Journal of Medical Sciences
Derleme/ Review
https://doi.org/10.46629/JMS.2026.2

Symbiotic AI Diagnosis: A Reverse Mentorship Framework for Clinical Decision Support through Real-Time Physician Training by Machine Learning Models

Simbiyotik Yapay Zeka Teşhisi: Klinik Karar Destek Sistemlerinde Makine Öğrenmesi Modellerinin Hekimlere Gerçek Zamanlı Eğitimi İçin Bir Tersine Mentorluk Çerçevesi

Ufuk Burak KARCIOĞLU[1]

UBK: 0009-0006-9131-6407

[1]PhD Candidate İstinye University, Faculty of Health Sciences, Department of Health Management, İstanbul, Türkiye

Özet

Bu çalışma, yapay zekâ destekli klinik karar destek sistemlerinde (KKDS) “tersine mentorluk” ilkesine dayanan yeni bir kavramsal çerçeve olan Simbiyotik Yapay Zekâ Tanı (Symbiotic AI Diagnosis) modelini önermektedir. Geleneksel yaklaşımlar yapay zekâyı klinisyenlere yardımcı olan bir araç olarak konumlandırırken, bu çalışma yapay zekânın hekimler için gerçek zamanlı bir eğitsel aktör olarak da işlev görebileceğini savunmaktadır.

Açıklanabilir yapay zekâ (XAI) mekanizmaları aracılığıyla makine öğrenmesi modelleri yalnızca tahmin üretmekle kalmayıp, aynı zamanda yorumlanabilir açıklamalar da sunarak klinik karar destek sistemlerini aktif öğrenme sistemlerine dönüştürmektedir. Literatürün hedefe yönelik sentezine dayanan bu kavramsal çerçeve, yapay zekâ sistemleri ile klinisyenler arasında çift yönlü bir öğrenme döngüsü tanımlamaktadır. Bu döngü üç yinelemeli aşamadan oluşmaktadır: açıklamaların sunulması, klinisyenin yansıtıcı değerlendirmesi ve eylemin ve modelin geri bildirime dayalı olarak uyarlanması. Çerçeve, yapay zekânın klinik ortamlardaki eğitsel etkisini değerlendirmek amacıyla tanısal kalibrasyon, bilgi aktarımının etkililiği ve sezgisel muhakemenin gelişimi gibi yeni değerlendirme boyutları ortaya koymaktadır. Yapay zekânın pasif bir karar destek aracı yerine “öğretici ortak” olarak yeniden konumlandırılmasının sürekli mesleki gelişimi destekleyebileceği, klinik akıl yürütmeyi geliştirebileceği ve insan-yapay zekâ iş birliğini güçlendirebileceği ileri sürülmektedir.

Sonuç olarak, Simbiyotik Yapay Zekâ Tanı modeli, yetkinlik artırma odaklı yapay zekâ anlayışından öğrenme odaklı bir paradigmaya geçiş önermekte ve daha şeffaf, uyarlanabilir ve insan merkezli sağlık sistemlerinin geliştirilmesine katkı sağlama potansiyeli taşımaktadır.

Anahtar Kelimeler: Yapay zekâ, Sağlıkta dijital dönüşüm, Makine Öğrenmesi

Sorumlu Yazar / Corresponding Author: Ufuk Burak KARCIOĞLU
E posta: burakkarci23@gmail.com

Received/ Geliş Tarihi: 27.04.2026
Accepted/ Kabul Tarihi: 21.07.2026

Abstract

This study proposes a novel conceptual framework, Symbiotic AI Diagnosis, grounded in the principle of reverse mentorship within AI-driven clinical decision support systems (CDSS). While traditional approaches position artificial intelligence as an assistive tool for clinicians, this work argues that AI can also function as a real-time educational agent for physicians.

Through explainable AI (XAI) mechanisms, machine learning models not only generate predictions but also provide interpretable explanations, thereby transforming CDSS into active learning systems. Based on a targeted synthesis of the literature, the proposed framework defines a bidirectional learning loop between AI systems and clinicians, consisting of three iterative phases: explanation delivery, clinician reflection and action, and feedback-driven model adaptation. The framework introduces novel evaluation dimensions such as diagnostic calibration, knowledge transfer efficacy, and heuristic evolution to assess the educational impact of AI in clinical settings. It is argued that reframing AI as a “teaching partner” rather than a passive decision-support tool can enhance continuous professional development, improve clinical reasoning, and strengthen human–AI collaboration.

In conclusion, the Symbiotic AI Diagnosis model offers a shift from augmentation-oriented AI toward a learning-oriented paradigm, with the potential to foster more transparent, adaptive, and human-centered healthcare systems.

Key Words: Artificial Intelligence, Mentoring, Machine Learning

1.Introduction

While much discourse centers on AI as an adjunct to human expertise, where algorithms augment clinicians’ decisions, this perspective overlooks a burgeoning opportunity: AI’s potential to serve as a real-time educator for physicians, inverting the mentorship dynamic in clinical practice. This reverse mentorship framework, wherein ML models not only support but actively train healthcare providers through iterative feedback and explanatory mechanisms, could redefine continuous medical education (CME) and foster a symbiotic relationship between human intuition and algorithmic precision (1)

The global burden of disease, compounded by aging populations and rising chronic conditions, underscores the urgent need for innovative approaches to healthcare delivery. According to the WHO, non-communicable diseases account for 74% of all deaths worldwide, necessitating advanced tools to optimize resource allocation and clinical efficiency (2). In parallel, the OECD highlights that AI adoption in health systems could yield economic benefits exceeding USD 15 trillion by 2030, primarily through improved preventive care and personalized interventions (3). However, these projections hinge on addressing persistent challenges, such as clinician burnout and knowledge gaps in rapidly advancing medical fields. Surveys indicate that up to 40%

of physicians experience burnout, partly due to information overload and the pressure to stay abreast of evidence-based practices (4). Here, AI-driven CDSS can transcend passive support by embedding educational elements, such as real-time explanations of predictive outputs, thereby empowering clinicians to refine their diagnostic reasoning and adapt to novel scenarios.

