



Advancing Problem-Based Learning with Clinical Reasoning for Improved Differential Diagnosis in Medical Education

Yuansong Xu

School of Information Science and Technology
ShanghaiTech University
Shanghai, China
xuys2023@shanghaitech.edu.cn

Shaohan Shi

School of Information Science and Technology
ShanghaiTech University
Shanghai, Shanghai, China
shishh2023@shanghaitech.edu.cn

Yuheng Shao

School of Information Science and Technology
ShanghaiTech University
Shanghai, China
shaoyh2024@shanghaitech.edu.cn

Chang Jiang*

Department of Clinical Medicine
Shanghai Clinical Research and Trial Center
Shanghai, China
cjiang_fdu@yeah.net

Jiahe Dong

School of Information Science and Technology
ShanghaiTech University
Shanghai, China
dongjh@shanghaitech.edu.cn

Quan Li*

School of Information Science and Technology
ShanghaiTech University
Shanghai, Shanghai, China
liquan@shanghaitech.edu.cn

Abstract

Medical education increasingly emphasizes students' ability to apply knowledge in real-world clinical settings, focusing on evidence-based clinical reasoning and differential diagnoses. Problem-based learning (PBL) addresses traditional teaching limitations by embedding learning into meaningful contexts and promoting active participation. However, current PBL practices are often confined to medical instructional settings, limiting students' ability to self-direct and refine their approaches based on targeted improvements. Additionally, the unstructured nature of information organization during analysis poses challenges for record-keeping and subsequent review. Existing research enhances PBL realism and immersion but overlooks the construction of logic chains and evidence-based reasoning. To address these gaps, we designed *e-MedLearn*, a learner-centered PBL system that supports more efficient application and practice of evidence-based clinical reasoning. Through controlled study (N=19) and testing interviews (N=13), we gathered data to assess the system's impact. The findings demonstrate that *e-MedLearn* improves PBL experiences and provides valuable insights for advancing clinical reasoning-based learning.

Keywords

Problem-Based Learning, Medical Education, Differential Diagnosis

ACM Reference Format:

Yuansong Xu, Yuheng Shao, Jiahe Dong, Shaohan Shi, Chang Jiang, and Quan Li. 2025. Advancing Problem-Based Learning with Clinical Reasoning for Improved Differential Diagnosis in Medical Education. In *CHI Conference on Human Factors in Computing Systems (CHI '25)*, April 26–May 01, 2025, Yokohama, Japan. ACM, New York, NY, USA, 32 pages. <https://doi.org/10.1145/3706598.3713772>

*Quan Li and Chang Jiang are the corresponding authors.



This work is licensed under a Creative Commons Attribution 4.0 International License.
CHI '25, Yokohama, Japan

© 2025 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-1394-1/25/04

<https://doi.org/10.1145/3706598.3713772>

1 Introduction

Modern medical education emphasizes students' ability to utilize their knowledge to reason and solve problems in real-world contexts, as well as fostering self-directed and continuous learning skills [15, 23, 59, 61, 107]. However, traditional teaching methods tend to be “*teacher-centered and lecture-focused*” [76, 82], with students passively receiving information and focusing on memorization and repetition. This approach may hinder the development of students' active learning and practical problem-solving skills. To facilitate students' active learning and usage of knowledge in real clinical scenarios, Problem-Based Learning (PBL) was proposed and developed [9, 11, 53, 116]. Unlike traditional teacher-centered methods, PBL integrates learning into meaningful problem scenarios, promoting active participation, collaboration, and the solving of complex real-life problems. Specifically, the PBL process typically involves presenting a problem, discussing it, conducting research and analysis, solving it, and summarizing the findings to gain knowledge and insights [11, 18].

Despite PBL's effectiveness in simulating real clinical diagnostic scenarios and engaging students in problem analysis and resolution, its primary application remains largely confined to medical teaching contexts [10, 53]. This limits the ability of students to engage in self-directed, targeted learning and improvement through PBL. Typically, the cases used in PBL are chosen by the instructors based on their experience and available clinical cases, which restricts the opportunity for targeted learning. Students often review case data selectively based on their experience and preferences, but without additional guidance, inexperienced learners may struggle to find specific cases of interest within a vast clinical dataset. Moreover, in clinical diagnostics, physicians are often required to conduct *differential diagnosis*, which allows them to integrate and analyze multimodal data to construct a logical chain of reasoning [42, 57, 78]. As more evidence is gathered, the range of possible diagnoses narrows until a final diagnosis is determined. Senior physicians often rely on their existing knowledge to mentally organize this evidence. Although PBL aims to develop students' evidence-based reasoning through real-world scenarios, current practices often present data,

such as medical records, laboratory results, and radiological images, in a fragmented manner with slides and paper formats, making it inconvenient to organize and analyze comprehensively. Online platforms like *Lecturio*¹ and *eLfH*² offer some integration for viewing and downloading materials, but still lack contextual reference information and interactive features. This limits the efficiency of specific analyses, such as reference range checking and image analysis, and hampers the integrated analysis of diverse data modalities. Consequently, this restricts the ability to construct and reason through logical chains in differential diagnosis based on comprehensive case information. Additionally, the methods for recording information during these analysis processes are often ineffective. Notes and electronic documents are frequently unstructured, making them difficult to review and complicating collaborative discussions. Online documents like *Google Docs*³ also present challenges due to poor integration with original electronic medical records. These issues complicate the recording, review, and retrospective analysis, which are crucial for effective learning and improvement.

Existing research has explored various applications of PBL, including tools designed to enhance critical thinking [2, 58], and the implementation and challenges of PBL in online courses [31, 100]. However, these studies generally extend PBL's characteristics without altering its presentation method, leaving students facing ongoing challenges in targeted training and learning. Simulation-based learning, which uses tools to recreate clinical scenarios, offers a way to advance PBL beyond traditional classroom environments [46, 60, 94]. By leveraging technologies such as virtual reality (VR) and augmented reality (AR), these methods create immersive clinical settings [90], allowing students to engage in problem-solving with real-time interaction and feedback. Nevertheless, these studies often emphasize enhancing the realism of PBL simulations while overlooking the need for comprehensive analysis, summarization, and record-keeping essential to the learning process. Clinical Decision Support Systems (CDSS), developed using data-driven and AI techniques [32, 33, 54, 86, 119], are also employed for analyzing and diagnosing case data [28, 39, 117]. While CDSS aids physicians in clinical decision-making by detecting patterns and making predictions, it primarily focuses on data completeness and richness to enhance diagnostic accuracy and efficiency. However, it often overlooks the importance of forming a logical reasoning process. This contrasts significantly with the PBL approach, which emphasizes the development of metacognition and reasoning skills in educational contexts. Although existing tools such as *NEJM Healer* [52] provide case-based clinical reasoning support and training, they focus more on the reasoning and evaluation processes within the diagnostic workflow of specific cases. These tools lack considerations based on the PBL framework, such as actively exploring evidence through questioning (e.g. patient interviews) and explicitly presenting the logical reasoning process.

Drawing on these insights, our goal is to gain a deeper understanding of students' logical and cognitive processes in clinical diagnosis to enhance their ability to apply knowledge and skills in real-world clinical settings. Following the general process of PBL in the medical field, which includes four key elements: identifying

and defining the problem, generating hypotheses and learning issues, engaging in self-directed learning and research, and applying new knowledge to solve the problem [11, 18], we aim to develop a learner-centered PBL tool. This tool is designed to support students more effectively and facilitate the application of their knowledge and skills in clinical practice. To address existing research gaps and understand students' needs and barriers within the current PBL framework, we propose the following research questions (**RQ1–RQ2**): **RQ1:** *What are the needs and concerns of teachers and students during PBL practice and learning?* **RQ2:** *Under what conditions, and in what ways, do students learn and improve through the PBL process?*

To answer these research questions, we conduct an iterative design process in collaboration with a team comprising HCI researchers, senior physicians (teachers), and novice physicians (students). These physicians perform as teachers and students in their medical education activities within the PBL framework. Our study consists of three phases: In **Phase One**, we collaborate with physicians to leverage their domain knowledge and gain a thorough understanding of current PBL practices in medical education and the underlying clinical expectations such as clinical reasoning and differential diagnosis. We examine the experiences of both students and teachers throughout the PBL process, focusing on how students learn and improve, the enhancements teachers expect, and the perceived benefits of their current training. We also investigate students' actual experiences and perceptions, the challenges they face during the process, and their needs and concerns. In **Phase Two**, we brainstorm learner-centered design features and potential scenarios that align with the PBL steps of hypothesizing, gathering information, analyzing, and concluding based on case studies. We iterate on these features through design cycles, ultimately determining the essential features needed to assist users in completing the PBL process and engaging in targeted learning for continuous improvement. In **Phase Three**, we recruit 13 physicians (both students and teachers) involved in the PBL process to be involved in a user feedback session to gather their feedback on these features and gain insights into how these features can promote a learner-centered PBL process. In summary, the contributions are as follows:

- We collect perceptions of current PBL practices from both teachers and students, identifying their expectations, needs for improvement, and concerns related to PBL activities.
- We identify the clinical reasoning-based differential diagnosis process, which involves evidence supplementation and further reasoning to reach a conclusion. Based on this, we designed features through iterations to develop a system that supports learning and improvement throughout the PBL process.
- We collaborate with physicians involved in the PBL process to gather feedback, gaining insights into how these features can aid learners during the PBL process.

2 Background and Related Work

2.1 Problem-Based Learning in Clinical Education Practice

Traditional medical education predominantly occurs in large classroom settings where teachers deliver instruction, and students passively receive information [76]. This approach relies heavily on

¹<https://www.lecturio.com/>

²<https://www.wolterskluwer.com/en/solutions/uptodate>

³<https://docs.google.com/>

lectures and memorization, providing students with limited opportunities to engage and apply their knowledge to real-world problems. Consequently, there is often a disconnect between theoretical knowledge and practical application, hindering the development of students' active learning and problem-solving skills in real-life scenarios [10, 18, 93].

To enhance clinical practice in medical education, PBL is proposed as an educational approach that situates learning within the context of complex and meaningful problems [9, 11]. In PBL, students work collaboratively in small groups to solve practical problems, thereby acquiring embedded knowledge in these problems and developing problem-solving skills [9, 53]. According to the PBL process proposed by Bound and Feletti [18], students are presented with a problem, raise questions, discuss the problem, assign different questions to group members for research, reconvene to discuss and summarize new knowledge, and connect it to existing knowledge. Although specific instructional processes may vary among educators [11, 18, 97], they consistently involve problem identification, analysis, hypothesis formation, testing, and revision.

Research has shown that PBL significantly enhances clinical practice among learners, particularly in the development of critical thinking and self-directed learning skills [2, 68, 82]. By simulating real-world scenarios, PBL allows students to analyze and reason through clinical diagnostic processes. Building on the strengths of PBL, our work aims to bridge the gap between theory and practice in medical education. By focusing on learner-centered PBL practices, we seek to improve the training and development of students' clinical skills through these activities.

2.2 Design Medical Diagnostic Tools with Physicians

Existing research has detailed the process of collaborating with domain stakeholders for HCI design and development, highlighting the challenges involved, especially in the early stages or when stakeholders have limited experience in data science [26, 63, 89]. Previous collaboration has primarily focused on designing user interfaces, with less emphasis on in-depth aspects such as overall problem formulation and model pipelines [41, 113, 121]. With the rise of human-centered design [3, 12, 67, 108], research has begun to explore broader collaboration with domain stakeholders. The literature suggests that utilizing abstractions, sketches, and prototypes can significantly aid stakeholders in understanding relevant background knowledge, thereby enhancing communication and understanding [6, 118, 123].

In the field of medical diagnostics, current efforts often involve extensive collaboration with physicians to understand their needs and design human-centered features to assist in the diagnostic process [28, 49, 95, 117, 118, 122]. Designing in the medical field with stakeholders requires considering multiple factors, such as promoting shared decision-making (SDM) by involving both doctors and patients [49, 50, 70, 87], and enhancing communication between radiologists and clinical physicians to support radiology-based diagnoses [111, 117, 122]. These collaborative design processes usually involve complex workflows and communication, highlighting the importance of humanistic medicine [35, 96]. For example, the clinical diagnostic process, especially for chronic diseases, involves

disease detection, health status determination, communication with patients about treatment options, and finalizing shared decision-making results [49]. The radiology workflow involves physicians requesting image studies, radiologists performing scans and examinations, recording findings in reports, necessary communication between physicians and radiologists, and physicians making treatment decisions based on the reports [122]. Furthermore, co-design with physicians often involves addressing their attitudes toward AI. A lack of knowledge and insufficient explanation may lead to mistrust in collaborating with AI [85, 118], while it is equally crucial to avoid physicians becoming over-reliant on AI based on heuristics and shortcuts [25, 65].

Recognizing the potential of PBL to enhance the application of knowledge in clinical settings, and addressing the limitation of current information presentation and organization methods for students in constructing differential diagnosis logic chains, we engaged in in-depth collaboration with physicians. Through this collaboration, we identified the existing clinical reasoning process and explored possible enhancements and features to facilitate student participation in this process during PBL activities. Existing studies show that frameworks such as User-Centered Design (UCD) enhance user satisfaction and learning outcomes in educational tool development [77], while Activity Theory provides valuable insights for analyzing human-computer interactions and system design in educational contexts [44]. Based on these insights, we conduct learner-centered design and consider users' interactions with features during the PBL process. By incorporating the principles of PBL, we encourage students to drive their exploration and analysis of evidence through questioning, while explicitly updating the logical chain throughout the analytical process. This distinguishes our approach from tools that primarily focus on clinical reasoning through case data analysis [52].

2.3 Human-Centred Medical AI

With the development and application of AI, researchers increasingly emphasize the human aspect in AI collaboration workflows, specifically focusing on human-centered AI research [108, 117, 120].

In the medical field, human-centered medical AI research encompasses the study of different stakeholders involved in the medical process. Specifically, this research explores physicians' collaboration in clinical diagnosis decision-making with AI [29, 30, 32, 81]. This includes utilizing AI assistance in classifying and explaining medical imaging diagnosis in breast cancer [29], incorporating ML explanations into physicians' decision-making workflows [32], and considering patient characteristics like history and treatment preferences to provide diagnostic recommendations [24, 27, 74]. Some studies focus on enhancing collaboration among physicians based on diagnostic processes [47, 110, 117, 122]. Xie et al. [117] clarified the understanding and needs of physicians and radiologists regarding AI-enabled medical analysis and interaction. Subsequent research further explored the radiology workflow, investigating the needs and expectations of radiologists and clinical physicians concerning AI-assisted radiology workflows [122]. Beyond collaboration among physicians, research also aims to improve communication and decision-making between physicians

and patients [4, 49, 98]. Studies on patient-centered decision-making [70, 87] demonstrate that active patient participation in clinical decisions leads to higher satisfaction and better treatment outcomes [64]. The integration of AI into shared decision-making has been explored, such as developing tools to assist patients in making decisions about prostate cancer treatment [4] and facilitating clinical decision-making for older cancer patients [49].

While human-centered AI for stakeholders in medical diagnosis has been widely researched, its application in medical education remains relatively unexplored. Despite current tools and platforms [52, 79, 115] being utilized for the learning process in medical education, these tools primarily focus on decision-making in clinical practice or general educational learning, leveraging AI to analyze and assess users' learning performance and make recommendations. Building on human-centered AI concepts and research in the medical field, our study focuses on learner experiences in PBL activities within medical education. We aim to design learner-centered features that leverage interactive elements and AI to facilitate evidence exploration and analysis, enhancing students' PBL processes and improving their clinical reasoning skills and diagnostic practice, bridging the gap between knowledge and practical application.

3 Overview of Clinical Reasoning-based Differential Diagnosis

Differential diagnosis is a systematic approach employed by physicians to determine the disease or condition responsible for a patient's symptoms [56, 106]. As demonstrated in Figure 1, this process starts with data collection, including patient history and physical examination, to gather initial information. Physicians then organize this data to identify key features relevant to clinical diagnosis. During the differential diagnosis process, physicians use these features to formulate a problem representation: a concise summary using precise medical terminology that highlights the most relevant aspects of the patient's condition. This representation includes semantic qualifiers that frame the patient's symptoms in familiar terms, facilitating the integration of related findings into clinical syndromes. Finally, physicians apply a framework (e.g. anatomic or physiological) to generate a structured list of potential diagnoses.

For each potential diagnosis, physicians evaluate the likelihood of each condition based on the clinical presentation and the patient's background, including factors such as age and occupation. During the differential diagnosis process, physicians develop appropriate diagnostic test plans. The results of these tests are used to support or rule out potential diagnoses from the list. As shown in Figure 1, with the test results providing additional evidence, physicians integrate and synthesize the collected data, comparing clinical findings with known disease patterns and pathophysiology. This allows them to update their problem representation and diagnostic reasoning, leading to revised test or treatment plans. This iterative process incorporates new evidence and continuously refines the diagnostic reasoning. Through critical thinking and logical reasoning, physicians narrow down the differential diagnostic list to identify the most likely diagnosis. The process of confirming a diagnosis resembles a funnel-shaped curve, where the scope narrows as evidence accumulates, ultimately leading to a final conclusion [14, 47]. Diagnostic approaches can be categorized into two

types based on cognitive strategies: *heuristic* [72, 73] and *metacognitive* [36, 75]. *Heuristic* reasoning relying on past experiences and intuition to identify disease patterns quickly, while *metacognition* involves reflecting on and adjusting diagnostic decisions based on new information. In differential diagnosis, both *heuristic* reasoning and *metacognition* are used to identify potential diagnoses and refine decisions throughout the process [84].

PBL aims to train students to solve real-world problems in complex, practical contexts, similar to the process of differential diagnosis, by employing systematic thinking and evidence-based support clinical reasoning. This approach involves heuristic reasoning to generate initial hypotheses and metacognition to reflect on and adjust thinking processes. While much of the research in medical education focuses on case-based learning beyond clinical decision-making [81, 103, 104], there is relatively little research that considers design from the perspective of logical reasoning-based on differential diagnosis. Our study addresses these considerations by exploring design spaces that assist medical students in enhancing their clinical diagnostic reasoning through the PBL process.

4 Method

Our study is structured into three phases. In **Phase One**, we conduct discussion sessions with senior and novice physicians involved in the current PBL process. We specifically gather insights from their roles as teachers (senior physicians) and students (novice physicians), identifying needs and barriers from each perspective, and analyzing the findings. In **Phase Two**, we brainstorm learner-centered designs and scenarios, and iterate on these designs to develop features that effectively support students in completing the PBL process, with a focus on clinical reasoning-based differential diagnosis. In **Phase Three**, we propose a prototype for a learner-centered PBL process. We then conduct evaluations with participants involved in PBL activities to gather qualitative and quantitative feedback on the usability of the system features. All studies were conducted with the approval of the Institutional Review Board (IRB) and with informed consent from the participants.

5 Phase One: Identify Current Practices and Challenges in PBL

5.1 Procedures and Data Analysis

To answer **RQ1** and **RQ2**, we conducted discussion sessions in **Phase One** with physicians (they are not co-authors of this work) from our collaboration team. This includes senior physicians serving as teachers in PBL activities: a Neurosurgery Attending physician S1 (female, age:36) and an Orthopedic Associate Chief physician S2 (male, age: 47), and novice physicians participating as students: two Orthopedic Intern physicians N1 (male, age:26), N2 (male, age:27). We held four separate discussion sessions with each physician, each session involved semi-structured interviews featuring open-ended questions.

