

**Original Paper**

# Importance of Patient History in Artificial Intelligence-Assisted Medical Diagnosis: Comparison Study

Fumitoshi Fukuzawa<sup>1</sup>, MD; Yasutaka Yanagita<sup>1</sup>, MD, PhD; Daiki Yokokawa<sup>1</sup>, MD, PhD; Shun Uchida<sup>2</sup>, MD; Shiho Yamashita<sup>1</sup>, MD; Yu Li<sup>1</sup>, MD, PhD; Kiyoshi Shikino<sup>1</sup>, MD, MHPE, PhD; Tomoko Tsukamoto<sup>1</sup>, MD, PhD; Kazutaka Noda<sup>1</sup>, MD, PhD; Takanori Uehara<sup>1</sup>, MD, PhD; Masatomi Ikusaka<sup>1</sup>, MD, PhD

<sup>1</sup>Department of General Medicine, Chiba University Hospital, Chiba-shi, Japan

<sup>2</sup>Uchida Internal Medicine Clinic, Saitama-shi, Japan

**Corresponding Author:**

Fumitoshi Fukuzawa, MD

Department of General Medicine

Chiba University Hospital

1-8-1, Inohana, Chuo-ku

Chiba-shi, 260-8677

Japan

Phone: 81 43-222-7171

Email: [f.fukuzawa0919@ymail.ne.jp](mailto:f.fukuzawa0919@ymail.ne.jp)

## Abstract

**Background:** Medical history contributes approximately 80% to a diagnosis, although physical examinations and laboratory investigations increase a physician's confidence in the medical diagnosis. The concept of artificial intelligence (AI) was first proposed more than 70 years ago. Recently, its role in various fields of medicine has grown remarkably. However, no studies have evaluated the importance of patient history in AI-assisted medical diagnosis.

**Objective:** This study explored the contribution of patient history to AI-assisted medical diagnoses and assessed the accuracy of ChatGPT in reaching a clinical diagnosis based on the medical history provided.

**Methods:** Using clinical vignettes of 30 cases identified in *The BMJ*, we evaluated the accuracy of diagnoses generated by ChatGPT. We compared the diagnoses made by ChatGPT based solely on medical history with the correct diagnoses. We also compared the diagnoses made by ChatGPT after incorporating additional physical examination findings and laboratory data alongside history with the correct diagnoses.

**Results:** ChatGPT accurately diagnosed 76.6% (23/30) of the cases with only the medical history, consistent with previous research targeting physicians. We also found that this rate was 93.3% (28/30) when additional information was included.

**Conclusions:** Although adding additional information improves diagnostic accuracy, patient history remains a significant factor in AI-assisted medical diagnosis. Thus, when using AI in medical diagnosis, it is crucial to include pertinent and correct patient histories for an accurate diagnosis. Our findings emphasize the continued significance of patient history in clinical diagnoses in this age and highlight the need for its integration into AI-assisted medical diagnosis systems.

**JMIR Med Educ** 2024;10:e52674; doi: [10.2196/52674](https://doi.org/10.2196/52674)

**Keywords:** medical diagnosis; ChatGPT; AI in medicine; diagnostic accuracy; patient history; medical history; artificial intelligence; AI; physical examination; physical examinations; laboratory investigation; laboratory investigations; mHealth; accuracy; public health; United States; AI diagnosis; treatment; male; female; child; children; youth; adolescent; adolescents; teen; teens; teenager; teenagers; older adult; older adults; elder; elderly; older person; older people; investigative; mobile health; digital health

## Introduction

Over the past decade, medical knowledge and diagnostic techniques have expanded globally and have become more accessible with remarkable advancements in clinical

testing and useful reference systems. Despite these advancements, misdiagnosis significantly contributes to mortality, making it a significant public health issue [1,2]. Studies have shown discrepancies between clinical and postmortem autopsy diagnoses in at least 25% of patients [3-7]. One study

suggests that approximately 40,500 adult patients in intensive care units in the United States die of misdiagnoses annually, and the predicted prevalence of potentially lethal misdiagnoses is 6.3% [8]. Another report suggests that diagnostic errors contribute to approximately 10% of deaths and 6% to 17% of hospital adverse events, and are the leading cause of medical malpractice claims [7]. Considering the operative characteristics of clinical investigations combined with the inherent variability in disease presentation, it is often challenging to diagnose patients correctly—an issue that has concerned physicians perennially. Decades ago, a pivotal study proposed that patient history contributes to approximately 80% of the diagnostic process [9,10]. Medical history remains crucial for diagnosis [11,12] and is vital in contemporary physicians' clinical diagnoses.