Historically, CDSS have evolved from simple alert systems to sophisticated platforms incorporating ML for predictive analytics. For instance, early implementations focused on drug interaction warnings, but contemporary systems leverage deep learning to forecast patient deterioration in emergency settings (5). Despite these advances, a critical limitation persists: most CDSS prioritize decision augmentation over knowledge transfer. Physicians often view AI as a “black box,” eroding trust and limiting its educational value (4). The OECD’s analysis of AI maturity across health systems reveals that only a handful of countries, such as the United Kingdom and the United States, have achieved integrated ecosystems where AI facilitates collaborative learning, while many remain at emergent stages. This disparity amplifies inequities, particularly in low-resource settings, where access to CME is constrained (6).

The concept of reverse mentorship, originally popularized in corporate sectors to bridge generational knowledge gaps, offers a compelling adaptation for healthcare AI (7). In this context, reverse mentorship reframes AI not as a subordinate tool but as a mentor that imparts insights derived from vast datasets, challenging clinicians to question assumptions and embrace evidence-based evolution. For example, AI models equipped with explainable modules can dissect diagnostic probabilities, highlight overlooked variables, and foster reflective practice (Kukreti, 2025). This approach aligns with the WHO’s ethical guidelines for AI in health, which emphasize transparency, accountability, and human-centered design to mitigate risks like algorithmic bias (8). By inverting the traditional hierarchy where physicians “train” AI through data labeling, reverse mentorship positions ML as an active educator, promoting lifelong learning in real-time clinical encounters.

Empirical evidence supports the viability of this framework. In oncology, AI systems have demonstrated superior pattern recognition in imaging, not only aiding diagnosis but also educating radiologists on subtle features through counterfactual explanations (9). Similarly, predictive models for sepsis detection provide real-time feedback loops, enabling physicians to calibrate their heuristics against algorithmic outputs (9). The OECD warns, however, of potential pitfalls, including over-reliance on AI (automation bias) and workforce disruptions, in which roles may shift toward higher-order tasks such as ethical oversight (1). Addressing these requires robust governance, as outlined in the WHO’s readiness assessment for AI integration across European health systems, which advocates for interdisciplinary training and data sovereignty (6).

This paper introduces the Symbiotic AI Diagnosis framework, a conceptual model that operationalizes reverse mentorship in CDSS. Drawing on existing literature, it posits AI as a continuous learning partner, analyzing feedback from CDS interactions to measure educational impact. Unlike conventional CDSS, which focus on outcome metrics like diagnostic accuracy, this framework evaluates AI’s role in enhancing physician competencies through metrics such as knowledge retention and decision confidence (10). By adapting reverse mentorship principles emphasizing mutual respect and iterative Exchange to AI-healthcare interfaces, the model envisions a future where ML not only supports but also elevates clinical acumen, ultimately advancing equitable, high-quality care (3).

While much discourse centers on AI as an adjunct to human expertise, in which algorithms augment clinicians’ decisions, this perspective largely overlooks a burgeoning opportunity: AI’s potential to serve as a real-time educator for physicians through reverse mentorship. Although several XAI and human-AI collaboration frameworks have been proposed to enhance transparency and trust in clinical decision support systems, most remain focused on improving clinician acceptance of algorithmic recommendations or using human oversight primarily for bias correction and model refinement. These approaches typically maintain a unidirectional dynamic in which the AI supports or is corrected by the physician.

In contrast, the Symbiotic AI Diagnosis framework introduces a novel reverse mentorship paradigm that fundamentally inverts this relationship. It positions advanced machine learning models as active educational mentors capable of delivering real-time, evidence-based guidance to physicians. Through explainable AI mechanisms, the framework establishes a bidirectional, closed-loop, symbiotic learning ecosystem comprising three iterative phases: explanation delivery, clinician reflection and action, and feedback-driven model adaptation. This approach differs from existing frameworks by emphasizing AI's explicit role as a teacher that fosters deliberate practice, heuristic evolution, and continuous professional development in clinicians. Furthermore, it introduces education-specific evaluation metrics such as diagnostic calibration, knowledge transfer efficacy, heuristic evolution, and symbiotic maturity, which go beyond traditional outcome measures like diagnostic accuracy or alert acceptance rates. By adapting reverse mentorship principles from organizational learning to the AI-healthcare interface, this framework represents a conceptual shift from mere augmentation-oriented AI toward a truly symbiotic, learning-oriented paradigm in clinical decision support.

2. Method

This conceptual paper is based on a targeted literature synthesis aimed at developing and supporting the Symbiotic AI Diagnosis framework. Although not a systematic review, the synthesis process was guided by an adapted version of the PRISMA 2020 statement for scoping reviews and conceptual syntheses (Page et al., 2021). The PRISMA framework was modified to fit the purpose of this study, which prioritizes conceptual integration, identification of theoretical gaps, and framework development rather than exhaustive evidence mapping or meta-analysis.