With the senior physicians, we explored their typical organization of PBL teaching activities, covering aspects such as case selection, process explanation and guidance, and post-learning summarization. We also delved into their expectations for student learning during the PBL process, asking questions like “*Which aspects do you think PBL can enhance in students?*” and “*What are the similarities*

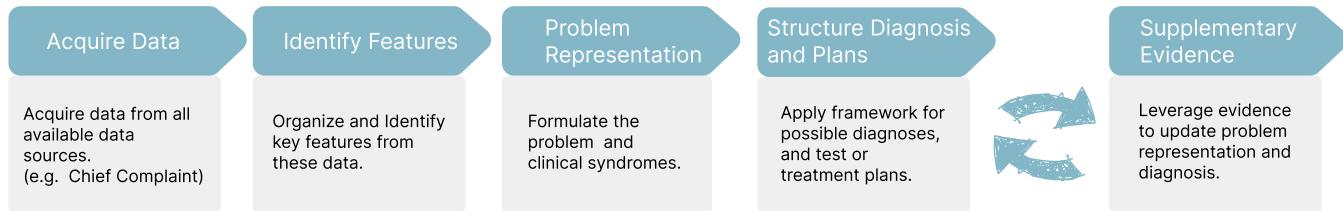


Figure 1: Overview of Differential Diagnosis workflow. The process consists of acquiring data, identifying key features from data, formulating problem representation from these features, developing diagnosis in a structured framework with the following plans for test and treatment, and iteratively updating diagnosis and plans with supplement evidence from test results.

and differences between PBL and clinical case reasoning, and how can PBL promote students' clinical reasoning skills?" Additionally, we gathered their insights and concerns on the current process.

For the novice physicians, we focused on their perspectives regarding learning and improvement during the PBL process. We posed questions like "How do you think you can achieve improvement in the current PBL process?" and "How does the current PBL process help you acquire clinical diagnostic skills?" We also collected their views on PBL, including their concerns and needs in the current practice, with questions like "Do you think the current information presentation and analysis process in PBL meets your learning needs, and what areas can be improved?"

We collected data by recording the interviews with the consent of the interviewees. After transcribing the recordings, we applied thematic analysis [19, 20] to analyze the data. Two researchers from the study team conducted the initial coding, generating labels based on the initial codes. Using an affinity diagram [38], we grouped these labels and organized them into themes with corresponding definitions. Following RQ1 and RQ2, which focus on perceptions and improvement in PBL, we organized the findings into two themes: **Perception of the PBL Process** and **Acquisition of Learning and Improvement**.

5.2 Perception of PBL Organization and Implementation

5.2.1 Teachers' case selection reflects their expectations for students to grasp the relevant knowledge. Senior physicians, serving as teachers in medical education activities, are responsible for selecting cases for PBL. Their case selection typically considers various factors, including alignment with the curriculum, the establishment of a learning gradient, and relevance to clinical scenarios.

PBL is typically implemented in classroom settings to enhance students' understanding and application of knowledge in real-world scenarios. Therefore, cases are often designed around specialized content that aligns with current diagnostic practices within specific disciplines. In designing these cases, senior physicians emphasize the importance of balancing complexity and familiarity. They recognize that medical learning can be categorized into two primary forms: reinforcing existing knowledge and acquiring new knowledge to expand expertise. As one senior physician (S1) noted, "For teaching, we usually start with classic cases to build a strong foundation, and then we mix [in] some complex cases to gradually increase the difficulty and help [develop] their problem-solving skills." Physicians acknowledge that in teaching scenarios, PBL primarily aims to

ensure mastery of fundamental concepts and common conditions, consistent with the philosophy of Occam's Razor in differential diagnosis [84] which prefers the simplest explanation and the most common conditions over rare ones, to reduce unnecessary complexity and avoid overdiagnosis. As another senior physician (S2) remarked, "We mainly design cases around classic examples because we've noticed that if we introduce too many complex [cases], students tend to jump to rare conditions when they see similar symptoms."

To promote the practice of clinical scenarios, physicians often integrate elements of real-world humanistic medicine into case presentations. As S2 noted, "Even though we mainly assess students' knowledge, we still present case descriptions with medical data in a storytelling format. We include elements of humanistic medicine, like communication during clinical decision-making and [ensuring] patient emotional stability, to help train students' overall skills in real clinical settings."

5.2.2 The teacher's role as a guide facilitates student discussion and encourages active exploration. In PBL activities, students take an active role, while teachers focus on organizing and guiding the learning process. Teachers typically introduce relevant research questions related to the cases to foster in-depth thinking and discussion among students. As S1 explained, "We usually start by sharing information about the patient's chief complaint, their medical history, and their current illness, along with their background and how their symptoms began. At this point, students need to make initial judgments based on the symptoms presented. We encourage their thinking and help guide their reasoning by posing questions and providing evidence throughout this process."

Senior physicians point out that PBL cases are typically presented in a hierarchical structure, mirroring the differential diagnosis process. After initial analysis, students receive additional evidence and engage in further analysis tasks, which helps sustain their engagement and focus. As S2 explained, "[...In the later stages,] when students get more information like radiology images and Complete Blood Count (CBC) results, they need to do a thorough analysis and reasoning to reach a diagnosis. I think this step-by-step approach keeps students engaged and interested, and it definitely helps them apply and master their knowledge better than traditional methods."

5.2.3 Students acknowledge that PBL increases their engagement and motivation, and they express a desire for more integrated information organization. In line with teachers' expectations for student knowledge acquisition, students recognized the benefits of PBL in promoting their participation. As N1 noted, "PBL really gets us more

engaged in learning. Just sitting [there] passively listening to lectures can get boring and tiring, but when we're involved in exploratory activities, it boosts our motivation and participation. Our thinking becomes more active and efficient during discussions."

Students also emphasized the need for integrated organization of diagnostic data to support their analysis and discussions, rather than relying on existing CDSS tools that merely present results. As one student N1 explained, "*I get that the whole point of PBL is to help us learn through analysis, like discussions. So, we're more focused on the analysis process [itself], not just on seeing the results. We need to connect and present different data to improve how we retrieve information and do hands-on analysis.*" Besides the diagnostic data, the organization of information during the analysis process is also crucial. Students expressed that properly structuring the analyzed information could significantly enhance their ability to analyze and discuss cases. One student (N2) noted, "*During our discussions, we often need to jot down and organize our thoughts to help [with] the analysis. For example, When we get a patient's symptoms, we list possible causes and then make a differential diagnosis as more clinical data comes [in]. Having everything structured really helps our analysis and makes us more efficient.*"

5.3 The Acquisition of Learning and Improvement

5.3.1 Teachers strive to impart a wide range of skills and improvements to their students, but they are often constrained by limited classroom time and the need to address the broader requirements of the majority. Senior physicians, serving as teachers in medical education, express a strong desire to share as much of their accumulated experience with students as possible. However, they face constraints due to limited classroom time, which necessitates compromises. As one physician (S2) explained, "*There are so many classifications and variations of diseases within the same specialty. For example, arthritis alone has over 100 [types], with osteoarthritis and rheumatoid arthritis being the most common. With the limited [classroom] time, we cannot cover and practice every single type, so we have to be selective about the cases we choose.*"

However, they also acknowledge that these selections are often guided by intuitive heuristic experience rather than systematic empirical evidence. As one physician (S2) noted, "*Currently, we do not have tools to help us systematically screen cases based on [things] like incidence rates and clinical manifestations. We usually just pick what we think are the important and classic clinical cases.*" This selection process may involve cognitive biases that could influence the final judgment and choice.

Additionally, senior physicians recognize that the teaching context often necessitates prioritizing the needs of the majority, which can result in limited individualized instructional support. As S1 noted, "*In classroom teaching, we always have to think about the needs of the majority, like whether most students have mastered the planned knowledge through the cases. Unfortunately, that means we don't give [enough] attention to individual [needs].*" This approach can leave certain student needs unmet, such as personalized supplementary training and opportunities for advanced skill enhancement.

5.3.2 Learning that is oriented toward clinical practice is effective, but the process of constructing clinical reasoning skills through PBL remains implicit and subtle. Both teachers and students acknowledge PBL as an effective approach for integrating learned knowledge with clinical practice. Teachers view it as a valuable way to pass on their accumulated experience to students. As S2 noted, "*Throughout the PBL process, we keep presenting evidence alongside the case introductions, which really helps illustrate the core of clinical diagnosis and treatment. This way, students can learn and improve through hands-on experience, which is much more effective than just telling them our conclusions.*"

They also recognize that the construction of clinical reasoning and logic in differential diagnosis is often implicit and subtle, a challenge mirrored in the PBL learning process. "*We might not explicitly guide students in constructing logical frameworks, like writing on the board or using other methods to make them aware of the reasoning and logical chains at play. This aspect still leaves room for students to explore. (S2)*" Students' feedback is consistent with this understanding: "*PBL activities help [improve] my clinical practice skills, but sometimes the process feels a bit inefficient. During the analysis and discussion, there can be multiple speculations or treatment suggestions for the same condition, and while discussions do refine and clarify things, the logical reasoning behind them isn't always clear. It often stays in our minds and gets expressed verbally. We might use sketches or mind maps to help, but there isn't enough support or guidance for that (N2).*" This highlights the need for more structured approaches in executing the clinical reasoning process.

5.3.3 Student express a desire for a more manageable and accessible PBL learning process. Students identified chances for improvement in their learning experience. Specifically, they expressed a desire for a more participatory and manageable learning process, including the ability to select targeted cases and establish more effective feedback mechanisms.

As noted, teachers typically select cases to help students master classic cases and basic knowledge, but this approach may not meet all students' needs. As N2 observed, "*I understand that the textbook symptoms are important, but I also know that in real clinical practice, diseases can present with different subtypes and manifestations. So, systematic learning should focus on identifying the gaps between my knowledge and practical scenarios and conducting targeted analysis and training.*" In this context, Hickam's Dictum [16], which counters Occam's Razor, is relevant. It emphasizes the importance of thorough analysis and evaluation rather than prematurely seeking simple correlations.

Senior physicians also emphasized the crucial role of reflection and review in the PBL process. S1 noted, "*Reflection and review are really key parts of the learning process. From experience, we know that learning from our mistakes is a powerful way to improve ourselves.*" However, feedback has highlighted some shortcomings in the current mechanism. As noted by N1, "*Right now, teachers typically give final summaries and comments after PBL [activities], but these are often more improvised than systematic. Even though we submit analysis reports, they usually don't follow a standardized format for reviewing the activity.*" This absence of a standardized approach complicates subsequent reviews, making it challenging

to trace specific content and insights, and diminishing the initiative for thorough retrospection and reflection.

6 Phase Two: Deriving Design Space and Solutions for Enhancement

6.1 From Findings to Design Strategies

Based on feedback from physicians, we analyze these findings to identify their key concerns, categorizing them into two themes: **What to Learn** and **How to Learn and Improve**. These themes are further summarized as *Content Setting* and *Analysis Process*.

Content Setting involves the selection of cases and data in PBL activities, focusing on factors such as choosing cases that align with the teaching material, establishing a progression in difficulty from simple to complex, covering key points within a limited timeframe, and balancing individual specific needs with the broader group requirements. *Analysis Process* consists of two aspects of PBL analysis: procedural organization and information organization. Procedural organization addresses the steps and execution of clinical reasoning and differential diagnosis during PBL analysis. Information organization involves the presentation, organization, and recording of information throughout the analysis process, including case details, emerging ideas and hypotheses, and insights and reflections. The codebook for these findings is presented in Table 1.

Building on the summarized differential diagnosis workflow (Figure 1), we focus on the application of clinical reasoning in differential analysis throughout the PBL process. This includes the presentation, analysis, and organization of information and analysis. The process involves clinical information such as the patient's chief complaints, physical examination findings, test results, and other diagnostic evidence, as well as analyzed data such as sketches, mind maps, annotations, and insights generated during the analysis. Proper presentation and organization of this information facilitate analytical discussions and enhance learning outcomes from PBL activities. Based on this analysis, we structured the entire PBL process into three stages: *Data Construction*, *Information Analysis*, and *Record and Review*. We then categorized the codes within these themes according to these stages, which informed the design features for each stage.

6.2 Design and Development of e-MedLearn

In the design of *e-MedLearn*, we adopted the UCD framework [77], conducting multiple rounds of interviews with medical students and teachers in our collaboration team to ensure that the system's design and features meet user needs. Throughout the design process, various ideas were explored and iteratively refined based on user feedback. Initial prototypes integrated the entire case analysis process into a single interface for analysis (Figure 12). However, user feedback revealed that this design created a complex interface, and despite through the navigation structure, it still imposed a significant cognitive load. Following further discussions, we decided to streamline the design by dividing the features into distinct steps, informed by the results of the thematic analysis. By reorganizing the content from Table 1, we developed a five-step presentation process designed to enhance students' learning and improvement throughout the PBL process. As shown in Figure 2, these steps are aligned with the three stages of the PBL process: *Data Construction*,

Information Analysis, and *Record and Review*. This structured approach ensures that students receive comprehensive support as they progress through the entire PBL learning process. A step-by-step demonstration of the system interface is provided in Figure 19.

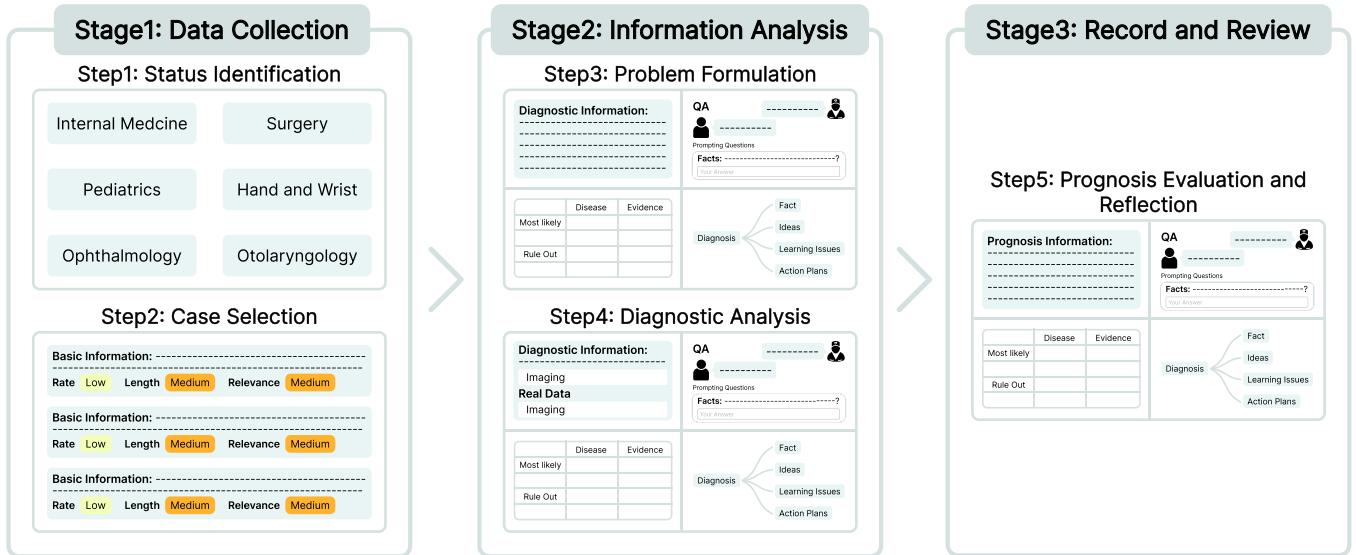
6.2.1 Stage One: Data Construction. In **Stage One**, users prepare for analysis based on the preliminary review of collected data. In *Step 1: Status Identification* (Figure 3), users assess their learning status by determining relevant knowledge areas, such as specialties and types of diseases. We adopt a multi-level classification mechanism that helps users identify their current learning status by navigating from general categories to more specific ones (e.g. Surgery > Orthopedic Surgery > Spine Surgery), which aids in selecting appropriate cases for further study. In *Step 2: Case Selection* (Figure 4), the system supports targeted case selection for PBL activities. Users, guided by their understanding of their learning status, choose relevant cases with system assistance. The interface presents cases categorized under the current classification along with brief descriptions. Cases were initially evaluated by two senior physicians in our team, who evaluated difficulty levels based on three dimensions: incidence rate, length of logical chain, and relevance to classic diseases (the detailed criteria are shown in the Appendix) for reference. Users can also use the search bar to find cases by matching keywords in the descriptions.

6.2.2 Stage Two: Information Analysis. In **Stage Two**, the system provides explicit support for analyzing and organizing information based on the clinical reasoning process of differential diagnosis. Leveraging GPT-3.5-turbo-16k model [22], we have integrated AI-driven QA features to facilitate information analysis. Specifically, we extracted data from each case and input it into the model. Using specifically designed structured prompts, the model generates responses to user inquiries about case details, enabling a more thorough and thoughtful exploration during analysis. The detailed prompt template is present in Table 12. Collaborating with physicians, we developed a set of predefined template questions across various steps (as shown in the Appendix) following the FILA (Facts, Ideas, Learning Issues, Action Plans) framework [34]. These questions are designed to prompt users in case analysis and reasoning.

In *Step 3: Problem Formulation* (Figure 5), users are guided to analyze and organize the preliminary case information, identify key ideas, and formulate subsequent action plans. This step focuses on helping users construct and evaluate logical chains based on initial evidence. The interface consists of (A) Preliminary case information; (B) A QA section with prompting questions; (C) A diagnostic list for reasoning updated with evidence; and (D) A FILA-based mind map template for analysis organization and record. The diagnosis list and mind map content are consistently shared across different steps to ensure accurate information integration and recording throughout the analysis process. We deliberately processed the case information to omit descriptions of "denial" symptoms, such as the denial of fever, within the case study. These unexpressed symptoms, termed as "*negative symptoms*", are identified through the physician's proactive questioning during diagnosis. They are crucial in reflecting the thoroughness of the physician's diagnostic reasoning. Users can obtain this information by asking relevant questions in the QA section, enabling them to practice their questioning and reasoning skills during the diagnostic process. When

Table 1: The codebook for findings derived from discussion sessions.

Theme	Code	Sub-Code	Definition
Content Setting	Case Selection	Related to instructional knowledge	Instructors select relevant cases for PBL based on the instructional content to help students reinforce and enhance their understanding.
		Gradual progression of difficulty/complexity levels	The learning process is gradual, with appropriate expansion upon mastering the foundational content.
		Adaptation to individual learner needs	Learners have specific needs for personalized training and development during the PBL process.
Analysis Process	Procedural Organization	Logical Construction during Diagnostic Reasoning	The diagnostic reasoning process involves structured development of logical chains.
		Evidence-based Reasoning and Differential Diagnosis	The reasoning and differential diagnosis are conducted based on evidence.
Information Organization	Consistent Presentation of Information		The case information needs to be presented consistently for contextual analysis and viewing.
		Recording and review of the Analysis Process	Relevant information from the analysis process, such as notes and annotations need to be organized and recorded to help review.