With the advent of artificial intelligence (AI) in recent years, numerous studies have focused on AI-assisted diagnoses, including cancer screening and treatment [13-15], diagnostic ultrasound imaging [16-19], x-ray imaging [20], computed tomography [21], magnetic resonance imaging [22], and endoscopy [15,23]. Other reports on AI-assisted imaging diagnoses include AI's applications in radiology, pathology, and dermatological imaging [13,24]. There have also been reports on the use of AI in diagnosing specific conditions [25-27]. However, while several studies have reported that AI is useful in screening, diagnosing, and even treating certain medical conditions, to the best of our knowledge, no study has examined the importance of patient history in AI-assisted medical diagnosis. In addition, the extent to which AI considers patient history in its diagnostic processes remains to be fully understood.

This study aimed to investigate the importance of patient history in an AI-assisted medical diagnostic process aided by ChatGPT (version 4.0; June 2, 2023), one of the most well-known large language models that was released on March 14, 2023, to better understand the future of diagnostic medicine where AI is predicted to play an increasingly

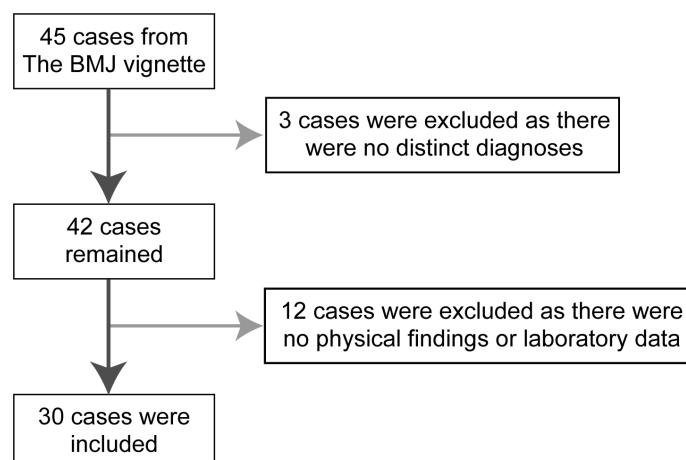
prominent role. Our study explored the contribution of patient history to AI-assisted medical diagnoses and assessed the accuracy of ChatGPT in reaching a clinical diagnosis based on the medical history that was provided. By reevaluating the significance of patient history, our study contributes to the ongoing discourse on optimizing diagnostic processes, both conventional and AI-assisted.

## Methods

### Study Design, Settings, and Participants

In our study, we used some of the 45 standardized clinical vignettes in *The BMJ* ([Multimedia Appendix 1](#)) to evaluate the diagnostic and triage accuracy of web-based symptom checkers [28]. These vignettes were published on June 5, 2015. They offer a balanced set of cases, with 15 cases requiring immediate attention, 15 cases requiring consultation but not immediately, and 15 cases not requiring immediate attention or consultation. They were identified from various clinical sources, including materials used to educate health professionals as well as a medical resource website, with content provided by a panel of physicians. Researchers have used these clinical vignettes to evaluate the usefulness of web-based symptom checkers and self-triage [28-31]. We chose these vignettes because of their varied severity levels, their origins from multiple resources rather than just 1 resource, and their credibility, having been used in prior studies. They also include some of the most commonly observed conditions in outpatient settings. Of the 45 cases, we selected those that included physical examination findings, test data, and medical history and provided a single distinct diagnosis. As illustrated in [Figure 1](#), we excluded patients with no distinct diagnoses within the vignettes to serve as a reference (3 cases) and those who did not undergo any physical examination or laboratory tests (12 cases). Finally, the remaining 30 cases were used in this study.

**Figure 1.** Inclusion and exclusion criteria.



### Data Collection and Measurements

We assigned the correct diagnosis for each of these 30 cases to "Answer." We then used the AI model, ChatGPT,

to generate 2 diagnoses: the first, labeled "History," was obtained by inputting only the medical history into ChatGPT; the second set, labeled "All," was produced by inputting the medical history and all the other additional information

in the clinical vignettes. Each time ChatGPT was prompted to generate a diagnosis, a separate chat window was used ([Multimedia Appendix 2](#)). Thus, we used 2 chat windows for each case—one for the “History” diagnosis and the other for the “All” diagnosis. Additionally, the patients’ information was not inputted incrementally.

The concordance rate was assessed among “Answer,” “History,” and “All.” To extract a diagnosis from ChatGPT, we ended each input session with the phrase “What is the most likely diagnosis?” For both the “History” and “All,” the session was deemed complete when the AI returned the single most likely diagnosis. If ChatGPT suggested multiple diagnoses or indicated that it did not provide the most likely diagnosis, we repeated the process under the same conditions for a maximum of 5 attempts. Cases for which a single diagnosis could not be obtained even after 5 attempts were excluded without making further attempts.

## Ethical Considerations

Our research does not involve humans, medical records, patient information, observations of public behaviors, or secondary data analyses; hence, it is exempt from ethical approval, the requirement of informed consent, and institutional review board approval. Additionally, as no identifying information was included, the data did not need to be anonymized or deidentified, and the need for compensation did not arise because no human participants were included in the study.