Relevant literature was identified through purposive searching of major databases, including PubMed, Google Scholar, and The Lancet archives. Combinations of the following keywords were used: "explainable AI", "XAI in healthcare", "clinical decision support systems", "reverse mentorship", "AI in medical education", "human-AI collaboration", "symbiotic AI", and "physician training with AI". Priority was given to publications from 2021 onward to capture recent developments in XAI technologies, CDSS implementations, and international policy guidance on AI in health.

Studies were included if they met the following criteria: (1) peer-reviewed articles, systematic reviews, or authoritative reports from recognized organizations (e.g., WHO, OECD); (2) focus on XAI mechanisms with potential educational or reflective value in clinical settings, reverse mentorship adaptations in healthcare or organizational contexts, or empirical studies on clinician learning from AI explanations; and (3) publication date 2021 or later (except for foundational works). Records were excluded if they were outdated (pre-2021 unless highly influential), non-peer-reviewed, unrelated to healthcare AI or medical education, or lacked sufficient emphasis on transparency, trust, or knowledge transfer.

The initial search identified approximately 150 records. After removal of 32 duplicates, 118 records were screened by title and abstract, resulting in the exclusion of 85 irrelevant studies. Full-text assessment was conducted on the remaining 33 records. Ten articles were further excluded due to low conceptual alignment, insufficient focus on educational impact, or redundancy with already selected high-impact sources. Ultimately, 25 high-quality sources were included in the final synthesis. These comprised peer-reviewed articles from reputable journals, systematic reviews, and key policy documents. The flow of information through the different phases of the literature synthesis is presented in Figure 1.

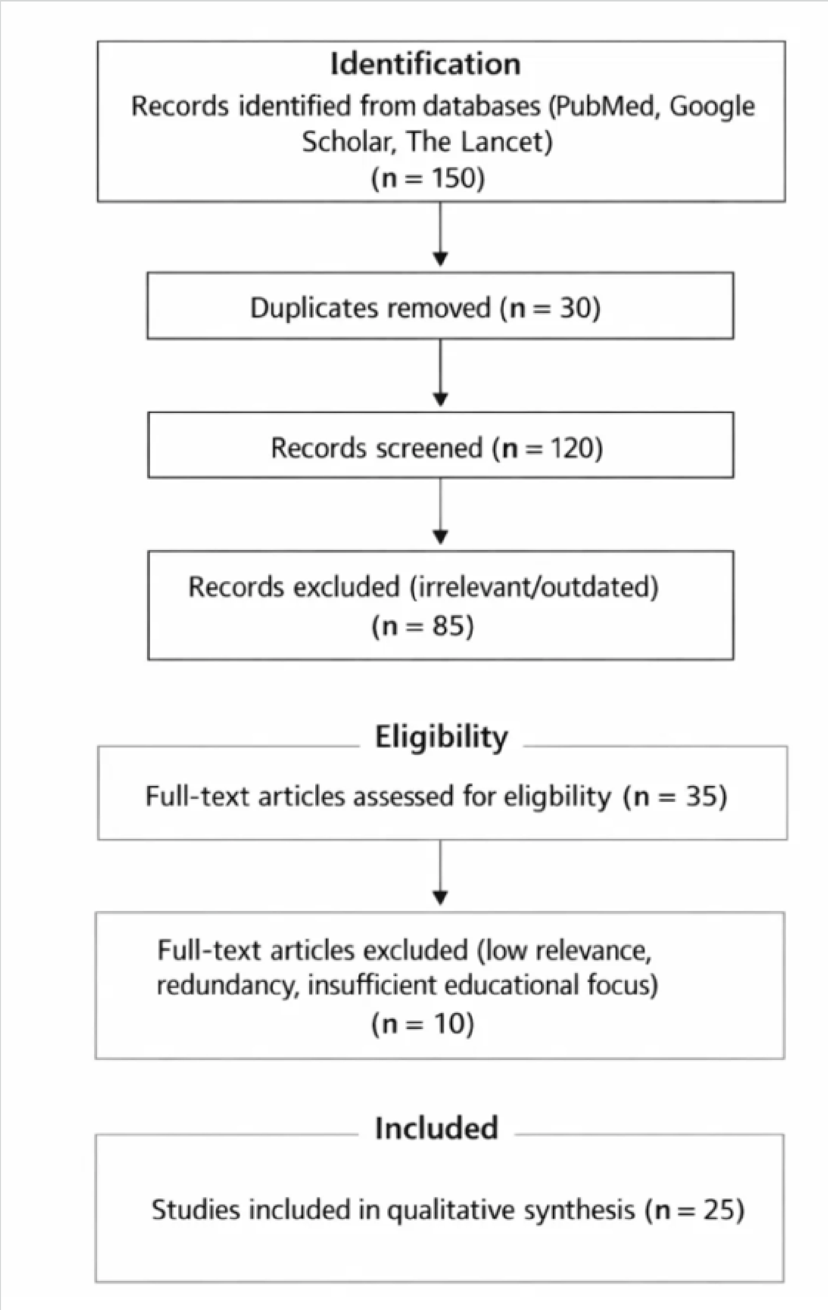

Figure 1. Targeted Literature Synthesis and Reference Selection for Conceptual Paper

3.Conceptual Framework: Reverse Mentorship in AI-Driven Clinical Decision Support

The Symbiotic AI Diagnosis framework is grounded in several established theories from education, organizational learning, and human-computer interaction. At its core lies Kolb's Experiential Learning Theory, which posits that effective learning occurs through a cyclical process of concrete experience, reflective observation, abstract conceptualization, and active experimentation. In this framework, the AI-generated explanations serve as concrete experience, clinician reflection corresponds to reflective observation, integration of insights into clinical reasoning represents abstract conceptualization, and overrides or annotations enable active experimentation. This creates a continuous, real-time learning cycle embedded directly in clinical workflows.