**Figure 2: The stages and steps of the e-MedLearn prototype system pipeline.**

asked about negative symptoms, the model is prompted to generate responses based on the patient's actual condition. If the case information includes relevant details, the system will display them

accordingly. Otherwise, it will indicate that the information is not relevant to the analysis. In the diagnosis list, users write down their initial thoughts on the case, we design the list into three categories:

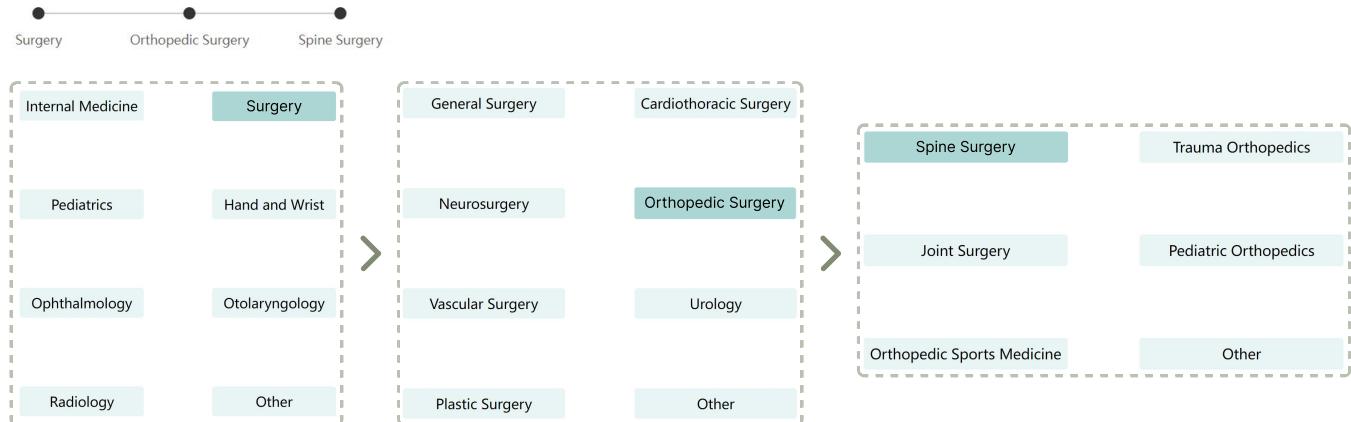


Figure 3: In Step 1: Status Identification, the *e-MedLearn* platform provides a top-down, multi-level selection process to help users determine the content of their learning.

Basic Information: a 49-year-old married female and professional technician, has been experiencing left-sided neck and shoulder pain for over 10 years without any obvious cause. She did not receive standardized treatment during this time. About one month ago, she developed numbness in the ulnar side of her left upper limb, including the back of her hand, palm, and one and a half fingers on the ulnar side. She also reports weakness when holding objects between her fingers.

Incidence Rate: Low Logical Chain Length: Medium Relevance to Classic Diseases: High

Figure 4: In Step 2: Case Selection, relevant cases are presented as cards, with each card displaying difficulty scores for reference. Users can also search cases by content.

Most likely, Need to Rule Out, and Not Likely. This design addresses the importance of evaluating conditions by likelihood, ensuring that critical diseases are not missed, while also considering the dynamic nature of diagnoses, where diseases may shift from improbable to probable. In the QA section, these prompting questions guide users' analysis. The responses to these questions are then integrated under the corresponding factor (Facts, Ideas, Learning Issues, Action Plans) in the mind map, where users can further edit and refine the information. The prompting questions cover: A) **Facts:** “What facts are described in the current case? How can these facts be contextualized in clinical syndrome descriptions?”; B) **Ideas:** “Given that the patient exhibits <symptoms>, what inferences and hypotheses do you propose?”; and C) **Action Plans:** “Based on these <inferences and hypotheses>, what should the next action plans be, such as conducting specific tests or treatments?” Additionally, users can proactively inquire about patient information not mentioned in the dialogue area such as surgical prognosis or allergy history to further support the differential diagnosis process.

In *Step 4: Diagnosis Analysis*, users perform diagnostic analysis using data from additional examinations or tests. This involves updating their initial problem formulation and refining their reasoning to plan subsequent tests or treatments. The system organizes and presents further case scenarios and narratives, through collapsible sections (Figure 6), displaying various types of information, such as physical examination results (text), radiology images (image), and test results (indicator). For imaging data, the system includes annotation features that allow users to draw, label, and add notes

to specific areas for marking. Test results are presented in a clear format, listing “item, result, normal range” for easy viewing and comparison. The QA section offers prompting questions to guide users through analyzing the results and making decisions. Example questions include: A) **Facts:** “What results does the <test> show, and what do they indicate?”; B) **Ideas:** “What inferences can be drawn from the <findings>, and how do they assist in further differentiation and discovery?”; C) **Learning Issues:** “Are the <descriptions> of the patient consistent with your problem representation, and do the <tests> address previous considerations?”, and D) **Action Plans:** “Based on this analysis, what additional tests or treatment plans should be considered?” Following the prompt, users can inquire about the tests and examinations, with the results displayed on cards for each test type, facilitating comparison with subsequent test data presented in the collapsible sections. Users update their thoughts and problem representation in the mind map provided during analysis. This iterative process, indicated by the updated data (Figure 1- ↗), continues until the differential diagnosis is complete. At the end of this process, users consolidate their reasoning based on the evidence to finalize the diagnosis and treatment plan.

6.2.3 Stage Three: Record and Review. During **Stage Three**, users engage in the discussion and evaluation of the prognosis, organizing and recording their analysis and insights for further review and reflection. The interface, similar to Figure 5 but with updated information, is detailed in the Appendix.

A

Diagnostic Information:
a 49-year-old married female and professional technician, has been experiencing left-sided neck and shoulder pain for over 10 years without any obvious cause. She did not receive standardized treatment during this time. About one month ago, she developed numbness in the ulnar side of her left upper limb, including the back of her hand, palm, and one and a half fingers on the ulnar side. She also reports weakness when holding paper objects between her fingers.

She denies any other limb discomfort. She denies blackout. She denies sudden falls. She denies dizziness.

B

The patient reports the following: She denies blackout.

The patient reports the following: She denies sudden falls.

The patient reports the following: She denies dizziness.

Type your message:

Is sudden falls ever happened?

Is there any mention of dizziness from the patient?

C

	Disease	✓ Supporting Evidence	X Excluding Evidence
Most Likely	Cervical Disc Herniation	left-sided neck and shoulder pain, numbness in the ulnar side of	
	Cervical Spondylosis	left-sided neck and shoulder pain, numbness in the ulnar side of	
	Nerve Injury	numbness in the ulnar side of her left upper limb, weakness when	without any obvious cause
Need to Rule Out	Stroke	weakness when holding paper objects between her fingers	numbness in the ulnar side of her left upper limb
Not Likely	Multiple Sclerosis	numbness in the ulnar side of her left upper limb, weakness when	

D

Differential Diagnosis

- Fact: left-sided neck and shoulder pain
- Fact: numbness in the ulnar side of her left upper limb
- Fact: weakness when holding paper objects between...
- Ideas: Could be CDM, CS, or Nerve Injury
- Ideas: Need to rule out Stroke
- Learning Issues: Identify action plans from ideas
- Action Plan: Physical Examination
- Action Plan: MRI examination

Figure 5: Step 3: The Problem Formulation interface includes: (A) Case information card. (B) The prompting QA section. (C) The template mind map to organize and record analyzed information.

Physical Examination >

Imaging >

Tests >

Imaging >

+ Label

C5/6
C5/6 intervertebral disc herniation.

Need to evaluate the nerve root or spinal cord compression.

Physical Examination >

Vital Signs: Temperature 36.5°C, respiratory rate 18 breaths/min, pulse 80 beats/min, blood pressure 120/70 mmHg.

Orthopedic-Related Measurements: Weight 70 kg, height 168 cm, waist circumference not examined, hip circumference not examined.

BMI: 24.8

General Physical Examination: The patient is alert, with stable mental status and smooth breathing. Nutrition is moderate, facial expressions are normal, and development is appropriate. The patient is in a self-supported position, responds fluently, and cooperates well during the examination. The skin is without jaundice, no palmar erythema or spider angiomas are observed. No enlargement of superficial lymph nodes is detected. The head is normal in shape, sclera is not jaundiced, no exophthalmos, pupils are equal and reactive to light. Hearing is normal, no discharge from the external auditory canal, no tenderness in the auricle or mastoid process. The nasal septum is not deviated, no nasal flaring, no tenderness in the sinus areas. Lips are red and moist, the oral cavity has no unusual odors, tongue is midline, tonsils are not enlarged, and the parotid glands are normal. The neck is soft, trachea is...

Tests >

Cardiac Markers

Item	Result	Normal Range
Tropponin T	0.014 ng/mL	<0.01 ng/mL
CK-MB	5.1 ng/mL	0-49 ng/mL
Myoglobin	224.2 ng/mL	25-72 ng/mL
NT-proBNP	3927.0 pg/mL	<300 pg/mL
Procalcitonin	0.07 ng/mL	<0.5 ng/mL

Figure 6: In Step 4: Diagnostic Analysis, the case information is presented in a collapsible section, including imaging (images), physical examination (text), and test results (indicator).

In *Step 5: Prognosis Evaluation and Reflection*, users examine the case's details regarding diagnosis, treatment, and prognosis, including diagnostic results, treatment methods, and recovery outcomes, with consideration of medical humanities. The QA interface provides prompting questions designed to facilitate both learning about the prognosis and guiding users through the review and reflection of their diagnostic process based on the case narrative. For prognosis, the prompting questions include: A) **Facts**: “*What key prognosis details are highlighted in the case?*”; B) **Learning Issues**: “*What aspects, such as patient-family communication or social considerations in treatment decisions, were previously overlooked and needs attention?*” For reviewing and reflecting on the analysis, the prompting questions address: A) **Facts**: “*Are the facts related to diagnosis and treatment consistent with your previous analysis? What discrepancies or overlooked areas are present?*” B) **Ideas**: “*What caused any inconsistencies or oversights? Where did the issues arise in the mind map's analytical pathway?*” C) **Learning Issues**: “*What lessons can be drawn from this case analysis? What areas of improvement were identified, and which areas need further enhancement?*” Users use these prompts to review their analysis process, identify errors or deficiencies through mind maps and notes, and reflect on their diagnostic thought processes. With traceable and accessible case data and analysis notes, users can systematically record insights, reflections, and summaries. Additionally, they can export these notes and summaries for further review and retrospection.

7 Phase Three: Stimulating User Feedback

In this phase, we engaged physicians experienced in PBL to participate in feedback sessions. The experimental design was centered on the system's goal of enhancing students' clinical reasoning, knowledge acquisition, and overall improvement within the PBL process. Instead of replacing traditional PBL education in classroom settings, our approach focuses on the advancement of clinical reasoning within the PBL framework to facilitate targeted learning and improvement.

We conducted a two-stage study to evaluate the effectiveness and usability of *e-MedLearn*, gathering both quantitative and qualitative data. First, a controlled study was carried out to evaluate the impact of integrating clinical reasoning into the PBL process on improving differential diagnosis. Second, a series of testing interviews focused on the PBL process and user feedback were conducted. This approach provided insights from participants, helping us identify the system's strengths, areas for improvement, and its overall effectiveness and usability in advancing PBL within medical education.

7.1 Controlled Study

7.1.1 Participants. We recruited 19 participants (11 males, 8 females, average age: 27.5) through mailing lists and snowball sampling within a local hospital, facilitated by our collaborating team. The participants are not co-authors of this work and did not participate in the discussion sessions during **Phase One**. They are novice physicians with prior PBL experience as part of their continuous medical education. Detailed demographic information about the participants (S1-S9) is provided in Table 2. We also involved two senior physicians (2 males, average age: 42) as teachers in the PBL process to evaluate the participants' performance in the study.

The demographic information of the senior physicians (T1, T2) is provided in Table 3.

Table 2: Demographic information of participants in the controlled study: PBL participation encompasses various levels of engagement, including regular involvement (Routine) and participation based on individual needs (On-demand).

ID	Gender	Age	Title	PBL Role	PBL Participation	Clinical Domain
S1	Male	26	Resident	Student	Routine	Orthopedics
S2	Male	24	Intern	Student	Routine	Orthopedics
S3	Male	26	Resident	Student	Routine	Orthopedics
S4	Male	28	Resident	Student	Routine	Orthopedics
S5	Female	27	Resident	Student	On-demand	Orthopedics
S6	Male	26	Resident	Student	Routine	Orthopedics
S7	Female	27	Resident	Student	Routine	Orthopedics
S8	Male	32	Resident	Student	On-demand	Orthopedics
S9	Female	28	Resident	Student	Routine	Orthopedics
S10	Male	29	Resident	Student	On-demand	Orthopedics
S11	Male	31	Resident	Student	Routine	Orthopedics
S12	Female	25	Intern	Student	On-demand	Orthopedics
S13	Male	27	Resident	Student	Routine	Orthopedics
S14	Female	26	Resident	Student	Routine	Orthopedics
S15	Male	28	Resident	Student	On-demand	Orthopedics
S16	Female	30	Resident	Student	Routine	Orthopedics
S17	Male	30	Resident	Student	On-demand	Orthopedics
S18	Female	29	Resident	Student	Routine	Orthopedics
S19	Female	23	Intern	Student	Routine	Orthopedics

Table 3: Demographic information of senior physicians in the controlled study: PBL participation encompasses various levels of engagement, including regular involvement (Routine) and participation based on individual needs (On-demand).

ID	Gender	Age	Title	PBL Role	PBL Participation	Clinical Domain
T1	Male	43	Associate Chief Physician	Teacher	Routine	Orthopedics
T2	Male	41	Attending	Teacher	On-demand	Orthopedics

7.1.2 Baseline. The baseline system was designed based on several key considerations. First, its application scenario needs to align with individual analysis in PBL, as opposed to traditional PBL teaching scenarios or retrospective learning that involves comprehensive knowledge review. Second, the analysis process should follow a structured PBL approach, rather than a “strawman” format that presents materials without organization.

Based on these criteria, our baseline system was developed by removing analysis features from *e-MedLearn*, such as the diagnosis list, QA section, and mind map. The baseline retains general fixed-format prompts like “*Patient Information, Inferences/Hypotheses, Proposed Actions, and Issues for Further Learning*” to guide the

analysis process. In **Stage 2: Information Analysis** and **Stage 3: Record and Review**, the baseline system provides fixed-format questions to guide users' analysis through the process and reflection. An example interface of the baseline system is shown in Figure 13.

7.1.3 Procedure. Prior to the study, we conducted a 15-minute group meeting to ensure all participants shared a clear understanding of PBL learning and clinical reasoning-based differential diagnosis (as described in Figure 1). We also outlined the experiment's tasks, which involved selecting cases through the PBL process, followed by case analysis and review. Participants first completed a pre-questionnaire to collect demographic information, including their specialty and prior PBL participation. They were then randomly assigned to one of two groups: 10 in the *e-MedLearn* condition and 9 in the baseline condition. Afterward, we provided system tutorials tailored to each group, along with sample case data, to help participants familiarize themselves with the system.

The study is based on the three-stage pipeline of *e-MedLearn*. In **Stage 1: Data Collection**, users begin by selecting orthopedic cases of interest for about 20 minutes. This step primarily evaluates the system's ability to identify and select relevant cases. To evaluate the system's application of clinical reasoning in specific case-based PBL analysis through comparison, we provided all users with the same case at the beginning of **Stage 2: Information Analysis**. Participants used the system for approximately 50 minutes to complete the case analysis and subsequent record and review process in **Stage 3: Record and Review**.

After completing the analysis, participants filled out a post-task questionnaire and conducted a 10-minute interview to share their views of the questions in the questionnaire. They also completed a templated case analysis report (Table 9), documenting their analysis and reasoning process following the Clinical Reasoning Cycle (CRC) framework [55]. In the *e-MedLearn* condition, participants could directly review the analyzed information within the system while preparing their case analysis report. In contrast, participants in the baseline condition relied on fixed-format prompting questions to review their notes. The reports were then evaluated by two senior physicians (T1, T2), who scored them blindly based on the completeness and organization of the clinical reasoning process. Detailed scoring criteria are provided in Table 10. Prior to scoring, a meeting was held to ensure the senior physicians have a consistent understanding of the criteria. The final score for each report was calculated as the average of the two physician's scores. This study was approved by the institutional IRB, and each participant received a \$20 compensation for their time and effort.

7.1.4 Data Collection and Measurement. We collected the questionnaires and case analysis reports from the participants and conducted a quality check on the data. One participant's data in the *e-MedLearn* condition was excluded due to an analysis duration of less than 30 minutes. In the end, we obtained 18 valid data sets, with 9 from the *e-MedLearn* condition and 9 from the baseline condition.

For the post-task questionnaire, we used a 7-point Likert scale (1: Strongly disagree, 7: Strongly agree) to collect participants' feedback on the analysis process and their perceptions of their own performance. First, under **Usability**, we adapted questions from the System Usability Scale (SUS) [21] to assess (1) *Ease of use*, (2) *Helpfulness for clinical reasoning*, (3) *Satisfaction*, (4) *Distraction*,

and (5) *Recommendation and future use*. Second, to assess **Workload**, we based questions on the NASA-TLX survey [51], focusing on (1) *Mental demand*, (2) *Physical demand*, (3) *Temporal demand*, (4) *Performance perception*, (5) *Effort*, and (6) *Frustration*. Third, we designed questions (Table 4) to evaluate the **Effectiveness** of the system, including (1) *Targeted case selection*, (2) *Thoughtful information identification*, (3) *Structured organization*, (4) *Evidence-based reasoning*, (5) *Reasoning-supported action plan*, (6) *Prognosis consideration*, (7) *Clinical reasoning reflection*. Due to the small sample size, we applied the nonparametric Mann-Whitney U test [71] to compare results between conditions, as it does not assume specific data distribution. For the case analysis reports, we performed Spearman's Rank Correlation [101] to assess consistency between the two raters' scores. It is important to note that the case analysis report primarily evaluates the user's ability to organize and reflect on the case analysis process based on clinical reasoning in the PBL context, rather than immediate knowledge acquisition.

7.2 Testing Interviews

We subsequently conducted testing interviews to collect feedback on participants' practice. To avoid any influence from the prior study, we re-recruited 10 novice physicians (7 males, 3 females, average age: 26.6) through snowball mailing lists. We also involved 3 senior physicians (2 males, 1 female, average age: 46.7), where two physicians scored the case analysis reports in the controlled study but had no direct exposure to the system. All of the participants were not co-authors of this work and did not participate in **Phase One**. Participant details are provided in Table 5.

7.2.1 Procedure. The process of the testing interviews was structured as follows: first, reviewing current PBL practices and participants' perceptions; second, introducing the system's design considerations and features aimed at enhancing diagnostic reasoning during the PBL process; and third, collecting feedback on the fully implemented system after its use in the PBL process. This study was approved by the institutional IRB, and each participant received \$20 in compensation for their time and participation.

We began by exploring participants' current practices and perceptions of PBL in educational settings, with a focus on the organization and knowledge acquisition processes. In PBL activities, novice physicians typically engage as students, using the sessions for learning, while senior physicians serve as instructors, guiding and facilitating the process. For novice physicians, we focused on their experience as students in the PBL process, as they represent the primary target users of *e-MedLearn*. For senior physicians, we examined their roles as instructors, particularly their perspectives on PBL teaching and their observation of students' progress throughout the process.

Next, we introduce the system's design considerations, highlighting how it supports users in applying knowledge to real-world scenarios and enhances their PBL skills, particularly in clinical reasoning and differential diagnosis. At this stage, we do not present the system itself.

Following the introduction, the system is presented to the users, and its features are explained. Users are then given 5 minutes to familiarize themselves with the system before proceeding with a full PBL analysis, which includes case selection, analysis, and reflection,

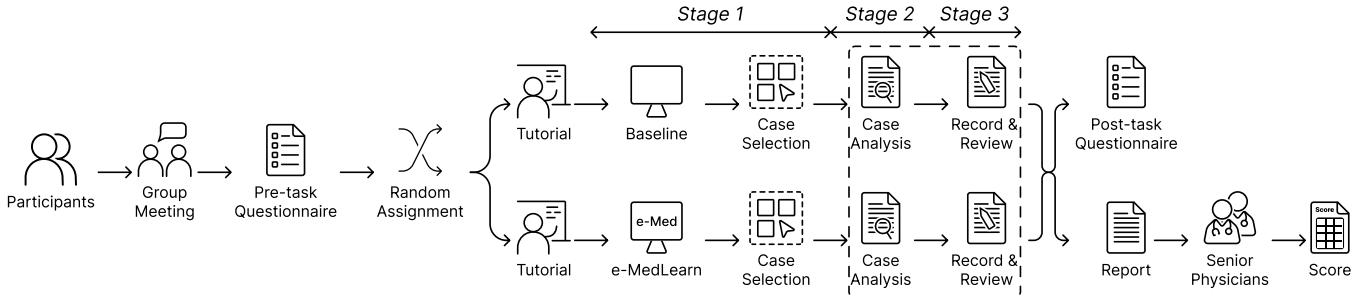


Figure 7: The pipeline of the controlled study. Participants were randomly assigned to two conditions: the baseline and *e-MedLearn*. In Stage 1: Data Collection, participants browsed and selected orthopedic cases of interest using the system. In Stage 2: Information Analysis, all participants analyzed the same case. In Stage 3: Record and Review, participants documented and reviewed their analyses. Following these stages, participants completed a post-task questionnaire and submitted a case analysis report, which was then evaluated by senior physicians.