## Data Analysis

Three board-certified physicians working in a medical diagnostic department at our facility assessed the concordance among the 3 AI-proposed diagnoses (“Answer,” “History,”

and “All”). Of the 3 physicians, 1 is general medicine board-certified, 1 is internal medicine board-certified, and 1 is internal medicine-, general internal medicine-, and family medicine board-certified; their postgraduate education spanned 7, 9, and 11 years, respectively. A diagnosis was considered to match if at least 2 of the 3 physicians agreed upon the correspondence. Distinguishing between acute pharyngitis and acute upper respiratory tract infection necessitated determining whether to consider diseases resulting from similar pathologies as correct diagnoses. In contrast, for diseases that are essentially the same but have different nomenclatures, such as oral ulcers and canker sores, we considered them correct diagnoses.

## Results

Among the 30 cases, 19 patients were male and 11 were female, with ages ranging from 18 months to 65 years. In total, 12 individuals were younger than 20 years.

The results are shown in [Table 1](#). Cases 1-15 of the original vignette represent those requiring emergent care, cases 16-30 represent those requiring nonemergent care, and cases 31-45 represent those that are appropriate for self-care. A comparison with the correct diagnosis listed in *The BMJ* vignettes (labeled as “Answer”) showed that “Answer” and “History” coincided 76.6% of the time, while “Answer” and “All” had a concordance rate of 93.3%. Five (16.7%) patients could not be diagnosed on the basis of medical history alone but were diagnosed when additional information was provided. In 1 (3.3%) case, the diagnosis was different and incorrect under both conditions (“History” and “All”). In 1 (3.3%) case, the incorrect diagnosis was the same under both conditions (“History” and “All”).

**Table 1.** List of answers and diagnoses made by ChatGPT<sup>a</sup>.

Case number of the original vignette	Original diagnosis (Answer)	Output from history only (History) <sup>b</sup>	Output from all information (All) <sup>c</sup>
1	Acute liver failure	Acute liver failure <sup>d</sup>	Acute liver failure <sup>d</sup>
2	Appendicitis	Acute gastroenteritis	Acute peritonitis, possibly secondary to a ruptured appendix (perforated appendicitis) <sup>d</sup>
5	Deep vein thrombosis	Deep vein thrombosis <sup>d</sup>	Deep vein thrombosis <sup>d</sup>
6	Heart attack	Acute myocardial infarction <sup>d</sup>	Acute anterior wall myocardial infarction <sup>d</sup>
7	Hemolytic uremic syndrome	Hemolytic uremic syndrome <sup>d</sup>	Hemolytic uremic syndrome <sup>d</sup>
9	Malaria	Malaria <sup>d</sup>	Malaria <sup>d</sup>
10	Meningitis	N/A <sup>e</sup> × 5 <sup>f</sup>	Meningitis <sup>d</sup>
11	Pneumonia	Community-acquired pneumonia <sup>d</sup>	Community-acquired pneumonia <sup>d</sup>
12	Pulmonary embolism	Pulmonary embolism <sup>d</sup>	Pulmonary embolism <sup>d</sup>
13	Rocky Mountain spotted fever	Tick-borne illness, such as Rocky Mountain spotted fever or ehrlichiosis <sup>d</sup>	Rocky Mountain spotted fever <sup>d</sup>
16	Acute otitis media	Viral upper respiratory tract infection	Acute otitis media <sup>d</sup>
17	Acute pharyngitis	Strep throat <sup>d</sup>	Streptococcal pharyngitis <sup>d</sup>
18	Acute pharyngitis	Streptococcal pharyngitis <sup>d</sup>	Streptococcal pharyngitis <sup>d</sup>
19	Acute sinusitis	Acute sinusitis <sup>d</sup>	N/A × 2 <sup>g</sup> ; acute bacterial sinusitis <sup>d</sup>