The framework also draws upon Argyris and Schön's Organizational Learning Theory, particularly the concepts of single-loop and double-loop learning. Single-loop learning occurs when clinicians adjust their actions based on AI explanations to improve performance, while double-loop learning emerges when they question and modify underlying assumptions and heuristics through sustained interaction with the system. Furthermore, the reverse mentorship dynamic is informed by Vygotsky's Zone of Proximal Development (ZPD), in which the AI serves as a more knowledgeable "mentor" that scaffolds the clinician's growth by providing guidance just beyond the clinician's current level of expertise.

From an organizational perspective, the model adapts reverse mentorship principles originally developed to bridge generational knowledge gaps in corporate settings (11). In healthcare, this inversion positions the AItrained on vast, continuously updated datasets as the source of contemporary, data-driven insights, while the clinician contributes contextual wisdom, ethical judgment, and real-world applicability. This bidirectional knowledge exchange fosters a symbiotic learning ecosystem rather than a hierarchical one.

These theoretical foundations collectively provide a robust conceptual basis for the Symbiotic AI Diagnosis framework, distinguishing it from purely technical XAI models by emphasizing educational processes, mutual adaptation, and long-term professional development.

The Symbiotic AI Diagnosis framework reimagines the role of machine learning (ML) within clinical decision support systems (CDSS) by adapting the concept of reverse mentorship originally developed to foster intergenerational knowledge exchange in organizational settings to the human-AI interface in healthcare (12). In traditional mentorship, senior clinicians impart experiential wisdom to juniors; reverse mentorship inverts this hierarchy, enabling less experienced individuals to share contemporary insights, such as digital fluency or novel perspectives, with veterans. Applied to AI-driven CDSS, reverse mentorship positions advanced ML models trained on expansive, heterogeneous datasets as the "mentor" that delivers real-time, evidence-based guidance to physicians, while physicians, through their interactions, feedback, and overrides, iteratively refine the system. This bidirectional dynamic cultivates a symbiotic learning ecosystem in which AI educates clinicians on subtle patterns and probabilistic reasoning, while human expertise grounds algorithmic outputs in contextual clinical judgment (1).

Central to this framework is the premise that ML models possess inherent educational capacity beyond mere predictive output. Explainable AI (XAI) techniques, such as SHAP (SHapley Additive exPlanations) values, feature importance rankings, and counterfactual reasoning, transform "black-box" predictions into interpretable narratives that elucidate why a particular diagnosis or intervention ranks highest (13). These explanations serve as pedagogical tools, prompting physicians to reflect on overlooked variables, recalibrate cognitive heuristics, and integrate novel evidence into practice. For instance, when an ML model flags early sepsis with high confidence based on temporal vital sign trends and biomarker trajectories, its XAI module can highlight the differential weighting of lactate elevation versus heart rate variability, thereby educating the clinician on emerging pathophysiological nuances that may not yet be codified in guidelines (9). Over repeated encounters, this process fosters deliberate practice akin to simulation-based training, seamlessly embedded in real-world workflows (4).

The reverse mentorship loop operates through three interdependent phases: explanation delivery, clinician reflection and action, and model adaptation via feedback. In the first phase, the CDSS presents not only a recommendation but also a structured explanation tailored to the clinician's query (e.g., "What features drove this prediction?"). This draws on clinician-informed XAI evaluation frameworks that prioritize transparency, relevance to decision-making, and alignment with clinical reasoning (13). In the reflection phase, physicians actively engage in accepting, overriding, or annotating outputs, generating rich metadata on trust, knowledge gaps, and contextual overrides. The adaptation phase closes the loop: aggregated, anonymized feedback retrains or fine-tunes the model (e.g., through reinforcement learning from human preferences or active learning queries), while longitudinal clinician-level analytics track improvements in diagnostic calibration, confidence scores, and knowledge retention (1).

This framework extends beyond passive decision augmentation by quantifying AI's educative impact. Proposed metrics include:

Explanatory fidelity, alignment between XAI outputs and clinician-endorsed rationales.

Knowledge transfer efficacy, pre/post-encounter changes in physician performance on similar cases (simulated or real).

Heuristic evolution shifts in override patterns, indicating reduced reliance on outdated biases.

Symbiotic maturity, mutual improvement indices, where AI accuracy rises with clinician expertise, and vice versa.

Such metrics differentiate this approach from conventional CDSS, which emphasize outcome metrics like sensitivity/specificity or alert acceptance rates (11). By contrast, the reverse mentorship paradigm aligns with WHO guidance on human-centered AI, emphasizing transparency, continuous learning, and empowerment of health professionals rather than automation displacement (8).

