Research Article Volume 01, Issue No. 02. 2025

The Medical Journal of Radiology and Oncology Radiation



Evaluation of Risk of Bias in Neuroimaging-Based Artificial Intelligence Models for Psychiatric Diagnosis

[1] Yunia, [1] Marsha

[1] Faculty of Medicine, Sebelas Maret University, Surakarta City, Central Java, Indonesia

Correspondence : dr.yunia39484@gmail.com

Article History :

Received date : 2024/07/11
Revised date : 2024/08/29
Accepted date : 2024/09/05
Published date : 2025/01/10

Copyright: *© 2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (BY NC) license (https://creativecommons.org/licenses/by-nc/4.0/).*

ABSTRACT

Background: The evaluation of risk of bias in neuroimaging-based artificial intelligence (AI) models for psychiatric diagnosis has garnered significant attention in recent scholarly discourse. This literature review synthesizes key findings from various studies that highlight the implications of bias in AI systems, particularly in the context of mental health diagnostics. **Literature Review:** Ethical considerations are further explored by (Goisauf & Cano Abadía, 2022), who highlight the societal implications of systemic biases in AI, particularly regarding gender and sex. This perspective is critical, as it addresses the potential for AI tools to perpetuate inequities in mental healthcare diagnostics. The comprehensive literature review by (Muley et al., 2023) categorizes the risks associated with AI in healthcare, including the risk of bias and perpetuation of inequities. Their call for collaboration among stakeholders to evaluate and regulate AI systems is vital for maximizing benefits while minimizing adverse impacts on patient care. The narrative review by (R. Gashkarimov et al., 2023) focuses on the application of machine learning techniques in diagnosing schizophrenia, acknowledging the challenges posed by data heterogeneity and the necessity for robust methodologies. This highlights the importance of addressing the variability in data sources to improve diagnostic accuracy. Further examination of bias in EHR-based AI models by (Chen et al., 2023) reveals the pervasive nature of implicit and selection biases, advocating for standardized frameworks to ensure fairness in AI applications. This insight is crucial for understanding how biases in healthcare data can affect AI outcomes. **Conclusion:** In conclusion, the literature reveals a nuanced understanding of the multifaceted challenges posed by bias in neuroimaging-based AI models for psychiatric diagnosis. There is a critical need for ongoing research, collaboration, and the development of standardized frameworks to ensure equitable and effective AI applications in mental healthcare. Addressing these challenges is essential for improving diagnostic accuracy and ensuring that AI technologies contribute positively to patient outcomes.

Keyword: Risk of Bias, Neuroimaging-Based Artificial Intelligence Models, Psychiatric Diagnosis

Downloaded from The Medical Journal of Radiology and Oncology Radiation
Copyright © International Medical Journal Corp. Ltd. All rights reserved

INTRODUCTION

The evaluation of risk of bias in neuroimaging-based artificial intelligence (AI) models for psychiatric diagnosis has garnered significant attention in recent scholarly discourse. This literature review synthesizes key findings from various studies that highlight the implications of bias in AI systems, particularly in the context of mental health diagnostics.

Starting with the work of (Singh et al., 2021), the authors systematically survey bias in facial analysis tasks, emphasizing the prevalence of biased predictions across AI applications in healthcare. Their findings underscore the necessity for the research community to develop mechanisms aimed at improving fairness and trustworthiness in AI systems. The authors categorize bias mitigation techniques and identify open challenges that require further exploration, setting a foundational understanding of bias in AI diagnostics.

(Jayakumar et al., 2022) build upon this foundation by assessing the quality of systematic reviews in AI diagnostic accuracy studies. Their meta-research reveals a concerning inconsistency in the evaluation of bias risks, particularly in studies utilizing photographic images. They highlight that inadequate representation of diverse demographics can lead to significant disparities in algorithm performance, thus amplifying existing biases—a critical consideration for neuroimaging applications.

(Wang et al., 2022) provide empirical evidence supporting the notion that careful training can significantly mitigate bias in machine learning models applied to neuroimaging data. Their study indicates that rigorous preprocessing and hyper-parameter optimization are essential for developing unbiased models, although they caution that certain biases may remain unaddressed if the sources lie beyond the data preprocessing stage. This insight is crucial for understanding the complexities involved in training AI models for psychiatric diagnosis.

Downloaded from The Medical Journal of Radiology and Oncology Radiation
Copyright © International Medical Journal Corp. Ltd. All rights reserved

In the realm of ethical considerations, (Goisauf & Cano Abadía, 2022) explore the societal implications of AI in radiology and oncology imaging. They argue that the intersection of AI tools with systemic biases has not been adequately examined, particularly concerning gender and sex biases. This lack of attention to demographic factors in AI algorithms poses a risk of perpetuating inequities in mental healthcare diagnostics.

