

# Artificial Intelligence in Medical Diagnosis

Prepared by: Lama Mansour Al-Subaie

Course: AI800 – Research Methodology in Computer Science

Date: July 2025

## Introduction

Artificial Intelligence (AI) is transforming the healthcare industry by providing advanced tools to aid physicians in the diagnosis and treatment of various medical conditions. Among its most significant applications is medical diagnosis, where AI systems analyze patient data, medical images, and clinical notes to suggest possible conditions. This report provides a detailed literature review of six recent studies focused on the role of AI in medical diagnostics. It explores the methodologies, results, strengths, and limitations of each study and offers a synthesis of findings to understand the overall impact and trends in this field.

## Methodology

To ensure a high-quality and relevant literature review, we established the following selection criteria for papers: (1) published between 2018 and 2025, (2) peer-reviewed or well-recognized preprints, (3) direct relevance to AI in diagnosis, and (4) inclusion of performance metrics or real-world implementation. A total of six papers were chosen: three systematic/meta-analyses providing high-level evidence and three real-world clinical or technical validations focusing on performance, applicability, and accuracy of AI diagnostic tools. The search was conducted through databases such as PubMed, Nature, and arXiv, with preference given to papers featuring rigorous evaluation or implementation in clinical settings.

## Literature Review

1. Takita et al. (2025) conducted a comprehensive meta-analysis comparing the diagnostic performance of generative AI (e.g., GPT-4) to that of physicians across 83

studies published between 2018 and 2024. The AI achieved an average diagnostic accuracy of 52.1%, statistically inferior to expert physicians by a mean difference of 15.8% ( $p = 0.007$ ). However, the AI performed similarly to non-expert physicians in many domains.

2. A MedRxiv study (2024) included 54 diagnostic datasets and evaluated AI performance with a mean accuracy of 56.9%. The analysis was promising but revealed concerns with heterogeneity and publication bias, as many included studies lacked peer review and standardized metrics.

3. A synthesis by Phillips (2025) categorized AI diagnostic performance across specialties. AI tools showed high efficacy in visual diagnosis (e.g., dermatology) but struggled with conditions requiring multi-modal reasoning. The data, though insightful, lacked formal peer-reviewed structure.

4. Aidoc's clinical validation (2024) focused on using AI for triaging CT scans to detect pulmonary embolism and intracranial hemorrhage. Across multi-center trials, the AI achieved sensitivities of 84–99% and specificities above 93%, showing strong performance in emergency radiology.

5. Jefferson Radiology (2024) conducted a real-world evaluation of Aidoc implementation. Their AI tool achieved high accuracy (sensitivity ~93.3%) in detecting intracranial hemorrhage and reduced reporting delays, although the published sensitivity figure had inconsistencies.

6. Wismüller & Stockmaster (2020) conducted a randomized clinical trial published in AJR. Their study showed AI reduced radiology turnaround time (TAT) from 132 minutes to 73 minutes while maintaining a sensitivity of 95% and specificity of 96.7%. The trial highlighted the potential for AI to improve workflow efficiency in high-volume settings.

## Synthesis and Discussion

The selected studies collectively highlight AI's growing role as a diagnostic aid. Meta-analyses show that generative AI tools are comparable to non-expert clinicians

but not yet competitive with experienced professionals. In contrast, specialized tools like Aidoc demonstrate real-world clinical value, especially in high-risk imaging contexts.

There is clear specialty-based variation. AI excels in domains reliant on visual pattern recognition (e.g., dermatology, radiology) but underperforms in complex, multi-factorial diagnostics. Moreover, while AI reduces diagnostic delays, full clinical adoption faces barriers such as interpretability, ethical approval, and trust.

Explainable AI (XAI) was emphasized as a vital component in recent literature. Without transparency, AI tools risk rejection by clinicians. Integrating XAI frameworks ensures models are interpretable and can complement human judgment, especially in life-critical applications.

Overall, the studies reflect a positive trend in AI performance, but they also call for more robust trials, better dataset standardization, and clearer regulatory pathways. AI should be seen as augmenting, not replacing, human intelligence in diagnostics.

## Conclusion

Artificial Intelligence in medical diagnosis offers promising advancements in healthcare. While tools like GPT-4 show potential, their diagnostic capabilities are still limited when compared to expert physicians. On the other hand, domain-specific applications such as Aidoc have demonstrated clinically validated effectiveness.

To fully harness AI's potential, future efforts must focus on explainability, clinical validation, integration into existing systems, and addressing regulatory and ethical challenges. If implemented wisely, AI can become an indispensable assistant to healthcare professionals, improving diagnostic accuracy, patient outcomes, and overall system efficiency.

## References

1. 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.

2. MedRxiv. (2024). Diagnostic Performance Comparison between Generative AI and Physicians: A Systematic Review and Meta-Analysis.

3. Phillips, V. (2025). Generative AI vs Physicians: A Systematic Review. LinkedIn.

4. Aidoc. (2024). AI-Powered Technology in Emergency Radiology. <https://www.aidoc.com>

5. Jefferson Radiology. (2024). AI-Powered Technology Revolutionizing Patient Care.

6. Wismüller, A., & Stockmaster, L. (2020). A Randomized Trial on Turnaround Time for AI-Based ICH Detection. AJR.