Table 4: Questions about the Effectiveness in the post-task questionnaire.

(1) Targeted case selection	I can select cases suitable for in-depth analysis based on my learning objectives.
(2) Thoughtful information identification	I can identify key diagnostic information through thoughtful exploration.
(3) Structured organization	I can effectively organize and integrate the collected information for analysis.
(4) Evidence-based reasoning	My reasoning process is clear and logical based on evidence.
(5) Reasoning-supported action plan	My action plans are supported by reasoning with a clear rationale.
(6) Prognosis consideration	I can recognize the prognosis and consider relevant follow-up actions.
(7) Clinical reasoning reflection	I can reflect on my reasoning process and identify insights into its strengths and weaknesses.

taking approximately 50 minutes. Upon completing the analysis, interviews are conducted to collect feedback on the system.

Due to the senior physicians' experience and roles, they are not required to fully engage in the entire PBL analysis process. Instead, they focus on understanding the system's high-level design considerations and provide feedback based on their interactions and perspectives during the subsequent interviews.

7.2.2 Data Collection and Analysis. Each interview session was conducted via video conference software. With the participants' knowledge and consent, we audio and video recorded the entire testing interview process. The qualitative data from these recordings were analyzed using thematic analysis [19, 20] and affinity diagram[38].

First, the audio recordings were transcribed into text using Otter.ai [80], an automated speech transcription tool. Subsequently, the first author manually verified the transcripts to ensure accuracy. The transcripts were then organized preliminarily. Two team members conducted an inductive analysis to identify various themes, sub-themes, and relevant codes.

The analysis includes reflections from both students and instructors on the current PBL approach, as well as their thoughts and feedback on the system's design and features. These insights highlight opportunities for *e-MedLearn* to enhance clinical reasoning-based differential diagnosis within the PBL process. Additionally, in-task questionnaires were collected to evaluate the feature effectiveness from participants with the role of students role in the PBL process,

Table 5: Demographic information of participants in the testing interviews: PBL participation encompasses various levels of engagement, including regular involvement (Routine), participation based on individual needs (On-demand), and temporary involvement for specific purposes (Ad-hoc).

ID	Gender	Age	Title	PBL Role	PBL Participation	Clinical Domain
P1	Male	26	Resident	Student	Routine	Orthopedics
P2	Female	25	Intern	Student	On-demand	Neurosurgery
P3	Male	28	Resident	Student	Routine	Neurosurgery
P4	Female	25	Intern	Student	On-demand	Neurosurgery
P5	Male	27	Resident	Student	Routine	Neurosurgery
P6	Male	26	Resident	Student	Routine	Orthopedics
P7	Female	28	Resident	Student	On-demand	Neurosurgery
P8	Male	26	Resident	Student	Routine	Orthopedics
P9	Female	28	Resident	Student	Routine	Neurosurgery
P10	Male	27	Resident	Student	On-demand	Neurosurgery
P11	Male	43	Associate Chief Physician	Teacher	Routine	Orthopedics
P12	Female	38	Attending	Teacher	Ad-hoc	Neurosurgery
P13	Male	41	Attending	Teacher	On-demand	Orthopedics

as the **PBL Role** shown in Table 5. The questions in the in-task questionnaires are shown in the Appendix.

8 Results

8.1 Results of Controlled Study

We present the results of the controlled study, including feedback from the post-task questionnaire and scores from the case analysis report. *Mean* and *SD* represent the average value and standard deviation for each question within the respective groups. *M* indicates the Mann-Whitney U value, and *p* denotes the level of statistical significance.

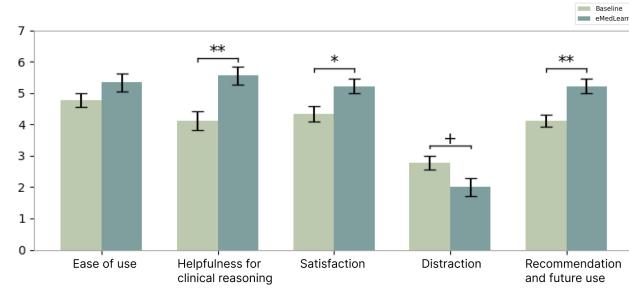


Figure 8: The results of usability perceived by users (*: $p < 0.05$; **: $p < 0.01$; *: $p < 0.001$).**

Usability: The usability results are shown in Figure 8. Compared to the baseline system, participants generally provided more positive feedback regarding the usability of *e-MedLearn*. Specifically, users rated the *Ease of use* of *e-MedLearn* ($Mean = 5.33$, $SD = 0.86$) higher compared to the baseline ($Mean = 4.78$, $SD = 0.67$). “*The design of steps in the e-MedLearn system is reasonable and aligns with our diagnostic analysis process, so it didn’t add any complexity that would make it feel hard to use.* (S3)” Furthermore, participants in *e-MedLearn* condition scored significantly higher on *Helpfulness for Clinical Reasoning* ($M = 11$, $p < 0.01$) and *Recommendation for Future Use* ($M = 10$, $p < 0.01$). For the scores of *Distraction*, *e-MedLearn* ($Mean = 2$, $SD = 0.87$) also received lower scores compared to the baseline ($Mean = 2.78$, $SD = 0.67$), while this difference was not statistically significant. Additionally, *e-MedLearn* demonstrated significantly higher scores for *Satisfaction* ($Mean = 5.22$, $SD = 0.67$) compared to the baseline system ($Mean = 4.33$, $SD = 0.71$) ($M = 16$, $p < 0.05$), indicating a higher level of overall usability satisfaction. Users acknowledged that the system for helping “*make their analysis more concrete, with each reasoning step based on the current evidence and grounded in a clear rationale.* (S7)”

Workload: The workload results are shown in Figure 9. Overall, participants in the *e-MedLearn* condition reported significantly lower workload perceptions across several dimensions, including *Mental Demand* ($M = 70$, $p < 0.01$), *Temporal Demand* ($M = 65$, $p < 0.05$), and *Effort* ($M = 68$, $p < 0.05$). Participant noted that “*I provided a low mental demand score because I think the system made it easier for me to form evidence-based hypotheses with corresponding rationales, thereby supporting reasoning process.* (S8)” Participants in *e-MedLearn* condition also reported lower *Physical Demand* (baseline: $Mean = 5.56$, $SD = 0.88$; *e-MedLearn*: $Mean = 4.11$, $SD = 0.93$), with no statistically significant difference. Additionally, *e-MedLearn* participants reported significantly lower *Frustration* scores ($Mean = 3.78$, $SD = 0.67$) compared to the baseline condition ($Mean = 4.89$,

$SD = 0.60$) ($M = 71$, $p < 0.01$). “*The fixed questions in the <baseline> system didn’t really help with my exploratory thinking, so I felt unsure and not very satisfied with the analysis process.* (S4)”

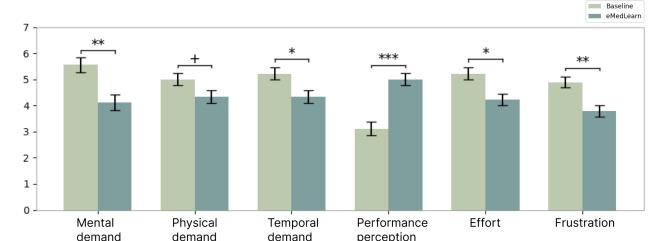


Figure 9: The results of workload perceived by users (*: $p < 0.05$; **: $p < 0.01$; *: $p < 0.001$).**

Effectiveness: Figure 10 presents the results on effectiveness. Overall, the *e-MedLearn* condition achieved significantly higher scores. In **Stage 1: Data Collection**, *e-MedLearn* ($Mean = 4.89$, $SD = 0.60$) outperformed the baseline ($Mean = 3.67$, $SD = 0.71$) in *Targeted case selection* with a significant difference. In **Stage 2: Information Analysis**, *e-MedLearn* demonstrated significantly better performance in *Thoughtful information identification* ($M = 19$, $p < 0.05$), *Structured organization* ($M = 3$, $p < 0.001$), *Evidence-based reasoning* ($M = 13$, $p < 0.05$), and *Reasoning-based action plans* ($M = 18.5$, $p < 0.05$). “*Problem-based analysis helps me think critically, like understanding why I make certain decisions and what evidence backs them up. I believe a good diagnosis really needs this kind of thinking to figure out both the disease and the action plan.* (S9)” In the **Stage 3: Record and Review**, *e-MedLearn* ($Mean = 5.22$, $SD = 0.67$) showed improved facilitation of *Prognosis consideration* compared to the baseline ($Mean = 4.67$, $SD = 0.87$), with feedback indicated that “*the prompting questions helped me think through relevant issues, making sure I didn’t miss any important prognostic details and keeping the diagnostic process thorough.*”. Additionally, *e-MedLearn* ($Mean = 5.11$, $SD = 0.60$) significantly outperformed the baseline ($Mean = 4.0$, $SD = 0.70$) in promoting *Clinical reasoning reflection* ($M = 10.5$, $p < 0.01$). As the participants indicated, “*After the analysis, I mainly reflect on the reasoning process to see where things went wrong. The diagnosis list and mind map help me figure this out easily.* (S3)”

We compared the case analysis reports submitted by users across eight dimensions, using the criteria in Table 10. For each dimension, we performed a Mann-Whitney U test to compare the scores between the baseline and *e-MedLearn* conditions. As shown in Table 6, the report scores revealed significant differences in different dimensions of clinical reasoning. These results highlight the effectiveness of *e-MedLearn* in facilitating clinical reasoning during PBL case analysis.

8.2 Results of Testing Interviews

For the in-task questionnaire in the **testing interviews**, we used a seven-point Likert scale (1: Strongly disagree, 7: Strongly agree) to assess the perceived usefulness of system features. As shown in Figure 11, most features in *e-MedLearn* were recognized as useful, though the prompting questions received relatively lower scores ($Mean = 5.08$, $SD = 0.64$).

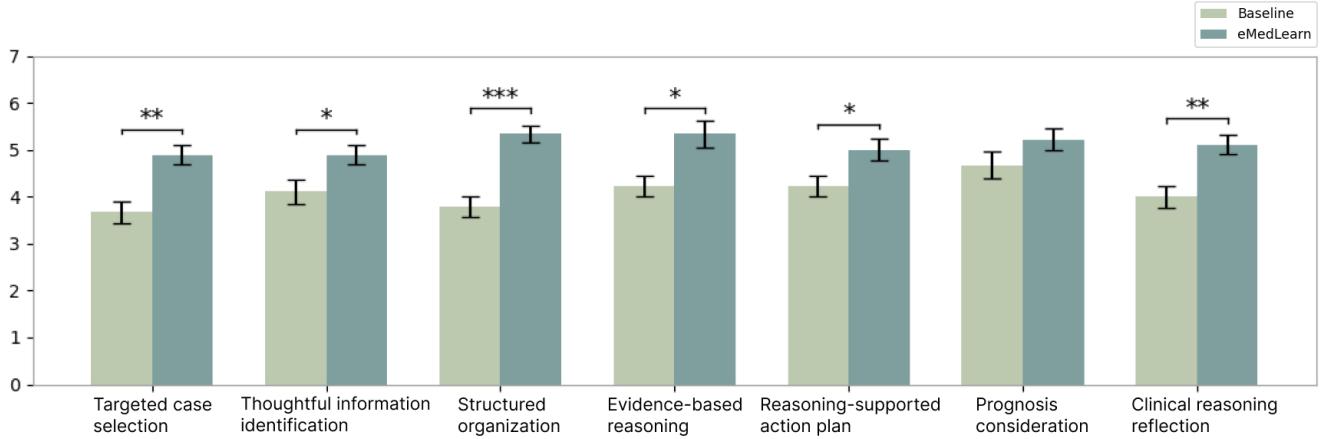


Figure 10: The results of effectiveness perceived by users (*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$).

Table 6: The average scores between Baseline and e-MedLearn condition in different dimensions of case analysis report criteria. "U" indicates Mann-Whitney U test results, and "p" indicates p-values for statistical significance (*: $p < 0.05$; **: $p < 0.01$; *: $p < 0.001$).**

	Baseline	e-MedLearn	U	p
Consider the Patient Situation	3.11	5.33	2.00	***
Collect Cues and Information	4.00	5.67	7.50	**
Process Information	3.22	4.78	7.50	**
Identify Problems/Issues	3.22	4.44	17.00	*
Establish Goals	3.78	5.78	8.50	**
Take Action	3.78	5.00	15.00	*
Evaluate Outcomes	3.67	5.22	9.50	**
Reflect on the Process and New Learning	3.00	5.22	4.00	**

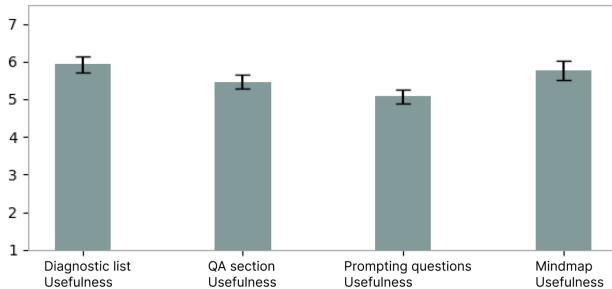


Figure 11: The usefulness of different features in e-MedLearn.

For the qualitative data in the **testing interviews**, we employed a thematic analysis approach [19, 20], categorizing the findings into three key themes: motivation and considerations in transitioning from knowledge to practice, the application of evidence-based reasoning in analysis, and the benefits observed in development and improvement. A detailed discussion of each theme is presented below.

Theme 1: “So gaps exist between problem-solving and patient diagnosis.” The motivation and considerations from knowledge to practice

8.2.1 *Awaken self-directed identification and awareness.* Regarding the role of e-MedLearn in bridging knowledge and practice, participants recognized that the system offers a solution for selecting PBL cases tailored to their learning status. They highlighted that this approach awakens their awareness during self-directed learning, including the identification of both knowledge mastery and learning intentions as they transition from theory to practice through PBL.

“Sometimes we try to simulate PBL for practice based on our own needs, but we find it challenging to identify the right cases without proper guidance. I think one of the key things about the system is that it helps us identify our learning progress and gives us the support to make informed case choices based on our specific interests and focus areas.” (P5)

The feedback also reflected the system’s effectiveness in clarifying learning intentions during the case selection process. Participants noted that previous challenges in selecting cases during PBL often stemmed from an unclear understanding of learning objectives. Features such as case display by incidence rate, logical chain, and relevance to classic diseases have been instrumental in enhancing users’ awareness, thereby providing a solid foundation for a self-directed PBL process.

“When I saw this <difficulty reference for each case>, it struck me as exactly what I need for real clinical practice. In the early stages of learning, we concentrate on classic disease cases to build a solid foundation. As we advance, we shift to more complex cases with longer logical chains to sharpen our diagnostic reasoning skills with that knowledge.” (P9)

Theme 2: “My analysis process fails due to unsubstantiated reasoning.” The application of evidence-based reasoning in analysis

8.2.2 The acquisition and extension of evidence. Participants emphasized that the key to applying learned knowledge in real-world diagnosis lies in reasoning ability. Rather than merely analyzing all available evidence, this reasoning process involves making initial judgments based on presented evidence and then uncovering additional information. This approach mirrors real clinical scenarios, where, as one participant noted, “patients typically start by describing their symptoms, rather than providing all relevant information upfront (P11).” It is through the interaction between physician and patient that further information is gathered, which is then used for exclusion and differential diagnosis. The ability to effectively gather this information is a crucial aspect of diagnostic training. Participants acknowledged that the explicit logical reasoning features of *e-MedLearn* assist them significantly in this process.

“I understand that in real clinical practice, unlike solving problems where you have all the information upfront, physicians gather more information through conversation and questioning with the patient, and make judgments based on this information. The system, by providing a question-and-answer feature that reveals information gradually as inquiries are made, rather than displaying all information at once, simulates this process, making my thinking more aligned with clinical diagnostic reasoning. (P7)”

The physicians further elaborated that the practice of expanding and analyzing based on initial evidence is integral to developing students’ reasoning and analytical skills. Senior physicians, while conducting patient interviews, simultaneously build logical connections in their minds—a skill refined through experience and repeated practice. Notably, when they document ‘negative symptoms’—those that the patient did not exhibit in the medical record, it demonstrates what one participant referred to as a “medical way of thinking” (P4).

“These denied symptoms mean that the patient was asked about them, but they said no, rather than the doctors overlooking them. The system simulates this process much like how a teacher guides responses in PBL teaching, helping students develop their medical thinking and reasoning skills. (P13)”

Additionally, the senior physician pointed out that students may tend to swing to the opposite extreme, asking about numerous symptoms one by one without considering the reasoning behind their questions. A clinical reasoning exemplar should strike a balance by “*avoiding redundancy while ensuring all essential aspects are covered*” in the inquiry process (P12). Moving forward, we aim to analyze users’ questioning patterns to guide them toward more systematic and logical approaches to patient evaluation.

8.2.3 Evidence prompts clinical reasoning and differential diagnosis. Almost all participants emphasized the importance of evidence-based reasoning in diagnosis. They noted that the diagnostic process entails forming initial hypotheses based on available evidence and subsequently updating these hypotheses as new information emerges, ultimately leading to a final diagnosis. Feedback indicates that these considerations are highly valued in enhancing the PBL process.

“In PBL activities, I often found that my analysis process failed due to unsubstantiated reasoning, as I relied too much on intuition and lacked systematic thinking. With this system, the structured analytical framework it provides is both clear and practical, helping me become more systematic and logical in my approach. (P9)”

As noted by several participants, systematically engaging users in explicit, evidence-based reasoning enhances their medical literacy on both cognitive and methodological levels. One participant remarked, “*When you consistently record and organize evidence and thoughts in a structured way, such as following a framework like <FILA>, your analysis process becomes more orderly and coherent*” (P8). Additionally, participants observed that this evidence-based approach aids in adopting the patient’s perspective, thereby improving the efficiency of information gathering during patient interactions and enhancing the clarity of explanations throughout the communication process.

“Patients sometimes provide irrelevant information during conversations with physicians. The evidence-based understanding during diagnosis helps us guide the patient to focus on their condition, facilitating our diagnostic reasoning. It also helps us reasonably explain our requests and decisions when the patient is confused or in doubt. (P5)”

Theme 3: “What reflections should I make after practice?” The benefits gained in development and improvement

8.2.4 Post-diagnosis review and reflection. In the interviews, participants noted that *e-MedLearn* supports learning about prognosis following diagnosis, an essential aspect of a physician’s responsibilities. They highlighted that the system’s features enhance the comprehensiveness of diagnostic analysis and help address the “*potential neglect of prognosis considerations due to an excessive focus on correctness and reasoning once the diagnosis is revealed*” (P11).