(Muley et al., 2023) further delineate the risks associated with AI in healthcare, categorizing them into seven major areas, including the risk of bias and perpetuation of inequities. Their comprehensive literature review emphasizes the need for collaboration among stakeholders to ensure that AI systems are rigorously evaluated and regulated, thus maximizing their potential benefits while minimizing negative impacts on patient care.

(R. Gashkarimov et al., 2023) delve into the application of machine learning techniques in diagnosing schizophrenia. Their narrative review highlights the growing body of research focused on utilizing various AI technologies to enhance diagnostic accuracy, though they also acknowledge the challenges posed by data heterogeneity and the need for robust methodologies in psychiatric evaluations.

(Chen et al., 2023) conduct a systematic review on bias detection and mitigation strategies in electronic health record (EHR)-based AI models. Their findings reveal the pervasive nature of biases in healthcare AI, particularly implicit and selection biases, and underscore the necessity for standardized frameworks to ensure fairness and equity in AI applications.

(Yang et al., 2024) survey recent methods for addressing AI fairness and bias in biomedicine, emphasizing the importance of understanding biases that may arise during model development. Their exploration of debiasing methods provides valuable insights into potential solutions for mitigating bias in psychiatric diagnostics.

Downloaded from The Medical Journal of Radiology and Oncology Radiation
Copyright © International Medical Journal Corp. Ltd. All rights reserved

(PERETS et al., 2024) contribute to the discourse by examining inherent biases in EHRs, highlighting the challenges posed by missing or poorly logged information. Their scoping review points to the need for more comprehensive research that addresses the complexities of bias in EHR data and its implications for AI model development.

Finally, (M Alhuwaydi, 2024) explores current trends and future directions in the role of AI in mental healthcare. The review discusses the capabilities of AI models in diagnosing mental disorders and emphasizes the importance of integrating large-scale data to enhance diagnostic accuracy. However, it also acknowledges the limitations posed by data heterogeneity and the need for further research to optimize AI applications in psychiatric care.

Through this literature review, a nuanced understanding of the multifaceted challenges posed by bias in neuroimaging-based AI models for psychiatric diagnosis emerges, underscoring the critical need for ongoing research and collaboration to ensure equitable and effective AI applications in mental healthcare.

LITERATURE REVIEW

The article "Anatomizing Bias in Facial Analysis" by (Singh et al., 2021) provides a comprehensive examination of bias within artificial intelligence systems, particularly in the context of facial analysis tasks. The authors articulate the critical implications of biased AI systems in healthcare diagnostics, emphasizing the heightened risk of unfair treatment for certain demographic groups. This is particularly relevant in the realm of psychiatric diagnosis, where neuroimaging-based AI models may inadvertently perpetuate existing biases if not carefully evaluated.

The authors present multiple instances demonstrating biased predictions in AI applications, notably in commercial gender classifiers that exhibit a marked bias against darker-skinned females. This finding is alarming, as it underscores the potential for AI systems to misdiagnose or misinterpret data based on demographic

Downloaded from The Medical Journal of Radiology and Oncology Radiation
Copyright © International Medical Journal Corp. Ltd. All rights reserved

characteristics, which could lead to significant disparities in psychiatric care. The systematic survey conducted by the authors highlights the need for a deeper understanding of the sources of bias in model predictions, an aspect that resonates with the necessity for rigorous evaluation of risk of bias in neuroimaging-based AI models.

Furthermore, (Singh et al., 2021) discuss various algorithms designed to mitigate bias, which is crucial for enhancing the fairness and trustability of AI systems. The development of these algorithms is a step towards addressing the ethical implications of AI in healthcare, particularly in psychiatric contexts where the consequences of misdiagnosis can be severe. The article effectively outlines the importance of creating diverse and representative datasets to train AI models, as well as the need for ongoing analysis of model performance across different demographic groups.

The authors also identify open challenges that require attention from the research community, such as the need for more robust frameworks for assessing and mitigating bias in AI systems. This call to action is particularly relevant for those working on neuroimaging-based AI models, as it emphasizes the importance of continuous evaluation and refinement of these technologies to ensure equitable outcomes for all patients.

The article titled "Unmasking Bias in AI: A Systematic Review of Bias Detection and Mitigation Strategies in Electronic Health Record-based Models" by (Chen et al., 2023) provides a comprehensive examination of biases inherent in AI models utilizing Electronic Health Record (EHR) data, particularly in the context of healthcare disparities. The authors systematically review existing literature to identify six major types of bias, emphasizing implicit and selection biases, which are critical in understanding how AI can exacerbate inequities in healthcare delivery.