Integration into clinical practice requires deliberate design choices. CDSS interfaces must embed "teachable moments" without cognitive overload using adaptive explanation depth (brief for routine cases, detailed for high-stakes or uncertain ones) and multimodal formats (text, visual heatmaps, natural language summaries). Governance structures, informed by OECD analyses of AI in health workforces, should ensure equitable access, bias mitigation in training data, and protection against automation bias, where over-trust in AI explanations erodes independent reasoning (1). Ethical safeguards, including audit trails for educational interactions and opt-out mechanisms, uphold physician autonomy while promoting accountability (8).

In essence, the Symbiotic AI Diagnosis framework envisions AI not as a subordinate assistant but as a dynamic, ever-evolving mentor embedded at the point of care. By inverting the mentorship hierarchy, it harnesses ML's scale and pattern-recognition prowess to accelerate lifelong learning, mitigate knowledge decay amid exponential medical growth, and cultivate resilient clinicians capable of thriving in an AI-augmented future (1). This conceptual shift from augmentation to symbiotic education offers a visionary pathway toward more adaptive, equitable, and human-centered healthcare delivery. Subsequent sections will outline methodological approaches to operationalizing and empirically evaluating this framework.

Figure 2 illustrates the conceptual framework of reverse mentorship in AI-driven clinical decision support systems, depicting the bidirectional symbiotic learning loop between machine learning models and physicians, with iterative phases of explanation delivery, clinician reflection, and action, and model adaptation via feedback

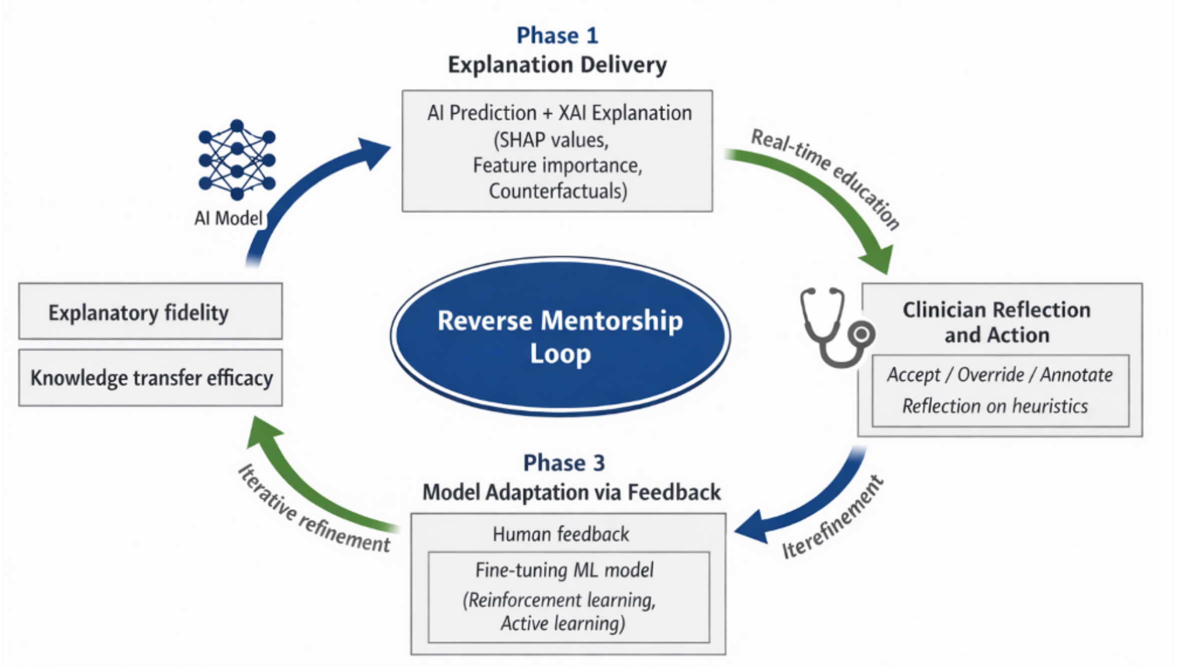

Figure 2. Conceptual framework of reverse mentorship in AI-driven clinical decision support systems

4.Empirical Insights and Case Studies

Empirical evidence from recent implementations of explainable AI (XAI) in clinical decision support systems (CDSS) provides growing support for the educational potential embedded in the Symbiotic AI Diagnosis framework. While direct applications of reverse mentorship remain emerging, analogous studies demonstrate how XAI mechanisms such as feature importance, SHAP values, and counterfactual explanations facilitate real-time clinician learning, reflective practice, and calibration of diagnostic heuristics, effectively inverting traditional knowledge flows by allowing algorithmic insights to educate physicians at the point of care (15-16).

In critical care settings, deep learning models for early sepsis prediction have shown not only improved patient outcomes but also indirect educational benefits through explanatory outputs. For instance, deployment of the COMPOSER model in hospital environments was associated with a 17% relative reduction in in-hospital sepsis mortality and enhanced bundle compliance, partly attributed to clinicians' increased awareness of key predictive features highlighted by the system (Boussina et al., 2024). Although primarily outcome-focused, these systems incorporate post-hoc explanations that prompt physicians to reconsider overlooked vital sign trajectories or biomarker patterns, fostering iterative knowledge refinement akin to reverse mentorship loops (17). Counterfactual explanations in sepsis therapy CDSS further enable clinicians to explore "what-if" scenarios, such as alternative interventions based on adjusted lactate or respiratory rates, promoting a deeper understanding of pathophysiological dynamics and supporting reflective decision-making (19).