Case number of the original vignette	Original diagnosis (Answer)	Output from history only (History) <sup>b</sup>	Output from all information (All) <sup>c</sup>
21	Cellulitis	N/A × 5	Cellulitis <sup>d</sup>
24	Mononucleosis	Infectious mononucleosis <sup>d</sup>	Infectious mononucleosis <sup>d</sup>
25	Peptic ulcer disease	Peptic ulcer disease <sup>d</sup>	Peptic ulcer disease <sup>d</sup>
26	Pneumonia	Pneumonia <sup>d</sup>	Community-acquired pneumonia <sup>d</sup>
27	<i>Salmonella</i> infection	<i>Campylobacter jejuni</i> infection	Acute gastroenteritis, likely due to food poisoning
30	Vertigo	Benign paroxysmal positional vertigo <sup>d</sup>	Benign paroxysmal positional vertigo <sup>d</sup>
31	Acute bronchitis	Acute bronchitis <sup>d</sup>	Acute bronchitis <sup>d</sup>
32	Acute bronchitis	Acute bronchitis <sup>d</sup>	Acute bronchitis <sup>d</sup>
33	Acute conjunctivitis	Viral conjunctivitis <sup>d</sup>	Viral conjunctivitis <sup>d</sup>
34	Acute pharyngitis	Viral upper respiratory tract infection	Upper respiratory tract infection
37	Bee sting without anaphylaxis	Pain of the sting	Localized allergic reaction to a bee sting <sup>d</sup>
38	Canker sore	Recurrent aphthous stomatitis <sup>d</sup>	Recurrent aphthous stomatitis <sup>d</sup>
39	Candida yeast infection	Vaginal candidiasis <sup>d</sup>	Vulvovaginal candidiasis <sup>d</sup>
42	Stye	Hordeolum <sup>d</sup>	Hordeolum <sup>d</sup>
43	Viral upper respiratory tract infection	Acute sinusitis <sup>d</sup>	Acute sinusitis <sup>d</sup>
44	Viral upper respiratory tract infection	Common viral illness, such as the common cold or influenza <sup>d</sup>	Viral upper respiratory tract infection <sup>d</sup>

<sup>a</sup>We repeated outputs until a single plausible diagnosis was made, with a maximum of 5 attempts.

<sup>b</sup>Matching answers between Answer and History: 23/30 (76.6%); median trial count 1 (Q1 1, Q2 1, Q3 1).

<sup>c</sup>Matching answers between History and All: 28/30 (93.3%); median trial count 1 (Q1 1, Q2 1, Q3 1).

<sup>d</sup>The output matched with that of “Answer.”

<sup>e</sup>N/A: not applicable.

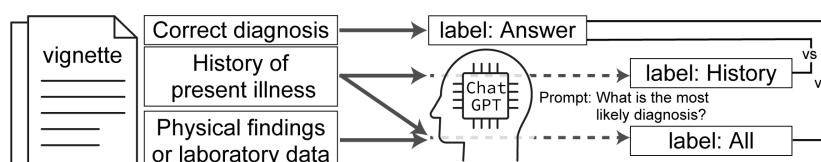
<sup>f</sup>We attempted to obtain a diagnosis 5 times but failed.

<sup>g</sup>We attempted to obtain a diagnosis twice but failed.

Figure 2 presents details regarding the number of attempts required. On average, 1.27 attempts were needed for inputs involving only medical history followed by the question “What is the most likely diagnosis?” When all possible information, including physical examination findings and laboratory data, were inputted, followed by the same

question, an average of 1.00 attempt was required. Regarding the 2 cases shown in Figure 2 that required 5 attempts, ChatGPT was unable to narrow down the diagnosis to the single most likely option. Consequently, these cases were counted as mismatches with the correct diagnoses listed in *The BMJ* vignettes.

**Figure 2.** Data collection and measurements.



## Discussion

### Principal Findings

Despite the advancements in medical knowledge and diagnostic techniques, misdiagnosis remains a significant issue. AI has shown promise in the diagnosis and treatment of medical conditions; however, there is limited understanding of how AI uses patient history for diagnostic purposes. Our study aimed to investigate the extent to which AI (ChatGPT) can use information from medical history to accurately diagnose common diseases, which are frequently encountered in general outpatient, emergency, and ward

management settings. Although some studies have investigated the accuracy of AI-based medical diagnosis, our study is novel because it emphasizes the importance of patient history. We compared the diagnostic accuracy of diagnoses made on the basis of only patient history and those made using all the information; this makes our study unique. To the best of our knowledge, no previous research has been conducted on this topic.

Our study investigated the role of patient history in AI-assisted medical diagnoses using ChatGPT. We analyzed 30 standardized patient vignettes from *The BMJ* to assess the concordance rates between AI-proposed diagnoses based

on medical history only and those based on both medical history and additional information. Our results showed high concordance rates of 76.6% between the “Answer” and “History” groups, suggesting the importance of patient history in AI-assisted diagnoses and highlighting the potential of AI in improving diagnostic accuracy. This result is similar to that of a previous study that involved actual physicians instead of ChatGPT [9,10].

Characteristics of cases that did not lead to appropriate diagnoses based on history alone include, for instance, the following: an appendicitis case (case 2 in [Multimedia Appendix 1](#)) for which there was no documentation of pain migration in the medical history, a meningitis case (case 10 in [Multimedia Appendix 1](#)) wherein only headache and fever were documented, an otitis media case (case 16 in [Multimedia Appendix 1](#)) wherein only upper respiratory symptoms were recorded with no mention of ear-related symptoms, errors in identifying the causative agent in a case of acute gastroenteritis (case 27 in [Multimedia Appendix 1](#)), and an acute pharyngitis case (case 34 in [Multimedia Appendix 1](#)) that lacked the necessary medical history to determine the Centor score. Such omissions in the medical history could be considered contributing factors to the misdiagnoses. When physical findings and test data were added, an accurate diagnosis was achieved in 28 out of 30 cases (93.3%), showing a 16.7% increase in the accuracy rate. These two cases were of acute pharyngitis diagnosed as acute upper respiratory tract infection and *Salmonella* enteritis diagnosed as acute gastroenteritis. While we considered these incorrect diagnoses for the purpose of this study, they could have been deemed correct under certain criteria. Of the 7 cases that did not match between “Answer” and “History,” 6 were of infectious diseases (21 of 30 cases were of infectious diseases). These included cases where appendicitis was mistaken for acute gastroenteritis, acute otitis media and acute pharyngitis were mistaken for upper respiratory infections, and a *Salmonella* infection was mistaken for a *Campylobacter* infection. Physical examinations or tests may help identify the site of infection or pathogen in cases of intra-abdominal or head and neck infections.

