

AI Decision Tools versus Traditional Methods: A Narrative Review of Diagnostic Accuracy and Clinical Reasoning in Senior Medical Students

Jesiya Liju

Dubai Medical University

Sofiya Mujeeb

Dubai Medical University

Jawahir Alemadi

Dubai Medical University

Aya Al Imami

Dubai Medical University

Mariam Elsayed

Dubai Medical University

Youness Zidoun

`youness.zidoun@gmail.com`

Dubai Medical University

Systematic Review

Keywords:

Posted Date: October 5th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10901704/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

AI Decision Tools versus Traditional Methods: A Narrative Review of Diagnostic Accuracy and Clinical Reasoning in Senior Medical Students

Jesiya Liju^{1*}, Sofiya Mujeeb^{1*}, Jawahir Alemadi^{1*}, Aya Al Imami^{1*}, Mariam Elsayed¹,
Youness Zidoun^{1,2*}

¹College of Medicine, Dubai Medical University, Dubai, United Arab Emirates

*These authors contributed equally to this work

Abstract

The aim of this narrative review was to explore whether artificial intelligence (AI)-based decision tools and clinical decision support systems (CDSS) improve diagnostic accuracy and clinical reasoning in senior medical students compared to traditional diagnostic methods. A synthesis of eight empirical studies published from 2020 to 2025 was found through PubMed and Google Scholar. AI-supported tools brought consistent benefits for diagnostic accuracy in standardised, imaging-dependent tasks. Systems with guided reflection and structured feedback supported diagnostic confidence and reflective reasoning. However, evidence on broader clinical reasoning was mixed. Effectiveness was mostly dependent on the quality of the inputs generated by the students and the students' ability to critically appraise the system recommendations for the less structured, information intensive clinical tasks. Some studies reported no diagnostic benefit and possible adverse effects including automation bias, decreased diagnostic efficiency and decreased trust in either the system or students own judgement. Overall, AI and CDSS tools seem to work best as adjuncts to, not replacements for, clinical reasoning. The educational value of them depends not only on the technological ability but also on a conscious pedagogical integration, which promotes reflection, critical evaluation and a proper use of the recommendations provided by the system. Additional longitudinal and mixed methods research is needed to assess their long-term impact on clinical reasoning and diagnostic performance.

Corresponding author:

Youness Zidoun PhD,
Youness.zidoun@gmail.com

Introduction and Background

Competencies and critical skills in medical practice, namely diagnostic accuracy and clinical reasoning, empower healthcare professionals to make reliable clinical decisions. As fundamental objectives of undergraduate medical education, they support data collection and analysis, leading to the formulation of solid differential diagnoses. Conventionally, lectures, textbooks, case discussions, and hands-on clinical training were primarily responsible for clinical reasoning development. However, concerns were raised regarding the adequacy and sufficiency of traditional teaching methodologies in equipping and preparing students for independent diagnostic decision-making due to the increasing complexity of clinical cases, time restrictions, limited diversity of case exposures, and a rise in student enrolment [1].

There has been an increase in the integration of Clinical Decision Support Systems (CDSS) and other artificial intelligence (AI)-based tools into the realm of medical education because of the recent developments in artificial intelligence and the sudden increase in its availability. The application of machine learning and algorithms helps provide unique diagnoses, investigation guidance, and improvements in clinical decision making. Medical students during their clinical training period have access to these AI-based decision-making tools, and this calls for a discussion on the effectiveness of such an approach towards education [2], [3].

Evidence indicating AI-driven tools' significant impact on diagnostic accuracy improvement in specific educational settings has been collected through research, particularly in those/tasks involving image-based and complex technical skills. Medical students' diagnostic capabilities were thus found to be effectively enhanced in that domain by the AI-assisted systems. For instance, with the use of radiographic analysis and point-of-care ultrasound tools, students' performance in areas like fracture identification and cardiac function evaluation was comparable to that of experienced clinicians, reaching similar diagnostic accuracy levels [4], [5], [6]. Moreover, some studies suggested that the use of CDSSs, which showcase advantages in structured tasks like appropriate imaging selection, can lead to lasting learning benefits [7]. Nevertheless, when compared to diagnostic outcomes of traditional methods for common diseases, other research observed no notable improvement; hence, the evidence pertaining to CDSS is not consistent [8].