“After the analysis of case data <in PBL process>, I spontaneously became curious about how my analysis matched the actual results and focused on the outcome. Looking back, I realize that I often neglected the subsequent <post-diagnosis> content. The prompting questions provided by the system help remind me to continue paying attention to the case after the diagnosis and to record key points as part of my ongoing learning.” (P2)

Regarding the benefits derived from the analysis, both students and teachers underscored the importance of review and reflection following the learning process, highlighting that these aspects are as crucial as the analysis itself. Feedback revealed that less experienced students may experience confusion after completing the analysis and struggle with the question of, “*What reflections should I make after practice?*” (P6).

“After finishing the diagnostic analysis, the answers usually get revealed, but the feedback from teachers is often just a general summary of the whole process. It rarely zeroes in on individual analysis and errors. So, I used to struggle with finding ways to improve through reflection, especially when I wasn’t very skilled at the analysis process.” (P2)

Participants observed that *e-MedLearn* effectively bridges the gap between practice and reflection in the PBL process. They noted that the system offers a structured workflow that guides users from analysis to documentation, allowing them to review their reasoning based on the thought process rather than just the outcome. This approach facilitates high-level, structured reflection and fosters continuous improvement beyond the specifics of individual cases.

"In e-MedLearn, we can see the previous diagnostic list that records all hypotheses and corresponding thoughts, as well as the mind map that organizes and records information. This allows me to review the entire traceable analysis from the perspective of my thought process. Correcting thinking patterns rather than merely focusing on the outcome is undoubtedly more valuable." (P1)

From a long-term development perspective, the interviews indicate reflection on methodology and thought processes is more beneficial for enhancing learners' abilities. With the previous perception of improvement in current PBL activities, participants consider the structured organization through *e-MedLearn* could better facilitate valuable reflections for improvement and also is "*consistent with the intention of PBL activities that transition knowledge into clinical practice*" (P9).

9 DISCUSSION

To promote the application of theoretical knowledge in practical settings, we explored the integration of clinical reasoning and differential diagnosis methodologies into PBL. Our work focused on advancing medical students' skills throughout the PBL process, with a strong emphasis on evidence-based discovery, exploration, and reasoning. Based on these findings, we identified design implications for enhancing the PBL experience, particularly in the application of diagnostic methodologies, while considering both educational and clinical scenarios. Additionally, we explored the potential for fostering continuous clinical development and comprehensive diagnostic decision-making within medical education.

9.1 Implication for Design in PBL Practice

9.1.1 Focusing on Clinical Diagnostic Methodologies. Our study's findings suggest that *e-MedLearn* effectively bridges the gap between current PBL practices and methodology-driven clinical reasoning and differential diagnosis. These insights provide valuable guidance for HCI and AI-based design considerations in the PBL process, facilitating the transition from theoretical knowledge to practical application.

Promoting evidence-based divergent thinking rather than uncritical usage. Participants acknowledged that *e-MedLearn* encourages users to actively explore and gather evidence by avoiding the presentation of all case data at once. This approach fosters the development of students' medical reasoning through directed exploration. However, some users reported experiencing a "cold start" during the data exploration process, where the expected divergent thinking based on initial evidence did not unfold as intended.

A key concern raised by participants is the system's step-by-step inquiry of evidence exploration limits in providing sufficient methodological guidance for informed exploration. While general prompts such as "*What symptoms should be inquired about*

next?" may encourage convergent analysis [112], they are sometimes viewed as inadequate for promoting creative divergent thinking [7, 92]. A potential solution is to introduce prompts more closely tied to the data, such as "*Based on the symptom, you may need to inquire about <related aspects>?*" However, this should be approached carefully to avoid explicit instructions that could reveal the content, which may increase learners' cognitive load in an information-rich context [40] and hinder the directed analysis.

Emphasizing problem-driven reasoning. Interview feedback emphasized the critical role of problem-based reasoning during the analysis process. Senior physicians (P11, P13) noted that "*the essence of case analysis in PBL is the continuous process of raising questions and solving problems based on those questions*", which aligns with the process of reasoning and differential diagnosis in clinical practice [37]. Although framework-based prompting questions have proven effective in advancing problem-motivated analysis through diagnostic methodologies, some participants expressed a desire for more flexible and targeted prompting questions to further enhance the analysis process. Unlike lecture-based learning [5], which focuses on content delivery, and case-based learning [114], which involves retrospective case reviews, PBL is characterized by its emphasis on problem-driven reasoning that bridges knowledge and practice. This necessitates careful attention to system design, not only to support problem analysis but also to foster students' skills in recognizing and formulating problems, a key aspect echoed in tools designed for academic analysis [88, 105, 124]. In this context, AI's role extends beyond facilitating evidence exploration to guiding reasoning. A potential solution is leveraging large language models (LLMs) to suggest possible areas of exploration based on user input, thereby enhancing the problem-driven reasoning process.

9.1.2 Bridging Medical Education with Clinical Practice. The design considerations of *e-MedLearn* demonstrate a theoretical approach to promoting the application of PBL in clinical diagnostics through clinical reasoning. However, in addition to this theoretical methodology, it is essential to account for various real-world factors present in actual clinical environments. Feedback obtained from the study highlights the importance of integrating educational considerations with practical diagnostic processes.

Integrating design consideration with clinical workflow. While participants recognized the system's support in facilitating the PBL analysis process, they also highlighted the gap between theoretical design considerations and practical clinical scenarios. A common observation was the need to adapt to real-world clinical environments, where factors such as patient interaction, emotional stress, and distractions complicate communication and shared decision-making (SDM) [49, 50, 66, 102]. Physicians emphasized that in clinical practice, these factors play a critical role. While classroom-based PBL methods offer foundational learning, formats such as Virtual Patients (VP) and Standardized Patients (SP) [45, 69] are crucial for preparing students to navigate real-world complexities. However, placing too much emphasis on practicality could limit the understanding of underlying methodologies. Therefore, integrating design considerations with clinical workflows involves striking a balance between practical experience and the mastery of theoretical approaches. Future efforts should focus on how to seamlessly embed methodological training into clinical environments.

Advanced technologies like AI and LLM, combined with VR/MR simulations, could play a significant role in enhancing immersive learning and practice, offering varied patient role simulations and enriching real-world preparation.

Leveraging evidence as the explanation in diagnostic reasoning. Participants in the study highlighted the importance of diagnostic updates informed by evidence. From a logical reasoning standpoint, this reflects the differential diagnosis process in evidence-based medicine [1], where reasoning and analysis are rooted in evidence. In this context, evidence can be seen as a form of explanation, helping physicians construct reasoning that supports their cognitive decisions. This approach, focused on learning to leverage self-reinforcing reasoning as a form of explanation, aligns with clinical practice as metacognition [36], which develops through experience and enhances the efficiency of outpatient processes. In PBL, utilizing the explanatory nature of evidence helps to connect educational theory with practical diagnostic applications.

A key element is the progress achieved through immediate feedback on evaluations of these explanations during the analysis process. While research has examined tools in medical learning [48, 81], medical education still largely depends on the traditional “*mentor-apprentice*” model to support students’ professional growth [13, 91]. A future direction for *e-MedLearn* is to integrate feedback into the analysis process, promoting the use of evidence as an explanation to enhance both student motivation and clinical practice.

9.2 Building Sustainable Medical Learning Ecosystem

The design of *e-MedLearn* has proven effective in supporting individual learners in enhancing their PBL practice. Feedback also highlighted the need for shared discussion and participation with the system to facilitate collaborative learning. This aligns with the awareness that the essence of learning, both in content and process, lies in the acquisition of accumulated experience [62]. These insights suggest opportunities to establish and sustain a medical education platform ecosystem that fosters continuous learner development and improvement. A key consideration is integrating past reflections into individual analyses and decision-making. Potential solutions include incorporating meta-information of cases, such as learner tags and case ratings, on the online learning platform to provide references and encourage active engagement.

Another key consideration is establishing a sustainable educational development process. Currently, medical cases are primarily compiled through casebook collections [8, 99], which poses challenges when manually integrating these cases into the system workflow. Enhancing sustainability could be achieved by enabling instructors to upload cases directly to the platform, alongside automated crawling or data extraction from provided sources. This would streamline case integration and significantly contribute to the ongoing training and development of medical students.

9.3 Incorporating Humanistic Factors Diagnosis for Comprehensive Diagnosis Consideration

The design considerations for *e-MedLearn* primarily emphasize practicing differential diagnosis within the PBL framework. While accurate diagnosis is crucial, medical practice also encompasses

clinical decision-making, including treatment and recovery planning [83, 109]. Additionally, in real-world medical practice, especially for chronic diseases, patient management should not solely depend on life expectancy. It is essential to consider the patient’s overall health, psychological state, and social context. Comprehensive medical care should therefore integrate both medical science and humanistic approaches. Training and expectations for physicians should align with the biopsychosocial model [17, 43], which addresses biological, psychological, and social factors. Consequently, the PBL process should evolve to include not just clinical reasoning and differential diagnosis, but also value-based treatment planning and decision-making. In response to practice-driven needs, we aim to incorporate features into *e-MedLearn* that encourage users to consider patient-centered factors during diagnosis and treatment. These enhancements will improve learners’ ability to apply their knowledge in a holistic medical practice.

9.4 Limitation and Future Work

Our work has several limitations. First, In **Phase One**, our discussion sessions involved four physicians from two specialties. While the primary goal was to explore the current practices and challenges in PBL activities from the perspectives of both teachers and students, we acknowledge that the process of identifying issues was limited by the small sample size and the specialty-specific constraints. Future research with a broader scope and sustained efforts could offer more comprehensive insights into potential system improvements in the PBL process. Second, the participants in this study were relatively few and primarily from the orthopedic and neurosurgical fields. Although limiting the study to specific medical specialties reduces variability and allows for a focused evaluation of the system, thereby establishing a foundational understanding of how *e-MedLearn* supports clinical reasoning in specialized medical contexts, the study does not account for potential differences between specialties that could influence the analysis and results. Future research should expand to include a broader range of clinical scenarios across a larger sample size and various medical specialties to provide a more comprehensive understanding of *e-MedLearn*’s generalizability and effectiveness. Third, as discussed in subsection 8.2, learning and improvement are gradual processes, and the development of methodological skills requires sustained practice for effective assessment. We used the system’s design concepts as a probe to promote targeted PBL practice among participants, offering further insights for design refinement. A longitudinal study is also planned to gather data on learners’ long-term usage and feedback. Lastly, our current data is limited to hospitals affiliated with the collaborating physicians. In real-world scenarios, challenges such as cross-institutional or cross-platform data sharing may arise, along with ethical concerns about the use of unreviewed data. While addressing these issues is beyond the scope of our system’s current design, we recognize their significance and are considering advocating for the establishment of institutions to supervise and review data-sharing practices, with the aim of fostering a sustainable health data ecosystem.

10 Conclusion

In this study, we explored the integration of clinical reasoning-based differential diagnosis to enhance targeted learning and improvement in PBL. The research followed a three-phase approach. In **Phase One**, we conducted interviews with physicians involved in PBL, gathering insights into their perceptions of the current PBL process, and areas for improvement. In **Phase Two**, we collaborated with experts to engage in design brainstorming. Through multiple design iterations, we identified key features that strengthen the application of clinical reasoning in PBL and developed the learner-centered *e-MedLearn* system to support targeted continuous improvement. In **Phase Three**, we conducted a controlled study and testing interviews to collect the system's feedback on clinical diagnostic reasoning in PBL. The results indicated that *e-MedLearn* effectively bridges the gap between knowledge acquisition and clinical practice, helping learners transition toward clinical diagnosis. Overall, *e-MedLearn* provides valuable insights into the application of clinical diagnostic reasoning in PBL, with opportunities for further improvement, such as the development of a collaborative online platform and the integration of clinical scenario factors and medical humanities into patient care and treatment.

Acknowledgments

We thank anonymous reviewers for their valuable feedback. This work is supported by grants from the National Natural Science Foundation of China (No. 62372298), Shanghai Engineering Research Center of Intelligent Vision and Imaging, Shanghai Frontiers Science Center of Human-centered Artificial Intelligence (ShangHAI), and MoE Key Laboratory of Intelligent Perception and Human-Machine Collaboration (KLIP-HuMaCo).

References

- [1] Anthony K Akobeng. 2005. Principles of evidence based medicine. *Archives of disease in childhood* 90, 8 (2005), 837–840.
- [2] Nazim Ali, Russell Crawford, and Margaux Horn. 2019. Critical thinking in PBL: Development of a bespoke tool for critical thinking. In *HEAD'19. 5th International Conference on Higher Education Advances*. Editorial Universitat Politècnica de València, 513–520.
- [3] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, et al. 2019. Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–13.
- [4] Anshu Ankolekar, Ben GL Vanneste, Esther Bloemen-van Gorp, Joep G van Roermund, Evert J van Limbergen, Kees van de Beek, Tom Marcelissen, Victor Zambon, Matthias Oelke, Andre Dekker, et al. 2019. Development and validation of a patient decision aid for prostate Cancer therapy: from paternalistic towards participative shared decision making. *BMC medical informatics and decision making* 19 (2019), 1–11.
- [5] Wolfram Antepohl and Stefan Herzig. 1999. Problem-based learning versus lecture-based learning in a course of basic pharmacology: a controlled, randomized study. *Medical education* 33, 2 (1999), 106–113.
- [6] Amid Ayobi, Jacob Hughes, Christopher J Duckworth, Jakub J Dylag, Sam James, Paul Marshall, Matthew Guy, Anita Kumaran, Adriane Chapman, Michael Boniface, et al. 2023. Computational notebooks as co-design tools: engaging young adults living with diabetes, family carers, and clinicians with machine learning models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–20.
- [7] John Baer. 2014. *Creativity and divergent thinking: A task-specific approach*. Psychology Press.
- [8] Ragavendra R Baliga. 2012. *250 Cases in Clinical Medicine E-Book*. Elsevier Health Sciences.
- [9] Howard S Barrows. 1986. A taxonomy of problem-based learning methods. *Medical education* 20, 6 (1986), 481–486.
- [10] Howard S Barrows. 1996. Problem-based learning in medicine and beyond: A brief overview. *New directions for teaching and learning* 1996, 68 (1996), 3–12.
- [11] Howard S Barrows, Robyn M Tamblyn, et al. 1980. *Problem-based learning: An approach to medical education*. Vol. 1. Springer Publishing Company.
- [12] Emma Beede, Elizabeth Baylor, Fred Hersch, Anna Iurchenko, Lauren Wilcox, Paisan Ruamviboonsuk, and Laura M Vardoulakis. 2020. A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–12.
- [13] Alan Bleakley, John Bligh, and Julie Browne. 2011. *Medical education for the future: Identity, power and location*. Vol. 1. Springer Science & Business Media.
- [14] Marsden S Blois. 1980. Clinical judgment and computers. *New England Journal of Medicine* 303, 4 (1980), 192–197.
- [15] Georges Bordage, Brian Carlin, and Paul E. Mazmanian. 2009. Continuing medical education effect on physician knowledge: effectiveness of continuing medical education: American College of Chest Physicians Evidence-Based Educational Guidelines. *Chest* 135 3 Suppl (2009), 29S–36S. doi:10.1378/chest.08-2515
- [16] Nathan Borden and Derek Linklater. 2013. Hickam's dictum. *Western Journal of Emergency Medicine: Integrating Emergency Care with Population Health* 14, 2 (2013).
- [17] Francesc Borrell-Carrió, Anthony L Suchman, and Ronald M Epstein. 2004. The biopsychosocial model 25 years later: principles, practice, and scientific inquiry. *The Annals of Family Medicine* 2, 6 (2004), 576–582.
- [18] David Boud and Grahame Feletti. 2013. *The challenge of problem-based learning*. Routledge.
- [19] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.
- [20] Virginia Braun and Victoria Clarke. 2012. *Thematic analysis*. American Psychological Association.
- [21] J Brooke. 1996. SUS: A quick and dirty usability scale. *Usability Evaluation in Industry* (1996).
- [22] Tom B Brown. 2020. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165* (2020).
- [23] L Maximilian Buja. 2019. Medical education today: all that glitters is not gold. *BMC medical education* 19 (2019), 1–11.
- [24] Wouter Bulten, Kimmo Kartasalo, Po-Hsuan Cameron Chen, Peter Ström, Hans Pinckaers, Kunal Nagpal, Yuannan Cai, David F Steiner, Hester Van Boven, Robert Vink, et al. 2022. Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge. *Nature medicine* 28, 1 (2022), 154–163.
- [25] Adrian Busone, Simone Stumpf, and Dympna O'Sullivan. 2015. The role of explanations on trust and reliance in clinical decision support systems. In *2015 international conference on healthcare informatics*. IEEE, 160–169.
- [26] Bill Buxton. 2010. *Sketching user experiences: getting the design right and the right design*. Morgan kaufmann.
- [27] Carrie J Cai, Samantha Winter, David Steiner, Lauren Wilcox, and Michael Terry. 2019. "Hello AI": uncovering the onboarding needs of medical practitioners for human-AI collaborative decision-making. *Proceedings of the ACM on Human-computer Interaction* 3, CSCW (2019), 1–24.
- [28] Francisco Maria Calisto, João Fernandes, Margarida Morais, Carlos Santiago, João Maria Abrantes, Nuno Nunes, and Jacinto C Nascimento. 2023. Assertiveness-based agent communication for a personalized medicine on medical imaging diagnosis. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–20.
- [29] Francisco Maria Calisto, Carlos Santiago, Nuno Nunes, and Jacinto C Nascimento. 2021. Introduction of human-centric AI assistant to aid radiologists for multimodal breast image classification. *International Journal of Human-Computer Studies* 150 (2021), 102607.
- [30] Francisco Maria Calisto, Carlos Santiago, Nuno Nunes, and Jacinto C Nascimento. 2022. BreastScreening-AI: Evaluating medical intelligent agents for human-AI interactions. *Artificial Intelligence in Medicine* 127 (2022), 102285.
- [31] James D Cheaney and Thomas Ingebrigtsen. 2005. Problem-based learning in an online course: A case study. *The International Review of Research in Open and Distributed Learning* 6, 3 (2005).
- [32] Furui Cheng, Dongyu Liu, Fan Du, Yanna Lin, Alexandra Zytek, Haomin Li, Huamin Qu, and Kalyan Veeramachaneni. 2021. Vbridge: Connecting the dots between features and data to explain healthcare models. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2021), 378–388. doi:10.1109/TVCG.2021.3114836
- [33] Edward Choi, Mohammad Taha Bahadori, Jimeng Sun, Joshua Kulas, Andy Schuetz, and Walter Stewart. 2016. Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. *Advances in neural information processing systems* 29 (2016).
- [34] Serene SY Choo. 2012. Scaffolding in problem-based learning. *One-day, one-problem: An approach to problem-based learning* (2012), 167–184.
- [35] Carol M Chou, Katherine Kellom, and Judy A Shea. 2014. Attitudes and habits of highly humanistic physicians. *Academic medicine* 89, 9 (2014), 1252–1258.
- [36] Elizabeth B Cloude, Megan D Wiedbusch, Daryn A Dever, Dario Torre, and Roger Azevedo. 2022. The role of metacognition and self-regulation on clinical reasoning: Leveraging multimodal learning analytics to transform medical education. In *The multimodal learning analytics handbook*. Springer, 105–129.