One of the key insights from the review is the acknowledgment of AI's growing role in healthcare, which, while promising, raises significant concerns regarding the

Downloaded from The Medical Journal of Radiology and Oncology Radiation
Copyright © International Medical Journal Corp. Ltd. All rights reserved

fairness and equity of these technologies. The authors highlight that biases in AI models can lead to skewed outcomes, thereby worsening existing disparities in healthcare access and quality. The systematic approach taken by the authors to categorize biases provides a valuable framework for future research and practical applications in the field.

The findings also indicate that while many studies employ fairness metrics to assess biases, there is considerable variability in methodologies. This methodological disparity complicates the comparison of results across studies and highlights the need for a more unified approach to evaluating bias in AI systems. The authors advocate for further investigation into the specific impacts of these biases on health disparities, particularly within the context of EHR data.

The article "Exploring the Role of Artificial Intelligence in Mental Healthcare: Current Trends and Future Directions – A Narrative Review for a Comprehensive Insight" by (M Alhuwaydi, 2024) provides a thorough examination of the integration of artificial intelligence (AI) in mental healthcare, particularly focusing on its applications in diagnosis and decision-making processes. The review synthesizes findings from various studies, highlighting both the potential and limitations of AI in this field.

A significant insight from the article is the discussion surrounding the decision support system AI model developed by Tutun et al., which reportedly diagnoses approximately 90% of mental disorders based on a concise questionnaire of 28 items. This finding underscores AI's potential to enhance clinical decision-making by providing reliable diagnostic support, thereby allowing mental healthcare providers to increase patient throughput ((M Alhuwaydi, 2024)). However, while the efficiency gains are promising, the review prompts a critical evaluation of the model's risk of bias, particularly in how it may generalize across diverse patient populations.

The evaluation of the EMPaSchiz model by Kalmady et al. introduces a critical perspective on diagnostic accuracy in AI applications. While the model shows

Downloaded from The Medical Journal of Radiology and Oncology Radiation
Copyright © International Medical Journal Corp. Ltd. All rights reserved

promise, the article highlights the necessity for further validation across varied demographic and clinical settings to ensure that findings are not skewed by inherent biases in the data used for training these models. The review calls attention to the heterogeneity of data, which can lead to differential performance of AI systems in real-world applications.

CONCLUSION

The evaluation of risk of bias in neuroimaging-based artificial intelligence (AI) models for psychiatric diagnosis is a critical area of research that addresses the implications of biased AI systems in mental healthcare. The literature reviewed provides a comprehensive overview of the various dimensions of bias, its sources, and the potential consequences for diagnostic accuracy and equity in psychiatric care.

The initial work by (Singh et al., 2021) systematically surveys bias in AI applications, particularly in facial analysis tasks, revealing significant disparities in algorithm performance across different demographic groups. This foundational understanding is essential, as it highlights the necessity for rigorous mechanisms to enhance fairness and trustworthiness in AI systems used for psychiatric diagnosis. The categorization of bias mitigation techniques presented in this study emphasizes the need for ongoing research to tackle open challenges in this domain.

Building on this foundation, (Jayakumar et al., 2022) critically assess the quality of systematic reviews in AI diagnostic accuracy studies, uncovering inconsistencies in bias risk evaluations, particularly in studies using photographic images. Their findings underscore the importance of diverse demographic representation in training datasets to avoid exacerbating existing biases in neuroimaging applications.

The empirical evidence provided by (Wang et al., 2022) supports the assertion that careful training and preprocessing can mitigate bias in machine learning models. However, the authors caution that biases may persist if they originate beyond the

Downloaded from The Medical Journal of Radiology and Oncology Radiation
Copyright © International Medical Journal Corp. Ltd. All rights reserved

data preprocessing phase, emphasizing the complexity of developing unbiased AI models for psychiatric diagnosis.

Ethical considerations are further explored by (Goisauf & Cano Abadía, 2022), who highlight the societal implications of systemic biases in AI, particularly regarding gender and sex. This perspective is critical, as it addresses the potential for AI tools to perpetuate inequities in mental healthcare diagnostics.

The comprehensive literature review by (Muley et al., 2023) categorizes the risks associated with AI in healthcare, including the risk of bias and perpetuation of inequities. Their call for collaboration among stakeholders to evaluate and regulate AI systems is vital for maximizing benefits while minimizing adverse impacts on patient care.