Concerns have been raised about possible limitations associated with the use of AI tools in supporting diagnostics during medical training. For instance, students may become prone to accepting incorrect AI recommendations, which can result in slower completion of tasks and the use of poor-quality clinical data, thus increasing the likelihood of automation bias and reducing diagnostic efficiency [9], [10]. According to existing literature, there appears to be an intricate interaction between human decision-making and decision support technologies since medical students tend to engage with AI recommendations selectively and not automatically [11], the use of artificial intelligence technology in

diagnosis will impact how students gather and evaluate the data, which largely depends on their ability to reason and the quality of data available from the clinical setting [1], [2].

Senior medical students experience a transitional phase in which they refine their clinical reasoning abilities and move towards increased diagnostic independence; hence, they represent the core demographic that AI-based decision-making tools' effectiveness can be assessed upon. Current research however lacks consistent and direct evidence focusing on senior students when comparing both AI-assisted and traditional means of diagnosis. Instead, it generally highlights AI-enhanced learning's advantages and drawbacks/ the advantages and drawbacks of AI-enhanced learning. Consequently, during this fundamental phase of medical education, a more comprehensive review of the literature is essential to clarify the influence of AI decision tools on diagnostic precision and clinical reasoning. Based on current literature, this review posits that AI-driven decision-making tools improve diagnostic accuracy for senior medical students, particularly in structured and image-based tasks. Nonetheless, ability of students in critical evaluation of AI-generated information determines the effect of these tools in clinical reasoning and distinguishes their impact from traditional diagnostic methods [12].

Methods

This review draws on literature identified through a search of PubMed and Google Scholar for studies published between 2020 and 2025, combining terms for artificial intelligence, machine learning, and clinical decision support systems with terms for diagnostic accuracy, clinical reasoning, and medical students. Eligible studies were empirical investigations in which clinical-year undergraduate medical students used an AI or CDSS tool directly in a diagnostic task, with an explicit comparison to traditional methods and a reported outcome relating to diagnostic accuracy or clinical reasoning. Studies involving non-medical or pre-clinical populations, tool-development work without student interaction, and non-empirical publications (reviews, commentaries, protocols) were excluded, as were studies without a direct AI-versus-traditional comparison.

