GMR
The Original

Clinical Evaluation of AI-Assisted Diagnostic Tools in Multispecialty Medical Practice

M. Gopi, Mr. Mohan, Dr. Daggupati Harith, Fazil Hasan, Dr Arti Muley, Dr. Nibedita Sahoo, Sunila Choudhary,

Professor, Department of Physiology, Kaloji Narayana Rao University of Health sciences Warangal, Telangana, Mamata Medical College, Khammam, India. drmgopi007@gmail.com. ORCID: https//orcid org/ 0000 0002 5770 3810
Department of Biochemistry, Aarupadai Veedu Medical College and Hospital, Vinayaka Missions Research Foundation (DU), India. mohan.sivanandham@avmc.edu.in 0000-0001-6905-1023
Assistant Professor, Department of Anaesthesiology and Critical Care Medicine, Mahatma Gandhi Medical College and Research Institute, Sri Balaji Vidyapeeth University, Puducherry, India. Email id- daggu.harith@gmail.com. Orchid id:0009-0003-6754-7748
Assistant Professor, Department of Agriculture, Noida International University, Uttar Pradesh, India. fazil@niu.edu.in. 0000-0003-0621-4248
Professor, Department of Pathology, Parul Institute of Medical Sciences & Research, Parul University, Vadodara, Gujarat, India, Email Id- arti.muley77892@paruluniversity.ac.in, Orcid Id- 0000-0002-0187-5728
Associate Professor, Department of Pathology, IMS and SUM Hospital, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India, Email Id- nibeditasahoo@soa.ac.in, Orcid Id- 0000-0002-3469-0594
Centre of Research Impact and Outcome, Chitkara University, Rajpura- 140417, Punjab, India. sunila.choudhary.orp@chitkara.edu.in. https://orcid.org/0009-0000-2253-5557

ABSTRACT

Background: In spite of the digital health development, diagnostic errors and clinician burnout have become serious issues in the multispecialty medical practice. Even though Artificial Intelligence (AI) demonstrated potential in the controlled in-silico setting, the most important gap in knowledge is its lack of clinical applicability in the in-vivo setting and the possibility of being implemented into various specialty practices at once. Goal: The present study will compare the accuracy in diagnoses, clinical efficacy, and the acceptance of a multispecialty AI-assisted diagnostic tool with the conventional clinical practice. Methods: This was a prospective observational study carried out in three different departments, namely Radiology, Cardiology, and Dermatology. The AI intervention became a part of the clinical routine as a second opinion decision support system. A multidisciplinary gold standard consensus diagnostic performance was used as a measure of diagnostic performance. Diagnostic accuracy (sensitivity and specificity) was the primary outcome, with time-to-diagnosis and the level of confidence in the clinician being the secondary outcomes, which were measured using Likert scales. Findings: 200 encounters of patients were studied. The AI + Clinician model was found to be more sensitive than clinicians alone; Radiology (95% vs 88%), Cardiology (93% vs 91%), and Dermatology (90% vs 82%). It is important to highlight that the AI-aided workflow decreased the average time-to-diagnosis by 2.2 minutes per encounter. Although there were more cases of clinician confidence in complicated cases, cases of automation bias were evident among junior residents, especially when conducting borderline dermatological tests. Conclusion: AI-based diagnostic systems have a considerable positive impact on the diagnostic sensitivity and efficiency in a multispecialty environment. Nevertheless, a specialty-based performance variance demonstrates the necessity to integrate the strategies in specialties. These results imply that although AI is already fit to support clinical work, it needs strong supervision in order to reduce over-dependence and provide high-quality human-in-the-loop decision-making.

Keywords: *Healthcare delivery, AI-Assisted Diagnostic Tools, Clinical Evaluation*

INTRODUCTION

The diagnostic error issue is a major obstacle to high-value healthcare delivery that leads to poor patient outcomes and increases system expenses [1]. The nature of the high volumes of patients, complicated clinical data, and the increasing burnout of clinicians existing in complex, multispecialty healthcare settings creates an environment where cognitive fatigue and misdiagnosis are likely to occur [6]. Although the conventional clinical practice is based on expert knowledge and the diagnostic guidelines, the Introduction of digital health technologies became necessary to support human performance and optimize working processes [2]. Present-day sources indicate that the mental burden on the professionals operating within the high-paced setting is often what results in the heuristic shortcuts, which again highlights the necessity of objective and data-driven assistance systems.

Artificial Intelligence (AI) has already become an effective change agent in the sphere of medical diagnostics, which shows impressive results in controlled, in-silico conditions [3]. In medical imaging patterns, as well as in the interpretation of complicated physiological signals, machine learning models often perform as well as experts in highly isolated, retrospective tasks. There is, however, a knowledge gap that is critical to the in-vivo clinical utility of these tools. The majority of the available studies are single-specialty application studies or offline validation, and the effects of AI on the ongoing, real-time processes of various departments, like Radiology, Cardiology, and Dermatology, are poorly understood. The shift of the experimental validation to the bedside application is in need of further insight into the workings of these tools when incorporated into the heterogeneous needs of everyday practice [7][8].