The narrative review by (R. Gashkarimov et al., 2023) focuses on the application of machine learning techniques in diagnosing schizophrenia, acknowledging the challenges posed by data heterogeneity and the necessity for robust methodologies. This highlights the importance of addressing the variability in data sources to improve diagnostic accuracy.

Further examination of bias in EHR-based AI models by (Chen et al., 2023) reveals the pervasive nature of implicit and selection biases, advocating for standardized frameworks to ensure fairness in AI applications. This insight is crucial for understanding how biases in healthcare data can affect AI outcomes.

The exploration of debiasing methods by (Yang et al., 2024) emphasizes the importance of understanding biases that arise during model development, providing valuable insights for mitigating bias in psychiatric diagnostics.

Additionally, the scoping review by (PERETS et al., 2024) highlights the challenges posed by incomplete or poorly logged information in EHRs, underscoring the need for comprehensive research addressing the complexities of bias in EHR data.

Downloaded from The Medical Journal of Radiology and Oncology Radiation
Copyright © International Medical Journal Corp. Ltd. All rights reserved

Finally, (M Alhuwaydi, 2024) discuss current trends and future directions in AI's role in mental healthcare, emphasizing the importance of integrating large-scale data to enhance diagnostic accuracy while acknowledging the limitations posed by data heterogeneity.

In conclusion, the literature reveals a nuanced understanding of the multifaceted challenges posed by bias in neuroimaging-based AI models for psychiatric diagnosis. There is a critical need for ongoing research, collaboration, and the development of standardized frameworks to ensure equitable and effective AI applications in mental healthcare. Addressing these challenges is essential for improving diagnostic accuracy and ensuring that AI technologies contribute positively to patient outcomes.

DISCLOSURE STATEMENT

- Disclosure Statement : The authors have no conflicts of Interest to declare
- Funding Sources : None
- Acknowledgements : -
- Author Contribution : All authors discussed and contributed the final content for journal submission and publication

REFERENCES

1. Singh, R., Majumdar, P., Mittal, S., & Vatsa, M., 2021. Anatomizing Bias in Facial Analysis. [PDF]
2. Jayakumar, S., Sounderajah, V., Normahani, P., Harling, L., R. Markar, S., Ashrafian, H., & Darzi, A., 2022. Quality assessment standards in artificial intelligence diagnostic accuracy systematic reviews: a meta-research study. ncbi.nlm.nih.gov
3. Wang, R., Chaudhari, P., & Davatzikos, C., 2022. Bias in Machine Learning Models Can Be Significantly Mitigated by Careful Training: Evidence from Neuroimaging Studies. [PDF]

Downloaded from The Medical Journal of Radiology and Oncology Radiation
Copyright © International Medical Journal Corp. Ltd. All rights reserved

4. Goisauf, M. & Cano Abadía, M., 2022. Ethics of AI in Radiology: A Review of Ethical and Societal Implications. ncbi.nlm.nih.gov
5. Muley, A., Muzumdar, P., Kurian, G., & Prasad Basyal, G., 2023. Risk of AI in Healthcare: A Comprehensive Literature Review and Study Framework. [PDF]
6. R. Gashkarimov, V., I. Sultanova, R., S. Efremov, I., & R. Asadullin, A., 2023. Machine learning techniques in diagnostics and prediction of the clinical features of schizophrenia: a narrative review. ncbi.nlm.nih.gov
7. Chen, F., Wang, L., Hong, J., Jiang, J., & Zhou, L., 2023. Unmasking Bias in AI: A Systematic Review of Bias Detection and Mitigation Strategies in Electronic Health Record-based Models. [PDF]
8. Yang, Y., Lin, M., Zhao, H., Peng, Y., Huang, F., & Lu, Z., 2024. A survey of recent methods for addressing AI fairness and bias in biomedicine. ncbi.nlm.nih.gov
9. PERETS, O. R. I. E. L., STAGNO, E. M. A. N. U. E. L. A., BEN YEHUDA, E. Y. A. L., MCNICHOL, M. E. G. A. N., ANTHONY CELI, L. E. O., RAPPOPORT, N. A. D. A. V., & DOROTIC, M. A. T. I. L. D. A., 2024. Inherent Bias in Electronic Health Records: A Scoping Review of Sources of Bias. ncbi.nlm.nih.gov
10. M Alhuwaydi, A., 2024. Exploring the Role of Artificial Intelligence in Mental Healthcare: Current Trends and Future Directions – A Narrative Review for a Comprehensive Insight. ncbi.nlm.nih.gov

Downloaded from The Medical Journal of Radiology and Oncology Radiation
Copyright © International Medical Journal Corp. Ltd. All rights reserved