There are situations in which physical examination and clinical test information may not be available in clinical settings. For instance, digital patient encounters owing to the impact of the COVID-19 pandemic often preclude physical examinations and clinical tests. The widespread use of telemedicine approaches in COVID-19 management, from screening to follow-up, has demonstrated the community’s acceptance and interest in telehealth solutions [32]. Moreover, even in face-to-face consultations, there are scenarios, such as in clinics, where detailed clinical tests may not be feasible depending on the setting. Furthermore, we cannot perform all physical examinations and tests on all patients. Therefore, we should consider potential differential diagnoses and decide which pertinent physical examinations or tests are the most suitable and should be performed. Most importantly, it has been reported that one rarely makes a correct diagnosis when one cannot make a differential diagnosis based on history [11]. In addition, accurately predicting the diagnosis based

on medical history is associated with a higher diagnostic accuracy of the physical examination, whereas incorrect prediction of the diagnosis based on medical history is associated with a lower diagnostic accuracy of the physical examination [33]. Based on these findings and suggestions, medical diagnosis using ChatGPT is considered heavily dependent on history.

Using AI for diagnosis can enhance diagnostic accuracy by more efficiently collecting medical histories. For instance, diagnosing acute appendicitis is sometimes challenging. AI may face the same challenge as that observed when, in our study, AI mistakenly identified acute appendicitis as acute gastroenteritis. This misdiagnosis may have occurred because the case lacked specific medical histories characteristic of appendicitis, such as pain migration. By configuring AI systems to verify pain migration in patients with abdominal pain, especially for such common conditions, diagnostic precision may improve.

There are 2 possible limitations in our study. First, it remains unclear whether similar results could be obtained with other vignettes or actual patients. Unlike using preprovided vignettes, among which we included 30 cases, diagnosis can be more challenging in clinical settings because it requires taking a medical history from patients. We included 30 cases from among the vignettes, which include some of the most commonly observed conditions in the outpatient setting. Although covering all the existing conditions is not feasible, we do not know if the case volume in our study is sufficiently high. This study included relatively simple cases in which patients had very few comorbidities, potentially making the diagnosis less challenging. Moreover, patients with psychiatric conditions tend to present with complex and lengthy case histories, and the wording used by mental health clinicians may differ, be inconsistent, be vague, or fail to pinpoint a diagnosis. Our vignettes did not include a diagnosis of any mental illness. Due to the abovementioned reasons, our results may not apply to all clinical settings. Furthermore, when we consider what the patient reports, results may differ if languages other than English are used since ChatGPT does not recognize some languages, and each language may have its unique nuance. This highlights the importance of linguistic diversity and cultural context in AI applications, particularly in medical diagnoses where patient communication and history are critical. Future iterations of AI systems should aim to incorporate a broader range of languages and understand cultural nuances to ensure more accurate and inclusive diagnostic support. This idea is important in the context of health inequality. Furthermore, disparities in technology access may pose some challenges. Future research should address these barriers to ensure equitable access to AI-assisted diagnostic tools.

Second, we encountered cases where the input of medical history followed by the question, “What is the most likely diagnosis?” failed to yield a single most likely diagnosis even after 5 attempts, which could have introduced bias into our results, although we only had 2 such cases.

In the future, studies should focus on training AI by implementing evidence-based medical information, enabling it to present the underlying reasons and guidelines for diagnoses. In the event of a misdiagnosis, analyzing the process that led to the false diagnosis could be challenging in an AI-assisted medical diagnosis. Given the current situation where reflection on misdiagnoses is not always feasible, AI should be used as an auxiliary tool in medical diagnosis. This approach underscores the importance of AI, deeming it a support system rather than a definitive diagnostic solution. This area needs further investigation. Future studies should also verify our results with certain common conditions or

diseases, such as the top 10 diseases identified in the Global Burden of Diseases study [34], potentially leveraging the benefits and limitations of AI-assisted medical diagnosis.

## Conclusions

Relevant patient history is essential for AI-assisted diagnosis. The input of relevant patient history or the development of AI systems capable of obtaining comprehensive medical histories is vital for AI-assisted medical diagnosis. Furthermore, even in the modern era of advanced medical knowledge and clinical testing, the significance of patient history in diagnosis remains crucial.