Moreover, the issue of introducing AI into live clinical practices provokes some critical questions about the human-in-the-loop process and the psychological effect on the practitioner. Although these tools are associated with better efficiency and sensitivity, the threat of automation bias, a phenomenon when clinicians use the suggestions of an algorithm too much, especially with less experienced practitioners, needs to be thoroughly researched [9]. The integration peculiarities of the specialty are also an issue, with the diagnostic needs of an imaging-focused specialty such as Radiology being vastly different from the morphological analysis in Dermatology [4]. Given that AI has had to deal with the uncertainty of real-world patient interactions across multiple disciplines at once, there is an urgent need to find out whether AI is able to sustain its functionality and deliver meaningful utility [10]. The current research assesses a multispecialty AI-assisted diagnostic tool in terms of the accuracy of the diagnosis, clinical efficiency, and adoption by clinicians, as opposed to traditional clinical practice. The proposed goal is to quantify human expertise and algorithmic support synergy in the real world by providing a prospective observational study across three separate departments. This study aims to offer a model on how AI should be used responsibly and effectively in complex medical ecosystems through the analysis of 200 patient encounters, which would guarantee that the growth in technology would be reflected in the corresponding increase in patient safety and provider efficiency.

The paper is divided into five main segments, as the logical flow of clinical evidence is ensured. The next step is the Introduction, as it provides the theoretical background and aims of the study. The section on Materials and Methods outlines the perspective of observational design and the implementation of the second opinion of AI in Radiology, Cardiology, and Dermatology. The results section presents a quantitative comparison of the diagnostic accuracy and efficiency measures, with a table specific to each specialty of the performance. These results are critically discussed in the Discussion, which covers the implications of increased sensitivity and the risks of automation bias of junior practitioners as seen. Lastly, the Conclusion summarizes the contributions of the study and provides feasible advice on the safe implementation of AI in multispecialty clinical settings.

MATERIALS AND METHODS

Study Design and Setting

The observational study to assess AI-assisted diagnostic tool integration in a multispecialty academic medical center was a prospective one, a six-month study. Three separate departments of the clinic, namely Radiology, Cardiology, and Dermatology, were studied. These specialties have been chosen as a wide spectrum of diagnostic inputs, such as cross-sectional images, electrophysiological information, and morphological visual analysis.

Participant and Encounter Selection

Two hundred patient encounters were recruited according to a consecutive sampling technique. The inclusion criteria were that the patients should be receiving routine diagnostic assessment in the participating departments at the time they were being observed as part of the study. In order to have an in-depth evaluation of the effect of the AI tool on various levels of expertise, senior attending physicians, fellows, and junior residents were included. All taking clinicians were informed and gave consent to participate in the study, and the protocol of the study was approved by the Institutional Review Board (IRB).

The AI Intervention

The AI system applied in this research was implemented into the current Electronic Health Record (EHR) and Picture Archiving and Communication System (PACS) to be a second opinion support tool. The workflow was constructed in the following way:

- Primary Evaluation: The clinician conducted a direct analysis of the clinical data (e.g., ECG strips, dermoscopic images, or CT scans).
- AI Activation: After the initial evaluation, the AI device gave a diagnostic recommendation and a rating of trustworthiness.
- Final Decision: The AI output was examined by the clinician, who made a final diagnosis on whether the AI input changed the initial clinical impression or not.

Outcome Measures

The major result was accuracy in diagnosis (sensitivity and specificity) with respect to a multidisciplinary consensus (gold standard) of diagnosis. This had been determined on a panel of two senior experts who were independent and had no knowledge related to the AI and initial clinicians' outcomes in each of the cases. Secondary outcomes were:

- Clinical Efficiency: Measured as the average time-to-diagnosis (in minutes), starting with the acquisition of data, until the completion of the diagnostic report.
- Clinician Confidence: Measured at the end of the encounter on a 5-point Likert Scale (1 = Not sure, 5 = Always sure).
- Automation Bias: There were cases of a clinician following an erroneous AI recommendation, which were recorded and classified according to the level of experience.

Statistical Analysis

A statistical test was conducted to determine the comparative diagnostic performance of the clinicians who were operating separately with that of the AI-augmented. The sensitivities and specificities were done with 95% confidence interval. Paired t-tests were conducted to test the continuous variables like time-to-diagnosis, and the Wilcoxon signed-rank test was employed to test data on Likert scales. The p-value ≤ 0.05 was taken as statistically significant.

