

Artificial Intelligence in Medical Diagnosis

Prepared by: Lama Mansour Al-Subaie

Course: AI800 – Research Methodology in Computer Science

Date: July 2025

Problem Statement

This research addresses the application of Artificial Intelligence (AI) in medical diagnosis, particularly in the fields of medical imaging and clinical decision-making.

Despite rapid advancements in AI development, diagnostic accuracy remains inconsistent across medical specialties. For instance, generative models like GPT-4 achieved an average accuracy of only 52.1%, significantly lower than expert physicians. On the other hand, domain-specific tools such as Aidoc showed strong clinical performance in critical emergency scenarios like intracranial hemorrhage and pulmonary embolism.

This issue is of high importance due to its direct impact on patient safety and healthcare outcomes. AI has the potential to reduce diagnostic delays, improve operational efficiency in emergency departments, and alleviate physician workload. However, unreliable or non-transparent tools may result in severe diagnostic errors.

Current solutions include the development of explainable and domain-specific AI models like Aidoc, which have shown clinically validated effectiveness with sensitivities up to 99% and specificities over 93%.

Although some AI tools have proven effective, further research is essential to evaluate their real-world applicability across diverse clinical domains. Understanding why AI performs better in some specialties than others remains a critical and underexplored area.

“AI achieved an average diagnostic accuracy of 52.1%, statistically inferior to expert physicians... In contrast, specialized tools like Aidoc demonstrate real-world clinical value.”

Research Objectives

The main objective of this research is to evaluate the effectiveness, accuracy, and real-world applicability of AI-based tools in medical diagnosis. This includes understanding both the potential and limitations of such tools across various clinical domains.

The specific objectives are:

1. To assess the diagnostic accuracy of generative AI models like GPT-4 compared to expert and non-expert physicians.
2. To evaluate the performance of domain-specific AI tools such as Aidoc in real-world clinical settings.
3. To identify the key factors contributing to the variance in AI performance across medical specialties.
4. To investigate the role of explainable AI (XAI) in improving clinical adoption and trust in AI-based diagnosis.
5. To propose recommendations for integrating AI diagnostic tools into healthcare systems effectively and ethically.

Literature Review

This section summarizes key findings from six recent studies focusing on AI in medical diagnosis:

1. Takita et al. (2025): A meta-analysis comparing AI and physicians' diagnostic performance across 83 studies. GPT-4 showed 52.1% accuracy, significantly lower than expert physicians.
2. MedRxiv (2024): Analyzed 54 diagnostic datasets. Average AI accuracy was 56.9%, but heterogeneity and publication bias were noted.
3. Phillips (2025): Reviewed AI performance by specialty. AI excelled in visual domains (e.g., dermatology), but underperformed in multi-modal diagnoses.
4. Aidoc Clinical Validation (2024): High accuracy in detecting pulmonary embolism and intracranial hemorrhage, with sensitivities between 84–99% and specificities above 93%.
5. Jefferson Radiology (2024): Found Aidoc reduced reporting delays and maintained ~93.3% sensitivity.

Wismüller & Stockmaster (2020): AI reduced radiology turnaround time from 132 to 73 minutes while maintaining high sensitivity (95%) and specificity (96.7%). 6.

Collectively, these studies highlight both the promise and the limitations of AI tools in diagnostics.

Conclusion

Artificial Intelligence in medical diagnosis presents a transformative opportunity for modern healthcare. While generative models like GPT-4 exhibit promise, their diagnostic accuracy still lags behind expert physicians. In contrast, specialized AI tools such as Aidoc have proven to be effective in specific clinical contexts like emergency radiology.

To fully utilize AI's potential, future efforts must focus on improving explainability, clinical validation, and integration with existing healthcare infrastructure. This research contributes to a deeper understanding of AI's role in diagnostic accuracy, the conditions for successful deployment, and the ethical frameworks necessary for widespread adoption.

References

- Takita, H., et al. (2025). A systematic review and meta-analysis of diagnostic performance comparison between generative AI and physicians. *npj Digital Medicine*, 8, 175. 1.
- MedRxiv. (2024). Diagnostic Performance Comparison between Generative AI and Physicians: A Systematic Review and Meta-Analysis. 2.
- Phillips, V. (2025). Generative AI vs Physicians: A Systematic Review. *LinkedIn*. 3.
- Aidoc. (2024). AI-Powered Technology in Emergency Radiology. <https://www.aidoc.com> 4.
- Jefferson Radiology. (2024). AI-Powered Technology Revolutionizing Patient Care. 5.
- Wismüller, A., & Stockmaster, L. (2020). A Randomized Trial on Turnaround Time for AI-Based ICH Detection. *AJR*. 6.