---

## Data Availability

All of our clinical vignettes, results, and prompts used are provided in [Multimedia Appendix 1](#).

---

## Authors' Contributions

FF conceptualized the study, designed the methodology, collected the data, and drafted the manuscript. YY, DY, and SU conceptualized the study, designed the methodology, and reviewed and edited the manuscript. SY, YL, KS, TT, KN, TU, and MI conceptualized the study and reviewed and edited the manuscript. No generative artificial intelligence was used in writing the manuscript.

---

## Conflicts of Interest

None declared.

---

## Multimedia Appendix 1

Clinical Vignettes used in our study.

[[PDF File \(Adobe File\), 159 KB-Multimedia Appendix 1](#)]

---

## Multimedia Appendix 2

Explanation of the prompts we used in our study.

[[PDF File \(Adobe File\), 49 KB-Multimedia Appendix 2](#)]

---

## References

1. Omron R, Kotwal S, Garibaldi BT, Newman-Toker DE. The diagnostic performance feedback “calibration gap”: why clinical experience alone is not enough to prevent serious diagnostic errors. *AEM Educ Train*. Oct 2018;2(4):339-342. [doi: [10.1002/aet2.10119](https://doi.org/10.1002/aet2.10119)] [Medline: [30386846](https://pubmed.ncbi.nlm.nih.gov/30386846/)]
2. Committee on Diagnostic Error in Health Care, Board on Health Care Services, Institute of Medicine, The National Academies of Sciences, Engineering, and Medicine. Balogh EP, Miller BT, Ball JR, editors. *Improving Diagnosis in Health Care*. National Academies Press; 2015. ISBN: 9780309377720
3. Friberg N, Ljungberg O, Berglund E, et al. Cause of death and significant disease found at autopsy. *Virchows Arch*. Dec 2019;475(6):781-788. [doi: [10.1007/s00428-019-02672-z](https://doi.org/10.1007/s00428-019-02672-z)] [Medline: [31691009](https://pubmed.ncbi.nlm.nih.gov/31691009/)]
4. Shojania KG, Burton EC, McDonald KM, Goldman L. Changes in rates of autopsy-detected diagnostic errors over time: a systematic review. *JAMA*. Jun 4, 2003;289(21):2849-2856. [doi: [10.1001/jama.289.21.2849](https://doi.org/10.1001/jama.289.21.2849)] [Medline: [12783916](https://pubmed.ncbi.nlm.nih.gov/12783916/)]
5. Schmitt BP, Kushner MS, Wiener SL. The diagnostic usefulness of the history of the patient with dyspnea. *J Gen Intern Med*. 1986;1(6):386-393. [doi: [10.1007/BF02596424](https://doi.org/10.1007/BF02596424)] [Medline: [3794838](https://pubmed.ncbi.nlm.nih.gov/3794838/)]
6. Kuijpers C, Fronczek J, van de Goot FRW, Niessen HWM, van Diest PJ, Jiwa M. The value of autopsies in the era of high-tech medicine: discrepant findings persist. *J Clin Pathol*. Jun 2014;67(6):512-519. [doi: [10.1136/jclinpath-2013-202122](https://doi.org/10.1136/jclinpath-2013-202122)] [Medline: [24596140](https://pubmed.ncbi.nlm.nih.gov/24596140/)]
7. Ball JR, Balogh E. Improving diagnosis in health care: highlights of a report from the National Academies of Sciences, Engineering, and Medicine. *Ann Intern Med*. Jan 5, 2016;164(1):59-61. [doi: [10.7326/M15-2256](https://doi.org/10.7326/M15-2256)] [Medline: [26414299](https://pubmed.ncbi.nlm.nih.gov/26414299/)]
8. Winters B, Custer J, Galvagno SM, et al. Diagnostic errors in the intensive care unit: a systematic review of autopsy studies. *BMJ Qual Saf*. Nov 2012;21(11):894-902. [doi: [10.1136/bmjqqs-2012-000803](https://doi.org/10.1136/bmjqqs-2012-000803)] [Medline: [22822241](https://pubmed.ncbi.nlm.nih.gov/22822241/)]
9. Hampton JR, Harrison MJ, Mitchell JR, Prichard JS, Seymour C. Relative contributions of history-taking, physical examination, and laboratory investigation to diagnosis and management of medical outpatients. *Br Med J*. May 31, 1975;2(5969):486-489. [doi: [10.1136/bmj.2.5969.486](https://doi.org/10.1136/bmj.2.5969.486)] [Medline: [1148666](https://pubmed.ncbi.nlm.nih.gov/1148666/)]
10. Peterson MC, Holbrook JM, Hales DV, Smith NL, Staker LV. Contributions of the history, physical examination, and laboratory investigation in making medical diagnoses. *Obstet Gynecol Surv*. Oct 1992;47(10):711-712. [doi: [10.1097/00006254-199210000-00013](https://doi.org/10.1097/00006254-199210000-00013)]