Broader meta-analyses and reviews reinforce these insights. A comprehensive synthesis of XAI in CDSS across domains, including ICU sepsis monitoring, oncology prognosis, and predictive analytics, highlights that SHAP-based and counterfactual methods improve explanatory fidelity and clinician trust, indirectly supporting knowledge transfer by making algorithmic reasoning accessible and actionable (15). Quantitative assessments indicate that XAI-augmented systems can enhance physician reliance calibration without inducing over-trust, a prerequisite for educational symbiosis (19). Table 1 summarizes key empirical examples, focusing on settings, XAI mechanisms, observed impacts on clinician behavior or learning, and alignment with reverse mentorship principles.

Table 1. Selected Empirical Insights from XAI-Enabled CDSS Demonstrating Educational or Reflective Impacts

Clinical Domain	*XAI Mechanism(s)*	*Key Empirical Findings*	*Alignment with Reverse Mentorship*
Sepsis prediction (ICU)	Deep learning alerts with implicit feature highlighting	17% relative mortality reduction; increased bundle compliance; clinician awareness of predictive cues	Real-time feedback on overlooked features educates on early warning signs
Sepsis therapy recommendation	Counterfactual explanations via federated learning	Improved interpretability; clinicians explore alternative paths, enhancing treatment rationale understanding	"What-if" scenarios foster reflective practice and heuristic refinement
Glaucoma referrals (primary care)	Intrinsic/post-hoc explanations addressing knowledge imbalance	Better assessment of AI predictions against clinician experience; improved referral decisions	AI bridges expertise gaps, acting as mentor to non-specialists
Multiple (sepsis, oncology, ICU)	SHAP, LIME, counterfactuals	Enhanced transparency, trust, and usability; indirect knowledge transfer via actionable insights	Quantifies explanatory fidelity as pathway to clinician upskilling

These cases, while not yet fully realizing bidirectional model adaptation from clinician feedback in all instances, illustrate the pedagogical promise of XAI-integrated CDSS. They underscore how real-time explanations can transform passive decision support into active educational encounters, aligning closely with the proposed symbiotic framework (1,8). Future prospective studies incorporating longitudinal tracking of clinician performance metrics, such as diagnostic calibration pre- and post-exposure, will be essential to quantify knowledge transfer efficacy more rigorously. The following section addresses the challenges and ethical considerations inherent in scaling this approach, positioning AI primarily as a passive decision-augmenting tool. Empirical insights from analogous XAI implementations ranging from sepsis prediction in intensive care to glaucoma referrals in primary settings demonstrate that explanatory mechanisms such as SHAP values, counterfactual reasoning, and feature attribution can indeed prompt reflective practice, enhance diagnostic calibration, and bridge knowledge gaps among clinicians (20-22). These findings align with the proposed bidirectional loop, in which AI explanations serve as real-time educational inputs and physician feedback drives model refinement, fostering a symbiotic dynamic that elevates both algorithmic accuracy and human expertise over time.

This approach addresses longstanding limitations in CDSS adoption, including clinician distrust of opaque “black-box” outputs and the risk of automation bias, by prioritizing transparency and active engagement (23). By inverting the mentorship hierarchy, the framework empowers AI to act as a continuous educator, mitigating knowledge decay in an era of exponential growth in medical evidence and supporting lifelong learning embedded directly into clinical workflows. Such integration resonates with broader policy imperatives: the WHO’s guidance on human-centered AI underscores the need for systems that empower rather than displace health professionals, while OECD analyses highlight the potential of AI to transform the workforce, including through continuous, context-specific education (1,8). In resource-constrained settings, this model could democratize access to advanced insights, narrowing diagnostic disparities and promoting equitable care delivery (6).

Nevertheless, realizing the framework’s full potential hinges on overcoming implementation barriers. Over-reliance on AI explanations risks eroding independent clinical reasoning, particularly if physicians defer excessively to algorithmic outputs without critical appraisal. Longitudinal studies are needed to assess whether repeated exposure to XAI-driven “teachable moments” genuinely improves heuristic evolution or inadvertently fosters dependency. Moreover, the bidirectional feedback mechanism raises data governance questions: aggregating clinician overrides and annotations for model retraining must safeguard privacy, mitigate bias amplification from imbalanced feedback, and ensure equitable representation across diverse clinical populations (1). Ethical alignment with international standards remains paramount, as emphasized by WHO frameworks that advocate accountability, inclusivity, and the mitigation of algorithmic harms (10).

The proposed metrics, explanatory fidelity, knowledge transfer efficacy, heuristic evolution, and symbiotic maturity, offer a pathway to quantify educational impact beyond traditional performance indicators like diagnostic accuracy or alert acceptance. These metrics could inform system redesign, clinician training programs, and regulatory evaluation, shifting the discourse from AI as an efficiency enhancer to AI as a collaborative learning partner. In essence, the reverse mentorship paradigm challenges the field to reframe AI not as a replacement for human judgment but as a catalyst for its continuous refinement, ensuring that technological advancement strengthens rather than supplants the art and science of medicine.

