on diabetic readmissions (*S12*) found that Naive Bayes performed well for patients under 40, but scored weak non-discrimination metrics for younger populations. The ACCEPT-AI framework highlights key ethical considerations for using pediatric data in AI/ML research (*S72*).

Healthcare AI systems can inadvertently sustain xenophobic biases, worsening health inequalities for migrants and ethnic minorities (*S85*). For example, using proxies for sensitive traits can intensify exclusion and link foreignness with disease (*S76*). A case study (*S38*) on a skin color detection algorithm showed a 16% error in detecting skin tone and a 4% error in recognizing white faces.

Literature on AI fairness in healthcare highlights how real-world gender biases are reinforced (*S60*) (*S42*), (*S9*). Women have been historically underrepresented in health data, limiting AI's impact on their healthcare (*S43*). A skin color detection case study (*S38*) found a 6% error in recognizing women over men, while Cardiovascular AI research (*S23*) shows women receive less care.

5.4 RQ4: Mitigation Strategies to Reduce Algorithmic Disparities in AI for Healthcare

The most frequently proposed mitigation solutions and recommendations from the selected papers are displayed in Figure 6.

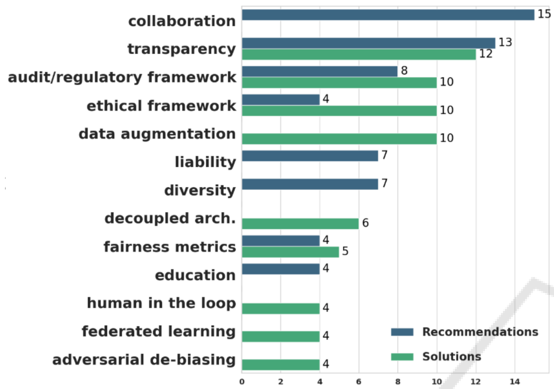

Figure 6: Distribution of mitigation strategies.

Ethical principles. Distinct approaches to fair AI include borrowing ethical principles from fields where they are successfully applied. Amugongo et al. (*S10*) suggest incorporating Ubuntu ethics—principles. Belenguer (*S83*) advocates applying ethics systems from the pharmaceutical industry, while Younas and Zeng (*S95*) propose using Central Asian ethics. Creating a perfect plan to address all ethical issues is unlikely, and debates continue on balancing algorithm transparency with data protection, as well as on certifying developers who meet ethical standards (*S53*, *S55*).

Frameworks. The most encountered bias mitigation solution stands in the form of a framework (*S14*, *S24*, *S26*, *S47*, *S48*, *S46*, *S38*, *S4*, *S21*, *S52*, *S54*, *S49*, *72*, *S81*, *S91*). Iabkoff et al. (*S52*) introduced one of the most comprehensive frameworks.

Transparency. Explainable AI (XAI) refers to AI systems that aim to provide human-understandable justifications for their decisions and predictions, enhancing transparency (*S35*, *S20*, *S46*, *S80*, *S96*, *S28*). Procedural justice emphasizes the importance of the decision-making plan (*S35*), while Gerdes (*S96*) and Kumar (*S41*) argue that AI models must be explainable in healthcare for better understanding. Legal constraints may enforce transparency (*S84*), and open data for XAI enhances accessibility and clarity, aiding informed decisions (*S20*). Yousefi (*S8*) suggests sharing data for public interest. The TWIX technique (*S73*) mitigates bias in surgeon skill assessments. Ziosi et al. (*S80*) evaluate XAI methods based on their ability to address fairness and transparency (*S30*).

Data handling. Although it is common practice to generate synthetic datasets (*S40*, *S58*, *S77*, *S32*) for drawing parallels between situations with balanced and unbalanced data (*S29*), or to generate datasets that are as diverse as possible, approaches for fairness include real-data datasets, or a middle solution that modifies existing data (*S51*, *S65*, *S77*). A proposal is to exclude demographic data (*S22*), as models achieve “locally optimal” fairness, but struggle in out-of-distribution (OOD) settings, suggesting less demographic encoding promotes “globally optimal” fairness (*S56*).

Human-in-the-Loop (HITL). The idea of HITL is essential for challenging the function of human knowledge and the interaction between algorithms and people. A human-guided approach guarantees that developments in technology are a reaction to real clinical requests (*S33*, *S34*, *S97*). Iniesta (*S71*) introduces a framework that emphasizes the HITL approach by several principles like accountability or patient education.

5.5 RQ5: Bias Assessment and Metrics to Identify Fairness in Healthcare Algorithms

Audit frameworks (*S59*, *S17*, *S77*, *S83*, *S94*, *S71*) aim at adapting the development stage to be bias-aware (*S26*), disparities to be acknowledged from data collection to decision design (*S46*) through bias identification solutions (*S38*) for assessing datasets against racial, age, or gender bias. Moghadasi et al. (*S21*) argue that identifying bias sources is key to mitigation. In order to guarantee a more thorough, customized, and impartial perspective, users should additionally employ a variety of bias measures (*S13*, *S31*, *S19*, *S29*).