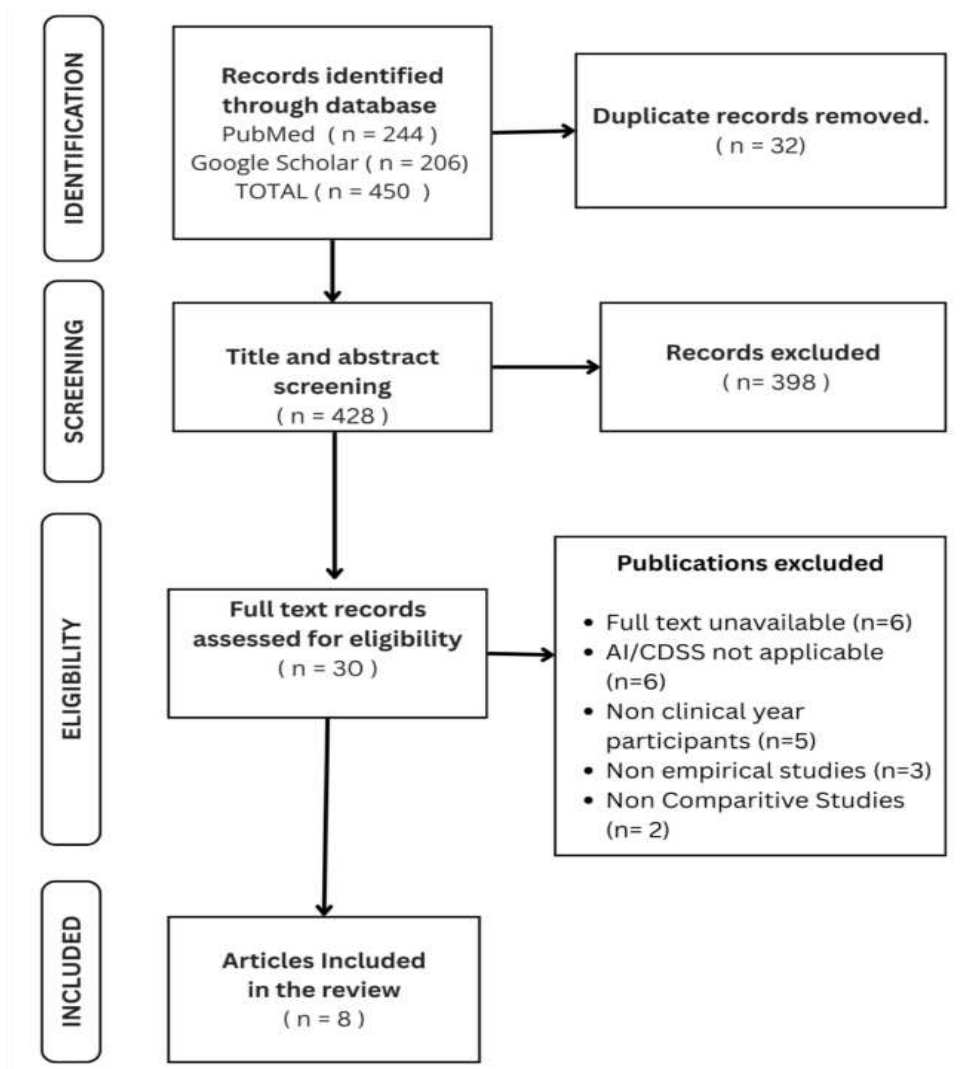


Fig.1 Screening and proceeding workflow

Rather than mapping the full breadth of this literature, the studies were read and synthesized thematically around the review's central question: how do AI decision tools compare with traditional methods on diagnostic accuracy and clinical reasoning in senior medical students, and what explains the variation in findings.

Findings

Table 1 shows the characteristics of the eight studies as they were broadly comparable in population — all involved clinical-year undergraduate medical students — but varied considerably in design (seven quantitative, one mixed-methods), sample size (range 8–166, median 40), and geography (half from Europe, the remainder from Asia and North America). This heterogeneity, discussed further below, limits how far any single conclusion

can be generalized and is one reason a narrative rather than a pooled quantitative synthesis is appropriate here.

Category	Characteristic	Frequency	Publication
Publication Year (2020 – 2025)	2020	2 (25%)	[2], [5]
	2021	3 (37.5%)	[1], [8], [9]
	2022	0	
	2023	2 (25%)	[4], [7]
	2024	0	
	2025	1 (12.5%)	[11]
Geographical Location (Continents)	Europe	4 (50%)	[1], [5], [7], [8]
	Asia	3 (37.5%)	[2], [4], [11]
	North America	1 (12.5%)	[9]
Study Design	Quantitative	7 (87.5%)	[2], [4], [5], [7], [8], [9], [11]
	Qualitative	0 (0%)	
	Mixed Methods	1 (12.5%)	[1]
Target Population	Undergraduate Clinical Year Medical Students	8 (100%)	[1], [2], [4], [5], [7], [8], [9], [11]
Sample Size	Range	8-166	
	Median	40	

Table 1: Characteristics of Included Studies

Table 2 shows the type of digital technology used, categorizing all included studies. Four articles (50%) [1], [2], [4], [8] encompassing applications such as HipGuide, LVivo EF (DiA Imaging Analysis), eCREST, and Diagnostic Decision Support Systems (DDSS) used approaches based on AI. Correspondingly, 4 studies (50%) [5], [7], [9], [11] including symptom checker tools such as Ada® (versions 3.3.0 to 3.7.0) developed by Ada Health GmbH and ESR eGUIDE used CDSS. An equal representation of AI-driven technologies and traditional CDSS platforms distribution is indicated among the included studies.