RESULTS

The 200 patient encounters analysis showed that there were considerable changes in the diagnostic performance and operational metrics after the adoption of AI-assisted support. The results of the comparative data of the clinician-only baseline and the AI-enhanced workflow are presented below.

Diagnostic Performance and Accuracy

Test-retest, Inter-Rater, Stability, and Content-Construct Grade 1 2 3. The AI + Clinician model showed statistically significant sensitivity improvement in all the specialties involved. Specificity was relatively high, but the greatest improvements were in Radiology and Dermatology, where the use of complex pattern recognition is a major part of the diagnostic process.

Table 1: Comparative Diagnostic Performance by Specialty

Specialty	Approach	Sensitivity (%)	Specificity (%)	Mean Time-to-Diagnosis (min)
Radiology	Clinician Only	88	92	14.2
	AI + Clinician	**95**	91	**11.5**
Cardiology	Clinician Only	91	94	8.8
	AI + Clinician	**93**	94	**7.2**
Dermatology	Clinician Only	82	89	10.5
	AI + Clinician	**90**	87	**8.1**

Operational Efficiency

Workflow with AI support has led to a significant decrease in average time-to-diagnosis. In the multispecialty cohort, the tool saved a mean of 2.2 minutes in every encounter. The greatest reduction in time was in Radiology (2.7 minutes), and this is possibly because the AI can quickly identify areas of interest in huge radiologic data, thus saving time on early scans.

Clinician Confidence and Behavioral Observations

The level of clinician confidence, measured through a 5-point Likert scale, tended to be on the increase with the general progression, especially in those instances when the panel was classified in the gold standard as either complex or borderline. The average score of the confidence scale went up to 4.4 ($p < 0.05$).

Nevertheless, it was found that the use of AI suggestions varied depending on professional experience when qualitative analysis was used:

• Attending Physicians: The AI was used as the main safety net to confirm findings or to check the existence of unusual subtleties.

Junior Residents: A more frequent occurrence of automation bias. Residents in about 7% of borderline dermatological exams reversed their original correct impressions to match an incorrect AI recommendation, which is a weakness of the human-in-the-loop model for the trainees.

DISCUSSION

The present study implies that the implementation of AI-aided diagnostic devices into multispecialty workflows has a notable quantifiable advantage to diagnostic sensitivity and working throughput. When the shift is made between the in-silico performance and in-vivo application, AI is proven capable of supporting human expertise, especially when it comes to pattern-intensive areas like Radiology and Dermatology. Nevertheless, the perceived difference in performance as well as the development of automation bias among the lower employees requires a subtle perspective of the observed results.

Interpretation of Diagnostic Gains

This improvement in AI-enhanced groups (Radiology 95% and, Cardiology 93%, and Dermatology 90% sensitivity) implies that AI is a successful diagnostic safety net. The tool reduces the possibility of false negatives through the detection of small irregularities that otherwise could be missed because of cognitive load or large volumes. The slight trend towards over-diagnosis when clinicians recommend algorithmic suggestions is indicated by the marginal loss in the specificity of Dermatology (89% to 87). Clinical skepticism, as demonstrated in this trade-off, is crucial in order to prevent unnecessary follow-up procedures or treatments [5].

Efficiency and Workflow Integration

The decrease in the mean time-to-diagnosis by an average of 2.2 minutes per encounter is an important operational improvement. These cumulative savings can be of great importance in high-volume settings to reduce the clinician workload and wait times of patients. The difference in efficiency in the Radiology domain was notably significant, probably due to the ability of the AI tool to pre-process massive imaging data, as the clinician does not need to conduct a voluminous manual search and thus can focus their attention on the pre-identified areas of interest.

The "Human-in-the-Loop" and Automation Bias

The main observation of this study is the difference in the use of AI by older physicians and younger residents. When the AI was applied as a secondary validation tool by more advanced clinicians, junior residents exhibited some measurable tendency to automation bias. The 7% error rate of borderline cases of dermatological cases among residents indicates that less experienced practitioners might not have the clinical ground truth to critically assess an algorithmic output. The realization highlights how dangerous it is to make AI a prop instead of a partner and suggests that the emergence of independent thinking in medical trainees may be compromised.

Limitations

The sample size of 200 encounters and the prospective nature of the study can be seen as a weakness of the study since it might not be representative of the entire spectrum of rare clinical pathologies. Furthermore, the research was carried out in one academic medical facility; thus, the findings cannot be well applicable to community-based practices and resources, as well as the patient base. The system of depending on a consensus panel of gold standard analysts, though strong, is still subject to the visual constraints of the human expert system.