11. Gruppen LD, Palchik NS, Wolf FM, Laing TJ, Oh MS, Davis WK. Medical student use of history and physical information in diagnostic reasoning. *Arthritis Care Res.* Jun 1993;6(2):64-70. [doi: [10.1002/art.1790060204](https://doi.org/10.1002/art.1790060204)] [Medline: [8399428](#)]
12. Tsukamoto T, Ohira Y, Noda K, Takada T, Ikusaka M. The contribution of the medical history for the diagnosis of simulated cases by medical students. *Int J Med Educ.* Apr 2012;3:78-82. [doi: [10.5116/ijme.4f8a.e48c](https://doi.org/10.5116/ijme.4f8a.e48c)]
13. Chen ZH, Lin L, Wu CF, Li CF, Xu RH, Sun Y. Artificial intelligence for assisting cancer diagnosis and treatment in the era of precision medicine. *Cancer Commun (Lond).* Nov 2021;41(11):1100-1115. [doi: [10.1002/cac2.12215](https://doi.org/10.1002/cac2.12215)] [Medline: [34613667](#)]
14. Mitsala A, Tsalikidis C, Pitiakoudis M, Simopoulos C, Tsaroucha AK. Artificial intelligence in colorectal cancer screening, diagnosis and treatment. A new era. *Curr Oncol.* Apr 23, 2021;28(3):1581-1607. [doi: [10.3390/curroncol28030149](https://doi.org/10.3390/curroncol28030149)] [Medline: [33922402](#)]
15. Ochiai K, Ozawa T, Shibata J, Ishihara S, Tada T. Current status of artificial intelligence-based computer-assisted diagnosis systems for gastric cancer in endoscopy. *Diagnostics (Basel).* Dec 13, 2022;12(12):3153. [doi: [10.3390/diagnostics12123153](https://doi.org/10.3390/diagnostics12123153)] [Medline: [36553160](#)]
16. Calisto FM, Santiago C, Nunes N, Nascimento JC. Breastscreening-AI: evaluating medical intelligent agents for human-AI interactions. *Artif Intell Med.* May 2022;127:102285. [doi: [10.1016/j.artmed.2022.102285](https://doi.org/10.1016/j.artmed.2022.102285)] [Medline: [35430044](#)]
17. Zhou LQ, Wang JY, Yu SY, et al. Artificial intelligence in medical imaging of the liver. *World J Gastroenterol.* Feb 14, 2019;25(6):672-682. [doi: [10.3748/wjg.v25.i6.672](https://doi.org/10.3748/wjg.v25.i6.672)] [Medline: [30783371](#)]
18. Peng S, Liu Y, Lv W, et al. Deep learning-based artificial intelligence model to assist thyroid nodule diagnosis and management: a multicentre diagnostic study. *Lancet Digit Health.* Apr 2021;3(4):e250-e259. [doi: [10.1016/S2589-7500\(21\)00041-8](https://doi.org/10.1016/S2589-7500(21)00041-8)] [Medline: [33766289](#)]
19. Drukker L, Noble JA, Papageorgiou AT. Introduction to artificial intelligence in ultrasound imaging in obstetrics and gynecology. *Ultrasound Obstet Gynecol.* Oct 2020;56(4):498-505. [doi: [10.1002/uog.22122](https://doi.org/10.1002/uog.22122)] [Medline: [32530098](#)]
20. Guermazi A, Tannoury C, Kompel AJ, et al. Improving radiographic fracture recognition performance and efficiency using artificial intelligence. *Radiology.* Mar 2022;302(3):627-636. [doi: [10.1148/radiol.210937](https://doi.org/10.1148/radiol.210937)] [Medline: [34931859](#)]
21. Zhang K, Liu X, Shen J, et al. Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography. *Cell.* Jun 11, 2020;181(6):1423-1433. [doi: [10.1016/j.cell.2020.04.045](https://doi.org/10.1016/j.cell.2020.04.045)] [Medline: [32416069](#)]
22. Gore JC. Artificial intelligence in medical imaging. *Magn Reson Imaging.* May 2020;68:A1-A4. [doi: [10.1016/j.mri.2019.12.006](https://doi.org/10.1016/j.mri.2019.12.006)] [Medline: [31857130](#)]
23. Okagawa Y, Abe S, Yamada M, Oda I, Saito Y. Artificial intelligence in endoscopy. *Dig Dis Sci.* May 2022;67(5):1553-1572. [doi: [10.1007/s10620-021-07086-z](https://doi.org/10.1007/s10620-021-07086-z)] [Medline: [34155567](#)]
24. Ramesh AN, Kambhampati C, Monson JRT, Drew PJ. Artificial intelligence in medicine. *Ann R Coll Surg Engl.* Sep 2004;86(5):334-338. [doi: [10.1308/147870804290](https://doi.org/10.1308/147870804290)] [Medline: [15333167](#)]
25. Revilla-León M, Gómez-Polo M, Barmak AB, et al. Artificial intelligence models for diagnosing gingivitis and periodontal disease: a systematic review. *J Prosthet Dent.* Dec 2023;130(6):816-824. [doi: [10.1016/j.prosdent.2022.01.026](https://doi.org/10.1016/j.prosdent.2022.01.026)] [Medline: [35300850](#)]
26. Chung H, Jo Y, Ryu D, Jeong C, Choe SK, Lee J. Artificial-intelligence-driven discovery of prognostic biomarker for sarcopenia. *J Cachexia Sarcopenia Muscle.* Dec 2021;12(6):2220-2230. [doi: [10.1002/jesm.12840](https://doi.org/10.1002/jesm.12840)] [Medline: [34704369](#)]
27. Uzun Ozsahin D, Ozgocmen C, Balcioglu O, Ozsahin I, Uzun B. Diagnostic AI and cardiac diseases. *Diagnostics (Basel).* Nov 22, 2022;12(12):2901. [doi: [10.3390/diagnostics12122901](https://doi.org/10.3390/diagnostics12122901)] [Medline: [36552908](#)]
28. Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. *BMJ.* Jul 8, 2015;351:h3480. [doi: [10.1136/bmj.h3480](https://doi.org/10.1136/bmj.h3480)] [Medline: [26157077](#)]
29. North F, Jensen TB, Stroebel RJ, et al. Self-triage use, subsequent healthcare utilization, and diagnoses: a retrospective study of process and clinical outcomes following self-triage and self-scheduling for ear or hearing symptoms. *Health Serv Res Manag Epidemiol.* 2023;10:2333928231168121. [doi: [10.1177/2333928231168121](https://doi.org/10.1177/2333928231168121)] [Medline: [37101803](#)]
30. Riboli-Sasco E, El-Osta A, Alaa A, et al. Triage and diagnostic accuracy of online symptom checkers: systematic review. *J Med Internet Res.* Jun 2, 2023;25:e43803. [doi: [10.2196/43803](https://doi.org/10.2196/43803)] [Medline: [37266983](#)]
31. Radionova N, Ög E, Wetzel AJ, Rieger MA, Preiser C. Impacts of symptom checkers for laypersons' self-diagnosis on physicians in primary care: scoping review. *J Med Internet Res.* May 29, 2023;25:e39219. [doi: [10.2196/39219](https://doi.org/10.2196/39219)] [Medline: [37247214](#)]
32. Khoshrounejad F, Hamednia M, Mehrjerd A, et al. Telehealth-based services during the COVID-19 pandemic: a systematic review of features and challenges. *Front Public Health.* 2021;9:711762. [doi: [10.3389/fpubh.2021.711762](https://doi.org/10.3389/fpubh.2021.711762)] [Medline: [34350154](#)]