- [37] Gino Roberto Corazza, Marco Vincenzo Lenti, and Peter David Howdle. 2021. Diagnostic reasoning in internal medicine: a practical reappraisal. *Internal and Emergency Medicine* 16 (2021), 273–279.
- [38] Juliet Corbin and Anselm Strauss. 2014. *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage publications.
- [39] Lorenzo Corti, Rembrandt Oltmans, Jiwon Jung, Agathe Balayn, Marlies Wijsenbeek, and Jie Yang. 2024. “It Is a Moving Process”: Understanding the Evolution of Explainability Needs of Clinicians in Pulmonary Medicine. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–21.
- [40] Ton De Jong. 2010. Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional science* 38, 2 (2010), 105–134.
- [41] Fernando Delgado, Stephen Yang, Michael Madaio, and Qian Yang. 2021. Stakeholder Participation in AI: Beyond “Add Diverse Stakeholders and Stir”. *arXiv preprint arXiv:2111.01122* (2021).
- [42] Arthur S Elstein and Alan Schwarz. 2002. Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. *Bmj* 324, 7339 (2002), 729–732.
- [43] George L Engel. 1977. The need for a new medical model: a challenge for biomedicine. *Science* 196, 4286 (1977), 129–136.
- [44] Yrjö Engeström. 2015. *Learning by expanding*. Cambridge University Press.
- [45] Octavia L Flanagan and Kristina M Cummings. 2023. Standardized patients in Medical Education: a review of the literature. *Cureus* 15, 7 (2023).
- [46] David M Gaba. 2004. The future vision of simulation in health care. *BMJ quality & safety* 13, suppl 1 (2004), i2–i10.
- [47] Hongyan Gu, Jingbin Huang, Lauren Hung, and Xiang’Anthony’ Chen. 2021. Lessons learned from designing an AI-enabled diagnosis tool for pathologists. *Proceedings of the ACM on Human-computer Interaction* 5, CSCW1 (2021), 1–25.
- [48] Hongyan Gu, Chunxu Yang, Mohammad Haeri, Jing Wang, Shirley Tang, Wenzhong Yan, Shujin He, Christopher Kazu Williams, Shino Magaki, and Xiang’Anthony’ Chen. 2023. Augmenting pathologists with NaviPath: design and evaluation of a human-AI collaborative navigation system. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19.
- [49] Yuexing Hao, Zeyu Liu, Robert N Riter, and Saleh Kalantari. 2024. Advancing Patient-Centered Shared Decision-Making with AI Systems for Older Adult Cancer Patients. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–20.
- [50] Ian Hargraves, Annie LeBlanc, Nilay D Shah, and Victor M Montori. 2016. Shared decision making: the need for patient-clinician conversation, not just information. *Health affairs* 35, 4 (2016), 627–629.
- [51] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183. doi:10.1016/S0166-4115(08)62386-9
- [52] NEJM Healer. 2023. NEJM Healer: An Interactive Clinical Reasoning Platform. <https://healer.lecturio.com/> Accessed: 2023-12-05.
- [53] Cindy E Hmelo-Silver. 2004. Problem-based learning: What and how do students learn? *Educational psychology review* 16 (2004), 235–266.
- [54] Maia Jacobs, Jeffrey He, Melanie F. Pradier, Barbara Lam, Andrew C Ahn, Thomas H McCoy, Roy H Perlis, Finalé Doshi-Velez, and Krzysztof Z Gajos. 2021. Designing AI for trust and collaboration in time-constrained medical decisions: a sociotechnical lens. In *Proceedings of the 2021 chi conference on human factors in computing systems*. 1–14.
- [55] Deepthy James. 2021. Levette Jones’ Clinical Reasoning Cycle: A Case Study. *Nursing Journal of India* 112, 5 (2021), 233–236.
- [56] Jerome P Kassirer. 1989. Diagnostic reasoning. *Annals of internal medicine* 110, 11 (1989), 893–900.
- [57] Jerome P Kassirer, John B Wong, and Richard I Kopelman. 1991. Learning clinical reasoning. (1991).
- [58] Megan Yih Chyn A Kek and Henk Huijser. 2011. The power of problem-based learning in developing critical thinking skills: preparing students for tomorrow’s digital futures in today’s classrooms. *Higher Education Research & Development* 30, 3 (2011), 329–341.
- [59] Jan Kiesewetter, Rene Ebersbach, Nike Tsallas, Matthias Holzer, Ralf Schmidmaier, and Martin R Fischer. 2016. Knowledge is not enough to solve the problems—The role of diagnostic knowledge in clinical reasoning activities. *BMC medical education* 16 (2016), 1–8.
- [60] Roger Kneebone. 2005. Evaluating clinical simulations for learning procedural skills: a theory-based approach. *Academic medicine* 80, 6 (2005), 549–553.
- [61] Malcolm S Knowles. 1975. Self-directed learning: A guide for learners and teachers. (1975).
- [62] David A Kolb. 2014. *Experiential learning: Experience as the source of learning and development*. FT press.
- [63] Sean Kross and Philip Guo. 2021. Orienting, framing, bridging, magic, and counseling: How data scientists navigate the outer loop of client collaborations in industry and academia. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–28.
- [64] Lotta Kuosmanen, Maija Hupli, Satu Ahtiluoto, and Elina Haavisto. 2021. Patient participation in shared decision-making in palliative care—an integrative review. *Journal of clinical nursing* 30, 23–24 (2021), 3415–3428.
- [65] Min Hun Lee and Chong Jun Chew. 2023. Understanding the effect of counterfactual explanations on trust and reliance on ai for human-ai collaborative clinical decision making. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–22.
- [66] Philip Ley. 1988. *Communicating with patients: Improving communication, satisfaction and compliance*. Croom Helm.
- [67] Tianyi Li, Mihaela Vorvoreanu, Derek DeBellis, and Saleema Amershi. 2023. Assessing human-AI interaction early through factorial surveys: a study on the guidelines for human-AI interaction. *ACM Transactions on Computer-Human Interaction* 30, 5 (2023), 1–45.
- [68] Sofie MM Loyens, Joshua Magda, and Remy MJP Rikers. 2008. Self-directed learning in problem-based learning and its relationships with self-regulated learning. *Educational psychology review* 20 (2008), 411–427.
- [69] Tayyaba Gul Malik, Usman Mahboob, Rehan Ahmed Khan, and Rabail Alam. 2024. Virtual patients versus standardized patients for improving clinical reasoning skills in ophthalmology residents. A randomized controlled trial. *BMC Medical Education* 24, 1 (2024), 429.
- [70] Dee Mangin, Cathy Risdon, Larkin Lamarche, Jessica Langevin, Abbas Ali, Jenna Parascandalo, Gabriele Stephen, and Johanna Trimble. 2019. ‘I think this medicine actually killed my wife’: patient and family perspectives on shared decision-making to optimize medications and safety. *Therapeutic Advances in Drug Safety* 10 (2019), 2042098619838796.
- [71] Henry B Mann and Donald R Whitney. 1947. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics* (1947), 50–60. doi:stable/2236101
- [72] Julian N Marewski and Gerd Gigerenzer. 2012. Heuristic decision making in medicine. *Dialogues in clinical neuroscience* 14, 1 (2012), 77–89.
- [73] Clement J McDonald. 1996. Medical heuristics: the silent adjudicators of clinical practice. 56–62 pages.
- [74] Scott Mayer McKinney, Marcin Sieniek, Varun Godbole, Jonathan Godwin, Natasha Antropova, Hutan Ashrafiyan, Trevor Back, Mary Chesnut, Greg S Corrado, Ara Darzi, et al. 2020. International evaluation of an AI system for breast cancer screening. *Nature* 577, 7788 (2020), 89–94.
- [75] Jerusalem Merkebu, Mario Veen, Shera Hosseini, and Lara Varpio. 2024. The case for metacognitive reflection: a theory integrative review with implications for medical education. *Advances in Health Sciences Education* (2024), 1–20.
- [76] Laurie Murphy, Nina B Eduljee, and Karen Croteau. 2021. Teacher-centered versus student-centered teaching: Preferences and differences across academic majors. *Journal of Effective teaching in Higher education* 4, 1 (2021), 18–39.
- [77] Don Norman. 2013. *The design of everyday things: Revised and expanded edition*. Basic books.
- [78] Geoffrey Norman. 2005. Research in clinical reasoning: past history and current trends. *Medical education* 39, 4 (2005), 418–427.
- [79] Osmosis. 2024. Osmosis. <https://www.osmosis.org/>.
- [80] Otter.ai. 2024. Otter.ai – AI Meeting Note Taker & Real-time AI Transcription. <https://otter.ai/>
- [81] Yang Ouyang, Yuchen Wu, He Wang, Chenyang Zhang, Furui Cheng, Chang Jiang, Lixia Jin, Yuanwu Cao, and Quan Li. 2023. Leveraging historical medical records as a proxy via multimodal modeling and visualization to enrich medical diagnostic learning. *IEEE Transactions on Visualization and Computer Graphics* (2023).
- [82] Candan Ozturk, Gonca Karayagiz Muslu, and Aklime Dicle. 2008. A comparison of problem-based and traditional education on nursing students’ critical thinking dispositions. *Nurse education today* 28, 5 (2008), 627–632.
- [83] Laura O’Grady and Alejandro Jadad. 2010. Shifting from shared to collaborative decision making: a change in thinking and doing. *Journal of Participatory Medicine* 2, 13 (2010), 1–6.
- [84] Clifford D Packer and Clifford D Packer. 2019. Approaches to Differential Diagnosis. *Presenting Your Case: A Concise Guide for Medical Students* (2019), 99–117.
- [85] Cecilia Panigutti, Andrea Beretta, Fosca Giannotti, and Dino Pedreschi. 2022. Understanding the impact of explanations on advice-taking: a user study for AI-based clinical Decision Support Systems. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–9.
- [86] Bhavik N Patel, Louis Rosenberg, Gregg Willcox, David Baltaxe, Mimi Lyons, Jeremy Irvin, Pranav Rajpurkar, Timothy Amrine, Rajan Gupta, Safwan Halabi, et al. 2019. Human-machine partnership with artificial intelligence for chest radiograph diagnosis. *NPJ digital medicine* 2, 1 (2019), 111.
- [87] Lia Patrício, Daniela Sangiorgi, Dominik Mahr, Martina Čaić, Saleh Kalantari, and Sue Sundar. 2020. Leveraging service design for healthcare transformation: Toward people-centered, integrated, and technology-enabled healthcare systems. *Journal of Service Management* 31, 5 (2020), 889–909.
- [88] Zhenhui Peng, Yuzhi Liu, Hanqi Zhou, Zuyu Xu, and Xiaojuan Ma. 2022. CReBot: Exploring interactive question prompts for critical paper reading. *International Journal of Human-Computer Studies* 167 (2022), 102898.
- [89] David Piorkowski, Soya Park, April Yi Wang, Dakuo Wang, Michael Muller, and Felix Portnoy. 2021. How ai developers overcome communication challenges

- in a multidisciplinary team: A case study. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–25.
- [90] Jack Pottle. 2019. Virtual reality and the transformation of medical education. *Future healthcare journal* 6, 3 (2019), 181–185.
- [91] Kate Rassie. 2017. The apprenticeship model of clinical medical education: time for structural change. *The New Zealand Medical Journal (Online)* 130, 1461 (2017), 66.
- [92] Olga M. Razumnikova. 2012. *Divergent Thinking and Learning*. Springer US, Boston, MA, 1028–1031. doi:10.1007/978-1-4419-1428-6_580
- [93] Peter Reuell. 2019. Study Shows Students in “Active Learning” Classrooms Learn More Than They Think. *The Harvard Gazette* (2019).
- [94] Kathleen R Rosen. 2008. The history of medical simulation. *Journal of critical care* 23, 2 (2008), 157–166.
- [95] Sheree May Saßmannshausen, Nazmun Nisat Ontika, Aparecido Fabiano Pinatti De Carvalho, Mark Rouncefield, and Volkmar Pipek. 2024. Amplifying Human Capabilities in Prostate Cancer Diagnosis: An Empirical Study of Current Practices and AI Potentials in Radiology. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–20.
- [96] Ami Schattner. 2020. The essence of humanistic medicine. *QJM: An International Journal of Medicine* 113, 1 (2020), 3–4.
- [97] Peter Schwartz. 2013. *Problem-based learning*. Routledge.
- [98] Berit Seljelid, Cecilia Varsi, Lise Solberg Nes, Kristin Astrid Øystese, and Elin Borosund. 2022. Feasibility of a Digital Patient-Provider Communication Intervention to Support Shared Decision-Making in Chronic Health Care, InvolveMe: Pilot Study. *JMIR Formative Research* 6, 4 (2022), e34738.
- [99] Parminder Singh and Catherine Swales. 2012. *100 Cases in Orthopaedics and Rheumatology*. CRC Press.
- [100] Ilse Johanna Sistermans. 2020. Integrating competency-based education with a case-based or problem-based learning approach in online health sciences. *Asia Pacific Education Review* 21, 4 (2020), 683–696.
- [101] Charles Spearman. 1904. The Proof and Measurement of Association between Two Things. *The American Journal of Psychology* 15, 1 (1904), 72–101. doi:10.2307/1412159
- [102] Trine Stub, Sara A Quandt, Thomas A Arcury, Joanne C Sandberg, and Agneta E Kristoffersen. 2017. Complementary and conventional providers in cancer care: experience of communication with patients and steps to improve communication with other providers. *BMC complementary and alternative medicine* 17 (2017), 1–14.
- [103] Nicole Sultanum, Michael Brudno, Daniel Wigdor, and Fanny Chevalier. 2018. More text please! understanding and supporting the use of visualization for clinical text overview. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.
- [104] Nicole Sultanum, Farooq Naeem, Michael Brudno, and Fanny Chevalier. 2022. ChartWalk: Navigating large collections of text notes in electronic health records for clinical chart review. *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2022), 1244–1254. doi:10.1109/TVCG.2022.3209444
- [105] Lu Sun, Aaron Chan, Yun Seo Chang, and Steven P Dow. 2024. ReviewFlow: Intelligent Scaffolding to Support Academic Peer Reviewing. In *Proceedings of the 29th International Conference on Intelligent User Interfaces*. 120–137.
- [106] Rand Swenson. 1999. Differential diagnosis: a reasonable clinical approach. *Neurologic clinics* 17, 1 (1999), 43–63.
- [107] Masami Tagawa. 2008. Physician self-directed learning and education. *The Kaohsiung Journal of Medical Sciences* 24, 7 (2008), 380–385. doi:10.1016/S1607-551X(08)70136-0
- [108] Anja Thieme, Maryann Hanratty, Maria Lyons, Jorge Palacios, Rita Faia Marques, Cecily Morrison, and Gavin Doherty. 2023. Designing human-centered AI for mental health: Developing clinically relevant applications for online CBT treatment. *ACM Transactions on Computer-Human Interaction* 30, 2 (2023), 1–50.
- [109] Julia CM van Weert, Barbara C Van Munster, Remco Sanders, René Spijker, Lotty Hooff, and Jesse Jansen. 2016. Decision aids to help older people make health decisions: a systematic review and meta-analysis. *BMC medical informatics and decision making* 16 (2016), 1–20.
- [110] Himanshu Verma, Jakub Mlynar, Roger Schaer, Julien Reichenbach, Mario Jreige, John Prior, Florian Evéquoz, and Adrien Depeursinge. 2023. Rethinking the role of AI with physicians in oncology: revealing perspectives from clinical and research workflows. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19.
- [111] Himanshu Verma, Roger Schaer, Julien Reichenbach, Mario Jreige, John O Prior, Florian Evéquoz, and Adrien Depeursinge. 2021. On improving physicians’ trust in AI: Qualitative inquiry with imaging experts in the oncological domain. *BMC Medical Imaging*, in review (2021).
- [112] Sarah Voss, Jerrid Kruse, and Isaiah Kent-Schneider. 2022. Comparing student responses to convergent, divergent, and evaluative nature of science questions. *Research in Science Education* 52, 4 (2022), 1277–1291.
- [113] Qiaosi Wang, Michael Madaio, Shaun Kane, Shivani Kapania, Michael Terry, and Lauren Wilcox. 2023. Designing responsible ai: Adaptations of ux practice to meet responsible ai challenges. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–16.
- [114] Brett Williams. 2005. Case based learning—a review of the literature: is there scope for this educational paradigm in prehospital education? *Emergency Medicine Journal* 22, 8 (2005), 577–581.
- [115] Wolters Kluwer. 2024. UpToDate. <https://www.wolterskluwer.com/en/solutions/upToDate>.
- [116] Diana F Wood. 2003. Problem based learning. *Brnj* 326, 7384 (2003), 328–330.
- [117] Yao Xie, Melody Chen, David Kao, Ge Gao, and Xiang’Anthony Chen. 2020. CheXplain: enabling physicians to explore and understand data-driven, AI-enabled medical imaging analysis. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [118] Qian Yang, Yuxing Hao, Kexin Quan, Stephen Yang, Yiran Zhao, Volodymyr Kuleshov, and Fei Wang. 2023. Harnessing biomedical literature to calibrate clinicians’ trust in AI decision support systems. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–14.
- [119] Qian Yang, Aaron Steinfeld, and John Zimmerman. 2019. Unremarkable AI: Fitting intelligent decision support into critical, clinical decision-making processes. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–11.
- [120] Nur Yildirim, Changhoon Oh, Deniz Sayar, Kayla Brand, Supritha Challa, Violet Turri, Nina Crosby Walton, Anna Elise Wong, Jodi Forlizzi, James McCann, et al. 2023. Creating design resources to scaffold the ideation of AI concepts. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. 2326–2346.
- [121] Nur Yildirim, Mahima Pushkarna, Nitesh Goyal, Martin Wattenberg, and Fernanda Viégas. 2023. Investigating how practitioners use human-ai guidelines: A case study on the people+ ai guidebook. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [122] Nur Yildirim, Hannah Richardson, Maria Teodora Wetscherek, Junaid Bajwa, Joseph Jacob, Mark Ames Pinnock, Stephen Harris, Daniel Coelho De Castro, Shruthi Bannur, Stephanie Hyland, et al. 2024. Multimodal healthcare AI: identifying and designing clinically relevant vision-language applications for radiology. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–22.
- [123] Nur Yildirim, Susanna Zlotnikov, Deniz Sayar, Jeremy M Kahn, Leigh A Bukowski, Sher Shah Amin, Kathryn A Rimani, Billie S Davis, John S Minturn, Andrew J King, et al. 2024. Sketching AI Concepts with Capabilities and Examples: AI Innovation in the Intensive Care Unit. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–18.
- [124] Kangyu Yuan, Hehai Lin, Shilei Cao, Zhenhui Peng, Qingyu Guo, and Xiaojuan Ma. 2023. CriTrainer: An Adaptive Training Tool for Critical Paper Reading.(2023). (2023).

A Criteria for the difficulty of cases

Table 7: The difficulty criteria of cases from three dimensions: Incidence Rate, Logical Chain Length, and Relevance to Classic Diseases.

Dimension	Low Difficulty	Medium Difficulty	High Difficulty
Incidence Rate	>100/100,000 annually	Medium 10-100/100,000 annually	<10/100,000 annually
Logical Chain Length	1-3 steps	4-7 steps	>7 steps
Relevance to Classic Diseases	Directly related	Somewhat related	Not related

B Prompting Questions Template

Table 8: Prompt questions template categorized by facts, ideas, learning issues, and action plan.

Category	Questions
Facts	<p>Step 3: What facts are described in the current case? How can the <facts> be described in clinical syndrome descriptions?</p> <p>Step 4: What results do the <updated data> present, and what facts and findings do they indicate?</p> <p>Step 5: What key points about prognosis are mentioned in the case? Are the <facts> related to diagnosis and treatment consistent with your previous analysis? What inconsistencies or overlooked aspects exist?</p>
Ideas	<p>Step 3: Based on the fact that the patient exhibits <symptoms>, what inferences and hypotheses do you have?</p> <p>Step 4: What can we infer from <findings>, and how do they help you update the problem representation and reasoning?</p> <p>Step 5: What caused the inconsistencies and overlooked of <facts>? Which part of the constructed logical chain caused these problems?</p>
Learning Issues	<p>Step 3: /</p> <p>Step 4: Are the <descriptions> of the patient consistent with your problem representation? Is the <updated data> consistent with your previous action plans?</p> <p>Step 5: What aspects were previously overlooked and needs consideration during <prognosis>? What was learned and strengthened from the analysis? What issues were identified and need further improvement?</p>
Action Plan	<p>Step 3: According to the <inferences and hypotheses>, what should the next action plans be? You may consider conducting specific examinations or treatments.</p> <p>Step 4: Based on the above analysis and reasoning, what further action plans should be considered? You may consider conducting specific examinations or treatments.</p> <p>Step 5: What are the targeted plans for further improvement and practice identified from the analysis?</p>

C Case Analysis Report Template

Table 9: Case analysis report template based on Clinical Reasoning Cycle (CRC).