Mienye et al. (*S19*) reviewed fairness metrics (*S62*) and categorized them into six types: *group fairness*, *individual fairness*, *equality of opportunity*, *demographic parity*, *equalized odds*, and *counterfactual fairness*, each with its own mathematical formula (*S61*, *S69*, *S32*, *S29*, *S56*, *S61*).

A framework proposed for improved and ethical patient outcomes (*S17*) implements fairness measures to evaluate the impact of AI algorithms on different demographic groups using the disparate impact metric for assessing fairness (*S69*, *S6*, *S29*, *S44*, *S56*).

Bowtie analysis is used to assess risks and ensure safety (*S32*). Similarly, Forward-Backward analysis

(*S43*) splits the process into forward (examining consequences) and backward (tracing causes) steps.

6 FINDINGS AND DISCUSSIONS

This section outlines the findings of this investigation, providing several perspectives regarding gaps, challenges, and open issues. Opportunities and recommendations are detailed at the end of the section.

6.1 Discussions of Results

Although bias can also arise during data acquisition by competent institutions or results interpretation by clinicians, most frameworks (*S14*, *S4*, *S26*) and approaches (*S20*, *S22*, *S37*, *S81*, *S83*) assign the task of working on bias mitigation to researchers and developers. Very few of them are addressed to medical personnel (*S48*, *S5*). Some scholars and practitioners posit that governance frameworks may serve as a viable solution (*S42*, *S90*, *S42*). Combinations of legal and audit frameworks are proposed as potential strategies (*S84*), (*S83*).

Multiple studies consider fairness and even propose approaches for specific tasks or branches of medicine (*S7*), such as ICU readmission (*S11*), diabetes (*S12*) (*S92*), lung ultrasound (*S14*), pulmonary embolism (*S18*), critical care (*S33*), liver allocation (*S35*, *S87*), cardiac sarcoidosis (*S21*), pain detection (*S45*), medical imaging (radiology, dermatology, ophthalmology) (*S56*), *S70*), virology (*S88*) (with emphasis on Covid-19 (*S61*) (*S82*), oncology (*S63*) (*S89*), cardiac imaging (*S23*), orthopedics (*S65*), thyroid ultrasound (*S67*), antibiotic prescription (*S68*), pediatrics (*S72*), surgeon skill assessment (*S73*), healthcare time series (*S77*), neuroscience (*S86*), point-of-care diagnostics (*S39*). Tandon et al. (*S49*) aim to provide help for developers to decide which AI-enabled strategies to use for certain designs in the medical field based on context-specific criteria.

6.2 Gaps, Challenges, Open Issues

Several gaps have been identified based on the investigated papers:

- Underrepresentation of minority groups in medical datasets.
- Many AI models are "black boxes", making it hard to identify and mitigate biases.
- Limited testing of AI systems on varied demographic and geographic populations.
- Insufficient collaboration between technologists, clinicians, and patients.
- AI systems can perpetuate and magnify existing systemic biases.

6.3 Opportunities and Recommendations

Several *Opportunities* to address healthcare bias in AI are identified and provided next: Enhanced data diversity by incorporating diverse datasets, Personalized medicine advancements by having customizable treatments, Collaborative multidisciplinary research through multiple experts, Development of fairness metrics that are tailored for healthcare applications, and Policy and regulation innovation.

The following *Recommentations* are provided to address healthcare bias in AI: regular bias audits, diversify training datasets, stakeholder engagement, implement eXplainable AI (XAI), and adopt fairness-centric regulations.

7 THREATS TO VALIDITY

Despite being thorough, this literature study has a limitation that should be noted, namely the speed at which AI is developing, which may have resulted in new discoveries after the research under consideration. Publication bias may have skewed the findings, as studies with positive results are more likely to be published, while null or negative findings often go unreported. To mitigate this, we aimed to include a diverse range of studies and explicitly sought those discussing limitations or failures.

8 CONCLUSIONS

Numerous ethical, political, and economic factors have influenced the application of AI in healthcare, improving the technology's fairness in this area. By taking these aspects into consideration, a suitable degree of control over AI's detrimental effects on the healthcare industry has been demonstrated. Still, many of the statements made in this study encourage academics to identify additional useful implications for practitioners and policymakers about the reliability of AI applications in healthcare, which will enhance the literature in the future.

ACKNOWLEDGEMENTS

The publication of this article was partially supported by the 2024 Development Fund of the Babes-Bolyai University.

REFERENCES

Author(s), A. (accessed January 2025). Healthcare bias in ai: A systematic literature review. = https://figshare.com/s/cc41c628a51a442181fd.

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(4):771–787.