5. Future Directions

Translating the Symbiotic AI Diagnosis framework from conceptual model to scalable clinical reality will require multifaceted, prospective research agendas. First, rigorous empirical validation through mixed-methods pilots in real-world CDSS environments is essential. Longitudinal cohort studies could track physician performance metrics, such as diagnostic calibration, confidence scores, and knowledge retention, pre- and post-implementation of reverse mentorship-enabled systems, and compare them with standard CDSS (20,24). Randomized controlled trials across diverse settings (e.g., emergency departments, primary care, oncology) would quantify the efficacy of knowledge transfer and the evolution of heuristics, while qualitative analyses would explore clinicians' perceptions of AI as a mentor versus a tool.

Technical advancements should focus on enhancing the pedagogical sophistication of XAI modules. Adaptive explanation engines that tailor depth and format to individual clinician expertise, case complexity, and cognitive load could optimize "teachable moments" without overwhelming users (25 Integration of reinforcement learning from human preferences (RLHF) or federated learning approaches would enable secure, privacy-preserving model adaptation from clinician feedback across institutions, addressing data sovereignty concerns (Düsing et al., 2024). Multimodal interfaces combining natural language summaries, visual heatmaps, and interactive counterfactuals hold promise for broader accessibility, particularly in low-resource contexts.

Workforce and educational implications warrant dedicated exploration. Incorporating reverse mentorship principles into medical curricula and continuing professional development programs could prepare future clinicians for symbiotic AI interactions, building AI literacy alongside clinical acumen (1,6,10). Interdisciplinary collaborations merging informatics, education science, and ethics should develop governance frameworks to prevent automation bias, ensure equitable access, and monitor long-term workforce shifts toward higher-order tasks like ethical oversight and patient-centered innovation.

Regulatory and policy pathways must evolve in parallel. Prospective health technology assessments could evaluate symbiotic CDSS against educational impact criteria, informing reimbursement models and certification standards (3). Global harmonization efforts, guided by WHO and OECD recommendations, should prioritize bias mitigation, transparency mandates, and inclusive data strategies to realize the framework's equity potential across diverse health systems (6,8).

Ultimately, these directions converge toward a vision of AI-augmented healthcare where continuous, bidirectional learning becomes the norm, cultivating resilient, adaptive clinicians and more intelligent, humane systems.

Despite its promising potential, implementing the Symbiotic AI Diagnosis framework faces several significant challenges. First, automation bias and over-reliance on AI explanations may erode independent clinical reasoning, particularly among less experienced physicians. Second, integrating bidirectional feedback loops raises important data governance and privacy concerns, especially when aggregating clinician overrides and annotations for model retraining across institutions. Third, algorithmic bias and inequities in training data could exacerbate existing disparities in healthcare delivery, particularly in low-resource settings. Technical challenges such as cognitive overload from overly detailed explanations and the computational cost of real-time model adaptation also require careful consideration.

The current study has several limitations inherent to its conceptual nature. It relies on a targeted literature synthesis rather than a comprehensive systematic review or empirical validation. The proposed framework and metrics have not yet been tested in real-world clinical environments, limiting claims about their practical effectiveness. Additionally, integrating reverse mentorship principles from corporate settings into healthcare AI requires further cultural and contextual adaptation.

From a practical standpoint, successful implementation will require thoughtful interface design (adaptive explanation depth, multimodal outputs), institutional support for clinician training, and robust governance structures aligned with WHO and OECD recommendations. If effectively realized, the framework could contribute to reduced knowledge decay, improved diagnostic calibration, enhanced clinician confidence, and more equitable access to advanced medical insights.

6. Conclusion

The Symbiotic AI Diagnosis framework introduces a visionary yet grounded paradigm: reverse mentorship in AI-driven clinical decision support, wherein machine learning models evolve from decision aids into real-time educators that actively train physicians through transparent explanations and iterative feedback. By inverting traditional hierarchies and embedding educational mechanisms at the point of care, this approach harnesses AI's analytical scale to accelerate lifelong learning, mitigate knowledge gaps, and foster reflective practice amid rapid medical advancement.

Supported by emerging empirical evidence from XAI applications in critical care, imaging, and primary diagnostics, the framework demonstrates tangible potential to enhance clinician calibration, trust, and performance while aligning with international imperatives for human-centered, equitable AI in health. Despite challenges in implementation, ethics, and validation, the proposed metrics and methodological pathways provide a roadmap for rigorous evaluation and refinement.

In an era where AI permeates healthcare, shifting the narrative from augmentation to symbiosis offers profound promise: empowering clinicians to thrive alongside intelligent systems, ensuring human judgment remains central while leveraging algorithmic insights for superior, patient-centered care. This reverse mentorship model invites the field to embrace a future where AI and physicians learn from one another continuously, collaboratively, and equitably, advancing the shared goal of healthier populations worldwide.

References

1. OECD. (2024a). Artificial intelligence and the health workforce. OECD Publishing. https://doi.org/10.1787/9a31d8af-en

2. World Health Organization. (2024). Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models. https://www.who.int/publications/i/item/9789240084759

3. OECD. (2024b). AI in health: Huge potential, huge risks. OECD Publishing.

4. Tun, H. M., Rahman, H. A., Naing, L., & Malik, O. A. (2025). Trust in artificial intelligence–based clinical decision support systems among health care workers: systematic review. Journal of Medical Internet Research, 27, e69678.https://doi.org/10.2196/69678

5. Yeo, H. J., Noh, D., Kim, T. H., Jang, J. H., Lee, Y. S., Park, S., ... & Kang, H. K. (2024). Development and validation of a machine learning-based model for post-sepsis frailty. ERJ Open Research, 10(5). https://doi.org/10.1183/23120541.00166-2024

6. World Health Organization Regional Office for Europe. (2025). [Readiness assessment and AI integration in European health systems].