Category	Description	Number of Studies	Examples
AI	Tools incorporating artificial intelligence to support diagnosis or learning through image analysis, simulated patients, or algorithmic reasoning.	4 (50%)	• HipGuide • LVivo EF (DiA Imaging Analysis) • eCREST • DDSS
CDSS	Decision support systems providing structured, rule-based or guideline-driven recommendations based on entered clinical data.	4 (50%)	• Symptom checker Ada®. Versions 3.3.0 until 3.7.0 from Ada Health GmbH • ESR eGUIDE.

Table 2: Classification of Included Studies by Type of Digital Technology

Across the reviewed literature, researchers used a variety of statistical analyses to analyze data, such as group comparison tests (t-test and ANOVA; n= 2 each) [2], [7], chi-square tests (n = 3) [1], [8], [11], and Fisher's exact test (n = 1) [9]. Other analyses methods included Pearson's correlation and regression analyses, which were reported in 3 studies [2], [4], [8]. Bland–Altman agreement, kappa reliability, Wilcoxon matched-pairs signed-rank testing, and Shapiro–Wilk normality testing was each reported once [2], [4], [5] as shown in table 3.

Types of tests	Number of tests done
Tests of normality	1
Parametric and non-parametric group comparisons	5
Association analyses	2
Categorical data analyses	4
Agreement and reliability measures	2
Regression and Pearson's based modelling approaches	3

Table 3: Statistical Methods Reported Across Included Studies

As illustrated in table 4, learning and confidence outcomes were most frequent among the eight studies (each 6 out of 8, 75%) [1], [2], [4], [5], [7], [11] followed by usability (5 out of 8, 62.5%) [1], [2], [4], [5], [11]. Usability mainly focused on ease of use, time efficiency, and support for task performance, while learning outcomes highlighted improvements in diagnostic accuracy and task performance. Confidence-related outcomes indicated greater diagnostic certainty and better management of clinical uncertainty. Trust was reported in 2 studies (25%) [1], [7], and perception and negative or other outcomes were each reported in 1 study (12.5%) [1]. Negative or other outcomes were reported in 2 studies (25%) [8], [9].

Negative outcomes related to AI-based decision support use were reported in 2 studies (2 out of 8; 25%) [8], [9]. Students in this study did not demonstrate improved diagnostic accuracy compared with controls or showed lower accuracy while using the diagnostic decision support system. No observed improvement in learning or diagnostic skill development was attributed to reported issues including reduced trust and willingness to use the system after practical exposure, increased task completion time, and acceptance of incorrect AI-generated suggestions.

Outcomes	Articles reporting outcome
Confidence	6 (75%)
Trust	2 (25%)
Usability	5 (62.5%)
Perception	1 (12.5%)
Learning	6 (75%)
Negative/other	2 (25%)

Table 4: Outcome Domains Reported Across Included Studies

Diagnostic Accuracy Gains in Structured, Image-Based Tasks

The most consistent and clearest benefit of AI assistance that emerges from the literature we reviewed is for assignments with a bounded, perceptual, or rule-governed structure. In the interpretation of hip fracture, AI-based education improved students' diagnostic performance and brought it closer to that of experienced readers [2]. Similarly, in point-of-care echocardiography, students using an AI-assisted tool achieved left ventricular ejection fraction measurements that closely matched expert assessments, improving the accuracy and precision of visual estimation alone [4]. In a randomised cross-over study of the ESR eGUIDE decision-support tool, students improved their ability to select appropriate imaging tests across scenarios, with performance holding up even after the tool was withdrawn – evidence of learning retention rather than mere dependence on the system while it was present [5].