Step	Description
1. Consider the Patient Situation	<p>Patient Background Information: - Age, gender, occupation, lifestyle, etc.</p> <ul style="list-style-type: none"> - Chief Complaint: The main symptoms and concerns presented by the patient. - Medical History: Relevant diseases, treatments, or surgeries. <p>Clinical Context: - Clinical background and the patient's current situation.</p> <ul style="list-style-type: none"> - Potential urgency or severity of the health problem.
2. Collect Cues and Information	<p>Clinical Data Collected: - Symptoms: Main complaints, nature of symptoms (pain location, intensity, etc.).</p> <ul style="list-style-type: none"> - Signs: Physical exam findings (vital signs, physical exam). - Laboratory and Imaging Data: Blood tests, imaging (CT, MRI, etc.). - Other Information: Psychological status, lifestyle, medications, etc. <p>Information Summary: - List the key information gathered and summarize its relevance to the diagnosis.</p>
3. Process Information	<p>Analysis of Information: - Analyze the data collected and identify key clinical cues.</p> <ul style="list-style-type: none"> - Interpret relationships among symptoms, signs, and test results. - Develop diagnostic hypotheses based on this analysis. <p>Reasoning Process: - Derive the most likely diagnoses from symptoms, signs, and test results.</p> <ul style="list-style-type: none"> - Compare different diagnostic hypotheses and explain the rationale for each one.
4. Identify Problems/Issues	<p>Core Issues: - Identify the main clinical problems or issues at hand.</p> <p>Differential Diagnosis: - List possible diagnoses, both common and rare.</p> <ul style="list-style-type: none"> - Provide evidence supporting each diagnosis and explain why others are excluded.
5. Establish Goals	<p>Short-term Goals: - Determine immediate issues to address (e.g., pain relief, stabilize vital signs).</p> <ul style="list-style-type: none"> - Set clear, actionable treatment goals (e.g., control blood pressure, further diagnostics). <p>Long-term Goals: - Set long-term health goals (e.g., stroke prevention, improve quality of life).</p>
6. Take Action	<p>Clinical Action Plan: - Based on the established goals, outline treatment plan including medications, further tests, and patient education.</p> <ul style="list-style-type: none"> - Provide evidence supporting each action (e.g., clinical guidelines, literature). <p>Treatment Plan: - Outline the treatment steps, including medication choices or interventions.</p> <ul style="list-style-type: none"> - Specify if further testing or referrals are required (e.g., neurology evaluation, imaging tests).
7. Evaluate Outcomes	<p>Evaluate Treatment Effectiveness: - Monitor patient responses to treatment and assess improvements or side effects.</p> <p>Assessment Criteria: - Evaluate based on symptom relief, lab results, and overall improvement.</p> <ul style="list-style-type: none"> - Reflect on whether the treatment plan achieved the desired goals and whether adjustments are necessary.
8. Reflect on the Process and New Learning	<p>Self-Reflection: - Reflect on potential biases or weaknesses in your reasoning process.</p> <ul style="list-style-type: none"> - Summarize new knowledge gained from the case and how it can be applied in future scenarios. <p>Future Improvements: - Suggest improvements for the reasoning process or decision-making, to increase accuracy and efficiency.</p>

D Case Analysis Report Criteria

Table 10: Case analysis report criteria based on Clinical Reasoning Cycle (CRC).

Dimension	Grading Criteria: Description
1. Consider the Patient Situation	<p>(1) Description of Patient Situation: Does the student describe the patient's condition, including chief complaint, symptoms, and background in a clinical context?</p> <p>(2) Extraction of Key Information: Does the student identify key issues, highlighting symptoms or signs relevant to the diagnosis?</p>
2. Collect Cues and Information	<p>(3) Completeness of Information Collection: Has the student gathered relevant clinical information (history, exam, labs)? Is it comprehensive?</p> <p>(4) Organization and Analysis of Information: Is the information well-organized and integrated for analysis?</p>
3. Process Information	<p>(5) Depth of Information Processing: Does the student analyze the information and extract diagnostic clues? Is reasoning clear?</p> <p>(6) Logical Consistency of Reasoning: Does the student explain relationships between data points in a logical manner?</p>
4. Identify Problems/Issues	<p>(7) Accuracy of Problem Identification: Does the student accurately identify the core problem and explain differential diagnoses?</p> <p>(8) Completeness of Differential Diagnosis: Has the student proposed multiple diagnoses based on available data?</p>
5. Establish Goals	<p>(9) Reasonableness of Goal Setting: Does the student set clear goals based on identified issues and the patient's needs?</p> <p>(10) Specificity and Feasibility of Goals: Are the treatment goals specific, actionable, and achievable (e.g., symptom improvement, ruling out conditions)?</p>
6. Take Action	<p>(11) Appropriateness of Actions Taken: Has the student developed an appropriate plan (e.g., treatment, further tests, follow-up)?</p> <p>(12) Evidence Support for Actions: Does the student provide strong supporting evidence for chosen actions (e.g., literature, guidelines)?</p>
7. Evaluate Outcomes	<p>(13) Comprehensiveness of Outcome Evaluation: Does the student evaluate the outcomes comprehensively, considering treatment effects and patient responses?</p> <p>(14) Reflection and Plan Adjustment: Does the student reflect on the outcomes and adjust the plan if necessary?</p>
8. Reflect on the Process and New Learning	<p>(15) Depth of Reflection and Self-assessment: Does the student self-reflect on strengths and weaknesses in their reasoning process?</p> <p>(16) Integration and Application of Learning: Can the student apply new learning to improve future clinical reasoning?</p>

- **7 points:** Completely meets the standard; information is collected thoroughly, analysis is logically clear, and reasoning is well-supported.
- **5-6 points:** Mostly meets the standard, but certain aspects could be more detailed or further developed.
- **3-4 points:** There are some omissions or insufficient areas; reasoning or analysis lacks comprehensiveness or rigor.
- **1-2 points:** Significant omissions or errors; reasoning is incoherent or lacks sufficient evidence.
- **0 points:** Does not meet the requirements at all; critical components are missing or clinical reasoning standards are not followed.

E In-task Questionnaire in Testing Interviews

Table 11: In-task questionnaire assessing feature usefulness.

Category	Question
Feature Usefulness	<ol style="list-style-type: none"> 1. The diagnostic list is useful for supporting evidence-based analysis. 2. The QA section helps me facilitate the exploration of evidence. 3. The prompting questions assist in structuring my clinical reasoning. 4. The mind map is useful for organizing my thoughts during analysis.

F Prompt Template

Table 12: Prompt template for the QA section in the *e-MedLearn* interface.

Prompt
You are a medical assistant. Based on the following patient case data, respond to the user's symptom inquiries by adhering to the following guidelines:
1. Single Symptom Evaluation:
- If the patient's case data includes the mentioned symptom, respond with "Yes" and provide relevant details (e.g., severity, duration, related measurements). - If the symptom is not present in the case data, respond with "Irrelevant".
2. Multiple Symptoms Evaluation:
- If the user mentions multiple symptoms, evaluate each symptom individually. - Present the results in a clear, organized list format. For example: Symptom 1: Yes Relevant Data: [Details] Symptom 2: Irrelevant Symptom 3: Yes Relevant Data: [Details]
3. Response Format:
- Ensure responses are clear, concise, and professionally formatted. - Use bullet points or numbered lists for multiple symptoms to enhance readability.
4. Language and Tone:
- Maintain a professional and objective tone. - Avoid subjective judgments or ambiguous language.
5. Error Handling:
- If the user's input is unclear or does not specify a symptom, respond with a clarifying question. For example: "Could you please specify the symptom you are inquiring about?"
Example Scenario:
- User Input: "Does the patient have a fever and headache?" - Possible Response: Fever: Yes Relevant Data: Temperature is 38.5°C, duration of 3 days. Headache: Irrelevant
Ensure that all responses strictly follow the above guidelines to maintain consistency and reliability.

G Initial Prototype

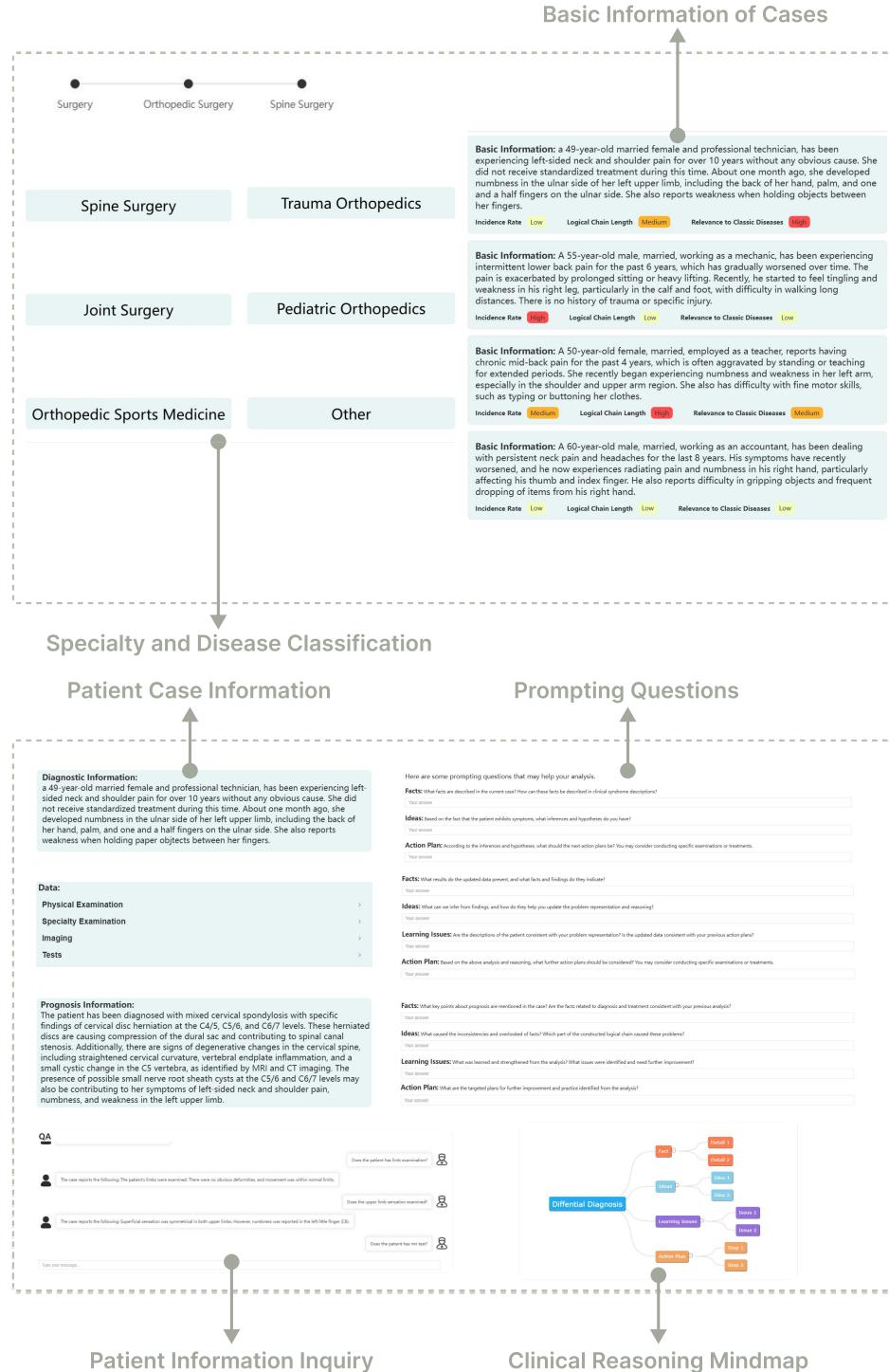


Figure 12: The initial version of the e-MedLearn prototype.

H Baseline System

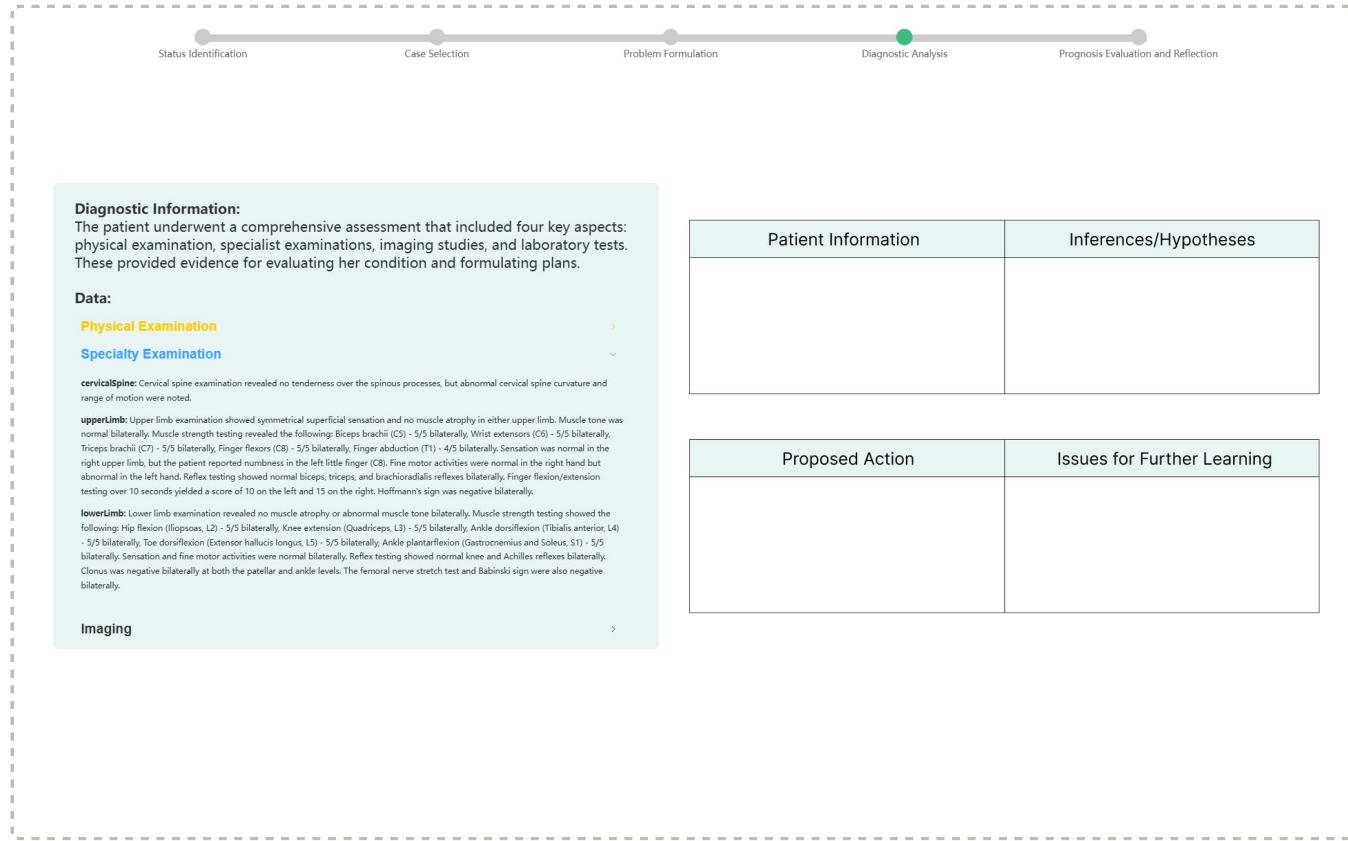


Figure 13: The interface of baseline system in the controlled study.

I System Interface

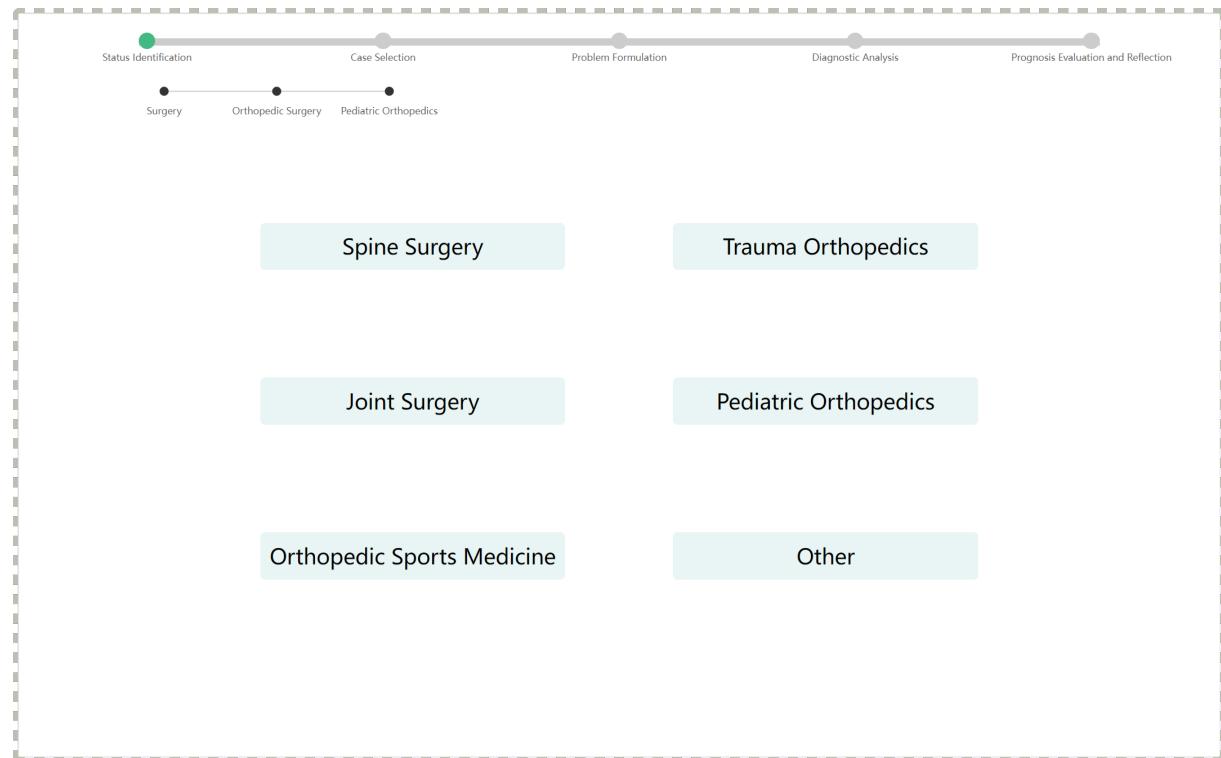


Figure 14: Step1: Status Identification

The interface shows a horizontal timeline with five stages: Status Identification, Case Selection, Problem Formulation, Diagnostic Analysis, and Prognosis Evaluation and Reflection. The 'Case Selection' stage is highlighted with a green dot. Below the timeline, there is a search bar and a 'Search' button.

Basic Information: A 49-year-old married female and professional technician, has been experiencing left-sided neck and shoulder pain for over 10 years without any obvious cause. She did not receive standardized treatment during this time. About one month ago, she developed numbness in the ulnar side of her left upper limb, including the back of her hand, palm, and one and a half fingers on the ulnar side. She also reports weakness when holding objects between her fingers.