CONCLUSION

The results of the present research confirm the clinical usefulness of AI-assisted diagnostic instruments in the framework of multispecialty and indicate significant improvement in diagnostic sensitivity and a tangible decrease in time-to-diagnosis. Through the effective shift between the experimental in-silico settings and the active in-vivo practice, the data help to justify the use of AI integration as a feasible approach to patient safety and responding to operational challenges in high-volume departments. The Introduction of automation bias among junior employees, however, points to the fact that AI should be a supplement instead of a substitute for clinical judgment. To make such technologies safely deployed, specialty-specific integration protocols should be adopted by the organization, and they should have so-called human-in-the-loop protection, i.e., independent human approval should be needed before AI is turned on. Besides, medical training programs have to be modified with the addition of skepticism algorithms to ensure that the skills of junior practitioners are preserved. To conclude, although AI has a transformative potential regarding diagnostic accuracy and efficiency, its success in the long-term requires a solid oversight and a balanced synergy between the human expertise and the support of the algorithms.

Reference

[1] Alshehri, S., Alahmari, K. A., & Alasiry, A. (2024). A comprehensive evaluation of AI-assisted diagnostic tools in ENT medicine: insights and perspectives from healthcare professionals. *Journal of Personalized Medicine*, *14*(4), 354.

[2] Qamar, R. (2024). The Role of Artificial Intelligence in Enhancing Diagnostic Accuracy in Clinical Medicine. *Multidisciplinary Journal of Healthcare (MJH)*, *1*(2), 55-64.

[3] Galdames, I. S. (2024). From Anatomy to Algorithm: Scope of AI-Assisted Diagnostic Competencies in Health Sciences Education. *International Journal of Medical and Surgical Sciences,(IJMSS)*, *11*(3), 1-24.

[4] Kalita, A. J., Boruah, A., Das, T., Mazumder, N., Jaiswal, S. K., Zhuo, G. Y., ... & Kao, F. J. (2024). Artificial intelligence in diagnostic medical image processing for advanced healthcare applications. In *Biomedical Imaging: Advances in Artificial Intelligence and Machine Learning* (pp. 1-61). Singapore: Springer Nature Singapore.

[5] Sun, D., Hadjiiski, L., Alva, A., Zakharia, Y., Joshi, M., Chan, H. P., ... & Matuszak, M. (2025, April). Observer study: impact of case complexities and physician characteristics on AI-assisted treatment response assessment in bladder cancer. In *Medical Imaging 2025: Computer-Aided Diagnosis* (Vol. 13407, pp. 363-368). SPIE.

[6] Hirani, H., Saboo, B., Modi, A., Modi, P., Samajdar, S. S., Saboo, B., ... & Kadam, P. (2025). Clinical Assessment of Large Language Models: A Comprehensive Multi-domain Performance Study for Healthcare Applications. *International Journal of Diabetes and Technology*, *4*(4), 159-165.

[7] Reddy, R., Pickhardt, P. J., Manrai, A., Summers, R. M., Kim, D., & Rajpurkar, P. (2025). NotifAI-OS: an AI framework for automated CT-based opportunistic screening in post-acute value-based care. *Nature Biomedical Engineering*, *9*(11), 1791-1796.

[8] Pattanayak, S. P. (2025). Artificial Intelligence in Early Disease Diagnosis: Transforming General Medicine through Predictive Analytics. *International Journal of Research in General Medicine and Health*, *1*(01), 19-28.

[9] Yacoub, B., Varga-Szemes, A., Schoepf, U. J., Kabakus, I. M., Baruah, D., Burt, J. R., ... & Emrich, T. (2022). Impact of artificial intelligence assistance on chest CT interpretation times: a prospective randomized study. *American Journal of Roentgenology*, *219*(5), 743-751.

[10] Rahim, R. (2025). Enhancing Sentiment Analysis: A Comparative Study of Rule-Based Discourse Models and Graph-Based Neural Networks. *Frontiers in Computational Science and Engineering*, 27-34.

[11] Ali, O. M., Wright, B., Goodhead, C., & Hampton, P. J. (2024). Patient-led skin cancer teledermatology without dermoscopy during the COVID-19 pandemic: important lessons for the development of future patient-facing teledermatology and artificial intelligence-assisted self-diagnosis. *Clinical and Experimental Dermatology*, *49*(9), 1056-1059.

[12] Tilavov, T., Alimova, R., Nizamkhodjaev, S., Gafforova, Z., Gorin, A., & Abdurahmonova, K. (2025). Innovative Health System Transformation for Environmentally Sustainable Communities through Environmental Determinants and Global Health Technologies. *Acta Innovations*, 1-8.

[13] Sharma, R., Pradhan, A. K., Karavadi, B., Mahapatra, S. K., Harinaiha, A., & Kakkar, P. H. (2026). Applying the species-area relationship model to predict biodiversity loss in deforested regions. *Natural and Engineering Sciences, 10*(2), 79–92. https://doi.org/10.28978/nesciences.1763892