The commonality of these three studies is a task structure that makes use of AI's strengths: a fixed input, a relatively determined correct answer, and an image or a guideline to anchor the recommendation. Most of the literature reviewed indicates that AI or CDSS are better than traditional methods for this type of task.

Confidence, Trust, and the Management of Uncertainty

Some research suggests AI and CDSS tools might be helpful for how confidently and reflectively students deal with uncertainty, not just raw accuracy. An experimental study on CDSS use found that the tools added value especially in complex or atypical cases, supporting diagnostic reasoning, trust and confidence, and reducing uncertainty [7]. In a mixed-methods study of virtual patient tools, reflective prompts prompted students to reconsider their initial judgements and identify cognitive prejudices, which was associated with greater confidence and better decision-making, especially in environments with limited case exposure [1].

The findings suggest a second, distinct role for decision-support tools in facilitating learning: not providing the correct answer but encouraging reflection and structuring the way students approach uncertain cases. Indeed, confidence and learning were the two most frequently reported outcomes across the eight studies, each reported in 6 of 8, followed by usability 5 of 8, and well ahead of trust and perception (1 and 2, respectively)—a pattern consistent with a literature more concerned with whether these tools change how students' reason than with whether students simply like using them.

Automation Bias and the Limits of System Support

In contrast, two studies show no diagnostic benefit at all, and their results show what can go wrong with AI and CDSS assistance. In a randomised controlled trial of a DDSS for three rheumatic diseases, diagnostic accuracy was no better than in controls; the system's recommendations were sometimes misleading, the time to complete the task increased,

efficiency decreased, and students' desire to use the system afterwards declined. In this trial diagnostic performance followed students' own clinical reasoning much more closely than it did system assistance [8].

A second randomised trial of a CDSS found that its benefit was highly dependent on how students worded the clinical information they entered. The relevant clinical features of a case were often identifiable by students, but they had difficulty in converting them into the semantic qualifiers required by the system and in correctly interpreting its output; when qualifiers did not match, the system did not provide a diagnostic advantage over traditional methods [11].

These two trials make the same point from different directions: the utility of a decision-support system depends on the quality of the interaction between the student and the system, not just the system's own ability. When that interaction fails, whether due to misleading output, mismatched input, or students simply overriding the tool with their own judgement, the tool can slow students down and erode confidence, with no corresponding gain in accuracy. The empirical face of automation bias is not blind acceptance of every AI recommendation, but an uneven and counterproductive reliance on it.

Learner Engagement and the Question of Appropriate Reliance

A recurring theme across the more positive studies is that benefit comes from a balanced, as opposed to passive, form of engagement. In the CDSS study that reported gains in confidence and trust, students neither ignored the system's recommendations nor accepted them uncritically [7]. A pattern of engagement notably absent from the two negative trials. A different study of clinical decision support for hospital discharge planning also found that uptake depended on the way the recommendation interacted with patients' and students' own preferences, rather than being adopted wholesale [9].

The collective findings of the eight studies suggest that the operative variable is not "AI versus traditional methods" as a binary, but the mode of engagement between student and tool: whether students regard system output as one input to be weighed against their own reasoning, or as a substitute for it.

Educational Implications

This narrative synthesis supports a qualified version of the opening thesis. AI decision tools do improve diagnostic accuracy for senior medical undergraduates, but only for structured, image-based, or otherwise perceptually bounded tasks — hip fracture interpretation, or echocardiographic measurement, or imaging-test selection — in which the tool's recommendation is tied to a comparatively determinate answer. In areas outside that zone, where tasks are more about gathering, weighing and correctly framing clinical information, the evidence is really mixed. Some tools supported reflection and confidence, and

reasoning under uncertainty; others provided no measurable benefit, and in two trials actively slowed students down or damaged their trust in the system.