Incidence Rate: Low | Logical Chain Length: Medium | Relevance to Classic Diseases: High

Basic Information: A 55-year-old male, married, working as a mechanic, has been experiencing intermittent lower back pain for the past 6 years, which has gradually worsened over time. The pain is exacerbated by prolonged sitting or heavy lifting. Recently, he started to feel tingling and weakness in his right leg, particularly in the calf and foot, with difficulty in walking long distances. There is no history of trauma or specific injury.

Incidence Rate: High | Logical Chain Length: Low | Relevance to Classic Diseases: Low

Basic Information: A 50-year-old female, married, employed as a teacher, reports having chronic mid-back pain for the past 4 years, which is often aggravated by standing or teaching for extended periods. She recently began experiencing numbness and weakness in her left arm, especially in the shoulder and upper arm region. She also has difficulty with fine motor skills, such as typing or buttoning her clothes.

Incidence Rate: Medium | Logical Chain Length: High | Relevance to Classic Diseases: Medium

Basic Information: A 60-year-old male, married, working as an accountant, has been dealing with persistent neck pain and headaches for the last 8 years. His symptoms have recently worsened, and he now experiences radiating pain and numbness in his right hand, particularly affecting his thumb and index finger. He also reports difficulty in gripping objects and frequent dropping of items from his right hand.

Incidence Rate: Low | Logical Chain Length: Low | Relevance to Classic Diseases: Low

Basic Information: A 48-year-old female, married, who works as a sales manager, has suffered from severe lower back pain and stiffness for the past 5 years. The pain has progressively spread to her hips and legs. Recently, she has noticed numbness in her left buttock and leg, along with a loss of strength in her left foot, making it challenging for her to walk or climb stairs.

Incidence Rate: High | Logical Chain Length: High | Relevance to Classic Diseases: Low

Basic Information: A 53-year-old male, married, employed as a construction worker, has experienced chronic pain in his upper back and shoulders for the past 7 years, with occasional radiating pain down his left arm. Recently, he developed numbness and a tingling sensation in his left hand and fingers, along with significant weakness in gripping tools and performing manual tasks.

Incidence Rate: Medium | Logical Chain Length: High | Relevance to Classic Diseases: Low

Figure 15: Step2: Case Selection

The interface shows a timeline with five stages: Status Identification, Case Selection, Problem Formulation, Diagnostic Analysis, and Prognosis Evaluation and Reflection. The 'Problem Formulation' stage is highlighted with a green dot.

Diagnostic Information:

a 49-year-old married female and professional technician, has been experiencing left-sided neck and shoulder pain for over 10 years without any obvious cause. She did not receive standardized treatment during this time. About one month ago, she developed numbness in the ulnar side of her left upper limb, including the back of her hand, palm, and one and a half fingers on the ulnar side. She also reports weakness when holding paper objects between her fingers.

She denies any other limb discomfort. She denies blackout. She denies sudden falls. She denies dizziness.

QA

- The patient reports the following: She denies blackout.
- The patient reports the following: She denies sudden falls.
- The patient reports the following: She denies dizziness.

Type your message: [Text input field]

Here are some prompting questions that may help your analysis.

Facts: What facts are described in the current case? How can these facts be described in clinical syndrome descriptions?

left-sided neck and shoulder pain, numbness in the ulnar side of her left upper limb, weakness when holding paper objects between her fingers

Ideas: Based on the fact that the patient exhibits symptoms, what inferences and hypotheses do you have?

Consider: Cervical Disc Herniation, Cervical Spondylosis, and Nerve Injury. Need to rule out Stroke. Not likely to be Multiple Sclerosis.

Action Plan: according to the inferences and hypotheses, what should the next action plan be? You may consider conducting specific examinations or treatments.

Physical Examination and MRI examination

	Disease	✓ Supporting Evidence	X Excluding Evidence
Most Likely	Cervical Disc Herniation	left-sided neck and shoulder pain, numbness in the ulnar side of her left upper limb	
	Cervical Spondylosis	left-sided neck and shoulder pain, numbness in the ulnar side of her left upper limb	
	Nerve Injury	numbness in the ulnar side of her left upper limb, weakness when holding paper objects between her fingers	without any obvious cause
Need to Rule Out	Stroke	weakness when holding paper objects between her fingers	numbness in the ulnar side of her left upper limb
Not Likely	Multiple Sclerosis		numbness in the ulnar side of her left upper limb, weakness when holding paper objects between her fingers

Differential Diagnosis

```

graph TD
    Fact --> Ideas
    Ideas --> LearningIssues
    Ideas --> ActionPlan
    LearningIssues --> PhysicalExam
    LearningIssues --> MRIExam
    ActionPlan --> PhysicalExam
    ActionPlan --> MRIExam
    
```

Fact: left-sided neck and shoulder pain, numbness in the ulnar side of her left upper limb, weakness when holding paper objects between her fingers

Ideas: Could be Cervical Disc Herniation, Cervical Spondylosis, or Nerve Injury. Need to rule out Stroke. Not likely to be Multiple Sclerosis.

Learning Issues: Identify action plans from ideas

Action Plan: Physical Examination, MRI examination

Figure 16: Step3: Problem Formulation

Diagnostic Information:
The patient underwent a comprehensive assessment that included four key aspects: physical examination, specialist examinations, imaging studies, and laboratory tests. These provided evidence for evaluating her condition and formulating plans.

Examination Analysis Outline:

- Physical Examination:** The patient's limbs were examined. There were no obvious deformities, and movement was within normal limits.
- Specialty Examination:** Superficial sensation was symmetrical in both upper limbs. However, numbness was reported in the left little finger (CB).
- Imaging:**
- Tests:** The hematology test is conducted and present below.

Data:

- Physical Examination**
- Specialty Examination**
- Imaging**

Disease	✓ Supporting Evidence	X Excluding Evidence
Cervical Disc Herniation	left-sided neck and shoulder pain, numbness in the ulnar side of	
Most Likely	Cervical Spondylosis	left-sided neck and shoulder pain, numbness in the ulnar side of
	Nerve Injury	numbness in the ulnar side of her left upper limb, weakness when holding paper objects between
		without any obvious cause
Need to Rule Out	Stroke	weakness when holding paper objects between her fingers
		numbness in the ulnar side of her left upper limb
Not Likely	Multiple Sclerosis	numbness in the ulnar side of her left upper limb, weakness when holding paper objects between

QA

The case reports the following: The patient's limbs were examined. There were no obvious deformities, and movement was within normal limits.

Does the patient have limb examination?

The case reports the following: Superficial sensation was symmetrical in both upper limbs. However, numbness was reported in the left little finger (CB).

Does the upper limb sensation examined?

Does the patient have MRI test?

Facts: What results do the updated data present, and what facts and findings do they indicate?
The patient has disc herniations at C4/5, C5/6, and C6/7 levels, along with cervical spine degeneration.

Ideas: What can we infer from findings, and how do they help you update the problem representation and reasoning?
Imaging findings are consistent with the patient's symptoms and physical signs, leading to the conclusion that the patient likely suffers from cervical disc herniation (CDH).

Learning Issues: Are the descriptions of the patient consistent with your problem representation? Is the updated data consistent with your previous action plans?
Updated data provide a clearer understanding of the underlying causes of the disease.

Action Plan: Based on the above analysis and reasoning, what further action plans should be considered? You may consider conducting specific examinations or treatments.
It is necessary to discuss whether the patient should proceed with surgical intervention. Different corresponding measures should be taken based on the decision between surgery and conservative treatment.

Differential Diagnosis

```

graph TD
    DD((Differential Diagnosis)) --> Fact[Fact]
    DD --> Idea[Idea]
    DD --> LI[Learning Issues]
    DD --> AP[Action Plan]
    
    Fact --> F1[Left-sided neck and shoulder pain]
    Fact --> F2[Numbness in the ulnar side of her left upper limb]
    Fact --> F3[Weakness when holding paper objects between]
    Fact --> F4[Disc herniations at C4/5, C5/6, and C6/7 levels]
    Fact --> F5[Could be CDH, CS, or Nerve injury]
    Fact --> F6[Need to rule out Stroke]
    Fact --> F7[The patient likely suffers from cervical disc herniation]
    
    Idea --> I1[Identify action plans from ideas]
    Idea --> I2[Updated data provide a clearer understanding]
    
    LI --> L1[More proactive surgical intervention should be considered]
    LI --> L2[Delaying treatment may further impair nerve function recovery]
    
    AP --> A1[Physical examination]
    AP --> A2[MRI examination]
    AP --> A3[Determine whether the patient should proceed with surgery]
  
```

Figure 17: Step4: Diagnostic Analysis

Prognosis Information:
The patient has been diagnosed with mixed cervical spondylosis with specific findings of cervical disc herniation at the C4/5, C5/6, and C6/7 levels. These herniated discs are causing compression of the dural sac and contributing to spinal canal stenosis. Additionally, there are signs of degenerative changes in the cervical spine, including straightened cervical curvature, vertebral endplate inflammation, and a small cystic change in the C5 vertebra, as identified by MRI and CT imaging. The presence of possible small nerve root sheath cysts at the C5/6 and C6/7 levels may also be contributing to her symptoms of left-sided neck and shoulder pain, numbness, and weakness in the left upper limb.

Given the chronic nature of Ms. Shao's condition and the recent onset of neurological symptoms (e.g., numbness and weakness in the ulnar side of the left upper limb), surgical intervention is likely necessary to alleviate the nerve compression and prevent further deterioration. If successful, the surgery may reduce her symptoms and improve her quality of life. However, the long-term prognosis will depend on the extent of the spinal degeneration and the success of the surgical intervention in relieving the compression. Postoperative rehabilitation will be crucial to maximize recovery and restore function. Without treatment, there is a risk of progressive neurological deficits and worsening pain, which could significantly impact her daily activities and overall quality of life. It is essential to monitor her response to treatment closely and adjust the management plan as needed to address any ongoing or emerging issues.

QA

The case reports the following: The hematology test is conducted and present below.

Does the patient have hematology test?

The plans include rehabilitation after surgery, monitoring recovery, and long-term follow-up.

What should be considered during the patient's prognosis?

Type your message.

Facts: What key points about prognosis are mentioned in the case? Are the facts related to diagnosis and treatment consistent with your previous analysis?
Diagnosis is consistent with analysis. While the prognosis showed no recovery from pain, numbness, and weakness.

Ideas: What caused the inconsistencies and overlooked facts? Which part of the constructed logical chain caused these problems?
The disc herniations, left untreated surgically, have resulted in increased compressions. Thus the prognosis is not well.

Learning Issues: What was learned and strengthened from the analysis? What issues were identified and need further improvement?
More proactive surgical intervention should be considered, as delaying treatment may further impair nerve function recovery.

Action Plan: What are the targeted plans for further improvement and practice identified from the analysis?
Provide the patient with a more specific prognosis regarding treatment options, taking into account individual factors.

Differential Diagnosis

```

graph TD
    DD((Differential Diagnosis)) --> Fact[Fact]
    DD --> Idea[Idea]
    DD --> LI[Learning Issues]
    DD --> AP[Action Plan]
    
    Fact --> F1[Left-sided neck and shoulder pain]
    Fact --> F2[Numbness in the ulnar side of her left upper limb]
    Fact --> F3[Weakness when holding paper objects between]
    Fact --> F4[Disc herniations at C4/5, C5/6, and C6/7 levels]
    Fact --> F5[Progress showed no recovery from pain, numbness, and weakness]
    Fact --> F6[Could be CDH, CS, or Nerve injury]
    Fact --> F7[Need to rule out Stroke]
    Fact --> F8[The patient likely suffers from cervical disc herniation]
    
    Idea --> I1[Identify action plans from ideas]
    Idea --> I2[Updated data provide a clearer understanding]
    
    LI --> L1[More proactive surgical intervention should be considered]
    LI --> L2[Delaying treatment may further impair nerve function recovery]
    
    AP --> A1[Physical examination]
    AP --> A2[MRI examination]
    AP --> A3[Determine whether the patient should proceed with surgery]
  
```

Figure 18: Step5: Prognosis Evaluation and Reflection

J System Walkthrough

Diagnostic Information:

a 49-year-old married female and professional technician, has been experiencing left-sided neck and shoulder pain for over 10 years without any obvious cause. She did not receive standardized treatment during this time. About one month ago, she developed numbness in the ulnar side of her left upper limb, including the back of her hand, palm, and one and a half fingers on the ulnar side. She also reports weakness when holding paper objects between her fingers.

She denies any other limb discomfort. She denies blackout. She denies sudden falls. She denies dizziness.

Physical Examination >

Imaging Tests >

Imaging >

C6/6 C5/6 intervertebral disc herniation.

Study Date: 2022-01-04 Study Time: 18:59:44

Need to evaluate the nerve root or spinal cord compression.

Physical Examination >

Visual Signs: Tachypnoea 30/10, respiratory rate 18 breath/min, pulse 80/min, blood pressure 120/70 mmHg.

Orthopedic Related Measurements: Weight 70 kg, height 168 cm, waist circumference not measured, hip circumference not measured.

BMI: 24.6

General Physical Examination: The patient is alert with stable mental status and smooth breathing. Nutrition is moderate, facial expressions are normal, and development is appropriate. The patient is in a self-supported position, responds fluently, and cooperates well during the examination. The skin is pink and warm, with no rashes or lesions. There is no edema in the lower extremities. The heart rate is 80 bpm, regular rhythm, and a single murmur is present. No bruits are heard in the abdomen. The lungs are clear. The abdomen is soft, non-tender, and no masses are palpable. The rectal examination is unremarkable. The genitalia are normal. The neurological examination shows normal reflexes in all four limbs, with no sensory deficits. The gait is normal. The patient can stand on one leg without support.

Tests >

Cardiac Markers

Item	Result	Normal Range
Troponin T	0.014 ng/mL	<0.04 ng/mL
CK-MB	5.1 ng/mL	0-45 ng/mL
Magnesium	204.2 ng/mL	25-75 ng/mL
Nt-proBNP	3037 pg/mL	<300 pg/mL
Procalcitonin	0.07 ng/mL	<0.5 ng/mL

Prognosis Information:

The patient has been diagnosed with mixed cervical spondylosis with specific findings of cervical disc herniation at the C4/5, C5/6, and C6/7 levels. These herniated discs are causing compression of the dural sac and contributing to spinal canal stenosis. Additionally, there are signs of degenerative changes in the cervical spine, including straightened cervical curvature, vertebral endplate inflammation, and a small cystic change in the C5 vertebra, as identified by MRI and CT imaging. The presence of possible small nerve root sheath cysts at the C5/6 and C6/7 levels may also be contributing to her symptoms of left-sided neck and shoulder pain, numbness, and weakness in the left upper limb.

Differential Diagnosis:

d1 disc herniations at the C4/5, C5/6, and C6/7 levels

Disease	✓ Supporting Evidence	X Excluding Evidence
Cervical Disc Herniation	left-sided neck and shoulder pain, numbness in the ulnar side of her left upper limb	
Most Likely	Cervical Spondylosis	left-sided neck and shoulder pain, numbness in the ulnar side of her left upper limb
	Nerve Injury	numbness in the ulnar side of her left upper limb, weakness when holding paper objects between her fingers
Need to Rule Out	Stroke	weakness when holding paper objects between her fingers
		numbness in the ulnar side of her left upper limb
Not Likely	Multiple Sclerosis	numbness in the ulnar side of her left upper limb, weakness when holding paper objects between her fingers

Diffential Diagnosis:

Fact: left-sided neck and shoulder pain, numbness in the ulnar side of her left upper limb, weakness when holding paper objects between her fingers, disc herniations at C4/5, C5/6, and C6/7 levels, without any obvious cause.

Ideas: Could be CDH, CS, or Nerve Injury, Need to rule out Stroke, the patient likely suffers from cervical disc herniation.

Learning Issues: Identify action plans from ideas, Updated data provide a clearer understanding of the patient's condition.

Action Plan: Physical Examination, MRI examination, discuss whether the patient should proceed with surgery.

```

graph TD
    DD[Differential Diagnosis] --> Fact[Fact]
    Fact --> Ideas[Ideas]
    Ideas --> LI[Learning Issues]
    LI --> AP[Action Plan]
    AP --> P[Physical Examination]
    AP --> MRI[MRI examination]
    AP --> D[disc herniations at the C4/5, C5/6, and C6/7 levels]
  
```

Figure 19: The process of conducting case analysis with e-MedLearn after Stage 1: Case Selection consists of (a) Presenting the preliminary case data, exploring the evidence, and formulating initial ideas of diagnosis. (b) Perform diagnostic analysis using data from examinations or tests and update the ideas of the diagnosis. (c) Consider prognosis and assess analysis results, reflect ideas to conclude learning issues. (d) Record and update the diagnosis in the diagnosis list. (e) Record and review the insights and thoughts of the analysis process in the mind map.

We present the system's interactive usage through a complete walkthrough. In this scenario, Jon, a medical intern, uses e-MedLearn for his learning process. Initially, he selected the orthopedic category with a focus on spine-related diseases as his area of interest. Jon then selected a case with a low *incidence rate* but significant *relevance to classic diseases* for analysis.

In Step 3: Problem Formulation, Jon first reviewed the initial diagnostic information of the case (Figure 19-a). He then engaged with the system in the QA section to further explore and collect evidence (Figure 19-a1). Through questioning, Jon confirmed that the patient *denies experiencing blackouts, sudden falls, dizziness, or any other limb discomfort*.

Following prompting questions, Jon organized the collected information: *the current facts indicate that the patient experiences left-sided neck and shoulder pain, numbness in the ulnar side of her left upper limb, and weakness when holding paper objects between her fingers*. Based on these facts, the ideas of possible hypotheses and inferences include: *Consider Cervical Disc Herniation, Cervical Spondylosis, Nerve Injury, and Stroke. While Multiple Sclerosis is not likely* (Figure 19-a2). The proposed action plans involve conducting a *Physical Examination and an MRI examination*.

Jon then updated the preliminary diagnoses in the diagnosis list (Figure 19-d) accordingly: *Based on supporting evidence, the most likely diagnoses are Cervical Disc Herniation and Cervical Spondylosis. In addition, Stroke needs to be ruled out, as weakness when holding paper objects between her fingers serves as supporting evidence for this possibility. Multiple Sclerosis is considered unlikely because the numbness in the ulnar side of her left upper limb and weakness when holding paper objects provide evidence against this diagnosis*.

In *Step 4: Diagnostic Analysis*, Jon further inquired in the QA section to identify the types of updated examination and tests (Figure 19-b1), and then reviewed the detailed results (Figure 19-b). Following the prompting questions, Jon analyzed the data and extracted the following facts: *The patient has disc herniations at the C4/5, C5/6, and C6/7 levels, along with cervical spine degeneration*.

These findings led him to the idea that *imaging results are consistent with the patient's symptoms and physical signs* (Figure 19-b2), leading to the conclusion that the patient likely suffers from *Cervical Disc Herniation*. Jon noted that *the updated data provides a clearer understanding of the underlying causes of the disease*. Following this, Jon updated the diagnosis list, adding the evidence provided by MRI examination for *Cervical Disc Herniation* (Figure 19-d1).

In *Step 5: Prognosis and Evaluation*, Jon reviewed the patient's prognosis information (Figure 19-c) and gathered key considerations regarding the prognosis through inquiry in the QA section (Figure 19-c1). Jon confirmed that the diagnosis results align with his previous analysis. Reflecting further, Jon realized that *more proactive surgical intervention should be considered, as delaying treatment may further impair nerve function recovery* (Figure 19-c2). The next action plan involves *providing the patient with a more specific prognosis regarding treatment options and taking into account individual factors*.