7. Tenajas, Rebeca, and David Miraut. "1) Sustainable Communication Workflows Through Smart Message Delegation in Public Healthcare." (2025).

8. World Health Organization. (2021). Ethics and governance of artificial intelligence for health: WHO guidance. https://www.who.int/publications/i/item/9789240029200

9. Huang, Junyan, et al. "Application of artificial intelligence in medical imaging for tumor diagnosis and treatment: a comprehensive approach." Discover Oncology 16.1 (2025): 1625. doi: 10.1007/s12672-025-03307-3

10. Elhaddad M, Hamam S. AI-Driven Clinical Decision Support Systems: An Ongoing Pursuit of Potential. Cureus. 2024 Apr 6;16(4):e57728. doi: 10.7759/cureus.57728

11. Marcinkus Murphy, Wendy. "Reverse mentoring at work: Fostering cross-generational learning and developing millennial leaders." Human resource management 51.4 (2012): 549-573.

12. Giebel GD, Raszke P, Nowak H, Palmowski L, Adamzik M, Heinz P, Tokic M, Timmesfeld N, Brunkhorst FM, Wasem J, Blase N. Improving AI-Based Clinical Decision Support Systems and Their Integration into Care From the Perspective of Experts: Interview Study Among Different Stakeholders. JMIR Med Inform. 2025 Jul 7;13:e69688. doi: 10.2196/69688.

13. Caterson J, Lewin A, Williamson E. The application of explainable artificial intelligence (XAI) in electronic health record research: A scoping review. Digit Health. 2024 Oct 30;10:20552076241272657. doi: 10.1177/20552076241272657.

14. Bragazzi, I., Fekih-Romdhane, F., Hallit, S., Loch, A. A., Glenn, J. M., Fessi, M. S., ... & Ben Saad, H. (2024). ChatGPT is not ready yet for use in providing mental health assessment and interventions. Frontiers

15. Athukorala, V. S., & Ilmini, W. M. K. S. (2026). Explainable AI for critical care: a systematic review of interpretable models for sepsis and ICU mortality prediction. BMC Medical Informatics and Decision Making. https://doi.org/10.1186/s12911-026-03344-0

16. Salimparsa, Mozhgan, et al. “Explainable AI for clinical decision support systems: literature review, key gaps, and research synthesis.” Informatics. Vol. 12. No. 4. MDPI, 2025.

17. Yang, X., Liu, J., Huang, X., Liang, H., Cui, P., He, S., ... & Huang, J. (2025). Machine learning-driven clinical decision support for low bone mineral density: a web-based prediction model with explainable AI integration. Bone, 117592. https://doi.org/10.1016/j.bone.2025.117592

18. Bienefeld, N., Boss, J. M., Lüthy, R., Brodbeck, D., Azzati, J., Blaser, M., ... & Keller, E. (2023). Solving the explainable AI conundrum by bridging clinicians’ needs and developers’ goals. npj digital medicine, 6(1), 94. https://doi.org/10.1038/s41746-023-00837-4

19. Duesing, C., Cimiano, P., Rehberg, S., Scherer, C., Kaup, O., Koester, C., ... & Borgstedt, R. (2024). Integrating federated learning for improved counterfactual explanations in clinical decision support systems for sepsis therapy. Artificial Intelligence in Medicine, 157, 102982. https://doi.org/10.1016/j.artmed.2024.102982

20. Abbas Q, Jeong W, Lee SW. Explainable AI in Clinical Decision Support Systems: A Meta-Analysis of Methods, Applications, and Usability Challenges. Healthcare (Basel). 2025 Aug 29;13(17):2154. doi: 10.3390/healthcare13172154.

21. Boussina A, Krishnamoorthy R, Quintero K, Joshi S, Wardi G, Pour H, Hilbert N, Malhotra A, Hogarth M, Sitapati AM, VanDenBerg C, Singh K, Longhurst CA, Nemati S. Large Language Models for More Efficient Reporting of Hospital Quality Measures. NEJM AI. 2024 Oct 24;1(11):10.1056/aics2400420. doi: 10.1056/aics2400420.

22. Garcia-Gomez JM, Blanes-Selva V, Alvarez Romero C, de Bartolomé Cenzano JC, Pereira Mesquita F, Pazos A, Doñate-Martínez A. Mitigating patient harm risks: A proposal of requirements for AI in healthcare. Artif Intell Med. 2025 Sep;167:103168. doi: 10.1016/j.artmed.2025.103168.

23. Rosenbacke R, Melhus Å, McKee M, Stuckler D. AI and XAI second opinion: the danger of false confirmation in human-AI collaboration. J Med Ethics. 2025 May 21;51(6):396-399. doi: 10.1136/jme-2024-110074

24. Yang, C. C. (2022). Explainable artificial intelligence for predictive modeling in healthcare. Journal of healthcare informatics research, 6(2), 228-239. https://doi.org/10.1007/s41666-022-00114-1

25. Yang, Y., & Liu, X. (2026). Application of explainable artificial intelligence integrating with electronic health record in oncology. Exploration of Targeted Anti-tumor Therapy, 7, 1002357. https://doi.org/10.37349/etat.2026.1002357