The boundary between these outcomes is not the technology category—AI-based and CDSS-based tools are on both sides of it—but the quality of the interaction between student and the system: whether the student's baseline reasoning and input data were sufficient, and whether the student considered the system's output a prompt for further reasoning rather than a final answer. "The use of these tools in training has direct implication on this. This literature does not support replacing traditional teaching with AI or CDSS support as a simple substitute; it does support deliberate pedagogical integration (e.g., pairing tool use with reflective prompts, guidance on how to frame clinical input, and explicit instruction on when and how to challenge a system's recommendation). Tools built around reflection, such as the virtual patient system studied by [1], and structured feedback, such as ESR eGUIDE [5], produced outcomes that persisted after the tool was removed; tools used without that scaffolding did not.

Limitations of the Evidence and Review

This synthesis is not without limitations that bear on the extent of credence to be given to its conclusions. The underlying evidence base is small: eight studies, published unevenly across the review period with no studies at all in 2022 or 2024, suggesting a field that is still exploratory rather than one that is systematically accumulating evidence. The studies are also methodologically lopsided. Most studies used a purely quantitative design and only one used qualitative method, which leaves the how and why of student engagement with these tools under-explored relative to the what. Sample sizes ranged from 8 to 166 participants with a median of 40. Half of the studies were conducted in Europe, which limits the confidence in the generalizability of these findings across educational contexts, tool types or regions. Because this review emphasises depth of argument over exhaustive coverage, it is a narrative rather than a systematic synthesis: it does not claim to have captured every relevant study published in the period, and studies in languages other than English, or not indexed in the two databases searched, are not represented. Finally, with the rapid development of artificial intelligence systems, studies conducted on tools assessed in the 2020-2023 period may not fully reflect the capabilities of current AI for decision support.

Conclusion

The studies reviewed show reliable diagnostic gains for senior medical undergraduates with AI-based decision tools for structured, image-based and technically bounded tasks, and that AI-based decision tools can foster confidence and reflective reasoning if designed around guided interaction. However, their impact on clinical reasoning more generally is not automatic but depends on students' baseline reasoning ability, the quality of the clinical information they provide and their willingness and ability to critically evaluate what the system tells them. For the least structured, or most cognitively demanding tasks, AI support did not always out-perform traditional methods. In some trials it introduced real costs:

slower task completion, lower efficiency and dwindling trust – without a gain in accuracy to compensate. The pattern that seems to emerge is one of complementarity rather than substitution. These tools tend to reinforce reasoning that is already reasonably sound and do less to repair reasoning that is not. Thus, their educational value does not lie in their technical capability but rather in how carefully they are embedded in a curriculum with training on how to use them, reflective guidance and safeguards against uncritical reliance. Longitudinal and mixed-methods research is still needed to determine how these effects develop across the course of training, and how these compare over time with traditional diagnostic learning.

References

- [1] R. Plackett, A. P. Kassianos, J. Timmis, J. Sheringham, P. Schartau, and M. Kambouri, "Using Virtual Patients to Explore the Clinical Reasoning Skills of Medical Students: Mixed Methods Study," *J. Med. Internet Res.*, vol. 23, no. 6, p. e24723, Jun. 2021, doi: 10.2196/24723.
- [2] C.-T. Cheng et al., "Artificial intelligence-based education assists medical students' interpretation of hip fracture," *Insights Imaging*, vol. 11, no. 1, p. 119, Nov. 2020, doi: 10.1186/s13244-020-00932-0.
- [3] E. H. H. Rincón, D. Jimenez, L. A. C. Aguilar, J. M. P. Flórez, Á. E. R. Tapia, and C. L. J. Peñuela, "Mapping the use of artificial intelligence in medical education: a scoping review," *BMC Med. Educ.*, vol. 25, no. 1, p. 526, Apr. 2025, doi: 10.1186/s12909-025-07089-8.
- [4] Z. Dadon et al., "Empowering Medical Students: Harnessing Artificial Intelligence for Precision Point-of-Care Echocardiography Assessment of Left Ventricular Ejection Fraction," *Int. J. Clin. Pract.*, vol. 2023, p. 5225872, 2023, doi: 10.1155/2023/5225872.
- [5] T. Diekhoff, F. Kainberger, L. Oleaga, M. Dewey, and E. Zimmermann, "Effectiveness of the clinical decision support tool ESR eGUIDE for teaching medical students the appropriate selection of imaging tests: randomized cross-over evaluation," *Eur. Radiol.*, vol. 30, no. 10, pp. 5684–5689, Oct. 2020, doi: 10.1007/s00330-020-06942-2.
- [6] S. Rasouli, D. Alkurdi, and B. Jia, "The Role of Artificial Intelligence in Modern Medical Education and Practice: A Systematic Literature Review," Jul. 26, 2024, *Medical Education*. doi: 10.1101/2024.07.25.24311022.
- [7] S. D. Kafke et al., "Can clinical decision support systems be an asset in medical education? An experimental approach," *BMC Med. Educ.*, vol. 23, no. 1, p. 570, Aug. 2023, doi: 10.1186/s12909-023-04568-8.