33. Shikino K, Ikusaka M, Ohira Y, et al. Influence of predicting the diagnosis from history on the accuracy of physical examination. *Adv Med Educ Pract.* 2015;6:143-148. [doi: [10.2147/AMEP.S77315](https://doi.org/10.2147/AMEP.S77315)] [Medline: [25759604](https://pubmed.ncbi.nlm.nih.gov/25759604/)]
34. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet.* Oct 17, 2020;396(10258):1204-1222. [doi: [10.1016/S0140-6736\(20\)30925-9](https://doi.org/10.1016/S0140-6736(20)30925-9)] [Medline: [33069326](https://pubmed.ncbi.nlm.nih.gov/33069326/)]

## Abbreviations

**AI:** artificial intelligence

*Edited by David Chartash, Gunther Eysenbach, Taiane de Azevedo Cardoso; peer-reviewed by Clarence Baxter, Hao Sun; submitted 12.09.2023; final revised version received 31.01.2024; accepted 15.02.2024; published 08.04.2024*

*Please cite as:*

Fukuzawa F, Yanagita Y, Yokokawa D, Uchida S, Yamashita S, Li Y, Shikino K, Tsukamoto T, Noda K, Uehara T, Ikusaka M. Importance of Patient History in Artificial Intelligence-Assisted Medical Diagnosis: Comparison Study. *JMIR Med Educ* 2024;10:e52674

URL: <https://mededu.jmir.org/2024/1/e52674>

doi: [10.2196/52674](https://doi.org/10.2196/52674)

© Fumitoshi Fukuzawa, Yasutaka Yanagita, Daiki Yokokawa, Shun Uchida, Shiho Yamashita, Yu Li, Kiyoshi Shikino, Tomoko Tsukamoto, Kazutaka Noda, Takanori Uehara, Masatomi Ikusaka. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 08.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on <https://mededu.jmir.org/>, as well as this copyright and license information must be included.