Bouderhem, R. (2024). Shaping the future of ai in healthcare through ethics and governance. *Humanities and Social Sciences Communications*, 11(1):1–12.

Burema, D., Debowski-Weimann, N., von Janowski, A., Grabowski, J., Maftei, M., Jacobs, M., Van Der Smagt, P., and Benbouzid, D. (2023). A sector-based approach to ai ethics: Understanding ethical issues of ai-related incidents within their sectoral context. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, pages 705–714.

Clark, P., Kim, J., and Aphinyanaphongs, Y. (2023). Marketing and us food and drug administration clearance of artificial intelligence and machine learning enabled software in and as medical devices: a systematic review. *JAMA Network Open*, 6(7):e2321792–e2321792.

Fraser, H., Coiera, E., and Wong, D. (2018). Safety of patient-facing digital symptom checkers. *The Lancet*, 392(10161):2263–2264.

Guidance, W. (2021). Ethics and governance of artificial intelligence for health. *World Health Organization*.

Haddaway, N. R., Page, M. J., Pritchard, C. C., and McGuinness, L. A. (2022). Prisma2020: An r package and shiny app for producing prisma 2020-compliant flow diagrams, with interactivity for optimised digital transparency and open synthesis. *Campbell systematic reviews*, 18(2):e1230.

Jain, A., Brooks, J. R., Alford, C. C., Chang, C. S., Mueller, N. M., Umscheid, C. A., and Bierman, A. S. (2023). Awareness of racial and ethnic bias and potential solutions to address bias with use of health care algorithms. In *JAMA Health Forum*, volume 4, pages e231197–e231197. American Medical Association.

Kamulegeya, L., Bwanika, J., Okello, M., Rusoke, D., Nassiwa, F., Lubega, W., Musinguzi, D., and Börve, A. (2023). Using artificial intelligence on dermatology conditions in uganda: A case for diversity in training data sets for machine learning. *African Health Sciences*, 23(2):753–63.

Kitchenham, B. and Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. 2.

Kumar, A., Aelgani, V., Vohra, R., Gupta, S. K., Bhagawati, M., Paul, S., Saba, L., Suri, N., Khanna, N. N., Laird, J. R., et al. (2024). Artificial intelligence bias in medical system designs: A systematic review. *Multimedia Tools and Applications*, 83(6):18005–18057.

Lecher, C. (2020). Can a robot decide my medical treatment?

Ledford, H. (2019). Millions affected by racial bias in health-care algorithm. *Nature*, 574(31):2.

Meline, T. (2006). Selecting studies for systemic review: Inclusion and exclusion criteria. *Contemporary issues in communication science and disorders*, 33(Spring):21–27.

Mienye, I. D., Obaido, G., Emmanuel, I. D., and Ajani, A. A. (2024). A survey of bias and fairness in healthcare ai. In *2024 IEEE 12th International Conference on Healthcare Informatics (ICHI)*, pages 642–650. IEEE.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., et al. (2021). The prisma 2020 statement: an updated guideline for reporting systematic reviews. *bmj*, 372.

Panch, T., Mattie, H., and Atun, R. (2019). Artificial intelligence and algorithmic bias: implications for health systems. *Journal of global health*, 9(2).

Simonite, T. (2020). How an algorithm blocked kidney transplants to black patients— wired.

Singh, B., Rahim, M. A. B. U., Hussain, S., Rizwan, M. A., and Zhao, J. (2023). Ai ethics in healthcare-a survey. In *2023 IEEE 23rd International Conference on Software Quality, Reliability, and Security Companion (QRS-C)*, pages 826–833. IEEE.

Sjoding, M. W., Dickson, R. P., Iwashyna, T. J., Gay, S. E., and Valley, T. S. (2020). Racial bias in pulse oximetry measurement. *New England Journal of Medicine*, 383(25):2477–2478.

Thompson, A. (2020). Coronavirus: Models predicting patient outcomes may be “flawed” and “based on weak evidence. *Yahoo! Retrieved from https://sg.style.yahoo.com/style/coronavirus-covid19-models-patientoutcomes-flawed-153227342.html*.

van Assen, M., Beecy, A., Gershon, G., Newsome, J., Trivedi, H., and Gichoya, J. (2024). Implications of bias in artificial intelligence: considerations for cardiovascular imaging. *Current Atherosclerosis Reports*, 26(4):91–102.

Wiggers, K. (2020a). Covid-19 vaccine distribution algorithms may cement health care inequalities.

Wiggers, K. (2020b). Google’s breast cancer-predicting ai research is useless without transparency, critics say.

Williamson, S. M. and Prybutok, V. (2024). Balancing privacy and progress: a review of privacy challenges, systemic oversight, and patient perceptions in ai-driven healthcare. *Applied Sciences*, 14(2):675.

Yfantidou, S., Sermpezis, P., Vakali, A., and Baeza-Yates, R. (2023). Uncovering bias in personal informatics. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 7(3):1–30.