- [8] J. Knitza et al., “Accuracy and usability of a diagnostic decision support system in the diagnosis of three representative rheumatic diseases: a randomized controlled trial among medical students,” *Arthritis Res. Ther.*, vol. 23, no. 1, p. 233, Sep. 2021, doi: 10.1186/s13075-021-02616-6.
- [9] J. C. Cox, I. L. Leeds, V. Sadiraj, K. E. Schnier, and J. F. Sweeney, “Effects of patients’ hospital discharge preferences on uptake of clinical decision support,” *PloS One*, vol. 16, no. 3, p. e0247270, 2021, doi: 10.1371/journal.pone.0247270.
- [10] R. A. M. Thompson, Y. B. Shah, F. Aguirre, C. Stewart, C. D. Lallas, and M. S. Shah, “Artificial Intelligence Use in Medical Education: Best Practices and Future Directions,” *Curr. Urol. Rep.*, vol. 26, no. 1, p. 45, Dec. 2025, doi: 10.1007/s11934-025-01277-1.
- [11] Y. gita et al., “Appropriate semantic qualifiers increase diagnostic accuracy when using a clinical decision support system: a randomized controlled trial,” *BMC Med. Educ.*, vol. 25, no. 1, p. 711, May 2025, doi: 10.1186/s12909-025-07294-5.
- [12] K. Shaw, M. A. Henning, and C. S. Webster, “Artificial Intelligence in Medical Education: a Scoping Review of the Evidence for Efficacy and Future Directions,” *Med. Sci. Educ.*, vol. 35, no. 3, pp. 1803–1816, Apr. 2025, doi: 10.1007/s40670-025-02373-0.

List of Abbreviations

CDSS Clinical Decision Support Systems

AI Artificial Intelligence

Declarations

Funding

This study received no funding.

Author information

Authors and Affiliations

Jesiya Liju, College of Medicine, Dubai Medical University, Dubai, UAE

Sofiya Mujeeb, College of Medicine, Dubai Medical University, Dubai, UAE

Jawahir Alemadi, College of Medicine, Dubai Medical University, Dubai, UAE

Aya Al Imami, College of Medicine, Dubai Medical University, Dubai, UAE

Mariam Elsayed, College of Medicine, Dubai Medical University, Dubai, UAE

Youness Zidoun, College of Medicine, Dubai Medical University, Dubai, UAE

Corresponding author

Correspondence to Youness Zidoun, Youness.zidoun@gmail.com

Authors' contributions

YZ conceptualized the study. JL, SM, JA, AA and ME searched the databases and collected the articles. All authors contributed to the screening rounds. JL, SM, JA and AA drafted the initial manuscript. YZ critically revised the manuscript. JL, SM, JA, AA and YZ edited the final manuscript. JL, SM, JA, AA and YZ contributed equally to this work. All authors approved the submitted version.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Human or animal rights

Not applicable.

Informed consent

Not applicable.

Clinical trial

Not applicable.

Competing interests

The authors declare no competing interests.

